Algebras and Orders

NATO ASI Series

Advanced Science Institutes Series

A Series presenting the results of activities sponsored by the NATO Science Committee, which aims at the dissemination of advanced scientific and technological knowledge, with a view to strengthening links between scientific communities.

The Series is published by an international board of publishers in conjunction with the NATO Scientific Affairs Division

A Life Sciences **B Physics**	Plenum Publishing Corporation London and New York
C Mathematical and Physical Sciences **D Behavioural and Social Sciences** **E Applied Sciences**	Kluwer Academic Publishers Dordrecht, Boston and London
F Computer and Systems Sciences **G Ecological Sciences** **H Cell Biology** **I Global Environmental Change**	Springer-Verlag Berlin, Heidelberg, New York, London, Paris and Tokyo

NATO-PCO-DATA BASE

The electronic index to the NATO ASI Series provides full bibliographical references (with keywords and/or abstracts) to more than 30000 contributions from international scientists published in all sections of the NATO ASI Series.
Access to the NATO-PCO-DATA BASE is possible in two ways:

– via online FILE 128 (NATO-PCO-DATA BASE) hosted by ESRIN, Via Galileo Galilei, I-00044 Frascati, Italy.

– via CD-ROM "NATO-PCO-DATA BASE" with user-friendly retrieval software in English, French and German (© WTV GmbH and DATAWARE Technologies Inc. 1989).

The CD-ROM can be ordered through any member of the Board of Publishers or through NATO-PCO, Overijse, Belgium.

Series C: Mathematical and Physical Sciences - Vol. 389

Algebras and Orders

edited by

Ivo G. Rosenberg

and

Gert Sabidussi

Département de mathématiques et de statistique,
Université de Montréal,
Montréal, Québec, Canada

Springer-Science+Business Media, B.V.

Proceedings of the NATO Advanced Study Institute and
Séminaire de mathématiques supérieures on
Algebras and Orders
Montréal, Canada
July 29 – August 9, 1991

ISBN 978-90-481-4243-9 ISBN 978-94-017-0697-1 (eBook)
DOI 10.1007/978-94-017-0697-1

Printed on acid-free paper

Table of Contents

Preface

In the summer of 1991 the Department of Mathematics and Statistics of the Université de Montréal was fortunate to host the NATO Advanced Study Institute "Algebras and Orders" as its 30th Séminaire de mathématiques supérieures (SMS), a summer school with a long tradition and well-established reputation. This book contains the contributions of the invited speakers.

Universal algebra – which established itself only in the 1930's – grew from traditional algebra (e.g., groups, modules, rings and lattices) and logic (e.g., propositional calculus, model theory and the theory of relations). It started by extending results from these fields but by now it is a well-established and dynamic discipline in its own right. One of the objectives of the ASI was to cover a broad spectrum of topics in this field, and to put in evidence the natural links to, and interactions with, boolean algebra, lattice theory, topology, graphs, relations, automata, theoretical computer science and (partial) orders. The theory of orders is a relatively young and vigorous discipline sharing certain topics as well as many researchers and meetings with universal algebra and lattice theory.

W. Taylor surveyed the abstract clone theory which formalizes the process of composing operations (i.e., the formation of term operations) of an algebra as a special category with countably many objects, and leading naturally to the interpretation and equivalence of varieties.

D. Schweigert presented a comprehensive survey of the related and very recent domain of hyperidentities and hypervarieties. Hyperidentities, in particular unary ones, can be used to characterize certain classes of algebras. Solid varieties and weak homomorphisms are also considered.

The lattice of clones on a finite universe with more than two elements is largely unknown. Even finding all the atoms, called minimal clones, seems to be a very hard problem. In their research paper, H. Machida and I. Rosenberg consider the following variant: characterize all inclusion-minimal elements of the set of all clones that are both essentially nonunary and nonminimal. A solution is given for such clones generated by a groupoid and satisfying a certain condition.

Every primal algebra generates an arithmetical variety; the best known example is provided by boolean algebras. A. Pixley surveys results on arithmetical algebras and varieties, affine completeness and locally affine completeness, shows new results for such algebras without proper subalgebras and presents some illuminating counter-examples.

B. Jónsson surveys boolean algebras with operators (i.e., sup-preserving operations) which evolved from Tarski's axiomatic framework for the calculus of binary relations. This leads to discriminator varieties, dualities, modal and monadic algebras, dynamic algebras and algebras of programs.

Algebraic duality theory grew from Stone's duality for boolean algebras and the dualities for abelian groups, distributive lattices and semilattices, and became a standard tool often used for the representation of algebras arising from various logics. B. Davey provides an introduction and survey of duality in the context of universal algebra presented in a noncategorical way and mostly applied to subdirect powers of a finite algebra endowed with the discrete topology. Full and strong dualities are discussed as well as those for 2-element clones.

Partial algebras form a part of universal algebra applicable in theoretical computer science. However, several concepts which are both natural and crisp for full algebras, allow many variants in partial algebras. P. Burmeister surveys the universal algebra part of partial algebras (various notions of subalgebras, homomorphisms, congruences, free algebras, etc.) and model-theoretic aspects with a particular emphasis on many-sorted algebras and applications.

Lattice theory not only occupies a central place in universal algebra but has also many applications in other sciences. In a sense, free lattices are the most general lattices. R. Freese provides an introduction to this difficult and fascinating area which, although some 50 odd years old, recently witnessed significant breakthroughs.

Algebraic ordered sets naturally generalize algebraic lattices and apply to algebra, topology and theoretical computer science. M. Erné surveys the order-theoretical, algebraic and topological aspects of compact generation in ordered sets; in particular, the notions of inductive closure and $\mathcal{Z}$-compactness, sober spaces and spatial frames.

I. Rival provides an introduction to the drawing of Hasse diagrams of orders on various surfaces in an easily readable way. In decision-making choices must be made among alternatives ranked by precedence or preference, and so a graphical presentation of the data still plays a decisive role. The paper considers various aspects like slopes, bending and stretching, the covering graph decision problem, planarity and representations on various surfaces.

M. Pouzet presented a series of lectures on orders, graphs and automata from the metric point of view which are not included in this volume. They were based on his concept of generalized metric, and are intrinsically linked to several other lectures given at this ASI.

Boolean algebras and relational structures also appear as the key ingredients in I. Fleischer's proposed formalization of predicate calculus. He sketches such a formalization and indicates some ideas for a proof of the Gödel completeness theorem in this context.

It is fair to say that we have met our objective of bringing together specialists and ideas in two neighbouring and closely interacting fields. To all who helped to make this ASI a success, lecturers and participants alike, we wish to express our sincere thanks. Special thanks go to Aubert Daigneault, the director of the ASI, and Ghislaine David, its efficient and charming secretary, for the high quality and smoothness with which they handled the organization of the meeting.

Funding for this ASI was provided in large part by NATO, with additional support from the Natural Sciences and Engineering Research Council of Canada, and the Université de Montréal. To all three organizations we would like to express our gratitude for their support. For their efforts on behalf of this ASI we are especially grateful to the Scientific and Environmental Affairs Division of NATO, particularly to Dr. Luis V. da Cunha, the Director of the ASI program.

Ivo G. Rosenberg
Scientific Director of the ASI

Participants

Paolo AGLIANO
Dipartimento di Matematica
Università di Siena
Via del Capitano 15
I-53100 Siena
Italy

Safak ALPAY
Department of Mathematics
Middle East Technical University
06531 Ankara
Turkey

Cecilia M.V.C.C.AZEVEDO
Departamento de matemática
Universidade do Minho
P-4700 Braga
Portugal

Stephen L. BLOOM
Department of Pure & Applied Mathematics
Stevens Institute of Technology
Hoboken, NJ 07030
USA

Ferdinand BOERNER
Karl-Weierstrass-Institut
Mohrenstrasse 39
Pf. 1304
DO-1086 Berlin
Germany

Pierre BOUTIN
Département de mathématiques
Cégep de Sorel-Tracy
3000, bd de la Marie
Tracy, Qué., J3R 5B9
Canada

Ivan CHAJDA
Dept. of Algebra & Geometry
Palacky University
Svobody 26
77146 Olomouc
Czechoslovakia

Marc CHASTAND
I.A.E.
Université Lyon III
15, quai Claude Bernard
B.P. 638
F-69239 Lyon Cédex 02
France

John COLEMAN
Department of Mathematics
University of Colorado
Boulder, CO 80309-0426
USA

Julien CONSTANTIN
Département de mathématiques
Université de Sherbrooke
Sherbrooke, Qué., J1K 2R1
Canada

Ernesto COROMINAS
Université Lyon I
(deceased)

Hong X. DANG
Department of Mathematics
SUNY at Geneseo
Geneseo, NY 14454
USA

Elias DAVID
23 Lingwood Road
E5 9BN London
Great Britain

Christian DELHOMME
École Normale Supérieure de Lyon
46, allée d'Italie
F-69364 Lyon Cédex 07
France

Klaus DENECKE
Fachbereich Mathematik
Brandenburgische Landeshochschule
Am Neuen Palais
DO-1571 Postdam
Germany

Josée DESHARNAIS
Département de mathématiques
et de statistique
Université de Montréal
C.P. 6128, Succ. A
Montréal, Qué., H3C 3J7
Canada

Kevin EWACHA
Department of Mathematics
University of Ottawa
585 King Edward Ave.
Ottawa, Ont., K1N 6N5
Canada

Anne FEARNLEY
Département de mathématiques
et de statistique
Université de Montréal
C.P. 6128, Succ. A
Montréal, Qué., H3C 3J7
Canada

Roland FRAÏSSÉ
Le Cheverny - 1
Parc de la Cadenelle
122, rue Commandant Rolland
F-13008 Marseille
France

Reiner FRITZSCHE
Sektion Mathematik
Martin-Luther-Universität
DO-4010 Halle-Wittenberg
Germany

T. Simonthy FUAD
Department of Mathematics
Iowa State University
400 Carver Hall
Ames, IA 50010
USA

Kazimierz GLAZEK
Mathematical Institute
University of Wroclaw
Pl. Grunwaldzki 2/4
PL-50-384 Wroclaw
Poland

Daniel GLUSCHANKOF
Équipe de logique mathématique
Université de Paris VII-CNRS
Tour 45-55 - 5e étage
2, place Jussieu
75251 Paris Cédex 05
France

Jonathan GOLAN
Department of Mathematics
University of Haifa
IL-31905 Haifa
Israel

Matthew GOULD
Department of Mathematics
Vanderbilt University
Nashville, TN 37235
USA

Ewa GRACZYŃSKA
Institute of Mathematics
Polish Academy of Sciences
ul. Kopernika 18
PL-51-617 Wroclaw
Poland

Gerd HARTUNG
FB Mathematik - AG 1
Technische Hochschule Darmstadt
Schlossgartenstr. 7
DW-6100 Darmstadt
Germany

Simone HAZAN
Department of Mathematics
University of California
Berkeley, CA 94720
USA

Lutz HEINDORF
Karl-Weierstrass-Inst. für Mathematik
Mohrenstrasse 39
Pf. 1304
DO-1086 Berlin
Germany

Ou-Henry HO
Department of Mathematics
Rutgers University
New Brunswick, NJ 08903
USA

Jennifer HYNDMAN
Department of Mathematics
University of Colorado
Boulder, CO 80303-0426
USA

Pierre ILLE
CNRS Marseille, URA 225
Dép. de Mathématiques-Informatique
Faculté des Sciences de Luminy
Case 901
F-13288 Marseille Cédex 9
France

Peter JIPSEN
Department of Mathematics
Vanderbilt University
1326 Stevenson Ctr.
Nashville, TN 37240
USA

Peter JOHNSON
Department of Mathematics
University of Manitoba
Winnipeg, Manitoba, R3T 2N2
Canada

Gérard KIENTEGA
Département de mathématiques
et de statistique
Université de Montréal
C.P. 6128, Succ. A
Montréal, Qué., H3C 3J7
Canada

Volker KLEINSCHMIDT
FB4 AG1
Technische Hochschule Darmstadt
Schlossgartenstr. 7
DW-6100 Darmstadt
Germany

Benoît LAROSE
Département de mathématiques
et de statistique
Université de Montréal
C.P. 6128, Succ. A
Montréal, Qué., H3C 3J7
Canada

François LAVIOLETTE
Département de mathématiques
et de statistique
Université de Montréal
C.P. 6128, Succ. A
Montréal, Qué., H3C 3J7
Canada

Renato LEWIN
Facultad de Matemáticas
Pontifica Universidad Catolica de Chile
Casilla 6177
Santiago
Chile

Weixuan LI
Department of Computer Science
University of Ottawa
Ottawa, Ont., K1N 6N5
Canada

Otfried LÜDERS
Fakultät Mathematik
Universität Kaiserslautern
Pf. 3049
DW-6750 Kaiserslautern
Germany

Gerasimos C. MELETIOU
Faculty of Agricultural Technology
TEI/M at Arta
P.O. Box 110
GR-47100 Arta
Greece

Irene MIKENBERG
Facultad de Matemáticas
Pontifica Universidad Catolica de Chile
Casilla 6177
Santiago
Chile

Driss MISANE
Département de mathématiques
Faculté des Sciences
Université Mohammed V
B.P. 1014
Rabat
Morocco

Jutta MITAS
FB4-AG1
Technische Hochschule Darmstadt
Schlossgartenstr. 7
DW-6100 Darmstadt
Germany

Norbert NEWRLY
FB4-AG1
Technische Hochschule Darmstadt
Schlossgartenstr. 7
DW-6100 Darmstadt
Germany

Katarzyna PALASINSKA
Department of Mathematics
400 Carver Hall
Iowa State University
Ames, IA 50010-2066
USA

Péter Pál PÁLFY
Mathematical Institute
Hungarian Academy of Sciences
P.O. Box 127
H-1364 Budapest
Hungary

Jon PHILLIPS
Department of Mathematics
400 Carver Hall
Iowa State University
Ames, IA 50011-2066
USA

Norbert POLAT
I.A.E.
Université Lyon III
15, quai Claude Bernard
B.P. 638
F-69239 Lyon Cédex 02
France

Reinhard PÖSCHEL
Technische Universität Dresden
Abteilung Mathematik
Institut für Algebra
Mommsenstrasse 13
DO-8027 Dresden
Germany

Maurice POUZET
Département de Mathématiques
Université Claude Bernard
43, boul. du 11 Novembre 1918
F-69622 Villeurbanne Cédex
France

Hilary PRIESTLEY
Mathematical Institute
University of Oxford
24-29 St. Giles
OX1 3LB Oxford
Great Britain

Robert QUACKENBUSH
Department of Mathematics
University of Manitoba
Winnipeg, Manitoba, R3T 2N2
Canada

Jiri RACHUNEK
Department of Algebra & Geometry
University Palacky
Svobody 26
77146 Olomouc
Czechoslovakia

John RAFTER
Department of Mathematics
Vanderbilt University
1326 Stevenson Ctr.
Nashville, TN 37240
USA

Jürgen REINHOLD
Institut für Mathematik
Universität Hannover
Welfengarten 1
DW-3000 Hannover 1
Germany

Vladimir REPNITZKII
Department of Mathematics
Sverdlovsk Pedagogical State University
ul. Kosmonavtov 26
620219 Sverdlovsk
Russia

Serge RIBEYRE
I.A.E.
Université Lyon III
15, quai Claude Bernard
B.P. 638
F-69239 Lyon Cédex 02
France

Michael S. RODDY
Department of Mathematics
Brandon University
Brandon, Manitoba, R7A 6A9
Canada

Barbara ROSZKOWSKA
Institut Matematyki
Politechnika Warszawska
Plac Politechniki 1
PL-00-661 Warszawa
Poland

Bernard ROUX
Centre de Mathématiques
Inst. National des Sci. Appl. de Lyon (INSA)
20, ave Albert Einstein
F-69621 Villeurbanne
France

Gert SABIDUSSI
Département de mathématiques
et de statistique
Université de Montréal
C.P. 6128, Succ. A
Montréal, Qué., H3C 3J7
Canada

Faouzi SAÏDANE
Institut de Mathématiques et d'informatique
Université Claude Bernard
43, boul. du 11 Novembre 1918
F-69622 Villeurbanne Cédex
France

Arturo SANGALLI
Mathématiques
Champlain Regional College
Lennoxville, Qué., J1M 2A1
Canada

Marcin SCHROEDER
Department of Mathematics
Southern Illinois University
Carbondale, IL 62901
USA

Steven SEIF
Department of Mathematics
University of Louisville
Louisville, KY 40292
USA

Lev M. SHNEERSON
Department of Mathematics
Ivanova State University
ul. Yermak 39
153377 Ivanova
Russia

Claudette TABIB
Mathématiques
Collège Édouard-Montpetit
945, ch. de Chambly
Longueuil, Qué., J4H 3M6
Canada

Claude TARDIF
Département de mathématiques
et de statistique
Université de Montréal
C.P. 6128, Succ. A
Montréal, Qué., H3C 3J7
Canada

Michael TISCHENDORF
FB4 AG1
Technische Hochschule Darmstadt
Schlossgartenstr. 7
DW-6100 Darmstadt
Germany

Silke UMBREIT
Sektion Mathematik
Martin-Luther-Universität
DO-4010 Halle
Germany

Matt VALERIOTE
Department of Mathematics & Statistics
McMaster University
1280 Main Street West
Hamilton, Ont., L8S 4K1
Canada

Frank VOGT
FB4 AG1
Technische Hochschule Darmstadt
Schlossgartenstr. 7
DW-6100 Darmstadt
Germany

Michail VOLKOV
Department of Mathematics & Mechanics
Ural State University
620083 Sverdlovsk
Russia

Édouard WAGNEUR
École des Mines de Nantes
3, rue Marcel Sembat
F-44049 Nantes Cédex 04
France

Linda WALD
Department of Mathematics
UCLA
Los Angeles, CA 90024-1555
USA

Nikolai WEAVER
Department of Mathematics
University of California
Berkeley, CA 94720
USA

Shelly L. WISMATH
Department of Mathematical Sciences
University of Lethbridge
4401 University Dr.
Lethbridge, Alberta, T1K 3M4
Canada

Nejib ZAGUIA
Département d'informatique
Université d'Ottawa
Ottawa, Ont., K1N 6N5
Canada

Monika ZICKWOLFF
FB4 AG1
Technische Hochschule Darmstadt
Schlossgartenstr. 7
DW-6100 Darmstadt
Germany

Contributors

Peter BURMEISTER
FB4 AG1
Technische Hochschule Darmstadt
Schlossgartenstr. 7
DW-6100 Darmstadt
Germany

Brian DAVEY
Department of Mathematics
La Trobe University
Bundoora
3083 Victoria
Australia

Marcel ERNÉ
Institut für Mathematik
Universität Hannover
Welfengarten 1
DW-3000 Hannover 1
Germany

Isidore FLEISCHER
Dept. of Mathematics and Statistics
University of Windsor
Windsor, Ont., N9B 3P4
Canada

Ralph FREESE
Department of Mathematics
University of Hawaii
2565 The Mall
Honolulu, HI 97822
USA

Bjarni JÓNSSON
Department of Mathematics
Vanderbilt University
P.O. Box 1541, Station B
Nashville, TN 37235
USA

Hajime MACHIDA
Department of Mathematics
Hitotsubashi University
2-1 Naka, Kunitachi
Tokyo 186
Japan

Alden F. PIXLEY
Department of Mathematics
Harvey Mudd College
Claremont, CA 91711
USA

Ivan RIVAL
Département d'informatique
Université d'Ottawa
34, rue George Glinski
Ottawa, Ont., K1N 6N5
Canada

Ivo ROSENBERG
Département de mathématiques
et de statistique
Université de Montréal
C.P. 6128, Succ. A
Montréal, Qué., H3C 3J7
Canada

Dietmar SCHWEIGERT
Fakultät Mathematik
Universität Kaiserslautern
Pf. 3049
DW-6750 Kaiserslautern
Germany

Walter TAYLOR
Department of Mathematics
University of Colorado
Boulder, CO 80309-0426
USA

Partial Algebras — An Introductory Survey

Peter BURMEISTER *
Fachbereich Mathematik, Arbeitsgruppe Allgemeine Algebra
Technische Hochschule Darmstadt
Schloßgartenstraße 7
DW-6100 Darmstadt
Germany

Abstract

Partial algebras are among the basic mathematical structures implemented on computers. Many-sorted algebras are basically partial algebras, too. These notes are meant to introduce into a theory of and a language for partial algebras in such a way that also a specification of (many-sorted) partial algebras as abstract data types can easily be performed. Besides the terminology and constructions from universal algebra (homomorphisms, generalized recursion theorem, epimorphism theorem, free partial algebras) also such from logic (existence equations and elementary implications), model theory (preservation and reflection of formulas by mappings) and from (elementary) category theory (factorization systems) prove to be quite useful for a good description of the arising concepts, as is shown at the end by the formulation of a "Meta Birkhoff Theorem".

Motivation

With these notes we want to provide a somewhat easier access to a theory of partial algebras than the one of Burmeister [B86] which treats in a parallel way partialness, many-sortedness and possibly infinitary operations. Although we are well aware of the fact that a great part of the possible readers will consist of computer scientists who rather need a theory of many-sorted than of one-sorted partial algebras, we do not want to burden this introduction too much. Moreover, as we shall indicate in the appendix — which should be read by an interested reader after (closed) homomorphisms have been introduced —, there is a relatively easy access to main parts of the theory of many-sorted (partial) algebras, which is based on the category of homogeneous — i.e. one-sorted — partial algebras with (closed) homomorphisms. For an extended application of this approach see [B86].

Let us present — together with some historical remarks — some motivations why it is or could be interesting to study a theory of partial algebras and, should the occasion arise, why to teach their theory in particular to computer scientists.

*Special thanks are due to Norbert Newrly, who did the typing and assisted also with the more difficult drawings and some corrections. Further corrections are due to W. Bartol, B. Wojdylo and F. Rosselló.

I. G. Rosenberg and G. Sabidussi (eds.), Algebras and Orders, 1–70.

Partial functions have been used in mathematics for a long time, e.g.

- partially defined functions in analysis,
- partial subtraction for natural numbers,
- partial division for integers,
- partial multiplicative inversion in arbitrary fields,
- partial recursive functions in computability theory.

However, this did not lead to an independent general theory for structures with partial operations. One rather tried in most cases to complete the structures as far as possible, and field theory gave great impulses to the development of model theory. However, in Kleene [Kl52] there already appeared in connection with partial recursive functions three kinds of semantics of "equality" for partial (recursive) functions.

Yet finally, in connection with an enforced development of universal algebra and parallel to it of category theory in the 1960's with growing tendencies to greatest generality, some authors — e.g. J.Słomiński (starting about 1964) and J.Schmidt (starting about 1965) — investigated in their papers also the universal algebraic properties of partial algebras, what was then continued, not only by their "schools" (B.Wojdyło at Toruń, P.Burmeister and H.Höft at Bonn and Houston, respectively), but also by R.Kerkhoff, V.Poythress and later by H.Andréka, I.Németi, I.Sain, A.Pasztor, H.-J.Hoehnke, H.Reichel, H.Kaphengst and some others in the following decade. Some other sources of motivation at that time (may) have been

- the result of T.Evans in 1951 (see [Ev51]) about the equivalence of the solvability of the word problem for varieties of total universal algebras and of the decidability of the embeddability problem w.r.t. (that means "with respect to") the variety under consideration for finite partial algebras satisfying — w.r.t. special semantics (see subsection 3.3) — the equations in the equational theory of the variety;
- the use of partial algebras — and of some results about them — for the proof of the result of G.Grätzer and E.T.Schmidt in [GSchm63] (and later by W.Lampe e.g. in [La69]) that each algebraic lattice is isomorphic to the congruence lattice of some (multi-unary) universal algebra;
- the use of partial algebras as starting structures for the presentation of total algebraic structures (by generators and relations) and for the construction of "relatively free algebras over partial relative substructures";
- the possibility to describe the structure of relatively free algebras by first considering only some part of the structure in order to generate the elements as special terms in a term algebra — and here one often needs a partial algebraic structure — and then defining the rest of the structure by using the defining equations (identities); this method is now known in computer science as "canonical term representation" or — in another context — as "specification by constructors and relators" (if one can just partition the set of fundamental operations for that purpose);

- the fact that (small) categories are also partial algebras and this time with possibly much "wilder" partiality than that observed e.g. for fields (where there is only one exceptional case).

After these first investigations there seemed to be some "motivation crisis": Since most fundamental notions from universal algebra for total algebras split into at least three relevant concepts in the case of partial algebras, "one could not see the forest among the trees". That is to say, one did not see a starting point for a "nice" and unifying theory (and of important and attractive results). And this lack of a "good" theory also hindered new and interesting applications (there were some e.g. from quantum mechanics, cf. S.Kochen and E.P.Specker, in [KoSp68], but the "weak equality" used there was not transitive (!)).

New motivation for studying partial algebras then came in connection with the software crisis in computer science and the beginning awareness of computer scientists for the fact that universal algebra provides a good language and theory for theoretical computer science, e.g. for dealing with abstract data types and with programming languages and their semantics (see e.g. H.Reichel [Re84]). And in this connection one also observed that many — or even most — structures in computer science are partial, even when considered as many-sorted structures. In particular, since in a computer only finite parts of a — usually infinite — structure can be realized and computed, almost every implementation of a computer program represents a partial algebraic structure.

Here we think it to be a task for mathematicians to teach computer scientists and their students already universal algebra and in particular a theory of partial algebras. Namely it should not continue that compilers like TURBO PASCAL implement natural numbers in such a way that

$$32767 + 1 = -32768$$

instead of an (external) error message informing the user that his calculations got out of range.

We also do not think it adequate to specify originally partial data types in such a way that the exceptional cases get meaningful values, e.g. to read "0" (the integer zero) from an empty stack of integers as it is often proposed in books on specifications.

New motivation for a further development of a theory of partial algebras also came from a more general category theoretic investigation of first order logic and a meta theorem for Birkhoff type results concerning implicationally definable classes (see the papers of H.Andréka, I.Németi and I.Sain, e.g. [AN82], and [NSa82]). This result, of which we shall formulate a more algebraic version by the end of these notes, indicated that the concept of equality which we now call "existence equality" forms a good basis for an equational and implicational theory and in general for an expressive model theory for partial algebras, as we shall realize during these notes.

Once a reasonable theory of partial algebras has been developed new motivation grows from the general principle that new points of view give new insights, which may be applied among others in the following cases:

- As the theory of partial algebras with homomorphisms as basic structure preserving mappings and existence equations as basic model theoretic concept is designed, it lies between the theories of relational structures and the one of total algebras, which it comprises in a natural way as a subcategory. It is distinguished from the theory of relational systems in that it allows more easily to speak about generation — and we think generation to be one of the most important concepts in algebra, since it allows to describe large entities by singling out a small subset and giving operations and axioms telling how to get (=generate) the rest of the structure. In addition our theory of partial algebras also allows to speak very easily about definedness and undefinedness of terms.

However, the theory of partial algebras also comprises the one of relational systems in a relatively natural way: Given an n-ary relation R on a set A, R may be considered as an n-ary partial operation on A, say f_R, which has R as its domain and acts, say as partial first projection: $f_R(a_1, \ldots, a_n) := a_1$ for all $(a_1, \ldots, a_n) \in R$ (and undefined otherwise). The definition of f_R can be expressed by a first order sentence for an n-ary function symbol f as

$$(\forall x_1) \ldots (\forall x_n)(fx_1 \ldots x_n \stackrel{e}{=} fx_1 \ldots x_n \Rightarrow fx_1 \ldots x_n \stackrel{e}{=} x_1).$$

This shows that the class of all relational systems of similarity type τ can be considered as an axiomatic subclass of the class of all partial algebras of type τ, where even the axioms are elementary implications of a very simple structure.

The considerations above may also help to understand, why any theory of partial algebras has to be so rich and full of important fundamental concepts (cf. e.g. our discussion of substructures below, where in addition to the concepts of relative and weak relative substructures (also used in the theory of relational systems) we have in addition the one of subalgebra (=relative substructure on a closed subset) which is closely connected with the algebraic concept of generation and the only one which is usually considered for algebras).

Our observation not only gives new insights on the side of partial algebras, but it is also useful e.g. on the side of (total) algebras with relations, the theory of which is now also embeddable in the one of partial algebras. Thus it unifies and extends the theories of relational systems as well as of total algebras.

- Another application of the above principle may be the case of many-sorted (partial) algebras. Their theory is usually presented in such a way that the carrier sets of different sorts may be assumed without loss of generality (or only by simple modifications) to be disjoint. Then, on the disjoint union of the carrier sets of different sorts, the original many-sorted structure establishes in a canonical way a partial algebraic structure induced by the specification of the many-sorted similarity type; and on the set, say S, of all sorts, it establishes what we shall call the corresponding — partial — sort-algebra. The mapping, which assigns to each element of the carrier set of such a partial algebra its sort, then becomes a homomorphism, which is closed iff (meaning: "if and only if") the original many-sorted algebra is total. In this way the category of many-sorted partial algebras is in a natural way isomorphic to a comma category of the category of all partial algebras of the corresponding similarity type. We shall discuss this in more detail in the appendix, which should be read after homomorphisms and closed homomorphisms have been discussed in these notes. This approach helps to investigate partial algebras and many-sorted (partial) algebras to a great extent in a parallel way (this has been done e.g. in P.Burmeister [B86]).

- There is a further unifying effect by using partial algebras: By extending the structure under consideration "in a most general way" adjoint situations resulting from forgetting (part of) the structure, can be considered within one category by using either inclusion functors or the principle of universal solutions (relatively free constructions), what makes the resulting constructions in general more easily understandable; and universal solutions have usually to be studied anyway.

In these notes we want to combine some "different mathematical languages". This shall help us to keep track of the wealth of (possibly) relevant concepts arising e.g. from the concepts in the universal algebra of total algebras, as mentioned above. Therefore we will consider in particular

- the language of universal algebra (generation, freeness, algebraic "constructions", etc.)
- the language of first-order logic based on a concept of existence equations $t \stackrel{e}{=} t'$, which comprises relatively easy ways to speak about
  - the scopes and limitations of generation (by term existence statements $t \stackrel{e}{=} t$) including the possibility to express definedness and undefinedness
  - properties of mappings between partial algebras by
    - preservation of first-order formulas (e.g. homomorphisms as preserving all existence-equations)
    - reflection of first order formulas (e.g. closedness as reflection of all term-existence statements $t \stackrel{e}{=} t$, injectivity as reflection of $x \stackrel{e}{=} y$ for distinct variables x and y, etc.)
- the elementary language of category theory, e.g. the part on factorization systems as an additional tool to formulate properties of structure preserving mappings between partial algebras.

We will present the basic definitions, facts and aspects of the theory of partial algebras, yet in general we shall not include any proofs — these have to be taken from [B86]. We restrict considerations to the finitary one-sorted case, although in computer science abstract data types are usually many-sorted (partial) algebras, in order not to burden the presentation too much. However, as mentioned before, the reader interested in many-sorted (partial) algebras should read the appendix after (closed) homomorphisms have been introduced and then reread the notes and translate all definitions and statements to many-sorted (partial) algebras. As long as he does not allow empty carrier sets for his sorts, no problems should arise even in connection with the existence-equational theory or further model theoretic concepts and results. If, however, empty carrier sets are allowed, then he should rather refer to other literature, e.g. [B86], where even infinitary (partial) operations are treated.

1 Universal algebra of partial algebras

1.1 Similarity types and partial algebras

Let Ω be any set whose elements will be called *operation symbols.* In addition, let $\tau : \Omega \to \mathbb{N}_0$ be a mapping from Ω into the set of natural numbers including zero; for $\varphi \in \Omega$, $\tau(\varphi)$ will be interpreted as the *arity* of the operation symbol φ. If

- $\tau(\varphi) = 0$, then φ will be called a *nullary operation symbol* or a (*nullary*) *constant,*
- $\tau(\varphi) = 1$, then φ will be called a *unary operation symbol,*
- $\tau(\varphi) = 2$, then φ will be called a *binary operation symbol,*
- $\tau(\varphi) = n$, then φ will be called a *n-ary operation symbol.*

τ – or more precisely the pair (Ω, τ) – will be called a *similarity type* or briefly a *type* (in computer science it is usually called a *signature*). If not explicitly stated differently, in what follows we shall always assume that we are given an arbitrary but fixed similarity type (Ω, τ), and that all partial algebras under consideration are of this (same) type.

Let A be any set, and let n be any natural number including zero. An *n-ary partial operation* ψ on A is a function $\psi : dom\ \psi \to A$ – where "*dom* ψ" designates the *domain* of ψ – such that $dom\ \psi \subseteq A^n := \{(a_1, \ldots, a_n) | a_1, \ldots, a_n \in A\}$; i.e. ψ is a partial function out of A^n into A, and we denote this fact by $\psi : A^n \cdot\!\!\to A$. Observe that, for $n = 0$, $A^0 = \{\emptyset\}$ contains just the *empty sequence*, and therefore a partial nullary constant ψ on A is either empty or distinguishes exactly one element of A – and it is then usually identified with this element. By $PO^n(A)$ we designate the set of all partial n-ary operations on A, i.e. $PO^n(A) = \bigcup_{D \subseteq A^n} A^D$, and we set $PO(A) := \bigcup_{n=0}^{\infty} PO^n(A)$ to be the set of all finitary partial operations an A, while $O^n(A) := A^{(A^n)}$ and $O(A) := \bigcup_{n=0}^{\infty} O^n(A)$ designate the sets of all total n-ary and the set of all total finitary operations on A, respectively.

Let (Ω, τ) be any similarity type. Then a *partial algebra* $\underline{A}$ of type τ is an ordered pair $(C, (\mathcal{J}_{\underline{A}}(\varphi))_{\varphi \in \Omega})$, where C is any set, called the *carrier set* of $\underline{A}$, and $\mathcal{J}_{\underline{A}} : \Omega \to PO(C)$, $\mathcal{J}_{\underline{A}}(\varphi) \in PO^{\tau(\varphi)}(C)$ for $\varphi \in \Omega$, provides the interpretation of the $\tau(\varphi)$-ary partial operation symbol φ as a $\tau(\varphi)$-ary partial operation on the carrier set C with respect to $\underline{A}$. $\mathcal{J}_{\underline{A}}(\varphi)$ is called the *fundamental operation* of $\underline{A}$ corresponding to the operation symbol φ. In general – but not always – we shall use the letter A as the name for the carrier set of the partial algebra $\underline{A}$ (i.e. the so-called *forgetful functor*, which maps each partial algebra to its carrier set, i.e. which forgets the partial algebraic structure, is indicated by "forgetting" the understroke). Moreover, $\mathcal{J}_{\underline{A}}(\varphi)$ will be abbreviated by $\varphi^{\underline{A}}$. This yields a notation for the partial algebra $\underline{A}$ as $(A, (\varphi^{\underline{A}})_{\varphi \in \Omega})$, which seems to be recursive or to contain a self-reference, but which is just meant to be suggestive. Once this has been understood the above notation should not cause any confusion. Moreover, for binary operations we shall often use infix notation, i.e. we write $a\varphi^{\underline{A}}b$ instead of $\varphi^{\underline{A}}(a, b)$.

As the first example of partial algebras we consider *small categories* as one-sorted partial algebras of similarity type

$$(\{Dom, Cod, \circ\}, \{(Dom, 1), (Cod, 1), (\circ, 2)\}).$$

Then a small category $\underline{M}$ is a partial algebra $(M, (Dom^{\underline{M}}, Cod^{\underline{M}}, \circ^{\underline{M}}))$[1] such that

(C 1) $Dom^{\underline{M}}$ and $Cod^{\underline{M}}$ are total unary operations on M and $\circ^{\underline{M}} : M \times M \cdot\!\!\to M$ is a binary partial operation on M such that, for $f, g \in M$, $g \circ^{\underline{M}} f$ is defined in M iff $Cod^{\underline{M}}(f) = Dom^{\underline{M}}(g)$.

(C 2) For each $f \in M$, $f \circ^{\underline{M}} Dom^{\underline{M}}(f)$ and $Cod^{\underline{M}}(f) \circ^{\underline{M}} f$ always exist and yield f as value.

(C 3) Whenever $g \circ^{\underline{M}} f$ is defined in $\underline{M}$, then $Dom^{\underline{M}}(g \circ^{\underline{M}} f) = Dom^{\underline{M}}(f)$ and $Cod^{\underline{M}}(g \circ^{\underline{M}} f) = Cod^{\underline{M}}(g)$.

(C 4) If, for any $f, g, h \in M$, $g \circ^{\underline{M}} f$ and $h \circ^{\underline{M}} g$ are defined, then $(h \circ^{\underline{M}} g) \circ^{\underline{M}} f$ and $h \circ^{\underline{M}} (g \circ^{\underline{M}} f)$ are defined and equal: $(h \circ^{\underline{M}} g) \circ^{\underline{M}} f = h \circ^{\underline{M}} (g \circ^{\underline{M}} f)$.

Observe that a *category* in general is defined similarly without requiring that M be a set, i.e. M may then be a proper class; however it is then required that for all $f, g \in M$ $\{h | h \circ^{\underline{M}} f \text{ and } g \circ^{\underline{M}} h \text{ exist }\}$ has to be a set.

Extremal examples of partial algebras are on the one side the so-called *discrete partial algebras*, where all fundamental operations are empty: $\underline{A}_{discrete} := (A, (\emptyset)_{\varphi \in \Omega})$. On each set there is exactly one discrete partial algebra — which can be identified with its carrier set — of a given similarity type.

On the other side all (*total universal*) *algebras* are special partial algebras; here all fundamental operations are "everywhere defined", i.e. $(A, (\varphi^{\underline{A}})_{\varphi \in \Omega})$ is total, iff $dom\ \varphi^{\underline{A}} = A^{\tau(\varphi)}$ for each $\varphi \in \Omega$.

Thus all total algebras like groupoids, semigroups, monoids, groups, rings, semilattices, lattices, Boolean lattices and Boolean algebras (Boolean rings) are special examples of partial algebras. Fields are proper partial algebras where the multiplicative inversion "$^{-1}$" is the only proper partial operation (all other fundamental operations are total) (similarity type $(\{+, \underline{0}, -, \cdot, \underline{1}, ^{-1}\}, \{(+, 2), (\underline{0}, 0), (-, 1), (\cdot, 2), (\underline{1}, 0), (^{-1}, 1)\}))$.

Observe that if we write occasionally in connection with some partial algebra $\underline{A}$ that $\varphi^{\underline{A}}(a_1, \ldots, a_{\tau(\varphi)}) = a$, then we mean that $\varphi^{\underline{A}}(a_1, \ldots, a_{\tau(\varphi)})$ exists and has the value a.

1.2 Substructures, generation

While one uses for total algebras only one concept of subobjects – namely the one of subalgebras –, the use of partial algebras in different contexts or kinds of applications justifies three different basic concepts of subobjects.

[1]This is the so-called one-sorted description, where M stands for the set of morphisms of the category under consideration and the subset $Cod^{\underline{M}}(M) = Dom^{\underline{M}}(M)$ represents the set of objects via their identity morphisms.

1. The principle of *generation of structures* corresponds to the concept of a (*closed*) *subalgebra*, which corresponds to the one of a subalgebra in connection with total algebras. It is based on the concept of a *closed subset* of a partial algebra:

A subset M of the carrier set A of some partial algebra $(A, (\varphi^{\underline{A}})_{\varphi\in\Omega})$ is called *closed* if it is closed with respect to applications of all fundamental operations of $\underline{A}$ to sequences formed only by elements of M, i.e.

(C) for all $\varphi \in \Omega$ and for all $(a_1, \ldots, a_{\tau(\varphi)}) \in M^{\tau(\varphi)} \cap dom\ \varphi^{\underline{A}}$ one also has $\varphi^{\underline{A}}(a_1, \ldots, a_{\tau(\varphi)}) \in M$.

A partial algebra $\underline{B} = (B, (\varphi^{\underline{B}})_{\varphi\in\Omega})$ is then called a *(closed) subalgebra* of the partial algebra $\underline{A} = (A, (\varphi^{\underline{A}})_{\varphi\in\Omega})$ of the same similarity type, iff

(c 1) B is a closed subset of $\underline{A}$;

(c 2) for each $\varphi \in \Omega$, $\varphi^{\underline{B}}$ is the *restriction* to $B^{\tau(\varphi)}$ of the fundamental operation $\varphi^{\underline{A}}$ (in symbols $\varphi^{\underline{B}} = \varphi^{\underline{A}}\ |_{B^{\tau(\varphi)}}$), i.e. one has $dom\ \varphi^{\underline{B}} = dom\ \varphi^{\underline{A}} \cap B^{\tau(\varphi)}$ and if $(b_1, \ldots, b_{\tau(\varphi)}) \in dom\ \varphi^{\underline{B}}$, then $\varphi^{\underline{B}}(b_1, \ldots, b_{\tau(\varphi)}) = \varphi^{\underline{A}}(b_1, \ldots, b_{\tau(\varphi)})$.

These two properties are combined in writing

$$graph\ \varphi^{\underline{B}} = graph\ \varphi^{\underline{A}} \cap (B^{\tau(\varphi)} \times A)$$

with $graph\ f := \{(c, f(c)) | c \in C\}$ designating the *graph* of a partial function $f : C \cdot\!\!\to D$.

This definition shows that not every subset of a partial algebra $\underline{A}$ can be the carrier set of a subalgebra of $\underline{A}$. We shall explain the connection to generation immediately after these definitions.

2. The principle of *restriction of structure* to subsets of an algebra corresponds to the one of a *relative subalgebra*:

A partial algebra $\underline{B} = (B, (\varphi^{\underline{B}})_{\varphi\in\Omega})$ is called a *relative subalgebra* of a partial algebra $\underline{A}$ of the same similarity type, iff

(r 1) B is some (arbitrary) subset of A,

(r 2) for each $\varphi \in \Omega$, $\varphi^{\underline{B}}$ is the *total restriction* of $\varphi^{\underline{A}}$ to B (in symbols: $\varphi^{\underline{B}} = \varphi^{\underline{A}}||_B$), what means that $dom \varphi^{\underline{B}} \subseteq dom\ \varphi^{\underline{A}}$, and for any $(b_1, \ldots, b_{\tau(\varphi)}) \in B^{\tau(\varphi)}$ if $(b_1, \ldots, b_{\tau(\varphi)}) \in dom\ \varphi^{\underline{A}}$ and $\varphi^{\underline{A}}(b_1, \ldots, b_{\tau(\varphi)}) \in B$, then $(b_1, \ldots, b_{\tau(\varphi)}) \in dom\ \varphi^{\underline{B}}$ and $\varphi^{\underline{B}}(b_1, \ldots, b_{\tau(\varphi)}) = \varphi^{\underline{A}}(b_1, \ldots, b_{\tau(\varphi)})$. This is briefly indicated by writing:

$$graph\ \varphi^{\underline{B}} = graph\ \varphi^{\underline{A}} \cap (B^{\tau(\varphi)} \times B).$$

This definition shows that every subset B of A is the carrier set of a relative subalgebra $\underline{B}$ of $\underline{A}$, and that the structure of $\underline{B}$ is totally defined by the structure of $\underline{A}$ and the specification of the subset B (as in the case of subalgebras with the only difference that for a subalgebra the carrier set has to be closed).

3. The principle of "*approximation*" (or "*exhaustion*") of algebraic structures by "very small" (e.g by finite) pieces (as it is somewhat done e.g. in computers) leads to the concept

of a weak relative subalgebra: A partial algebra $\underline{B} = (B, (\varphi^{\underline{B}})_{\varphi \in \Omega})$ is called a *weak relative subalgebra* of some partial algebra $\underline{A} = (A, (\varphi^{\underline{A}})_{\varphi \in \Omega})$, iff

(w 1) B is some (arbitrary) subset of A

(w 2) for each $\varphi \in \Omega$, the graph of $\varphi^{\underline{B}}$ is contained in the graph of $\varphi^{\underline{A}}$: *graph* $\varphi^{\underline{B}} \subseteq$ *graph* $\varphi^{\underline{A}}$, i.e. if $\varphi^{\underline{B}}(b_1, \ldots, b_{\tau(\varphi)}) = b$, then $\varphi^{\underline{A}}(b_1, \ldots, b_{\tau(\varphi)}) = b$ in $\underline{A}$. Observe that now each subset B of a partial algebra $\underline{A}$ is the carrier set of at least one but in general of quite a lot of weak relative subalgebras of $\underline{A}$.

In particular one has that every subalgebra of $\underline{A}$ is a relative subalgebra of $\underline{A}$, and every relative subalgebra of $\underline{A}$ is a weak relative subalgebra of $\underline{A}$. Moreover, every relative subalgebra on a closed subset is a subalgebra. As in the case of total algebras the process of generation is now based on

Proposition 1.1 *The intersection of an arbitrary set $\mathcal{G}$ of closed subsets of a partial algebra is again a closed subset of $\underline{A}$, e.g. $A = \bigcap \emptyset$ is a closed subset of A.*

Corollary and Definition *For each partial algebra $\underline{A}$ the system $cs(\underline{A})$ of all closed subsets of $\underline{A}$ is a closure system on A, and the corresponding closure operator is denoted by $\mathcal{C}_{\underline{A}}$, i.e. for each subset M of A one has:*

$$\mathcal{C}_{\underline{A}}M := \bigcap \{H | M \subseteq H \in cs(\underline{A})\}.$$

Let us recall that a closure operator $\mathcal{C}$ on a set A always has the following properties for all $M, N \subseteq A$:

(C 1) $M \subseteq \mathcal{C}M$ (extensity)

(C 2) $M \subseteq N \Rightarrow \mathcal{C}M \subseteq \mathcal{C}N$ (monotonicity)

(C 3) $\mathcal{C}\mathcal{C}M = \mathcal{C}M$ (idempotency)

Since we deal with finitary partial operations only, we have in addition the

Proposition 1.2 *The operators $\mathcal{C}_{\underline{A}}$ on partial algebras are* algebraic, *i.e. they satisfy*

(CA) $\quad \mathcal{C}_{\underline{A}}M = \bigcup \{\mathcal{C}_{\underline{A}}F | F \subseteq M \text{ finite}\}$

for every subset M of any given partial algebra $\underline{A}$.

In addition one gets that the finitely generated closed subsets are exactly the (lattice theoretically) *compact* elements of the complete lattice of all closed subsets of a given partial algebra $\underline{A}$ where the supremum of a given set $\mathcal{G}$ of closed subsets is computed as $\mathcal{C}_{\underline{A}} \bigcup \mathcal{G}$ as usual.

The observation of Proposition 1.1 can be generalized, if we introduce for arbitrary weak relative subalgebras $\underline{B}, \underline{C}$ of a partial algebra $\underline{A}$:

$\underline{B} \subseteq \underline{C}$ iff $B \subseteq C$ and, for each $\varphi \in \Omega$: *graph* $\varphi^{\underline{B}} \subseteq$ *graph* $\varphi^{\underline{C}}$. If $\mathcal{G}$ is a system of weak relative subalgebras of a partial algebra $\underline{A}$, then we define its intersection $\bigcap \mathcal{G}$ as

$$(\bigcap\{H|\underline{H} \in \mathcal{G}\}, (\bigcap\{\varphi^{\underline{H}}|\underline{H} \in \mathcal{G}\})_{\varphi\in\Omega})$$

with $graph \bigcap\{\varphi^{\underline{H}}|\underline{H} \in \mathcal{G}\} := \bigcap\{graph\, \varphi^{\underline{H}}|\underline{H} \in \mathcal{G}\}$. This is motivated by the

Proposition 1.3 *Let $\underline{A}$ be any partial algebra, and let $\mathcal{G}$ be a system of subalgebras, of relative subalgebras or of weak relative subalgebras, respectively, of $\underline{A}$. Then $\bigcap \mathcal{G}$ is a subalgebra, a relative subalgebra or a weak relative subalgebra of $\underline{A}$, respectively, with carrier set $B := \bigcap\{H|\underline{H} \in \mathcal{G}\}$.*

While relative and weak relative substructures can also be defined for relational systems, the concept of subalgebras is specific for partial algebras and total algebras and allows non-trivial generation. Since $\mathcal{C}_{\underline{A}}M$ requires the knowledge of all closed subsets of $\underline{A}$ containing M, one introduces more-step closure operators exhausting $\mathcal{C}_{\underline{A}}M$ in order to be able to "generate locally".

Definition Let $\underline{A}$ be a partial algebra and let M be an arbitrary subset of A.

– Then $\mathcal{C}_{\underline{A}}M$ is called the closed subset *generated by M*, and $\underline{\mathcal{C}}_{\underline{A}}M$ – the subalgebra of $\underline{A}$ with carrier set $\mathcal{C}_{\underline{A}}M$ – is called the *subalgebra* of $\underline{A}$ *generated by M*, and M is said to be a *generating subset* of $\underline{\mathcal{C}}_{\underline{A}}M$.

– Define

$$\mathcal{D}_{\underline{A}}M := M \cup \bigcup\{\varphi^{\underline{A}}(a_1,\ldots,a_{\tau(\varphi)})|\varphi \in \Omega \text{ and } (a_1,\ldots,a_{\tau(\varphi)}) \in M^{\tau(\varphi)} \cap dom\, \varphi^{\underline{A}}\}$$

and for each natural number n define recursively

$$\begin{aligned} \mathcal{D}^0_{\underline{A}}M &:= M, \\ \mathcal{D}^{n+1}_{\underline{A}}M &:= \mathcal{D}_{\underline{A}}(\mathcal{D}^n_{\underline{A}}M), \\ \mathcal{B}^0_{\underline{A}}M &:= M, \\ \mathcal{B}^{n+1}_{\underline{A}}M &:= \mathcal{D}^{n+1}_{\underline{A}} \setminus \mathcal{D}^n_{\underline{A}}M. \end{aligned}$$

$\mathcal{B}^n_{\underline{A}}M$ is called the n-th *Baire-class* of M.

The *principle of generation from below* is then described by

Proposition 1.4 *Let $\underline{A}$ be any partial algebra, and let $M \subseteq A$ be any subset. Then*

(i) $M \subseteq \mathcal{D}^n_{\underline{A}}M \subseteq \mathcal{D}^m_{\underline{A}}M \subseteq \mathcal{C}_{\underline{A}}M$ *for any natural numbers $n \leq m$.*

(ii) $\mathcal{C}_{\underline{A}}M = \bigcup_{n=0}^{\infty} \mathcal{D}^n_{\underline{A}}M = \bigcup_{n=0}^{\infty} \mathcal{B}^n_{\underline{A}}M$.

(iii) $\mathcal{C}_{\underline{A}}M = \bigcup_{n=0}^{k} \mathcal{D}^n_{\underline{A}}M$, *iff* $\mathcal{D}^{k+1}_{\underline{A}}M = \mathcal{D}^k_{\underline{A}}M$, *iff* $\mathcal{B}^{k+1}_{\underline{A}}M = \emptyset$.

Observe that for the proof of (ii) one mainly has to show that $B := \bigcup_{n=0}^{\infty} \mathcal{D}^n_{\underline{A}}M$ is a closed subset of $\underline{A}$.

Immediately from the definition of Baire-classes one gets the following

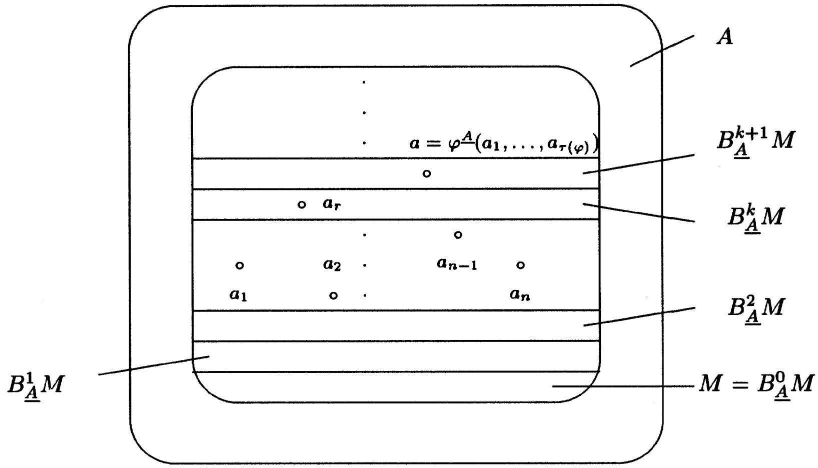


Figure 1: Layer model of generation from below

Lemma 1.5 *Let $\underline{A}$ be any partial algebra, let M be any subset of A and let $a \in \mathcal{C}_{\underline{A}}M$.*

(i) *If $a \in \mathcal{B}^n_{\underline{A}}M$ for some $n \geq 1$, i.e. e.g. $a \in \mathcal{C}_{\underline{A}}M \setminus M$, then there are $\varphi \in \Omega$, $a_1, \ldots, a_{\tau(\varphi)} \in \mathcal{D}^{n-1}_{\underline{A}}M$ such that $a = \varphi^{\underline{A}}(a_1, \ldots, a_{\tau(\varphi)})$. In particular, if $\tau(\varphi) \geq 1$, then there is at least one $r \in \{1, \ldots, \tau(\varphi)\}$ such that $a_r \in \mathcal{B}^{n-1}_{\underline{A}}M$.*

(ii) *$\varphi^{\underline{A}} \in \mathcal{D}_{\underline{A}}M$ for every non-empty nullary constant $\varphi^{\underline{A}}$ of $\underline{A}$.*

Figure 1 shows this *layer model of generation from below* in some details.

Examples

(i) In the (total) algebra $(\mathrm{N}_0; (\underline{0}, '))$ of similarity type $(\{\underline{0}, '\}, \{(\underline{0}, 0), (', 1)\})$ with $n' := n + 1$ one has $\mathcal{B}^n_{\underline{A}}\emptyset = \{n - 1\}$, i.e. the process of generation does not stop for any finite natural number n.

(ii) Consider vector spaces over a field $\underline{F}$ as one-sorted total algebras $\underline{V} := (V, +^{\underline{V}}, \underline{0}^{\underline{V}}, -^{\underline{V}}, (f\cdot^{\underline{V}})_{f\in F})$ of type $(\{+, \underline{0}, -\}\dot{\cup}\{f\cdot \mid f \in F\}, \{(+, 2), (\underline{0}, 0), (-, 1)\} \cup \{(f\cdot, 1) \mid f \in F\})$, where $f \cdot^{\underline{V}} (v) := fv$ for every scalar f and every vector v. Let $M := \{v_1, \ldots, v_k\}$ be a finite basis (i.e. linearly independent generating subset) of $\underline{V}$. Then $\mathcal{D}^n_{\underline{V}}M = V$ for $n = \lceil log_2 k \rceil + 1$, where $\lceil . \rceil$ are the upper Gaussian brackets ($\lceil r \rceil$ is the smallest integral number $n \geq r$ for any real number r).

Closely connected with the concept of generation is the proof principle of *algebraic induction* (in computer science it is called *structural induction*):

Let $\underline{A} = (A, (\varphi^{\underline{A}})_{\varphi\in\Omega})$ be any partial algebra of some similarity type (Ω, τ), let $\mathcal{P}$ be any property relevant for elements of A, and let M be any subset of A.

(i) If every element of M has the property $\mathcal{P}$ (*basis of induction*) and

(ii) if for any $\varphi \in \Omega$ and any $a_1, \ldots, a_{\tau(\varphi)} \in A$ with $(a_1, \ldots, a_{\tau(\varphi)}) \in dom\ \varphi^{\underline{A}}$ the fact that all of $a_1, \ldots, a_{\tau(\varphi)}$ have the property $\mathcal{P}$ (*induction hypothesis*) implies that also $\varphi^{\underline{A}}(a_1, \ldots, a_{\tau(\varphi)})$ has the property $\mathcal{P}$ (*induction step*),

then all elements of $C_{\underline{A}}M$ have the property $\mathcal{P}$ (*induction conclusion*).

This is a generalization of the method of complete induction: apply it to the total algebra $(\mathrm{N}_0, ')$ and to $M := \{0\}$ ($'$ designating the successor function $n \mapsto n+1$).

An example of this method of proof will be given in connection with Lemma 1.25, and the proof of Lemma 2.2 is also an excellent application of it.

1.3 Homomorphisms, monomorphisms and epimorphisms

For the *comparison* of partial algebras we use the concept of homomorphisms, which means "structure preserving" mappings which are defined on all of the start object. This concept has proven to be most useful for our purposes, e.g. of the one of designing a theory of partial algebras. Even on those occasions where partial mappings seem to be necessary — see e.g. below the valuation mappings — in connection with our development of the theory of partial algebras, the arising concepts can be formulated within the category of partial algebras with homomorphisms as morphisms (see the remarks at the end of subsection 3.5).

Homomorphisms and isomorphisms

Definition Let $\underline{A}$ and $\underline{B}$ be any partial algebras of the same similarity type (Ω, τ), and let $f : A \to B$ be any mapping (e.g. *dom* $f = A$). We call f a *homomorphism* from $\underline{A}$ into $\underline{B}$, if for all $\varphi \in \Omega$ and for all $a_1, \ldots, a_{\tau(\varphi)} \in A$ we have

(HC) $\varphi^{\underline{A}}(a_1, \ldots, a_{\tau(\varphi)}) = a$ implies $\varphi^{\underline{B}}(f(a_1), \ldots, f(a_{\tau(\varphi)})) = f(a)$.

This means that whenever $\varphi^{\underline{A}}(a_1, \ldots, a_{\tau(\varphi)})$ is defined, then $\varphi^{\underline{B}}(f(a_1), \ldots, f(a_{\tau(\varphi)}))$ has to be defined and has to be equal to $f(\varphi^{\underline{A}}(a_1, \ldots, a_{\tau(\varphi)}))$:

$$\varphi^{\underline{B}}(f(a_1), \ldots, f(a_{\tau(\varphi)})) = f(\varphi^{\underline{A}}(a_1, \ldots, a_{\tau(\varphi)})).$$

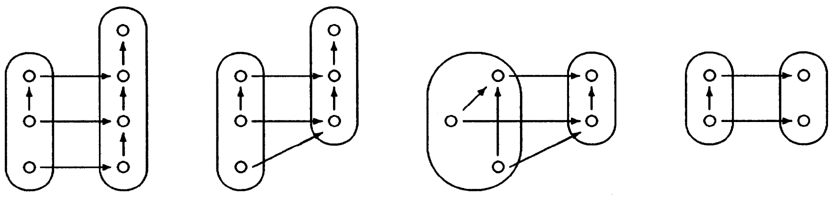
We express the fact that f is a homomorphism from $\underline{A}$ into $\underline{B}$ by writing $f : \underline{A} \to \underline{B}$.

Two properties of homomorphisms are needed quite often and therefore we introduce them right here:

A homomorphism $f : \underline{A} \to \underline{B}$ is called

- *full*, if for every $\varphi \in \Omega$ and for all $a_1, \ldots, a_{\tau(f)}, a \in A$ we have:

  if $\varphi^{\underline{B}}(f(a_1), \ldots, f(a_{\tau(f)})) = f(a)$, then there is $(a'_1, \ldots, a'_{\tau(\varphi)}) \in dom\ \varphi^{\underline{A}}$ such that $f(a'_i) = f(a_i)$ for all $1 \leq i \leq \tau(\varphi)$.



f_1	f_2	f_3	f_4
neither full nor closed	full, but not closed	closed (and therefore full)	not even a homomorphism

Figure 2: Examples of homomorphisms

– *closed*, if for every $\varphi \in \Omega$ and for all $a_1, \ldots, a_{\tau(\varphi)} \in A$ one has:
if $(f(a_1), \ldots, f(a_{\tau(\varphi)})) \in dom\ \varphi^{\underline{B}}$, then $(a_1, \ldots, a_{\tau(\varphi)}) \in dom\ \varphi^{\underline{A}}$.

Observe that every closed homomorphism is also full. As examples consider those in Figure 2.

Definition As usual and e.g. as in category theory we define an *isomorphism* to be any homomorphism $f : \underline{A} \to \underline{B}$ such that there exists a homomorphism $g : \underline{B} \to \underline{A}$ satisfying $g \circ f = id_A$ and $f \circ g = id_B$ (where $g \circ f$ and $f \circ g$ are the composed mappings $g \circ f : A \to A$, $f \circ g : B \to B$, $g \circ f(a) = g(f(a))$ etc., and where $id_A : A \to A$ is the identity mapping $id_A(a) = a$ for all $a \in A$ and id_B is defined correspondingly). g is called the *inverse* of f and often denoted by f^{-1}.

Proposition 1.6 *Let $\underline{A}, \underline{B}, \underline{C}$ be partial algebras of the same similarity type (Ω, τ). Then:*

(i) *The identity mapping $id_A : A \to A$ is always a homomorphism and even an isomorphism $id_A : \underline{A} \to \underline{A}$ with id_A being its inverse.*

(ii) *The composition of homomorphisms is always a homomorphism: If $f : \underline{A} \to \underline{B}$ and $g : \underline{B} \to \underline{C}$ are homomorphisms, then $g \circ f : \underline{A} \to \underline{C}$, $a \mapsto g(f(a))$ is also a homomorphism. If both, f and g are isomorphisms, then so is $g \circ f$.*

(iii) *If $f : A \to B$ is any mapping, and if $\underline{A}$ is a discrete partial algebra ($\varphi^{\underline{A}} = \emptyset$ for each $\varphi \in \Omega$), then $f : \underline{A} \to \underline{B}$ is a homomorphism.*

Statements (i) and (ii) above show that the class $Alg(\tau)$ of all partial algebras of type (Ω, τ) as object class together with all homomorphisms between elements of $Alg(\tau)$ forms a category, denoted by $\underline{Alg(\tau)}$. The subclass $TAlg(\tau)$ of all total algebras of type τ together with all homomorphisms between them forms the category which is well known from the universal algebra of total algebras.

While it is also well known for total algebras that every bijective homomorphism between total algebras is an isomorphism, this is not the case for partial algebras: Let $\underline{B}$ be any total algebra with non-empty carrier set B, and let $\underline{A}$ be the discrete partial algebra with carrier set B. Then $id_B : \underline{A} \to \underline{B}$ is a homomorphism while it is not a homomorphism from $\underline{B}$ into $\underline{A}$: for any φ and any $b_1, \ldots, b_{\tau(\varphi)} \in B$ $\varphi^{\underline{B}}(b_1, \ldots, b_{\tau(\varphi)})$ exists, while $\varphi^{\underline{A}}(b_1, \ldots, b_{\tau(\varphi)})$ does not exist (see e.g. f_4^{-1} in Figure 2). The situation is comparable with the one in general topology for continuous mappings; the total algebras here correspond to compact Hausdorff spaces there.

The corresponding result for partial algebras is contained in

Proposition 1.7 *Let $\underline{A}, \underline{B}$ be partial algebras of the same similarity type (Ω, τ), and let $f : \underline{A} \to \underline{B}$ be any homomorphism. Then the following statements are equivalent:*

(i) *f is an isomorphism.*

(ii) *f is bijective, and the inverse map $f^{-1} : B \to A$ — $f^{-1}(b) = a$ iff $f(a) = b$ — is also a homomorphism.*

(iii) *f is a full and bijective homomorphism.*

(iv) *f is a closed and bijective homomorphism.*

Monomorphisms and epimorphisms

In category theory there are (at least) two further concepts which are usually considered.

Definition Let $\mathcal{C} = (Ob(\mathcal{C}), Mor(\mathcal{C}), (\circ, Dom, Cod, 1))$ be an arbitrary category. Let $A, B \in Ob(\mathcal{C})$ and $f : A \to B$ be an arbitrary morphism in $\mathcal{C}$.

a) f is called a *monomorphism* of $\mathcal{C}$, if for any object $C \in Ob(\mathcal{C})$ and for any two morphisms $g, h : C \to A$, $f \circ g = f \circ h$ always implies $g = h$.

b) f is called an *epimorphism* of $\mathcal{C}$, if for any object $D \in Ob(\mathcal{C})$, and for any two morphisms $u, v : B \to D$, $u \circ f = v \circ f$ implies $u = v$.

Observe that an isomorphism is always both, a monomorphism and an epimorphism, and in the category $\underline{Alg(\tau)}$ of all partial algebras of type τ as objects and with all homomorphisms as morphisms each bijective homomorphism is as well a monomorphism as an epimorphism (see the statements below).

A detailed characterization of monomorphisms and epimorphisms is contained in:

Proposition 1.8 (Characterization of monomorphisms)
In $\underline{Alg(\tau)}$ a homomorphism $f : \underline{A} \to \underline{B}$ is a monomorphism iff f is an injective homomorphism.

As a preparation for the characterization of epimorphisms we state the following "*principle of unique homomorphic extension*":

Proposition 1.9 *Let $\underline{A}, \underline{B}$ be partial algebras of type (Ω, τ) and let $f, g : \underline{A} \to \underline{B}$ be any homomorphisms.*

(i) $\{a \in A | f(a) = g(a)\}$ *is always a closed subset of* $\underline{A}$.

(ii) *If $M \subseteq A$ is a generating subset of $\underline{A}$ such that f and g have the same restriction to M (i.e. $f|_M = g|_M$), then $f = g$.*

Proposition 1.10 (Characterization of epimorphisms)
In $\underline{Alg(\tau)}$ a homomorphism $f : \underline{A} \to \underline{B}$ is an epimorphism iff f is dense, *i.e. iff the image of f generates $\underline{B}$: $C_{\underline{B}} f(A) = B$.*

Proof

For the nontrivial part of the proof assume that f is not dense: Let $E := B \setminus C_{\underline{B}} f(A)$, and let E' be a set disjoint from B such that there exists a bijective mapping $d : E \to E'$. Define $D := B \cup E'$, and $v : B \to D$ by

$$v(b) := \begin{cases} b, & \text{if } b \in C_{\underline{B}} f(A) \\ d(b), & \text{if } b \in E. \end{cases}$$

Then v is an injective mapping onto $E' \cup C_{\underline{B}} f(A)$.

Finally we provide D with an algebraic structure as follows: For $\varphi \in \Omega$ set

$$dom\ \varphi^{\underline{D}} := dom\ \varphi^{\underline{B}} \cup \{(v(b_1), \ldots, v(b_{\tau(\varphi)})) \mid (b_1, \ldots, b_{\tau(\varphi)}) \in dom\ \varphi^{\underline{B}}\}.$$

And if $(d_1, \ldots, d_{\tau(\varphi)}) \in dom\ \varphi^{\underline{D}}$, then set:

$$\varphi^{\underline{D}}(d_1, \ldots, d_{\tau(\varphi)}) := \begin{cases} \varphi^{\underline{B}}(d_1, \ldots, d_{\tau(\varphi)}), & \text{if } d_1, \ldots, d_{\tau(\varphi)} \in B \\ v(\varphi^{\underline{B}}(v^{-1}(d_1), \ldots, v^{-1}(d_{\tau(\varphi)}))), & \text{otherwise.} \end{cases}$$

Finally, set $u = id_{BD} : \underline{B} \to \underline{D}$, $u(b) := b$ and consider v as a homomorphism from $\underline{B}$ into $\underline{D}$. Then, obviously $u \neq v$ ($u(e) \neq v(e) = d(e)$ for $e \in E$), while $u \circ f = v \circ f$, showing that f is not an epimorphism and that density is necessary for epimorphisms. □

Notation In what follows we will denote by

Hom the class of all *homomorphisms* in the category $\underline{Alg(\tau)}$,

Epi the class of all *epimorphisms* in the category $\underline{Alg(\tau)}$,

Mono the class of all *monomorphisms* in the category $\underline{Alg(\tau)}$,

Iso the class of all *isomorphisms* in the category $\underline{Alg(\tau)}$.

Recall that in category theory a subobject of an object A is a pair (B, m) consisting of an object B and a monomorphism m, and two subobjects (B, m) and (B', m') of A are equivalent, iff there exists an isomorphism $i : B \to B'$ such that $m = m' \circ i$. It is easy to see that each subobject of a partial algebra $\underline{A}$ in $\underline{Alg(\tau)}$ is in this way equivalent to (B, id_{BA}), where $\underline{B}$ is a weak relative subalgebra of $\underline{A}$. However, we have that

- $\underline{B}$ is a weak relative subalgebra of $\underline{A}$ iff id_{BA} is a monomorphism;
- $\underline{B}$ is a relative subalgebra of $\underline{A}$ iff id_{BA} is a full monomorphism;
- $\underline{B}$ is a subalgebra of $\underline{A}$ iff id_{BA} is a closed monomorphism.

1.4 Congruence relations, factor algebras, diagram completion I

The formation of subobjects is one way to get new algebraic structures from a given one. The dual way is to form factor algebras – as representatives of homomorphic images. The key for an "internal description" of homomorphic images is the concept of congruence relation.

Definition Let $\underline{A}$ be a partial algebra. A binary relation $\theta \subseteq A \times A$ is called a *congruence relation* on $\underline{A}$ iff

– θ is an equivalence relation on A, and

– θ is compatible with the fundamental operations of $\underline{A}$, i.e. for all $\varphi \in \Omega$ and for all $(a_i, b_i) \in A \times A$, $1 \leq i \leq \tau(\varphi)$ one has:
if $(a_i, b_i) \in \theta$ for all $i \in \{1, \ldots, \tau(\varphi)\}$, and if both $(a_1, \ldots, a_{\tau(\varphi)}), (b_1, \ldots, b_{\tau(\varphi)}) \in dom\ \varphi^{\underline{A}}$ then $(\varphi^{\underline{A}}(a_1, \ldots, a_{\tau(\varphi)}), \varphi^{\underline{A}}(b_1, \ldots, b_{\tau(\varphi)})) \in \theta$.
Written in matrix notation (or as a rule)

$$\begin{array}{rcl} (a_1, b_1) & \in & \theta \\ & \vdots & \\ (a_{\tau(\varphi)}, b_{\tau(\varphi)}) & \in & \theta \\ \hline (\varphi^{\underline{A}}(a_1, \ldots, a_{\tau(\varphi)}), \varphi^{\underline{A}}(b_1, \ldots, b_{\tau(\varphi)})) & \in & \theta \end{array}$$

Examples

– Observe that always $\Delta_A := \{(a, a) | a \in A\}$ and $\nabla_A := A \times A$ are congruence relations on $\underline{A}$.

– Let $\underline{G} = (G; (\circ^{\underline{G}}, e^{\underline{G}}, {}^{-1^{\underline{G}}}))$ be a group, then $\theta \subseteq G \times G$ is a congruence relation on $\underline{G}$ iff $N_\theta := \{g \in G | (g, e_{\underline{G}}) \in \theta\}$ is a normal subgroup of $\underline{G}$ and one has $(g, h) \in \theta$ iff $g \circ^{\underline{G}} h^{-1^{\underline{G}}} \in N_\theta$.

– For a ring $\underline{R} = (R, (+^{\underline{R}}, 0^{\underline{R}}, -^{\underline{R}}, \cdot^{\underline{R}}))$ a relation $\theta \subseteq R \times R$ is a congruence relation on $\underline{R}$ iff $\mathcal{J}_\theta := \{r \in R | (r, 0^{\underline{R}}) \in \theta\}$ is an ideal of $\underline{R}$ and one has $(r, s) \in \theta$ iff $r -^{\underline{R}} s \in \mathcal{J}_\theta$.

Lemma and Definition 1.11 *Let $\underline{A}, \underline{B}$ be any partial algebras of the same similarity type (Ω, τ). Then*

(i) Δ_A *and* ∇_A *are congruence relations of* $\underline{A}$.

(ii) *If $\mathcal{G}$ is a set of congruence relations on $\underline{A}$, then its intersection $\bigcap \mathcal{G}$ is also a congruence relation.*

(iii) *If $f : \underline{A} \to \underline{B}$ is any homomorphism, then*

$$ker\, f := \{(a, a') \in A \times A | f(a) = f(a')\}$$

is a congruence relation on $\underline{A}$, called the kernel *of f.*

The statement of (ii) means that the set of congruence relations of a partial algebra $\underline{A}$ is a closure system on $A \times A$. The corresponding closure operator will be denoted by $Con_{\underline{A}}$, i.e. for any $\theta \subseteq A \times A$ we have

$$Con_{\underline{A}}\theta = \bigcap\{\Psi | \Psi \text{ is a congruence relation on } \underline{A} \text{ and } \theta \subseteq \Psi\}.$$

The role of kernels of homomorphisms is described by the following definition and results:

Definition Let $\underline{A}$ be any partial algebra, and let θ be any congruence relation on $\underline{A}$. Let $A/\theta := \{[a]_\theta | a \in A\}$ be the set of all equivalence classes $[a]_\theta := \{b \in A | (a, b) \in \theta\}$ of elements of A, the so-called *quotient set* or *factor set* of A with respect to θ. The mapping $nat_\theta : A \to A/\theta$, $a \mapsto [a]_\theta$ is called the *natural mapping* with respect to θ. The fact that θ is not only an equivalence but also a congruence relation allows to define on A/θ a partial algebraic structure $(\varphi^{\underline{A}/\theta})_{\varphi\in\Omega}$, making nat_θ a full and surjective homomorphism:

For $\varphi \in \Omega$ set $dom\ \varphi^{\underline{A}/\theta} := \{([a_1]_\theta, \ldots, [a_{\tau(\varphi)}]_\theta) \in (A/\theta)^{\tau(\varphi)} |$ there is $(a'_1, \ldots, a'_{\tau(\varphi)}) \in dom\ \varphi^{\underline{A}}$ such that $(a_i, a'_i) \in \theta$ for $1 \leq i \leq \tau(\varphi)\}$, and if $(a_1, \ldots, a_{\tau(\varphi)}) \in dom\ \varphi^{\underline{A}}$, then set $\varphi^{\underline{A}/\theta}([a_1]_\theta, \ldots, [a_{\tau(\varphi)}]_\theta) := [\varphi^{\underline{A}}(a_1, \ldots, a_{\tau(\varphi)})]_\theta$.

Lemma and Notation 1.12 *With the notation and definitions from above we have:*

(i) *The value of $\varphi^{\underline{A}/\theta}([a_1]_\theta, \ldots, [a_{\tau(\varphi)}]_\theta)$ is independent from the choice of the representing sequence, i.e. $\varphi^{\underline{A}/\theta}$ is indeed a $\tau(\varphi)$-ary partial operation on A/θ.*

(ii) *$nat_\theta : A \to A/\theta$ is a full and surjective homomorphism $nat_\theta : \underline{A} \to \underline{A}/\theta$.*

$\underline{A}/\theta$ is called the factor algebra *or* quotient algebra *of the partial algebra $\underline{A}$ with respect to the congruence relation θ, and nat_θ is called the* natural homomorphism *or* natural projection *of $\underline{A}$ onto $\underline{A}/\theta$.*

More about the importance of the factor algebras of $\underline{A}$ as distinguished representatives of full homomorphic images of $\underline{A}$ (i.e. of full and surjective homomorphisms starting from $\underline{A}$) can be derived from the following *Diagram Completion Lemma for full and surjective homomorphisms*:

Lemma 1.13 (First Diagram Completion Lemma)

(i) *Let $f : A \to B$ be a surjective and $g : A \to C$ an arbitrary mapping. Then there exists a mapping $h : B \to C$ such that $g = h \circ f$, iff $ker\, f \subseteq ker\, g$ (see Figure 3).*

(ii) *If, in addition, f is a full (and surjective!) homomorphism, $f : \underline{A} \to_{full} \underline{B}$, and if $g : \underline{A} \to \underline{C}$ is a homomorphism, then h is a homomorphism, $h : \underline{A} \to \underline{C}$, whenever it exists. Let h exist (i.e. let $ker\, f \subseteq ker\, g$ be true). Then:*

$$\begin{tikzcd} A \arrow[r, two heads, "f"] \arrow[d, "g"'] & B \arrow[dl, "(\exists h) h \circ f = g \Leftrightarrow ker\, f \subseteq ker\, g"] \\ C & \end{tikzcd}$$

Figure 3: First Diagram Completion Lemma

(iii) *h is uniquely determined by f and g.*

(iv) *h is injective, iff $ker\, f = ker\, g$.*

(v) *h is surjective, iff g is surjective.*

Let $f : \underline{A} \to \underline{B}$ be a full and surjective homomorphism, let $g : \underline{A} \to \underline{C}$ be a homomorphism, and let h exist. Then:

(vi) *h is full, iff g is full.*

(vii) *If g is closed, then h is closed. If f is closed, then h is closed iff g is closed.*

(viii) *h is an isomorphism iff $ker\, f = ker\, g$ and g is a full and surjective homomorphism.*

Lemma 1.13 (viii) now shows that the natural homomorphism $nat_\theta : \underline{A} \to \underline{A}/\theta$ with respect to a congruence relation θ on $\underline{A}$ is determined by θ up to unique isomorphism as a full and surjective homomorphism $f : \underline{A} \to \underline{B}$ with $ker\, f = \theta$ (this fact is often called the *Homomorphism Theorem*). Of great importance is also the following consequence of Lemma 1.13:

Lemma 1.14 (Factorization Lemma for full and surjective homomorphisms and monomorphisms)
Let $f : \underline{A} \to \underline{B}$ be any homomorphism. Then there are up to unique isomorphism a full and surjective homomorphism $q : \underline{A} \to \underline{D}$ and a monomorphism $m : \underline{D} \to \underline{B}$ such that $f = m \circ q$ (cf. Figure 4).

Definition Let $\underline{A}$ and $\underline{B}$ be partial algebras of the same type (Ω, τ), Then $\underline{B}$ is called a

– (*weak*) *homomorphic image* of $\underline{A}$, if there exists a surjective homomorphism $f : \underline{A} \to \underline{B}$ from $\underline{A}$ onto $\underline{B}$;

– *full* (or *strong*) *homomorphic image* of $\underline{A}$, if there exists a full and surjective homomorphism $f : \underline{A} \to \underline{B}$ from $\underline{A}$ onto $\underline{B}$;

– *closed homomorphic image* of $\underline{A}$, if there exists a closed and surjective homomorphism $f : \underline{A} \to \underline{B}$ from $\underline{A}$ onto $\underline{B}$.

In the case of total algebras the three kinds of homomorphic images defined above coincide, since then each homomorphism is closed and therefore full.

$$\begin{tikzcd}
\underline{A} \arrow[rr, "f"] \arrow[dr, "q"', "\text{full, surj.}"] \arrow[ddr, "q'"', "\text{full, surj.}"] & & \underline{B} \\
& \underline{A}/\mathit{ker}\, f \arrow[ur, "m", "\text{mono}"'] \arrow[d, "i", "\cong"'] & \\
& D' \arrow[uur, "m'"', "\text{mono}"] &
\end{tikzcd}$$

Figure 4: Factorization Lemma

The First Diagram Completion Lemma now allows us to characterize full homomorphic images. However, this also includes — in connection with Proposition 1.15 below — a characterization of closed homomorphic images, once we have defined closed congruence relations:

Definition Let $\underline{A}$ be a partial algebra. A congruence relation θ of $\underline{A}$ is called *closed*, iff it has the following property:

(ccr) for each $\varphi \in \Omega$ and for any $a_1, \ldots, a_{\tau(\varphi)}, b_1, \ldots, b_{\tau(\varphi)} \in A$:
if $(a_i, b_i) \in \theta$ for $1 \leq i \leq \tau(\varphi)$, and if $(a_1, \ldots, a_{\tau(\varphi)}) \in dom\ \varphi^{\underline{A}}$,
then $(b_1, \ldots, b_{\tau(\varphi)}) \in dom\ \varphi^{\underline{A}}$.

(and clearly $(\varphi^{\underline{A}}(a_1, \ldots, a_{\tau(\varphi)}), \varphi^{\underline{A}}(b_1, \ldots, b_{\tau(\varphi)})) \in \theta$, since θ is assumed to be a congruence relation on $\underline{A}$).

This denotation is justified by the following

Proposition 1.15 *Let $\underline{A}$ be any partial algebra.*

(i) *For every closed homomorphism $f : \underline{A} \to \underline{B}$ its kernel $ker\, f$ is a closed congruence relation on $\underline{A}$.*

(ii) *A congruence relation θ on $\underline{A}$ is closed, iff $nat_\theta : \underline{A} \to \underline{A}/\theta$ is a closed homomorphism.*

The importance of closed homomorphisms and therefore of closed congruence relations will become evident later in connection with model theoretic concepts and among others with the description of special implicational classes, e.g. of what we shall call ECE-varieties (i.e. varieties defined by existentially conditional existence equations (see section 2 and subsection 3.3 as well as the end of subsection 3.5)).

Observe that Δ_A is always a closed congruence relation on $\underline{A}$, while $\nabla_A = A \times A$ *is closed, iff for each* $\varphi \in \Omega$ $\varphi^{\underline{A}}$ *is either discrete (empty) or total.*

Recall now that for any binary relations R and R' on a set A one defines

$$R \circ R' := \{(a,c) \in A \times A \mid \text{ there exists } b \in A \text{ such that } (a,b) \in R' \text{ and } (b,c) \in R\},$$

$$R^{-1} := \{(b,a) | (a,b) \in R\},$$

and

$$R^0 := id_A, \quad R^{n+1} := R \circ R^n.$$

Moreover, for $R \subseteq A \times A$, $\bigcup_{n=0}^{\infty} R^n$ *yields the reflexive and transitive closure of* R. Thus $\bigcup_{n=0}^{\infty}(R \cup R^{-1})^n$ describes *the smallest equivalence relation containing* R (exercise).

It is well known (see also Proposition 1.16.(ii) below) that for total algebras $\underline{A}$ the supremum of a set of congruence relations R_i $(i \in I)$ is just the smallest equivalence relation containing $\bigcup_{i \in I} R_i$. The same is true for closed congruence relations, while it is not true in general for arbitrary congruence relations of a partial algebra:

Proposition 1.16 *Let* $\underline{A}$ *be any partial algebra of similarity type* (Ω, τ) *and let* θ_1, θ_2 *be any congruence relations of* $\underline{A}$. *Then*

(i) *If* θ_2 *is closed, and if* $\theta_1 \subseteq \theta_2$, *then* θ_1 *is closed.*

(ii) *If* θ_1 *and* θ_2 *are closed, then their supremum within the closure system of all congruence relations of* $\underline{A}$, $Con_{\underline{A}}(\theta_1 \cup \theta_2)$, *is closed, too, moreover* $Con_{\underline{A}}(\theta_1 \cup \theta_2) = \bigcup_{n=0}^{\infty}(\theta_2 \circ \theta_1)^n$.

(iii) *The set* $Cong_c(\underline{A})$ *of all closed congruence relations of* $\underline{A}$ *is inductively ordered and therefore contains a maximal element by Zorn's Lemma.*

(iv) $Cong_c(\underline{A})$ *contains a largest closed congruence relation, which will be denoted by* $\chi_{\underline{A}}$.

In order to show that 1.16.(ii) does not hold for arbitrary congruence relations consider the following congruence relations θ_1 and θ_2 on the partial algebra $\underline{A}$ as depicted in Figure 5: $A := \{a,b,c,d,e\}$, $\Omega := \{\varphi\}$, $\tau(\varphi) := 1$, *graph* $\varphi^{\underline{A}} := \{(a,d),(c,e)\}$.

$\theta_1 := \Delta_A \cup \{a,b\}^2$, $\theta_2 := \Delta_A \cup \{b,c\}^2$, $\bigcup_{n=0}^{\infty}(\theta_1 \circ \theta_2)^n = \Delta_A \cup \{a,b,c\}^2$, $Con_{\underline{A}}(\theta_1 \cup \theta_2) = \{a,b,c\}^2 \cup \{d,e\}^2 \neq \bigcup_{n=0}^{\infty}(\theta_1 \circ \theta_2)^n$.

1.5 Some constructions of partial algebras

There are several ways to construct new partial algebras from given ones. Subalgebras, relative subalgebras, weak relative subalgebras and (full) homomorphic images need only one given algebra. The constructions like products, coproducts, pullbacks, pushouts, direct limits and reduced products considered in this section usually start from several, in some cases infinitely many given partial algebras (and possibly needing some homomorphisms between them).

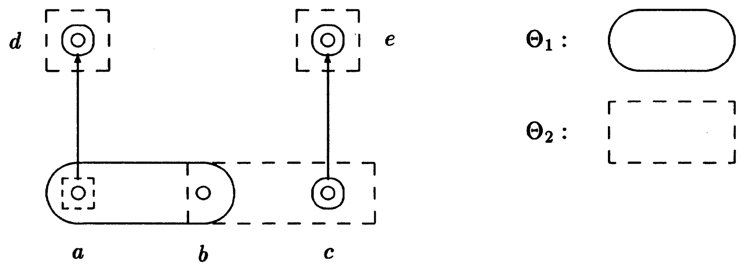


Figure 5: The equivalence generated by two congruences is in general no congruence

Products and reduced products

In category theory a *product* of a family $(C_i)_{i\in I}$ of objects is an object C together with a family $(p_i : C \to C_i)_{i\in I}$ of morphisms such that for each object B and each family $(f_i : B \to C_i)_{i\in I}$ of morphisms there is exactly one morphism $f : B \to C$ such that $p_i \circ f = f_i$ for every $i \in I$. In $\underline{Alg(\tau)}$ we define

Definition Let $(\underline{A_i})_{i\in I}$ be any family of partial algebras of the same similarity type (Ω, τ), where $\underline{A_i} = (A_i, (\varphi^{\underline{A_i}})_{i\in I})$ for each $i \in I$. Define

$$\bigtimes_{i\in I} A_i := \{a : I \to \bigcup_{i\in I} A_i \mid a(i) \in A_i \text{ for each } i \in I\}$$

to be the set of all so-called choice functions $a = (a_i)_{i\in I}$ of the family $(A_i)_{i\in I}$. And on $A := \bigtimes_{i\in I} A_i$ we define a partial algebraic structure $(\varphi^{\underline{A}})_{i\in I}$ as follows:

For $\varphi \in \Omega$ set

$$dom\ \varphi^{\underline{A}} := \{(a_1, \dots, a_{\tau(\varphi)}) \in A^{\tau(\varphi)} | (a_1(i), \dots, a_{\tau(\varphi)}(i)) \in dom\ \varphi^{\underline{A_i}} \text{ for each } i \in I\}$$

and if $(a_1, \dots, a_{\tau(\varphi)}) \in dom\ \varphi^{\underline{A}}$, then set

$$\varphi^{\underline{A}}(a_1, \dots, a_{\tau(\varphi)}) := (\varphi^{\underline{A_i}}(a_1(i), \dots, a_{\tau(\varphi)}(i)) \mid i \in I),$$

i.e. $(\varphi^{\underline{A}}(a_1, \dots, a_{\tau(\varphi)}))(i) := \varphi^{\underline{A_i}}(a_1(i), \dots, a_{\tau(\varphi)}(i))$ for each $i \in I$.

We define $\prod_{i\in I} \underline{A_i} := (\bigtimes_{i\in I} A_i, (\varphi^{\underline{A}})_{\varphi\in\Omega})$ to be the *direct product* of the family $(\underline{A_i})_{i\in I}$. For each $j \in I$ we denote by $pr_j := \bigtimes_{i\in I} A_i \to A_j$, $a \mapsto a(j)$ $(a \in \bigtimes_{i\in I} A_i)$ the j-the canonical projection.

Proposition 1.17 *With the assumptions and the notation from the above definition one has:*

(i) *For each* $j \in I$, pr_j *is a homomorphism* $pr_j : \prod_{i \in I} \underline{A_i} \to \underline{A_j}$.

(ii) $(\prod_{i \in I} \underline{A_i}, (pr_i)_{i \in I})$ *is a product of the family* $(\underline{A_i})_{i \in I}$ *in the category* $\underline{Alg(\tau)}$.

Proposition 1.18 *Let* I *be a non-empty set, let* $(f_i : \underline{B} \to \underline{A_i} | i \in I)$ *be a family of homomorphisms, and let* $f : \underline{B} \to \prod_{i \in I} \underline{A_i}$ *be the induced product homomorphism. Then*

$$ker f = \bigcap_{i \in I} ker f_i.$$

Observe that for each homomorphism $f : \underline{A} \to \underline{B}$, *graph* f is a closed subset of $\underline{A}\pi\underline{B}$ (in this way we denote the product object of the *two* factors $\underline{A}$ and $\underline{B}$), and $ker f$ is a closed subset of $\underline{A}\pi\underline{A}$, which is in addition an equivalence relation on A. However, a mapping $f : A \to B$, the graph of which is a closed subset of $\underline{A}\pi\underline{B}$, need not be a homomorphism; give an example of this fact (however, $\underline{B}$ has then to be really partial !).

We add some observations, which will be useful later in connection with statements about the extension of homomorphisms:

Lemma 1.19 *Let* $f : \underline{A} \to \underline{B}$ *be a homomorphism,* $g : dom\ g \to B$ *be any mapping out of* A *into* B *such that graph* $g \subseteq$ *graph* f, *and let* $\underline{dom}\ g$ *be the relative subalgebra of* $\underline{A}$ *on the domain of* g. *Then*

(i) $g : \underline{dom}\ g \to \underline{B}$ *is a homomorphism.*

(ii) *graph* g *is a closed subset of* $\underline{A}\pi\underline{B}$, *iff dom* g *is a closed subset of* $\underline{A}$.

Corollary 1.1 *For a homomorphism* $f : \underline{A} \to \underline{B}$ *and a subset* M *of* A *the following statements are equivalent:*

(i) M *generates* A, *i.e.* $C_{\underline{A}}M = A$.

(ii) *graph* $(f|_M)$ *generates graph* f *in* $\underline{A}\pi\underline{B}$:

$$C_{\underline{A}\pi\underline{B}}\ graph\ (f|_M) = graph\ f.$$

In connection with the description of classes of partial algebras which are describable by universal Horn formulas we shall need the concept of a reduced product:

Definition (Filters) Let I be a set and $\mathcal{F}$ a set of subsets of I. $\mathcal{F}$ is called a *filter* on I, if

(F1) $I \in \mathcal{F}$

(F2) $F_1, F_2 \in \mathcal{F}$ implies $F_1 \cap F_2 \in \mathcal{F}$

(F3) $F \in \mathcal{F}$ and $F \subseteq F'$ imply $F' \in \mathcal{F}$.

If one has in addition

(F4) $\emptyset \notin \mathcal{F}$,

then $\mathcal{F}$ is called a *proper filter* on I, and if $\mathcal{F}$ also satisfies

(F5) for all $F \subseteq I$ either $F \in \mathcal{F}$ or $I \setminus F \in \mathcal{F}$,

then $\mathcal{F}$ is called an *ultrafilter* on I. If $\mathcal{F} = \{N \subseteq I | M \subseteq N\}$ for some fixed subset M of I, then $\mathcal{F}$ is called a *principal filter* generated by M: $\mathcal{F} =: \uparrow M$.

Further important examples of filters are the so-called *generalized Fréchet filters* on I: Let c be any infinite cardinal number and let

$$\mathcal{F}_c := \{F \subseteq I | \#(I \setminus F) < c\}.$$

If $c \leq \#I$ ($\#M$ designating the cardinal number of the set M), then $\mathcal{F}_c$ is a non-principal proper filter on I.

Lemma 1.20 and Definition (Reduced product) *Let I be any set, let $(\underline{A}_i)_{i \in I}$ be a family of partial algebras of the same similarity type (Ω, τ), let $\mathcal{F}$ be some filter on I. Moreover, let $A := \bigtimes_{i \in I} A_i$, $\underline{A} := \prod_{i \in I} \underline{A}_i$, and for $a, b \in A$ define $I_{a,b} := \{i \in I | a(i) = b(i)\}$. Then we consider on A the following binary relation $\Theta_{\mathcal{F}} := \{(a,b) | a, b \in A$ and $I_{a,b} \in \mathcal{F}\}$ which is — as can be easily verified — an equivalence relation on A. Let $A_{\mathcal{F}} := A/\Theta_{\mathcal{F}}$ be the corresponding quotient set, and for $a \in A$ denote by $a_{\mathcal{F}} := [a]_{\Theta_{\mathcal{F}}}$ its equivalence class. For $\varphi \in \Omega$ define*

$$\begin{aligned} dom\, \varphi^{\underline{A}_{\mathcal{F}}} := \ & \{(a_{1\mathcal{F}}, \ldots, a_{\tau(\varphi)\mathcal{F}}) \in A_{\mathcal{F}}^{\tau(\varphi)} \mid \{i \in I | (a'_1(i), \ldots, a'_{\tau(\varphi)}(i)) \in dom\, \varphi^{\underline{A}_i}\} \in \mathcal{F} \\ & \text{for some sequence } (a'_1, \ldots, a'_{\tau(\varphi)}) \in A^{\tau(\varphi)} \text{ such that for } 1 \leq j \leq \tau(\varphi) \\ & \text{one has } a'_{j\mathcal{F}} = a_{j\mathcal{F}}\}. \end{aligned}$$

And if

$$\underline{a}_{\mathcal{F}} := (a_{1\mathcal{F}}, \ldots, a_{\tau(\varphi)\mathcal{F}}) \in dom\, \varphi^{\underline{A}_{\mathcal{F}}},$$

and if $(a'_1, \ldots, a'_{\tau(\varphi)}) \in A^{\tau(\varphi)}$ is a representative sequence causing $\underline{a}_{\mathcal{F}}$ to belong to dom $\varphi^{\underline{A}_{\mathcal{F}}}$, then define

$$\varphi^{\underline{A}_{\mathcal{F}}}(a_{1\mathcal{F}}, \ldots, a_{\tau(\varphi)\mathcal{F}}) =: a_{\mathcal{F}}$$

for some sequence $a \in A$ such that for $i \in I$ $a(i) := \varphi^{\underline{A}_i}(a'_1(i), \ldots, a'_{\tau(\varphi)}(i))$, if this exists, and $a(i)$ some fixed value of A_i, else. Then $(A_{\mathcal{F}}, (\varphi^{\underline{A}_{\mathcal{F}}})_{\varphi \in \Omega}) =: (\prod_{i \in I} \underline{A}_i)/\mathcal{F}$ is called the reduced product *of the family $(\underline{A}_i)_{i \in I}$ of partial algebras with respect to the filter $\mathcal{F}$.*

Observe, that $nat_{\mathcal{F}} := nat_{\Theta_{\mathcal{F}}} : \prod_{i \in I} \underline{A}_i \to (\prod_{i \in I} \underline{A}_i)/\mathcal{F}$ is a homomorphism, which is not necessarily full. E.g. if $\underline{A}_i$ is a total algebra for each $i \in I \setminus \{i_0\}$, and if $\underline{A}_{i_0}$ is a discrete partial algebra — all carrier sets are assumed to be non-empty —, and if there is some $F \in \mathcal{F}$ such that $i_0 \notin F$, then $(\prod_{i \in I} \underline{A}_i)/\mathcal{F}$ is a total algebra, while $\prod_{i \in I} \underline{A}_i$ is discrete.

Coproducts and direct limits

In category theory a coproduct is the dual concept of a product (see Figure 6), thus, for a family $(\underline{A}_j)_{j \in J}$ of similar partial algebras a *coproduct* is a partial algebra $\underline{A} =: \coprod_{j \in J} A_j$

— note that as in the case of products it is unique up to isomorphism — and a family $(i_j : \underline{A}_j \to \underline{A})_{j\in J}$ of homomorphisms, called *injections* (although they need not to be injective), such that for every family $(f_j : \underline{A}_j \to \underline{B})_{j\in J}$ of homomorphisms there is a unique homomorphism $f : \underline{A} \to \underline{B}$ such that $f \circ i_j = f_j$ for all $j \in J$.

1. In the case that the type τ does not specify any nullary constants the coproduct object $\underline{A} := \coprod_{j\in J} \underline{A}_j$ is easily described as follows:

Let A be the disjoint union of the sets A_j, e.g. $A := \bigcup_{j\in J} A_j \times \{j\}$. For $\varphi \in \Omega$ one has

$$dom\ \varphi^{\underline{A}} := \{(a_1, j), \ldots, (a_{\tau(\varphi)}, j)) | j \in J, (a_1, \ldots, a_{\tau(\varphi)}) \in dom\ \varphi^{\underline{A}_j}\},$$

and if $((a_1, j), \ldots, (a_{\tau(\varphi)}, j)) \in dom\ \varphi^{\underline{A}}$ for some $j \in J$, then

$$\varphi^{\underline{A}}((a_1, j), \ldots, (a_{\tau(\varphi)}, j)) := (\varphi^{\underline{A}_j}(a_1, \ldots, a_{\tau(\varphi)}), j),$$

i.e. $\varphi^{\underline{A}}$ can be considered as the disjoint union of the $\varphi^{\underline{A}_j}$ $(j \in J)$.

2. Now assume that τ specifies nullary constants, let $\Omega_0 := \{\varphi \in \Omega | \tau(\varphi) = 0\}$, and $\Omega' := \Omega \setminus \Omega_0$, $\tau' := \tau|_{\Omega'}$ the restriction of τ to Ω'. For any partial algebra $\underline{B} := (B, (\varphi^{\underline{B}})_{\varphi\in\Omega})$ we denote by $\underline{B}' := (B, (\varphi^{\underline{B}})_{\varphi\in\Omega'})$ what is called in general the *τ'–reduct* (or *(Ω', τ')–reduct*) of $\underline{B}$. Thus, for a given family $(\underline{A}_j)_{j\in J}$ of partial algebras of type (Ω, τ), let $(A^*, (\varphi^{\underline{A}^*})_{\varphi\in\Omega'}) = \underline{A}^{*\prime} := \coprod_{j\in J} \underline{A}'_j$ together with $(i'_j : \underline{A}'_j \to \underline{A}^{*\prime})$ be the construction from above for the (Ω', τ') reducts. Moreover let $R' \subseteq A^* \times A^*$ be the following relation:

$$R' := \{((\varphi^{\underline{A}_j}, j), (\varphi^{\underline{A}_k}, k)) \mid j, k \in J, \varphi \in \Omega_0, \varphi^{\underline{A}_j} \text{ and } \varphi^{\underline{A}_k} \text{ exist } \},$$

let $\Theta := Con_{\underline{A}^{*\prime}} R'$ be the congruence relation on $\underline{A}^{*\prime}$ generated by R', and define finally $\underline{A} := \coprod_{j\in J} \underline{A}_j := (A^*/\Theta, (\varphi^{\underline{A}^{*\prime}/\Theta})_{\varphi\in\Omega'} \cup (\varphi^{\underline{A}})_{\varphi\in\Omega_0})$, where for $\varphi \in \Omega_0$

$$\varphi^{\underline{A}} := \begin{cases} [(\varphi^{\underline{A}_j}, j)]_\Theta & \text{if } \varphi^{\underline{A}_j} \text{ exists for some } j \in J \\ \text{undefined,} & \text{otherwise;} \end{cases}$$

and $(i_j := nat_\Theta \circ i'_j : \underline{A}_j \to \underline{A})_{j\in J}$ is the family of corresponding canonical injections.

Proposition 1.21 *For any family $(\underline{A}_j)_{j\in J}$ the partial algebra $\underline{A} = \coprod_{j\in J} \underline{A}_j$ and the family $(i_j : \underline{A}_j \to \underline{A})_{j\in J}$ form a coproduct of the family $(\underline{A}_j)_{j\in J}$.*

While coproducts are one interesting instance of colimits, direct limits (more precisely: directed colimits) are another one.

Definition

(1) A *directed system* of partial algebras is any family $\mathcal{A}_{\underline{J}} := (\underline{A}_j, f_{jl} : \underline{A}_j \to \underline{A}_l | j, l \in J, j \leq l)$, where $\underline{J} := (J, \leq)$ is a *directed ordered set*, i.e. $\leq$ is a binary relation on J which is

(i) *reflexive*: $(\forall x) x \leq x$,

(ii) *antisymmetric*: $(\forall x)(\forall y)(x \leq y \wedge y \leq x \Rightarrow x = y)$,

$$\begin{tikzcd}
\underline{A}_i \arrow[r, "f_i"] \arrow[dr, "g_i"'] & \underline{A} \arrow[d, "\exists! g"] & \underline{A}_j \arrow[l, "f_j"'] \arrow[dl, "g_j"] \\
& \underline{B} &
\end{tikzcd}$$

Figure 6: Characteristic diagram for coproducts

(iii) *transitive*: $(\forall x)(\forall y)(\forall z)(x \leq y \wedge y \leq z \Rightarrow x \leq z)$,

(iv) *directed*: for any finite subset $\mathcal{F} \subseteq J$ there exists in J an upper bound, i.e. some $j \in J$ such that $i \leq j$ for all $i \in \mathcal{F}$;

moreover, for $j \leq l$ in $\underline{J}$, $f_{jl} : \underline{A}_j \to \underline{A}_l$ is a homomorphism such that

(a) $f_{jj} = id_{A_j}$ for all $j \in J$,

(b) for $i \leq j$ and $j \leq l$ in $\underline{J}$ one has $f_{jl} \circ f_{ij} = f_{il}$.

(2) Let $\mathcal{A}_{\underline{J}} := (\underline{A}_j, f_{jl} : \underline{A}_j \to \underline{A}_l | j \leq l \text{ in } \underline{J})$ be a directed system. A *direct limit* of $\mathcal{A}_{\underline{J}}$ is a pair $((f_j : \underline{A}_j \to \underline{A})_{j \in J}, \underline{A})$ consisting of the colimit (!) object $\underline{A}$ and the so-called *colimiting cocone* of homomorphisms f_j $(j \in J)$, such that

(C$_0$) for all $i \leq j$ in $\underline{J}$ one has $f_j \circ f_{ij} = f_i$, and

(CL) for any *compatible family* $((g_j : \underline{A}_j \to \underline{B})_{j \in J}, \underline{B})$ consisting of a partial algebra $\underline{B}$ and a family $(g_j : \underline{A}_j \to \underline{B})_{j \in J}$ of homomorphisms satisfying the condition
(C) for all $i \leq j$ in $\underline{J}$ one has $g_j \circ f_{ij} = g_i$,
there exists a unique homomorphism $g : \underline{A} \to \underline{B}$
with $g \circ f_j = g_j$ for all $j \in J$.

We write $\lim_{\to} \mathcal{A}_{\underline{J}} := ((f_j : \underline{A}_j \to \underline{A})_{j \in J}, \underline{A})$ for the direct limit of a directed system.

Proposition 1.22 *For any similarity type* (Ω, τ) *and for any directed system* $\mathcal{A}_{\underline{J}} = (\underline{A}_j, f_{jk} : \underline{A}_j \to \underline{A}_k | j \leq k \text{ in } \underline{J})$ *there exists the direct limit* $\lim_{\to} \mathcal{A}_j = ((f_j : \underline{A}_j \to \underline{A})_{j \in J}, \underline{A})$, *which can be defined as follows (cf. Proposition* 1.21*):*

Set $\underline{A}^{*\prime} := \coprod_{j \in J} \underline{A}'_j$, *where again* $\underline{A}'_j$ *is the* τ'*-reduct of* $\underline{A}_j$ *for* $\Omega' = \Omega \setminus \Omega_0$ $(\Omega_0 := \{\varphi \in \Omega \mid \tau(\varphi) = 0\})$, *let* $\Theta := \{((a, j), (b, k)) \in (\bigcup_{j \in J} A_j \times \{j\})^2 \mid$ *there is* $m \in J : j, k \leq m$ *and* $f_{jm}(a) = f_{km}(b)\}$, *then* Θ *is a congruence relation on* $\underline{A}^{*\prime}$. *Set* $A := A^*/\Theta$, $\varphi^{\underline{A}} := \varphi^{\underline{A}^{*\prime}/\Theta}$ *for* $\varphi \in \Omega'$, *and for* $\varphi \in \Omega_0$*:*

$$\varphi^{\underline{A}} := \begin{cases} [(\varphi^{\underline{A}_k}, k)]_\Theta, & \textit{if } \varphi^{\underline{A}_k} \textit{ exists for some } k \in J, \\ \text{undefined}, & \textit{if for no } k \in J\ \varphi^{\underline{A}_k} \textit{ exists.} \end{cases}$$

Finally set $f_j := nat_\Theta \circ i_j : \underline{A}_j \to \underline{A}$, *for each* $j \in J$, *where* $i_j(a) := (a, j)$ *for* $a \in A_j$.

Proposition 1.23 *Let* $\mathcal{A}_{\underline{J}} := (\underline{A}_j, f_{jk} : \underline{A}_j \to \underline{A}_k | j \leq k \text{ in } \underline{J})$ *be a directed system with* $\lim_{\to} \mathcal{A}_j := ((f_j : \underline{A}_j \to \underline{A})_{j \in J}, \underline{A})$, *let* $((g_j : \underline{A}_j \to \underline{B})_{j \in J}, \underline{B})$ *be a family compatible with* $\mathcal{A}_{\underline{J}}$ *and let* $g : \underline{A} \to \underline{B}$ *be the induced colimit homomorphism. Then one has:*

(i) *If all* f_{jk} $(j \leq k \text{ in } \underline{J})$ *are isomorphisms, injective, closed, surjective, epimorphisms or full, respectively, then all the* f_j *have the same property* $(j \in J)$.

(ii) *If all* g_j $(j \in J)$ *are isomorphisms, injective, closed, surjective, epimorphisms or full, respectively, then* g *has the same property.*

The following statement is of general interest:

Lemma 1.24 *Let* $\mathcal{A}_{\underline{J}} = (\underline{A}_i, f_{ij} : \underline{A}_i \to \underline{A}_j | i \leq j \text{ in } \underline{J})$ *be a directed system with direct limits* $\underline{f} := ((f_j : \underline{A}_j \to \underline{A})_{j \in J}, \underline{A})$ *and* $\underline{f}' := ((f'_j : \underline{A}_j \to \underline{A}')_{j \in J}, \underline{A}')$, *then there is a unique isomorphism* $g : \underline{A} \to \underline{A}'$ *such that* $g \circ f_j = f'_j$ *for each* $j \in J$.

With respect to the representation of a partial algebra as direct limit of "very small pieces" in Proposition 1.26 below we state the following:

Lemma 1.25 *Let* $\underline{A}$ *be any partial algebra and* M *a generating subset of* $\underline{A}$. *Then there exists for every* $a \in A$ *a* finite *weak relative subalgebra* $\underline{B}_a$ *of* $\underline{A}$ *such that*

(i) $a \in B_a = C_{\underline{B}_a}(B_a \cap M)$ *and*

(ii) *the structure of* $\underline{B}_a$ *is finite, i.e.* $\bigcup_{\varphi \in \Omega} \{\varphi\} \times graph\ \varphi^{\underline{B}_a}$ *is finite.*

Finite partial algebras $\underline{B}$ satisfying (ii) above will be called *totally finite.*

The proof of this lemma is a good application of the method of algebraic (=structural) induction, hence we give the details.

Proof (by structural induction): If $a \in M$, then set $\underline{B}_a := (\{a\}, (\emptyset)_{\varphi \in \Omega})$ which obviously satisfies (i) and (ii). If $a = \psi^{\underline{A}}$ is a constant, then set $\underline{B}_a := (\{a\}, (\varphi^{\underline{B}_a})_{\varphi \in \Omega})$ such that for $\varphi \in \Omega$

$$graph\ \varphi^{\underline{B}_a} = \begin{cases} ((), \psi^{\underline{A}}) & \text{if } \varphi = \psi \\ \emptyset & \text{if } \varphi \neq \psi. \end{cases}$$

Finally, assume that $a = \varphi^{\underline{A}}(a_1, \ldots, a_{\tau(\varphi)})$ for some $\varphi \in \Omega$ and $(a_1, \ldots, a_{\tau(\varphi)}) \in dom\ \varphi^{\underline{A}}$, and let the statement be true for each a_i, $1 \leq i \leq \tau(\varphi)$. Then define $B_a := \bigcup_{i=1}^{\tau(\varphi)} B_{a_i} \cup \{a\}$ and, for $\psi \in \Omega$,

$$graph\ \psi^{\underline{B}_a} := \begin{cases} \bigcup_{i=1}^{\tau(\varphi)} graph\ \varphi^{\underline{B}_{a_i}} \cup \{((a_1, \ldots, a_{\tau(\varphi)}), a)\}, & \text{if } \psi = \varphi, \\ \bigcup_{i=1}^{\tau(\varphi)} graph\ \varphi^{\underline{B}_{a_i}}, & \text{otherwise.} \end{cases}$$

Then it is obvious that $a \in B_a$, and that $\underline{B}_a$ satisfies the finiteness conditions in (ii). Moreover, for $M_a := B_a \cap M$ one has $a_i \in B_{a_i} = C_{\underline{B}_{a_i}}(B_{a_i} \cap M) \subseteq C_{\underline{B}_a} M_a$ and therefore also $a = \varphi^{\underline{A}}(a_1, \ldots, a_{\tau(\varphi)}) \in C_{\underline{B}_a} M_a$, i.e. $B_a = C_{B_a} M_a$. Thus $\underline{B}_a$ satisfies (i) and (ii). Since M generates $\underline{A}$ these arguments show the truth of the statement for every $a \in A$. □

Related to the three different kinds of subobjects which we have discussed at the beginning there are several directed systems "exhausting" a given partial algebra.

Assumption and notation Let $\underline{A}$ be any partial algebra of type (Ω, τ), and let M be any generating subset of $\underline{A}$. Then consider the following sets:

$I_1 := \{C_{\underline{A}} N \mid N \subseteq M$ and N is finite $\}$;

$I_2 := \{\underline{B} \mid \underline{B}$ is a relative subalgebra of $\underline{A}$, B is finite and $B = C_{\underline{B}}(B \cap M)\}$;

$I_3 := \{\underline{B} \mid \underline{B}$ is a totally finite weak relative subalgebra of $\underline{A}$, and $B = C_{\underline{B}}(B \cap M)\}$.

In each case define for $\underline{B}$, $\underline{B}' \in I_k$ and $1 \leq k \leq 3$ $\underline{B} \leq \underline{B}'$ if and only if $\underline{B}$ is a weak relative subalgebra of $\underline{B}'$; and if $\underline{B} \leq \underline{B}'$ then define $f_{\underline{B},\underline{B}'} := id_{BB'} : \underline{B} \to \underline{B}'$ to be the natural embedding of $\underline{B}$ into $\underline{B}'$. Finally, let for $1 \leq k \leq 3$ $\underline{I}_k := (I_k, \leq)$ and $\mathcal{A}_{\underline{I}_k} := \{\underline{B}, f_{\underline{B},\underline{B}'} : \underline{B} \to \underline{B}' \mid \underline{B} \leq \underline{B}'$ in $\underline{I}_k\}$.

Proposition 1.26 *With the above notation one has that each $\mathcal{A}_{\underline{I}_k}$ is a directed system with direct limit $((id_{BA} : \underline{B} \to \underline{A})_{\underline{B} \in I_k}, \underline{A})$ for $1 \leq k \leq 3$.*

Proposition 1.26 shows that the three kinds of subobjects of partial algebras allow "three different kinds of finiteness" which can be used in connection with generation and exhaustion of structures. In particular, the totally finite weak relative substructures used in $\mathcal{A}_{\underline{I}_3}$ are those which can usually at most be represented on computers, although finiteness is even much more restricted there, with the consequence that every infinite algebra will have — with respect to a generating system — parts, which will never be representable on a computer, while others will become representable when the capacity of computers increases.

Observation Direct limits provide another possibility to define reduced products: If $(\underline{A}_i)_{i \in I}$ is a family of partial algebras of type (Ω, τ) such that all $\underline{A}_i$ are non-empty ($i \in I$), if $\mathcal{F}$ is a filter on I, and if we define $\underline{A}_F := \prod_{i \in F} \underline{A}_i$ and for $F, F' \in \mathcal{F}$ with $F \supseteq F'$, $pr_{FF'} : \underline{A}_F \to \underline{A}_{F'}$, $(a_i \mid i \in F) \mapsto (a_i \mid i \in F')$ to be the corresponding projection, then

$$\mathcal{A}_{\mathcal{F}} := (\underline{A}_F, pr_{FF'} : \underline{A}_F \to \underline{A}_{F'} \mid F, F' \in \mathcal{F}, F \supseteq F'\}$$

is a directed system and $\lim_{\rightarrow} \mathcal{A}_{\mathcal{F}} \cong \prod_{i \in I} \underline{A}_i / \theta_{\mathcal{F}}$.
If, however, $A_i = \emptyset$ for at least one $i \in I$, and $I \setminus \{i \in I \mid A_i = \emptyset\} \in \mathcal{F}$, then $\prod_{i \in I} \underline{A}_i / \theta_{\mathcal{F}} = \underline{\emptyset}$, while

$$\lim_{\rightarrow} \mathcal{A}_{\mathcal{F}} = \lim_{\rightarrow} \mathcal{A}_{\{F \in \mathcal{F} \mid F \subseteq I \setminus \{i \in I \mid A_i = \emptyset\}\}}$$

is non-empty.
We shall refer to this construction of a reduced product as to the *category theoretic reduced product.*

Some operators derived from the constructions

Having at hand several ways to construct new partial algebras from given ones, these constructions lead to operators on the class of all partial algebras. These operators and their "behaviour with respect to other ones" will be of interest in connection with implicational model theory of partial algebras.

Definition Let $\mathfrak{K} \subseteq Alg(\tau)$ be any class of partial algebras of the same type τ. Then we define

$\mathcal{H}\mathfrak{K} := \mathcal{H}_w\mathfrak{K} := \{\underline{B} \in Alg(\tau) |$ there are $\underline{A} \in \mathfrak{K}$ and a *surjective* homomorphism $f : \underline{A} \to \underline{B}\}$ = class of all *(weak) homomorphic images* of $\mathfrak{K}$-algebras.

$\mathcal{H}_f\mathfrak{K} := \{\underline{B} \in Alg(\tau) |$ there are $\underline{A} \in \mathfrak{K}$ and a *full and surjective* homomorphism $f : \underline{A} \to \underline{B}\}$ = class of all *full homomorphic images* of $\mathfrak{K}$-algebras.

$\mathcal{H}_c\mathfrak{K} := \{\underline{B} \in Alg(\tau) |$ there are $\underline{A} \in \mathfrak{K}$ and a *closed and surjective* homomorphism $f : \underline{A} \to \underline{B}\}$ = class of all *closed homomorphic images* of $\mathfrak{K}$-algebras.

$\mathcal{I}\mathfrak{K} := \{\underline{B} \in Alg(\tau) |$ there are $\underline{A} \in \mathfrak{K}$ and an *isomorphism* $f : \underline{A} \to \underline{B}\}$ = class of all *isomorphic copies* of $\mathfrak{K}$-algebras.

$\mathcal{S}\mathfrak{K} := \mathcal{S}_c\mathfrak{K} := \{\underline{A} \in Alg(\tau) |$ there are $\underline{B} \in \mathfrak{K}$ and a *closed and injective* homomorphism $f : \underline{A} \to \underline{B}\}$ = class of all isomorphic copies of *(closed) subalgebras* of $\mathfrak{K}$-algebras.

$\mathcal{S}_f\mathfrak{K} := \{\underline{A} \in Alg(\tau) |$ there are $\underline{B} \in \mathfrak{K}$ and a *full and injective* homomorphism $f : \underline{A} \to \underline{B}\}$ = class of all isomorphic copies of *relative subalgebras* of $\mathfrak{K}$-algebras.

$\mathcal{S}_w\mathfrak{K} := \{\underline{A} \in Alg(\tau) |$ there are $\underline{B} \in \mathfrak{K}$ and an *injective* homomorphism $f : \underline{A} \to \underline{B}\}$ = class of all isomorphic copies of *weak relative subalgebras* of $\mathfrak{K}$-algebras.

$\mathcal{P}\mathfrak{K} := \{\underline{B} \in Alg(\tau) |$ there are a set I and a family $(\underline{A}_i)_{i\in I} \in \mathfrak{K}^I$ such that $\underline{B} \cong \prod_{i\in I} \underline{A}_i\}$ = class of all *products* of families of $\mathfrak{K}$-algebras.

$\mathcal{P}_+\mathfrak{K} := \{\underline{B} \in Alg(\tau) |$ there are a *non-empty* set I and a family $(\underline{A}_i)_{i\in I} \in \mathfrak{K}^I$ such that $\underline{B} \cong \prod_{i\in I} \underline{A}_i\}$ = class of all *products of non-empty families* of $\mathfrak{K}$-algebras.

$\mathcal{P}_r\mathfrak{K} := \{\underline{B} \in Alg(\tau) |$ there are a set I, a filter $\mathcal{F}$ on I and a family $(\underline{A}_i)_{i\in I} \in \mathfrak{K}^I$ such that $\underline{B} \cong (\prod_{i\in I} \underline{A}_i\})/\mathcal{F}\}$ = class of all isomorphic copies of *reduced products* of families of $\mathfrak{K}$-algebras.

$\mathcal{P}_{r+}\mathfrak{K} := \{\underline{B} \in Alg(\tau) |$ there are a *non-empty* set I, a filter $\mathcal{F}$ on I and a family $(\underline{A}_i)_{i\in I} \in \mathfrak{K}^I$ such that $\underline{B} \cong (\prod_{i\in I} \underline{A}_i\})/\mathcal{F}\}$ = class of all isomorphic copies of *reduced products of non-empty families* of $\mathfrak{K}$-algebras.

$\mathcal{P}_u\mathfrak{K} := \{\underline{B} \in Alg(\tau) |$ there are a set I, an *ultrafilter* $\mathcal{U}$ on I and a family $(\underline{A}_i)_{i\in I} \in \mathfrak{K}^I$ such that $\underline{B} \cong (\prod_{i\in I} \underline{A}_i\})/\mathcal{U}\}$ = class of all isomorphic copies of *ultraproducts* of families of $\mathfrak{K}$-algebras.

$o\mathfrak{K} := \mathfrak{K} \cup \underline{\emptyset} = \mathfrak{K}$ and the empty partial algebra.

$e\mathfrak{K} := \mathfrak{K} \cup \{\underline{T} | \underline{T}$ is a total algebra on a one-element set $\}$ = $\mathfrak{K}$ joint with the set of all products with an empty index set.

Let $\underline{\mathcal{O}}$ be the semigroup generated by the above operators with respect to composition. Then $\underline{\mathcal{O}}$ consists of finite sequences $\mathcal{Y}_1 \ldots \mathcal{Y}_n$ — for some natural number $n \geq 1$ —, where each $\mathcal{Y}$ is one of the above operators. We define for $\mathcal{Y}, \mathcal{Y}' \in \mathcal{O}$: $\mathcal{Y} \subseteq \mathcal{Y}'$ if and only if for every $\mathfrak{K} \subseteq Alg(\tau)$ one has $\mathcal{Y}\mathfrak{K} \subseteq \mathcal{Y}'\mathfrak{K}$. And we set $\mathcal{Y} = \mathcal{Y}'$, if $\mathcal{Y} \subseteq \mathcal{Y}'$ and $\mathcal{Y}' \subseteq \mathcal{Y}$. For convenience we add to the generating set of operators of $\underline{\mathcal{O}}$ the operator i: $i\mathfrak{K} := \mathfrak{K}$ for each $\mathfrak{K} \subseteq Alg(\tau)$. Thus $\underline{\mathcal{O}}$ is in fact a monoid.

Definition A class $\mathfrak{K} \subseteq Alg(\tau)$ of partial algebras is called

- *primitive*, if $\mathfrak{K} = \mathcal{HSP}\mathfrak{K}$,
- *quasiprimitive*, if $\mathfrak{K} = \mathcal{SP}\mathfrak{K} = \mathcal{ISP}\mathfrak{K}$.

Lemma 1.27 *Let $\mathcal{Y} \in \mathcal{O}$ be an operator on $Alg(\tau)$, then one has:*

(i) *$i\mathcal{Y} = \mathcal{Y}i = \mathcal{Y}$ and $\mathfrak{K} \subseteq \mathfrak{K}' \subseteq Alg(\tau)$ implies $\mathcal{Y}\mathfrak{K} \subseteq \mathcal{Y}\mathfrak{K}'$.*

(ii) *$i \subseteq \mathcal{Y}$, and, for $\mathcal{V}, \mathcal{W} \in \mathcal{O}$, $\mathcal{V} \subseteq \mathcal{W}$ implies $\mathcal{YV} \subseteq \mathcal{YW}$ and $\mathcal{VY} \subseteq \mathcal{WY}$.*

(iii) *$\mathcal{Y} \subseteq \mathcal{Y}$, and for $\mathcal{V}, \mathcal{W} \in \mathcal{O}$ one has that $\mathcal{Y} \subseteq \mathcal{V}$ and $\mathcal{V} \subseteq \mathcal{W}$ imply $\mathcal{Y} \subseteq \mathcal{W}$, i.e. "$\subseteq$" is a quasi-order relation on $\mathcal{O}$.*

(iv) *$\mathcal{IY} = \mathcal{YI}$.*

(v) *If $\mathcal{Y} \in \{\mathcal{H}, \mathcal{H}_f, \mathcal{H}_c, \mathcal{S}, \mathcal{S}_r, \mathcal{S}_w, \mathcal{I}, \mathcal{P}, \mathcal{P}_+, \mathcal{P}_r, \mathcal{P}_{r+}, \mathcal{P}_u\}$ or if $\mathcal{Y} \in \mathcal{O}$ is a sequence which somewhere contains one of these operators, then $\mathcal{IY} = \mathcal{Y}$.*

(vi) *If $\mathcal{Y}$ is one of the operators generating $\underline{\mathcal{O}}$, then $\mathcal{YY} = \mathcal{Y}$.*

(vii) *$\mathcal{P} = e\mathcal{P}_+ = \mathcal{P}_+e = e\mathcal{P} = \mathcal{P}e$ and $\mathcal{P}_r = e\mathcal{P}_{r+} = \mathcal{P}_{r+}e = e\mathcal{P}_r = \mathcal{P}_re$.*

(viii) $\mathcal{I} \subseteq \mathcal{H}_c \subseteq \mathcal{H}_f \subseteq \mathcal{H}$,
$\mathcal{I} \subseteq \mathcal{S} \subseteq \mathcal{S}_r \subseteq \mathcal{S}_w$,
$\mathcal{I} \subseteq \mathcal{P}_u \subseteq \mathcal{P}_r$,
$\mathcal{I} \subseteq \mathcal{P}$,
$\mathcal{P}_+ \subseteq \mathcal{P} \subseteq \mathcal{P}_r$,
$\mathcal{P}_+ \subseteq \mathcal{P}_{r+} \subseteq \mathcal{P}_r$.

(ix) *Let $\mathcal{Y} \in \{\mathcal{I}, \mathcal{H}, \mathcal{H}_f, \mathcal{H}_c\}$, $\mathcal{V} \in \{\mathcal{S}, \mathcal{S}_r, \mathcal{S}_w\}$, $\mathcal{W} \in \{\mathcal{P}, \mathcal{P}_+, \mathcal{P}_r, \mathcal{P}_{r+}, \mathcal{P}_u\}$, then one has: $\mathcal{VY} \subseteq \mathcal{YV}$, $\mathcal{WY} \subseteq \mathcal{YW}$, and $\mathcal{WV} \subseteq \mathcal{VW}$.*

(x) *With the notation from* (ix) *we have that all of the operators $\mathcal{Y}, \mathcal{V}, \mathcal{W}, \mathcal{YV}, \mathcal{YW}, \mathcal{VW}$ and $\mathcal{YVW}$ are closure operators on $Alg(\tau)$, i.e. they are monotonic, extensive and idempotent.*

(xi) $\mathcal{P}_{r+} \subseteq \mathcal{HP}_+$,
$\mathcal{P}_r \subseteq \mathcal{HP}$,
$\mathcal{P}_r \subseteq \mathcal{SPP}_u$,
$\mathcal{P}_{r+} \subseteq \mathcal{SP}_+\mathcal{P}_u$.

2 Free partial algebras and universal solutions

One of the most useful concepts in universal algebra, also with respect to computer science, is the one of a relatively free(ly generated) algebra (with respect to some class of algebras). In computer science free algebras are used in order to distinguish (specify) an algebra as (abstract) data type, e.g. the one freely generated by the empty set, in a class axiomatized by first order formulas. It is thus a tool of higher order applied to characterize algebras uniquely up to isomorphism as initial algebras. The methods and results carry over to partial algebras. As for total algebras free partial algebras play an important role in the (existence–) equational theory.

2.1 Free partial algebras

Definition Let $\underline{A}$ be a partial algebra of similarity type (Ω, τ), let $\mathfrak{K} \subseteq Alg(\tau)$ be any class of partial algebras, and let $M \subseteq A$ be any subset. Then we define:

(i) M is called a *$\mathfrak{K}$-independent ($\mathfrak{K}$-free) subset* of a partial algebra $\underline{A}$ (M is $\mathfrak{K}$-free in $\underline{A}$), if for every partial algebra $\underline{B}$ of $\mathfrak{K}$ and for every mapping $f_0 : M \to B$ there exists a homomorphic extension $f : C_{\underline{A}}M \to \underline{B}$, $f|_M = f_0$.

(ii) The subclass of $Alg(\tau)$, defined as

$$ind_{\underline{A}}M := \{\underline{B} \in Alg(\tau) | M \text{ is } \{\underline{B}\}\text{-free in } \underline{A}\}$$

is called the *independence class* of M with respect to the partial algebra $\underline{A}$.

(iii) If M is a generating subset of $\underline{A}$, if $\mathfrak{K} = \{\underline{A}\}$, and if M is $\mathfrak{K}$-free in $\underline{A}$, then we say that $\underline{A}$ is a *(relatively) free partial algebra, freely generated by M*, or — more briefly — a *free partial algebra* with *basis M*.

(iv) If M generates $\underline{A}$, if $\underline{A} \in \mathfrak{K}$, and if M is $\mathfrak{K}$-free in $\underline{A}$, then we say that $\underline{A}$ is a *$\mathfrak{K}$-free $\mathfrak{K}$-algebra, $\mathfrak{K}$-freely generated by M* (a *$\mathfrak{K}$-free $\mathfrak{K}$-algebra with $\mathfrak{K}$-basis M*). Since we will show below that a $\mathfrak{K}$-free $\mathfrak{K}$-algebra with $\mathfrak{K}$-basis M is determined by M and $\mathfrak{K}$ up to isomorphism, we shall denote it by $\underline{F}(M, \mathfrak{K})$, if it exists at all.

Remarks and examples

(i) A $\mathfrak{K}$-free $\mathfrak{K}$-algebra on some set M need not exist. E.g. in the class $\mathfrak{F}$ of all fields with arbitrary characteristics an $\mathfrak{F}$-free field does not exist. However, if p is any prime number or 0, and if $\mathfrak{F}_p$ designates the class of all fields of characteristic p, if $\mathbf{F}_p$ denotes the prime field of characteristic p (with a similarity type containing two nullary constants for 0 and 1), then $\mathbf{F}_p \cong \underline{F}(\emptyset, \mathfrak{F}_p)$, while for sets $M \neq \emptyset$ $\underline{F}(M, \mathfrak{F}_p)$ does not exist. (Observe that homomorphisms between fields have always to be injective.)

(ii) We shall see later that for any primitive or quasi-primitive class $\mathfrak{K}$ $\underline{F}(M, \mathfrak{K})$ belongs to $\mathfrak{K}$ for every set M. However, $\underline{F}(M, \mathfrak{K})$ will always be defined, even if it does not belong to $\mathfrak{K}$. Namely, it is more generally defined as the object of the $\mathfrak{K}$-universal solution of $M_{discrete}$ (see subsection 2.3).

(iii) If $\underline{A}$ is a nonempty partial algebra generated by the empty set, and if $\mathfrak{K}$ is any class of partial algebras of the same type, and if $\mathfrak{K}$ contains the empty partial algebra, then $\emptyset$

cannot be $\mathfrak{K}$-free in $\underline{A}$.

(iv) For every partial algebra $\underline{A}$ and for every set M the independence class $ind_{\underline{A}}M$ always contains all one-element total algebras, therefore it can never be empty.

(v) All total free algebras like e.g. free monoids (=word monoids), free groups, free rings (=polynomial rings over the ring $\underline{\mathbf{Z}}$ of integers) are also examples of (relatively) free partial algebras.

(vi) The discrete partial algebras are exactly the $Alg(\tau)$-free partial $Alg(\tau)$-algebras.

(vii) Let $\mathfrak{C}$ be the class of all small categories, M any set, then set $M_0 := \{(m,0) \mid m \in M\}$, $M_1 := \{(m,1) \mid m \in M\}$ and assume that $M \cap M_0 = M_0 \cap M_1 = M \cap M_1 = \emptyset$. Then $F := M \cup M_0 \cup M_1$ is the carrier set of the free category $\underline{F}$ $\mathfrak{C}$-freely generated by M, when we define for each $m \in M$:
$Dom^{\underline{F}}(m) := Dom^{\underline{F}}(m,0) := (m,0)$,
$Cod^{\underline{F}}(m) := Cod^{\underline{F}}(m,1) := (m,1)$,
$Dom^{\underline{F}}(m,1) := (m,1), Cod^{\underline{F}}(m,0) := (m,0)$,
$dom \circ^{\underline{F}} := \{(m,(m,0)),\ ((m,1),m),\ ((m,0),(m,0)),\ ((m,1),(m,1)) \mid m \in M\}$,
and $\circ^{\underline{F}}$ is then defined according to the rule (C 2) for small categories given in subsection 1.1.

Proposition 2.1 *For any partial algebra $\underline{A}$ and for every subset M of $\underline{A}$ the independence class $ind_{\underline{A}}M$ is always a primitive class: $\mathcal{HSP}ind_{\underline{A}}M = ind_{\underline{A}}M$.*
And if $M \neq \emptyset$, then trivially $\emptyset \in ind_{\underline{A}}M$.

Corollary 2.1 *Let $\underline{A} \in Alg(\tau)$ with some generating subset M. Then the following statements are equivalent:*

(i) *$\underline{A}$ is a free partial algebra with basis M.*

(ii) *There is some primitive class $\mathfrak{K}$ containing $\underline{A}$ such that $\underline{A}$ is $\mathfrak{K}$-freely generated by M.*

(iii) *$\underline{A} \in ind_{\underline{A}}M$.*

(iv) *$\mathcal{HSP}\{\underline{A}\} \subseteq ind_{\underline{A}}M$.*

2.2 Partial Peano algebras

A class of particular interesting partial algebras is the one of partial Peano algebras, which includes the class of all (global) term algebras.

Definition Let $\underline{P} \in Alg(\tau)$ be any partial algebra, and let X be any subset of P. We say that $\underline{P}$ is a *partial Peano algebra with Peano basis* X (briefly: *on* X), if the following *generalized Peano axioms* are valid in $\underline{P}$:

(P1) For every $\varphi \in \Omega$ and for every $(a_1, \ldots, a_{\tau(\varphi)}) \in dom\ \varphi^{\underline{P}}$ one has $\varphi^{\underline{P}}(a_1, \ldots, a_{\tau(\varphi)}) \notin X$ (i.e. $X \cap \bigcup_{\varphi \in \Omega} \varphi^{\underline{P}}(dom\ \varphi^{\underline{P}}) = \emptyset$).

(P2) For any $\varphi, \psi \in \Omega$, and for any $(a_1, \ldots, a_{\tau(\varphi)}) \in dom\ \varphi^{\underline{P}}$, $(b_1, \ldots, b_{\tau(\psi)}) \in dom\ \psi^{\underline{P}}$ one has that $\varphi^{\underline{P}}(a_1, \ldots, a_{\tau(\varphi)}) = \psi^{\underline{P}}(b_1, \ldots, b_{\tau(\psi)})$ implies $\varphi = \psi$ and $a_i = b_i$ for $1 \leq i \leq \tau(\varphi)$ (i.e. each partial operation is injective and any two distinct partial operations have disjoint ranges).

(P3) (Axiom of Induction) $\mathcal{C}_{\underline{P}}X = P$(, i.e. X generates $\underline{P}$).

Remarks

(i) A total Peano algebra (on X) is simply called a *Peano algebra* — often also *term algebra* or *word algebra* (see below).

(ii) A discrete partial algebra $\underline{D}$ is always a partial Peano algebra on D.

(iii) Let $\mathbf{N}_0 := \{0, 1, 2, \ldots\}$ be the set of natural numbers, and let $' : \mathbf{N}_0 \to \mathbf{N}_0$ be the successor function, i.e. $n' := n + 1$ for each $n \in \mathbf{N}_0$. Then $(\mathbf{N}_0, ')$ is a Peano algebra on $\{0\}$, and the generalized Peano axioms are just the usual Peano axioms for natural numbers.

Lemma 2.2 *Let $\underline{P}$ be any partial Peano algebra on X, and let $\underline{A}$ be any weak relative subalgebra of $\underline{P}$. Then $\underline{A}$ is a partial Peano algebra on $Y := A \setminus \bigcup_{\varphi\in\Omega} \varphi^{\underline{A}}(dom\ \varphi^{\underline{A}})$.*

The proof of this fact is also a good example for the principle of structural induction, here applied to X and P with respect to the property

$$a \in A \Rightarrow a \in \mathcal{C}_{\underline{A}}Y$$

All partial Peano algebras can be obtained up to isomorphism, by using this lemma, from the following total Peano algebras: Let X and Ω be disjoint, and let $(X \cup \Omega)^*$ be the free monoid on the alphabet $X \cup \Omega$, i.e. $W_0 := (X \cup \Omega)^* = \{y_1 \ldots y_n | y_1, \ldots, y_n \in X \cup \Omega, n \in \mathbf{N}_0\}$. On $(X \cup \Omega)^*$ define a total algebraic structure as follows: Let $w_1, \ldots, w_{\tau(\varphi)} \in W_0$, then $\varphi^{\underline{W}_0}(w_1, \ldots, w_{\tau(\varphi)}) := \varphi w_1 \ldots w_{\tau(\varphi)}$ (concatenation of sequences). Finally set $\underline{P} := \underline{\mathcal{C}_{W_0}X}$.

Theorem 2.1 *The algebra $\underline{P} := \underline{\mathcal{C}_{W_0}X}$ defined above is a (total!) Peano algebra on X of similarity type τ.*

One of the most important features of partial Peano algebras is the following

Theorem 2.2 (Recursion Theorem for Peano Algebras)
Let $\underline{P}$ be any partial Peano algebra with Peano basis X, and let $\underline{C}$ be any total algebra of the same similarity type. Then for any mapping $f : X \to C$ there exists a unique homomorphic extension $f' : \underline{P} \to \underline{C}$, i.e. $f'|_X = f$. In particular one has graph $f' = \mathcal{C}_{\underline{P}\times\underline{C}}$ graph f.

2.3 Free completions

Definition Let $\underline{A} \in Alg(\tau)$ be any partial algebra. $\underline{B} \in TAlg(\tau)$ is called an (*absolutely*) *free completion* of $\underline{A}$ if the following axioms hold:

(FC0) $A \subseteq B$ and for each $\varphi \in \Omega$ one has *graph* $\varphi^{\underline{A}} \subseteq$ *graph* $\varphi^{\underline{B}}$.

(FC1) For each $\varphi \in \Omega$ and for each sequence $\underline{b} = (b_1, \ldots, b_{\tau(\varphi)}) \in B^{\tau(\varphi)}$ one has $\varphi^{\underline{B}}\underline{b} \in A$ implies $\varphi^{\underline{B}}\underline{b} = \varphi^{\underline{A}}\underline{b}$ (thus in particular $\underline{b} \in dom\ \varphi^{\underline{A}} \subseteq A^{\tau(\varphi)}$).

(FC2) For all $\varphi, \psi \in \Omega$, for all $\underline{b} = (b_1, \ldots, b_{\tau(\varphi)}) \in B^{\tau(\varphi)}$, and for all $\underline{b}' = (b'_1, \ldots, b'_{\tau(\psi)}) \in B^{\tau(\psi)}$ one has: $\varphi^{\underline{B}}\underline{b} = \psi_{\underline{B}}\underline{b}' \notin A$ implies $\varphi = \psi$ and $\underline{b} = \underline{b}'$.

(FC3) (Axiom of Induction) $\mathcal{C}_{\underline{B}}A = B$ (i.e. the completion $\underline{B}$ of $\underline{A}$ is *minimal*).

Remarks

(i) The axioms (FC0) through (FC3) above are closely related to the generalized Peano axioms: (FC0) tells you that $\underline{A}$ is at least a weak relative subalgebra, while (FC1) says that $\underline{A}$ is even a *normal* relative subalgebra of $\underline{B}$ (normality means that no application of a fundamental operation to a sequence with at least one argument outside of A can have its value in A). This fact corresponds to (P1). (FC2) says that "outside" of A (P2) is satisfied, while (FC3) corresponds to (P3). This observation shows that if we forget for a free completion $\underline{B}$ of $\underline{A}$ the structure of $\underline{A}$ in $\underline{B}$, then we end with a partial Peano algebra on A.

(ii) Absolutely free completions of $\underline{A}$ are important in connection with a description of all minimal completions of $\underline{A}$. Therefore they are a useful tool in connection with *error handling* for the specification of abstract data types.

Corollary 2.2 *A total algebra $\underline{B}$ is a Peano algebra over some subset X iff the relative subalgebra $\underline{X}$ of $\underline{B}$ is discrete and $\underline{B}$ is the free completion of X.*

The existence of a free completion of any given algebra $\underline{A}$ can be based upon the existence of total Peano algebras: Let $\underline{P}$ be a total Peano algebra on A, and define for each $\varphi \in \Omega$ and for each $\underline{b} = (b_1, \ldots, b_{\tau(\varphi)})$ in $P^{\tau(\varphi)}$ on P:

$$\varphi^{\#}(\underline{b}) := \begin{cases} \varphi^{\underline{A}}(\underline{b}) & \text{, if } \underline{b} \in dom\ \varphi^{\underline{A}} \\ \varphi^{\underline{P}}(\underline{b}) & \text{, otherwise.} \end{cases}$$

Let $\underline{P}^{\#} := (P, (\varphi^{\#})_{\varphi\in\Omega})$ and $\underline{B} := \underline{C}_{\underline{P}^{\#}} A$.

Theorem 2.3 $\underline{B} := \underline{C}_{\underline{P}^{\#}} A$ *is a free completion of* $\underline{A}$.

Remark Consider again the Peano algebra $\underline{P}$ on A as above. Consider on $\underline{P}$ the congruence relation Θ, generated by the relation

$$R := \{(\varphi^{\underline{A}}(a_1, \ldots, a_{\tau(\varphi)}), \varphi^{\underline{P}}(a_1, \ldots, a_{\tau(\varphi)})) | \varphi \in \Omega, (a_1, \ldots, a_{\tau(\varphi)}) \in dom\ \varphi^{\underline{A}}\}.$$

Then there is a full and injective homomorphism $i : \underline{A} \to \underline{P}/\Theta$ such that $\underline{P}/\Theta$ is a free completion of the relative subalgebra $\underline{i(A)}$ of $\underline{P}/\Theta$ (and $\underline{i(A)}$ is isomorphic to $\underline{A}$ by i).

Observe that the second construction is the more canonical one, while the first one gives more directly insight into the structure of the result and can be compared with the modelling of algebras in computer science (see [GoTcWa78]).

In analogy to the Recursion Theorem for Peano algebras we have the

Theorem 2.4 (Recursion Theorem for Free Completions) *Let $\underline{A}$ be any partial algebra, $\underline{B}$ a free completion of $\underline{A}$, and $\underline{C}$ be any* total *algebra similar to $\underline{A}$. Then, for every homomorphism $f : \underline{A} \to \underline{C}$ there exists a unique homomorphism $f' : \underline{B} \to \underline{C}$ extending f, i.e. $f'|_A = f$.*

$$\begin{tikzcd} \underline{C} \arrow[rr, "f\,:\,\mathfrak{K}\text{-ext.}"] \arrow[dr, "g"'] & & \underline{D} \\ & \underline{E} \in \mathfrak{K} & \end{tikzcd} \quad \mapsto \quad \begin{tikzcd} \underline{C} \arrow[rr, "f\,:\,\mathfrak{K}\text{-ext.}"] \arrow[dr, "g"'] & & \underline{D} \arrow[dl, "h\,:\,h \circ f = g"] \\ & \underline{E} \in \mathfrak{K} & \end{tikzcd}$$

Figure 7: f is $\mathfrak{K}$-extendable

$$\begin{tikzcd} \underline{C} \arrow[rr, "f'\,:\,\mathfrak{K}\text{-ext. epi.}"] \arrow[dr, "f\,:\,\mathfrak{K}\text{-univ.}"'] & & \underline{D'} \\ & \underline{D} & \end{tikzcd} \quad \mapsto \quad \begin{tikzcd} \underline{C} \arrow[rr, "f'\,:\,\mathfrak{K}\text{-ext. epi.}"] \arrow[dr, "f\,:\,\mathfrak{K}\text{-univ.}"'] & & \underline{D'} \arrow[dl, "l\,:\,l \circ f' = f"] \\ & \underline{D} & \end{tikzcd}$$

Figure 8: f is a $\mathfrak{K}$-universal epimorphism.

Corollary 2.3 *Let $\underline{A} \in Alg(\tau)$ be any partial algebra. Then there is up to isomorphism (over id_A as restriction to A) exactly one free completion, say $\underline{B}$ of $\underline{A}$.*

Remarks and Definition

(i) Because of its "similarity" with a $TAlg(\tau)$-free partial algebra with $TAlg(\tau)$-basis A we denote the absolutely free completion of $\underline{A}$ by $\underline{F}(\underline{A}, TAlg(\tau))$. Indeed the free completion $\underline{F}(\underline{A}, TAlg(\tau))$ of $\underline{A}$ together with the canonical embedding

$$i = id_{A,F(\underline{A},TAlg(\tau))} : \underline{A} \to \underline{F}(\underline{A}, TAlg(\tau))$$

is a special instance of some more general concept (observe that i is an epimorphism because of (FC3)):

Let $\mathfrak{K} \subseteq Alg(\tau)$ be any class of partial algebras, and let $f : \underline{C} \to \underline{D}$ be any homomorphism in the category $\underline{Alg(\tau)}$.

(a) f is called *$\mathfrak{K}$-extendable*, and $\mathfrak{K}$ is said to be *injective w.r.t. f*, iff for every $\underline{E} \in \mathfrak{K}$ and for every homomorphism $g : \underline{C} \to \underline{E}$ there exists a homomorphism $h : \underline{D} \to \underline{E}$ such that $g = h \circ f$ (cf. Figure 7).

Observe that we can show later that the $\mathfrak{K}$-extendable epimorphisms represent implications, which are valid in $\mathfrak{K}$.

(b) The homomorphism $f : \underline{C} \to \underline{D}$ is called *$\mathfrak{K}$-universal*, if it is a $\mathfrak{K}$-extendable *epimorphism* (cf. the observation above), and if for every $\mathfrak{K}$-extendable epimorphism $f' : \underline{C} \to \underline{D'}$ there exists a homomorphism $l : \underline{D'} \to \underline{D}$ such that $l \circ f' = f$ (cf. Figure 8).

Observe, that then also f' is required to be an epimorphism.

(c) If $f : \underline{C} \to \underline{D}$ is a $\mathfrak{K}$-universal epimorphism, then the pair $(f, \underline{D})$ is called a *$\mathfrak{K}$-universal solution* of $\underline{C}$. And if, in addition, $\underline{D} \in \mathfrak{K}$, then $\underline{D}$ is called a *$\mathfrak{K}$-universal $\mathfrak{K}$-solution* of $\underline{C}$.

(ii) Observe, that for every partial algebra $\underline{A}$ $\mathfrak{K}$-freely generated by a subset M the homomorphism $id_{MA} : \underline{M}_{\text{discrete}} \to \underline{A}$ is a $\mathfrak{K}$-extendable epimorphism; and if $\underline{A} \in \mathfrak{K}$, then $(id_{MA}, \underline{A})$ is a $\mathfrak{K}$-universal $\mathfrak{K}$-solution of $\underline{M}_{\text{discrete}}$.

Similarly, it is easy to realize, that

$$(id_{A,F(\underline{A},TAlg(\tau))}, \underline{F}(\underline{A}, TAlg(\tau))$$

is a $TAlg(\tau)$-universal $TAlg(\tau)$-solution of $\underline{A}$.

(iii) From the definition it easily follows that, if $f : \underline{C} \to \underline{D}$ and $f' : \underline{C} \to \underline{D}'$ are $\mathfrak{K}$-universal, then there is a unique isomorphism $l : \underline{D}' \to \underline{D}$ such that $l \circ f' = f$. Thus $\mathfrak{K}$-universal solutions are determined up to unique isomorphism, and we write $\underline{D} =: \underline{F}(\underline{C}, \mathfrak{K})$ and $f =: r_{\underline{C},\underline{F}(\underline{C},\mathfrak{K})}$.

(iv) For those who know a bit more about category theory, let us observe that the fact that every partial algebra $\underline{A}$ has a $\mathfrak{K}$-universal $\mathfrak{K}$-solution is equivalent to the fact that the embedding functor $I : \underline{\mathfrak{K}} \to \underline{Alg(\tau)}$ of the full subcategory $\underline{\mathfrak{K}}$ of $\underline{Alg(\tau)}$ has a left adjoint.

(v) Our remarks show that a Peano algebra on X is exactly the $TAlg(\tau)$-algebra $TAlg(\tau)$-freely generated by X, i.e. the $TAlg(\tau)$-universal $TAlg(\tau)$-solution of $\underline{X}_{\text{discrete}}$.

Because of the close relationship between Peano algebras and free completions (cf. Corollary 2.2) one of the most useful tools for a model theory of partial algebras can be formulated and proved for free completions in general:

Theorem 2.5 (Generalized Recursion Theorem) *Let $f : \underline{A} \to \underline{B}$ be any homomorphism in $\underline{Alg(\tau)}$. Then there exists a closed homomorphic extension*

$$(f^{\sim})_{\underline{B}} =: f^{\sim} : \underline{dom}\ f^{\sim} \to \underline{B}$$

*of f such that $\underline{dom}\ f^{\sim}$ is an A-*generated *relative subalgebra of the* (absolutely) free completion $\underline{F}(\underline{A}, TAlg(\tau))$ *of $\underline{A}$. Moreover, we have:*

(i) *$graph\ f^{\sim} = C_{\underline{F}(\underline{A},TAlg(\tau))\pi\underline{B}}\, graph\ f$.*

(ii) *$f^{\sim}$ is the largest homomorphic extension of f to an A-*generated *relative subalgebra of $\underline{F}(\underline{A}, TAlg(\tau))$.*

(iii) *Let $\bar{f} : \underline{F}(\underline{A}, TAlg(\tau)) \to \underline{F}(\underline{B}, TAlg(\tau))$ be the homomorphic extension of f, which exists according to the Recursion Theorem for Free Completions. Then $f^{\sim} = \bar{f}|dom\ f^{\sim}$.*

Definition In honour of Jürgen Schmidt, who introduced these concepts in [Sch70], we call, for $f : \underline{A} \to \underline{B}$, $f^{\sim} : \underline{dom}\ f^{\sim} \to \underline{B}$ the *closed $\underline{A}$-initial extension* of f (in $\underline{F}(\underline{A}, TAlg(\tau))$); and e.g. $ker\ f^{\sim}$ is called the *S-kernel* (short for "*J.Schmidt-kernel*") of f (in symbols: $S\text{-}ker\ f\ (:= ker\ f^{\sim})$).

2.4 Terms and term operations

The Generalized Recursion Theorem 2.5 now allows to describe, what we want to understand by term operations in connection with partial algebras.

Definition As in the case of total algebras we denote for every set X the elements of the $TAlg(\tau)$-algebra $\underline{F}(X, TAlg(\tau))$, which is $TAlg(\tau)$-freely generated by X, as *(global) terms* with variables in X. $\underline{F}(X, TAlg(\tau))$ is therefore also called the *term algebra* with X as set of variables (arguments).

Observe that terms have an instructive representation by rooted trees (cf. e.g. [B86], subsection 5.7).

In order to define the concept of a subterm let us define more generally the *algebraic quasi-order* on any partial algebra $\underline{A}$.

Definition Let $\underline{A} \in Alg(\tau)$ be any partial algebra. Then define on A relations $\lhd_{\underline{A}}$ and $\preceq_{\underline{A}}$ as follows:

(i)

$$\lhd_{\underline{A}} := \{(a, a') \in A \times A \mid \text{there are } \varphi \in \Omega, i_0 \in \{1, \dots, \tau(\varphi)\}$$
$$\text{and } (a_1, \dots, a_{\tau(\varphi)}) \in dom\ \varphi^{\underline{A}} \text{ such that } a = a_{i_0} \text{ and } a' = \varphi^{\underline{A}}(a_1, \dots, a_{\tau(\varphi)})\}.$$

If $a \lhd_{\underline{A}} a'$, then we say that a is an *immediate predecessor* of a'.

(ii) $\preceq_{\underline{A}}$ is then defined as the *reflexive and transitive hull* of $\lhd_{\underline{A}}$. Thus, by definition, $\preceq_{\underline{A}}$ is a quasi-order on A which is called the *algebraic quasi-order* on $\underline{A}$.

(iii) If $M \subseteq A$ is any subset, then we denote by $\downarrow M$ the *initial segment of $\underline{A}$ generated by M*, i.e.

$$\downarrow M := \{a \in A \mid a \preceq_{\underline{A}} m \text{ for some } m \in M\},$$

$\downarrow \{m\} =: \downarrow m$ for singletons.

By $\underline{\downarrow M}$ we designate the relative subalgebra of $\underline{A}$ with carrier set $\downarrow M$.

(iv) If $\underline{A} = \underline{F}(X, TAlg(\tau))$, and if $t \in A$ is any term, then the elements of $\downarrow t$ are called *subterms* of t.

Proposition 2.3 *Let $\underline{A}$ be any partial Peano algebra on some set X, then:*

(i) *$\preceq_{\underline{A}}$ is a partial order on A.*

(ii) *$\preceq_{\underline{A}}$ only allows finite strictly descending chains.*

A useful observation is the following one:

Lemma 2.4 *Let $\underline{A}$ be any partial algebra, $\underline{B}$ a relative subalgebra of $\underline{F}(\underline{A}, TAlg(\tau))$ and $A \subseteq B$. Then the following statements are equivalent:*

(i) *$C_{\underline{B}}A = B$ (i.e. A generates B)*

(ii) *$\downarrow B = B$ (w.r.t. $\preceq_{\underline{F}(\underline{A}, TAlg(\tau))}$), i.e. B is an $\underline{A}$-initial segment of $\underline{F}(\underline{A}, TAlg(\tau))$.*

Definition Let X be any set[2], $\underline{A} \in Alg(\tau)$ any partial algebra, and $t \in F(X, TAlg(\tau))$

[2] We allow that X may possibly be an infinite set, although one will usually consider term operations only for finite argument sets or even only for finite *sequences*. However, our definition of a term operation does neither depend on the cardinality of the argument sequence nor on the fact that it is only a set. In any case a term operation will "depend only on finitely many of its arguments"!

any term. Then define $t^{\underline{A}}$ as a partial mapping out of A^X into A as follows:

$$dom\ t^{\underline{A}} := \{w \in A^X \mid t \in dom\ w^{\sim}\},$$

and $w \in dom\ t^{\underline{A}}$ implies $t^{\underline{A}}(w) := w^{\sim}(t)$.

$t^{\underline{A}}$ is called the (*global partial*) *term operation* on $\underline{A}$ induced by the term t. X is called the set of variables under consideration, while the mapping $w : X \to A$ is sometimes called a *valuation* of X in $\underline{A}$, or an *assignment* of values in A to the variables in X.

$t \in \underline{F}(X, TAlg(\tau))$ is said to be *evaluable* in $\underline{A}$ w.r.t. an assignment $w : X \to A$, iff $t \in dom\ w^{\sim}$ (iff $w \in dom\ t^{\underline{A}}$).

Example Let a similarity type τ, a term $t := \varphi\varphi'\varphi''x$ and a partial algebra $\underline{A}$ be given as sketched in Figure 9.

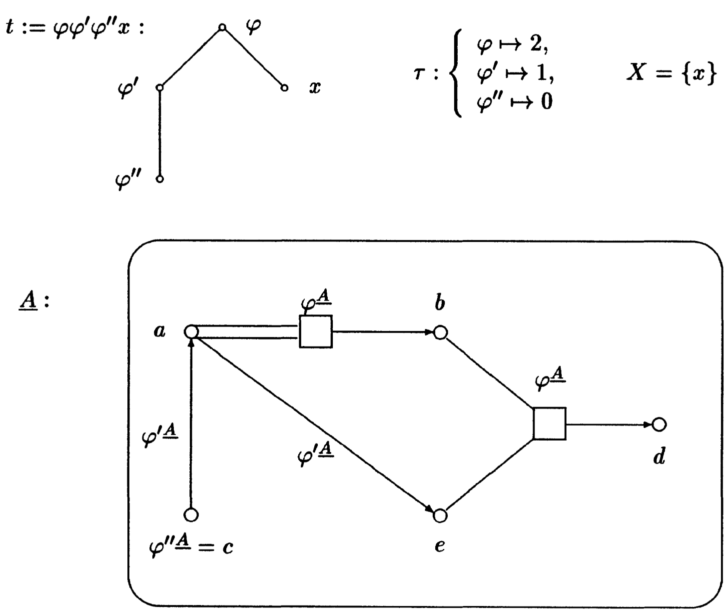


Figure 9: Example for the computation of a termoperation

Then $t^{\underline{A}}(a) = b$ exists, while $t^{\underline{A}}$ is undefined for any other assignment.

Lemma 2.5 *Let X be any set, $f : \underline{A} \to \underline{B}$ be any homomorphism, and $w : X \to A$ any valuation. Then*

$$dom\ w^{\sim} \subseteq dom\ (f \circ w)^{\sim}$$

(with equality, if f is closed), and

$$ker\ w^{\sim} \subseteq ker\ (f \circ w)^{\sim}.$$

Corollary 2.4 *Each homomorphism is compatible with each induced termoperation; i.e. if $f : \underline{A} \to \underline{B}$ is a homomorphism, $t \in F(X, TAlg(\tau))$, $w : X \to A$ and $a \in A$, then $t^{\underline{A}}(w) = a$ implies $t^{\underline{B}}(f \circ w) = f(a)$ (in particular, if $t^{\underline{A}}(w)$ exists, then so does $t^{\underline{B}}(f \circ w)$ and one has $f(t^{\underline{A}}(w)) = t^{\underline{B}}(f \circ w)$).*

The following proposition yields another description of the algebraic closure, and its proof is again a good application of structural induction:

Proposition 2.6 *Let $\underline{A}$ be any partial algebra and M any generating subset of $\underline{A}$.*

(i) *For every $a \in A$ there are a* finite *subset $M_a \subseteq M$ and a term $t \in F(M_a, TAlg(\tau))$ such that*

$$a = t^{\underline{A}}(id_{M_aM}).$$

(ii) *For every $a \in A$ there is a term $t \in F(M, TAlg(\tau))$ such that $a = t^{\underline{A}}(id_M)$.*

2.5 Diagram completion II, the Epimorphism Theorem

In the theory of partial algebras and e.g. in connection with their model theory the Epimorphism Theorem below is of much more importance than the homomorphism theorem, which, however, is an important tool for most of the proofs of the following results. As a preparation observe

Lemma 2.7 *The homomorphism $f : \underline{A} \to \underline{B}$ is an epimorphism, if and only if its greatest A-initial extension $f^{\sim}$ out of $\underline{F}(\underline{A}, TAlg(\tau))$ into $\underline{B}$ is surjective.*

Lemma 2.8 (Diagram Completion Lemma for Epimorphisms) *Let $f : \underline{A} \to \underline{B}$ be an epimorphism and $g : \underline{A} \to \underline{C}$ any homomorphism (see Figure 10).*

a) *Then the following statements are equivalent:*

(i) *There exists a unique homomorphism $h : \underline{B} \to \underline{C}$ such that $h \circ f = g$.*

(ii) *$ker f^{\sim} \subseteq ker g^{\sim}$ (as binary relations in $F(\underline{A}, TAlg(\tau))$).*

b) *If there exists $h : \underline{B} \to \underline{C}$ satisfying $h \circ f = g$, then one has:*

(iii) *h is an epimorphism if and only if g is an epimorphism.*

(iv) *h is surjective, if and only if $g^{\sim}|dom\ f^{\sim}$ is surjective (e.g. if g itself is surjective).*

(v) *h is injective, if and only if $ker f^{\sim} = ker g^{\sim} \cap (dom\ f^{\sim})^2$.*

(vi) *h is closed, if and only if $dom\ f^{\sim} = dom\ g^{\sim}$.*

(vii) *h is closed and injective, if and only if $ker f^{\sim} = ker g^{\sim}$.*

$$\begin{tikzcd}
\underline{D} := \underline{dom}\, f^{\sim} \arrow[rr, two heads, "f^{\sim}"] \arrow[dd, "(\exists\, id_{DE})\ldots"'] & & \underline{B} \arrow[dd, "(\exists\, h)(h \circ f = g) \text{ iff } ker\ f^{\sim} \subseteq ker\ g^{\sim}"] \\
& \underline{A} \arrow[ul, "id_{AD}"'] \arrow[ur, "Epi \ni f"'] \arrow[dl, "id_{AE}"] \arrow[dr, "(\forall)\ g"] & \\
\underline{E} := \underline{dom}\, g^{\sim} \arrow[rr, two heads, "g^{\sim}"'] \arrow[d, "id_{EF}"'] & & \underline{C} \\
\underline{F} := \underline{F}(\underline{A}, TAlg(\tau)) & &
\end{tikzcd}$$

Figure 10: Diagram completion for the epimorphism f

(viii) *h is an isomorphism, if and only if g is an epimorphism and $ker f^{\sim} = ker g^{\sim}$.*

(ix) *h is full, if and only if $g^{\sim}|dom\ f^{\sim}$ is full.*

As a corollary we get, what J.Schmidt has called "General Homomorphism Theorem" (see [Sch70]):

Theorem 2.6 (Epimorphism Theorem) *Let $f : \underline{A} \to \underline{B}$ and $g : \underline{A} \to \underline{C}$ be any two epimorphisms. Then the following statements are equivalent:*

(i) *There exists an isomorphism $h : \underline{B} \to \underline{C}$ such that $h \circ f = g$.*

(ii) *$ker f^{\sim} = ker g^{\sim}$.*

Corollary 2.5 *Let $f : \underline{A} \to \underline{B}$ be any homomorphism, $\underline{E} := \underline{dom f^{\sim}}$ in $\underline{F}(\underline{A}, TAlg(\tau))$, and $e : \underline{A} \to \underline{E}$ the natural injection.*

(i) *Then f is the composition of the $TAlg(\tau)$-extendable epimorphism e followed by the closed homomorphism $f^{\sim}$.*

(ii) *Assume that $f = g \circ e'$, where $e' : \underline{A} \to \underline{C}$ is any $TAlg(\tau)$-extendable epimorphism and $g : \underline{C} \to \underline{B}$ is a closed homomorphism. Then there exists an isomorphism (a unique one) $j : \underline{C} \to \underline{E}$ such that $j \circ e' = e$ and $f^{\sim} \circ j = g$.*

We will return later to this corollary (see Proposition 3.8) in order to realize that the $TAlg(\tau)$-extendable homomorphisms and the closed homomorphisms are two "partners" of a factorization system.

In Theorems 2.8 and 2.6 we have seen, that an epimorphism $f : \underline{A} \to \underline{B}$ induces a *surjective* homomorphism $f^{\sim} : \underline{dom f^{\sim}} \to \underline{B}$ out of $\underline{F}(\underline{A}, TAlg(\tau))$ onto $\underline{B}$. The converse is also true:

Proposition 2.9 *Let $\underline{A}$ be any partial algebra, $\underline{C}$ an A-generated relative subalgebra of $\underline{F}(\underline{A}, TAlg(\tau))$, θ a closed congruence relation on $\underline{C}$, and $nat_\theta : \underline{C} \to \underline{C}/\theta$ the natural projection. Then its restriction to A, i.e. $nat_\theta|_A : \underline{A} \to \underline{C}/\theta$ is an epimorphism.*

Definition Recall that the A-generated relative subalgebras of $\underline{F}(\underline{A}, TAlg(\tau))$ are called *$\underline{A}$-initial segments* of $\underline{F}(\underline{A}; TAlg(\tau))$. Hence the closed congruence relations on $\underline{A}$-initial segments of $\underline{F}(\underline{A}; TAlg(\tau))$ will be called *$\underline{A}$-initial congruences* of $\underline{F}(\underline{A}; TAlg(\tau))$.

Proposition 2.10 *Let $\underline{A} \in Alg(\tau)$ be any partial algebra. Then:*

(i) *The set of all $\underline{A}$-initial segments of the absolutely free completion $\underline{F}(\underline{A}; TAlg(\tau))$ of $\underline{A}$ is a closure system on $\underline{F}(\underline{A}; TAlg(\tau))$ with smallest element $\underline{A}$ (and largest element $\underline{F}(\underline{A}; TAlg(\tau))$).*

(ii) *The set of all $\underline{A}$-initial congruences of $\underline{F}(\underline{A}; TAlg(\tau))$ is a closure system on its square $\underline{F}(\underline{A}; TAlg(\tau))^2$ with smallest element Δ_A $(= \{(a,a) \mid a \in A\})$ and greatest element $\nabla_{F(\underline{A}, TAlg(\tau))}$ $(= F(\underline{A}, TAlg(\tau))^2)$.*

(iii) *For any set $\mathfrak{C}$ of $\underline{A}$-initial congruences, we have*

$$dom \bigcap \mathfrak{C} = \bigcap (dom\ \theta \mid \theta \in \mathfrak{C}),$$

where for any relation $R \subseteq B \times B$ we have

$$\begin{aligned} dom\ R &:= \{a, b \in B \mid (a,b) \in R\} \\ &= \{a \in B \mid \text{there is } b \in B \text{ such that } (a,b) \in R \text{ or } (b,a) \in R\} \end{aligned}$$

(iv) *The set of all $\underline{A}$-initial congruences is inductive.*

Basic for the investigations below is the following

Corollary 2.6 *Let $(f_i : \underline{A} \to \underline{B}_i)_{i \in I}$ be a family of homomorphisms, let $(f_i^\sim : \underline{dom f_i^\sim} \to \underline{B}_i)_{i \in I}$ be the family of their closed $\underline{A}$-initial extensions within $\underline{F}(\underline{A}, TAlg(\tau))$, and let $f : \underline{A} \to \prod_{i \in I} \underline{B}_i$ be the induced homomorphism.*

Then, for $f^\sim : \underline{dom f^\sim} \to \prod_{i \in I} \underline{B}_i$ we have

$$\underline{dom f^\sim} = \bigcap_{i \in I} \underline{dom f_i^\sim},$$

$$ker f^\sim = \bigcap_{i \in I} ker f_i^\sim.$$

For some proofs in what follows one will need estimates about the size of partial algebras generated by a given set. These can be derived from the following lemma:

Lemma 2.11 *Let $\underline{A}$ be any partial algebra and M be any generating subset of $\underline{A}$; by $\#M$ denote the cardinality of M. Then, with $\aleph_0$ being the smallest infinite cardinal:*

$$\#A \leq \#F(M, TAlg(\tau)) \leq max\{\#M, \#\Omega, \aleph_0\}.$$

2.6 On the existence of $\mathfrak{K}$-universal ($\mathfrak{K}$-)solutions and a characterization of primitive classes

Definition Let $\underline{A} \in Alg(\tau)$ be any partial algebra, and let $\mathfrak{K} \subseteq Alg(\tau)$ be any class of partial algebras. Then we define in analogy to the case of rings the *$\underline{A}$-characteristic* of $\mathfrak{K}$, $char_{\underline{A}}\mathfrak{K}$ as

$$char_{\underline{A}}\mathfrak{K} := \bigcap\{ker f^{\sim} \mid f : \underline{A} \to \underline{B} \text{ for some } \underline{B} \in \mathfrak{K}\},$$

i.e. as the intersection of all closed $\underline{A}$-initial congruences of $\underline{F}(\underline{A}, TAlg(\tau))$, which are the S-kernels of homomorphisms into $\mathfrak{K}$-algebras starting from $\underline{A}$ (see J.Schmidt [Sch62] and [Sch64] for his corresponding concept for total algebras).

Then we have the following

Theorem 2.7 (Characterization Theorem of $\mathfrak{K}$-universal Solutions) *Let $\underline{A}$ be any partial algebra of type τ and let $\mathfrak{K}$ be any class of partial algebras of type τ.*

(i) *Then*

$$(nat_{char_{\underline{A}}\mathfrak{K}}|_A, (\underline{dom}\, char_{\underline{A}}\mathfrak{K})/char_{\underline{A}}\mathfrak{K})$$

is a $\mathfrak{K}$-universal solution $(r_{\underline{A},\underline{F}(\underline{A},\mathfrak{K})}, \underline{F}(\underline{A}, \mathfrak{K}))$ of $\underline{A}$.

(ii) *$\underline{F}(\underline{A}, \mathfrak{K})$ is isomorphic to a subalgebra of a direct product of $\mathfrak{K}$-algebras, i.e. one always has $\underline{F}(\underline{A}, \mathfrak{K}) \in \mathcal{ISP}\,\mathfrak{K}$.*

(iii) *If $\underline{A} = A$ is discrete, then $\underline{dom}\, char_{\underline{A}}\mathfrak{K}$ is* freely *and $\mathfrak{K}$-*freely *generated by A.*

For the proof of the second statement of the above theorem one may realize, that there is only a *set* of possible $\underline{A}$-initial congruences. Hence one only needs a sub*set*, say $\mathfrak{K}_0$, of $\mathfrak{K}$ in order to compute $char_{\underline{A}}\mathfrak{K}$. Then it is easy to realize, using Corollary 2.6, that $\underline{F}(\underline{A}, \mathfrak{K})$ is isomorphic to the subalgebra of $\prod(\underline{B} \mid \underline{B} \in \mathfrak{K}_0, f : \underline{A} \to \underline{B}) =: \underline{B}_0$ generated by $\bar{f}(A)$, where $\bar{f} : \underline{A} \to \underline{B}_0$ is the unique homomorphism induced by the set (family)

$$\{f \mid f : \underline{A} \to \underline{B} \text{ and} \underline{B} \in \mathfrak{K}_0\} = \bigcup_{\underline{B}\in\mathfrak{K}_0} Hom(\underline{A}, \underline{B})$$

of homomorphisms.

Corollary 2.7 *For any quasiprimitive (i.e. $\mathcal{ISP}$- closed) class $\mathfrak{K}$ of partial algebras of type τ, $\mathfrak{K}$-universal $\mathfrak{K}$-solutions exist for all partial algebras of type τ and can be constructed as in the above theorem.*

If one has only $\mathfrak{K} = \mathcal{ISP}_+(\mathfrak{K})$, then the $\mathfrak{K}$-universal $\mathfrak{K}$-solution of $\underline{A}$ exists, if $\underline{A}$ allows at least one homomorphism into some $\mathfrak{K}$-algebra.

The above corollary implies among others, that quasiprimitive classes $\mathfrak{K} \subseteq Alg(\tau)$ with all homomorphisms are *epireflective subcategories* of the category $\underline{Alg(\tau)}$ of all partial algebras of type τ and homomorphisms as morphisms. In particular, in quasiprimitive classes $\mathfrak{K}$ *initial algebras* (which allow exactly one homomorphism into any other $\mathfrak{K}$-algebra) always exist within this class and are isomorphic to $\underline{F}(\emptyset, \mathfrak{K})$.

For a set M, $\underline{F}(M, \mathfrak{K})$ is the $\mathfrak{K}$-free partial algebra on M, which is "closest" to $\mathfrak{K}$, and it has a further description, using the concept of a fully invariant congruence relation:

Definition Let $\underline{A}$ be a partial algebra, θ any congruence relation on $\underline{A}$. Then θ is called a *fully invariant congruence relation*, if and only if for any endomorphism f of $\underline{A}$ (i.e. for any homomorphism $f : \underline{A} \to \underline{A}$) one has $(f(a), f(b)) \in \theta$ for every $(a, b) \in \theta$ (i.e. $(f \times f)(\theta) \subseteq \theta$).

Theorem 2.8 (i) *If $\underline{A}$ is a free partial algebra freely generated by M, and if θ is a fully invariant congruence relation on $\underline{A}$, then $\underline{A}/\theta$ is freely generated by $M/\theta := \{[m]_\theta \mid m \in M\}$; and if $\underline{A}/\theta$ has at least two elements* [3]*, then $nat_\theta|M : M \to M/\theta$ is bijective. Moreover, M is $\{\underline{A}/\theta\}$-free in $\underline{A}$.*

(ii) *If $f : \underline{A} \to \underline{B}$ is a homomorphism which maps some generating subset M of $\underline{A}$ injectively into some $\{\underline{B}\}$-free subset N of B, then $ker f$ is a fully invariant congruence on $\underline{A}$.*

(iii) *For any class $\mathfrak{K}$ of partial algebras of type τ and for any set X the X-initial congruence relation $char_X\mathfrak{K}$ of $\underline{dom}\, char_X\mathfrak{K}$ is closed and fully invariant, and $\underline{dom}\, char_X\mathfrak{K}$ is freely and $\mathfrak{K}$-freely generated by X.*

Corollary 2.8 *Let $\underline{A}$ be a free partial algebra of type τ and M a free generating subset of $\underline{A}$. Then the mapping*

$$m \mapsto [m]_{char_M\{\underline{A}\}}$$

induces an isomorphism between $\underline{A}$ and $(\underline{dom}\, char_M\{\underline{A}\})/char_M\{\underline{A}\}$.

Thus, up to isomorphism the free partial algebras $\underline{A}$ ($\{\underline{A}\}$-)freely generated by M are in one-to-one correspondence to the closed and fully invariant congruence relations in relative subalgebras of $\underline{F}(M, TAlg(\tau))$, which are freely generated by M.

Recall that independence classes are primitive, and that primitive classes are quasiprimitive:

Lemma 2.12 *If $\mathfrak{K}$ is any class, and if M is any at least countably infinite set, then*

$$ind_{\underline{F}(M,\mathfrak{K})}M = \mathcal{HSP}\mathfrak{K} \cup \{\underline{\emptyset}\} = \mathcal{HSP}o\mathfrak{K} = \mathcal{HSP}o\{\underline{F}(M, \mathfrak{K})\}.$$

This yields the following characterization theorem for primitive classes of partial algebras.

Theorem 2.9 *For a class $\mathfrak{K}$ of partial algebras of type τ and for any at least countably infinite set M the following statements are equivalent:*

(i) $\mathfrak{K} = \mathcal{HSP}\mathfrak{K} \cup \{\underline{\emptyset}\}$,

(ii) $\mathfrak{K} = ind_{\underline{F}(M,\mathfrak{K})}M$.

[3] In the many-sorted case one has to require that every phylum of $\underline{A}/\Theta$ has at least two elements (or at least as many elements as the corresponding phylum of M).

Moreover, the following statements are equivalent:

(a) $\mathfrak{K} = \mathcal{HSP}\mathfrak{K}$,

(b) $\mathfrak{K} = ind_{\underline{F}(M,\mathfrak{K})}M \cap ind_{\underline{F}(\emptyset,\mathfrak{K})}\emptyset$.

Lemma 2.13 *Let M and N be any sets with $N \subseteq M$, and let $\mathfrak{K}$ be any class of partial algebras. Then*

$$char_N\mathfrak{K} \subseteq char_M\mathfrak{K},$$

and except for $N = \emptyset$, where in addition $char_N\mathfrak{K} = \emptyset$ is possible, one even has[4]

$$char_N\mathfrak{K} = char_M\mathfrak{K} \cap (C_{\underline{F}(M,TAlg(\tau))}N)^2.$$

Observing the description of $\underline{F}(M;\mathfrak{K})$ as quotient algebra $(\underline{dom}\, char_M\mathfrak{K})/char_M\mathfrak{K}$ and the above results one may realize that, for any (at least) countably infinite set M one has:

Theorem 2.10 (Characterization Theorem for Primitive Classes) *Let M be any at least countably infinite set. Then each primitive class $\mathfrak{K}$ is characterized by a pair (θ_M, θ_0), where θ_0 and θ_M are closed and fully invariant congruence relations on their respective domains, which are relative subalgebras of $\underline{F}(M, TAlg(\tau))$ freely generated by $\emptyset$ and M, respectively, such that $\theta_0 = \emptyset$ or $\theta_0 = \theta_M \cap F(\emptyset, TAlg(\tau))^2$, and*

$$\mathfrak{K} = ind_{(\underline{dom}\,\theta_0)/\theta_0}\emptyset \cap ind_{(\underline{dom}\,\theta_M)/\theta_M}M,$$

and each such pair $\theta_0 \subseteq \theta_M$ characterizes a primitive class in this way.

Moreover, the pair (θ_M, θ_0) characterizing the primitive class $\mathfrak{K}$ in such a way is uniquely determined by $\mathfrak{K}$.

The reader knowing the equational theory of total algebras may realize the close relationship of the above results with the Birkhoff-Theorem and the characterization of equational theories for total algebras. The only difference here is that $char_M\mathfrak{K}$ (M infinite) only characterizes primitive classes containing $\underline{\emptyset}$, while for the characterization of arbitrary primitive classes $\mathfrak{K}$ one also needs $char_\emptyset\mathfrak{K}$.
We include the discussion concerning the empty partial algebra, since it is already some preparation for the case of many-sorted partial — and total — algebras, when empty phyla are allowed, as it is often done.

We now have similar to the total case (see [Mar58]) the following

Theorem 2.11 (Characterization Theorem for Free Partial Algebras) *Let $\underline{A}$ be a partial algebra generated by a set M such that the carrier set A has at least two elements. Then the following statements are equivalent:*

(i) *$\underline{A}$ is $\{\underline{A}\}$-freely generated by M.*

(ii) *$\underline{A} = \underline{F}(M, \mathfrak{K})$ for some non-trivial primitive class $\mathfrak{K}$ of partial algebras.*

[4] In the many-sorted case the situation is more complicated (cf. the footnote of Theorem 3.1).

(iii) $\underline{A}$ *is isomorphic to the partial algebra* $\underline{F}^M(\underline{A})$ *of all total term operations of arity* M *on* $\underline{A}$. *This isomorphism is induced by* $e_m^M \mapsto m (m \in M)$, *where* $e_m^M(a) := a(m)$ *for all* $a \in A^M$, *i.e.* e_m^M *is the m-th projection from* A^M *into* A.

(iv) *For every* $a \in A$ *there exists exactly one term* $t_a \in F(M, TAlg(\tau))$ *such that* $a = t_a^{\underline{A}}(id_M)$. *And for every term* $t \in F(M, TAlg(\tau))$ *the fact that* id_M *belongs to dom* $t^{\underline{A}}$ *implies that* $t^{\underline{A}}$ *is total.*

(v) *There exists a free* total *algebra* $\underline{B}$, *freely generated by* M, *such that* $\underline{A}$ *is an* M*-generated relative subalgebra of* $\underline{B}$, *and such that for every term* $t \in F(M, TAlg(\tau))$ *the fact that* $t^{\underline{B}}(id_M)$ *exists and belongs to* A *implies that the restriction of* $t^{\underline{B}}$ *to* A^M *is a total term operation on* A*:*

$$t^{\underline{B}}|_{A^M} = t^{\underline{A}} \in F^M(\underline{A}).$$

(vi) *Let* $\underline{F} = \underline{F}(M, TAlg(\tau))$. *Consider* $\beta : M \to A$, $\beta(m) := m$ *for* $m \in M$, *as a mapping out of* $\underline{F}$ *into* $\underline{A}$, *and let* $\beta^\sim := C_{\underline{F}\times\underline{A}}\beta$ *be the subalgebra of* $\underline{F} \times \underline{A}$ *generated by the graph of* β. *Then* $\beta^\sim$ *is the graph of a closed and surjective homomorphism* $\beta^\sim : \underline{F} \supseteq \underline{dom}\ \beta^\sim \to \underline{A}$, $ker\beta^\sim$ *is a (closed and) fully invariant congruence relation, and* $\underline{dom}\ \beta^\sim$ *is freely generated by* M.

3 Some model theoretic aspects of partial algebras

3.1 Existence equations and their theory

Since free partial algebras $\underline{F}$ with some basis M and in particular those generated by the empty set ($M = \emptyset$ yields the so-called initial partial algebras) are used respectively needed in computer science for the specification of data types, and since they correspond to primitive classes ($\mathfrak{K} = ind_{\underline{F}}M$), which are characterized by (pairs of) special sets of pairs of terms, one may take this as motivation of a corresponding special "equational theory" and for a first order logic based on it. As a matter of fact, the above characterization of free partial algebras and primitive classes has been the historic origin of what was later called existence equations.

Definition Let X be a set. In what follows the elements of X are called *variables.* A pair (t, t') of terms t and t' of $F(X, TAlg(\tau))$ is called *existence equation* (briefly *E-equation*) and written as $t \stackrel{e}{=} t'$ (or even as $t \stackrel{e}{=}_X t'$, if it has to be distinguished from $t \stackrel{e}{=}_\emptyset t'$, when already $t, t' \in F(\emptyset, TAlg(\tau))$). Moreover, we write Eeq_X for the set of *all* E-equations with variables in the set X. For any partial algebra $\underline{A}$ of type τ a mapping $v : X \to \underline{A}$ will be called a *valuation* (of X in $\underline{A}$) (an *assignment*) and $v^\sim : \underline{dom}\ v^\sim \to \underline{A}$ the *interpretation of terms induced by* v, i.e. a term $t \in F(X, TAlg(\tau))$ will be said to be *interpreted by* v, if and only if $t \in dom\ v^\sim$.

The denotation "existence equations" will become obvious from the following definition of their semantics.

Definition Let X be a set of variables, $\underline{A}$ a partial algebra of type τ, $v : X \to \underline{A}$ a valuation, $t \stackrel{e}{=} t'$ an E-equation in Eeq_X. We say that $\underline{A}$ *satisfies the E-equation* $t \stackrel{e}{=} t'$ *w.r.t. the valuation* v (in symbols: $\underline{A} \models t \stackrel{e}{=} t'[v]$), if and only if $(t, t') \in ker v^{\sim}$, i.e. if and only if

- the interpretation $v^{\sim}(t)$ of t by v *exists*, and
- the interpretation $v^{\sim}(t')$ of t' by v *exists*, and
- these interpretations are equal: $v^{\sim}(t) = v^{\sim}(t')$.

We say that the E-equation $t \stackrel{e}{=} t'$ *holds* (*is valid*) in the partial algebra $\underline{A}$ (in symbols: $\underline{A} \models t \stackrel{e}{=} t'$), if and only if $\underline{A}$ satisfies $t \stackrel{e}{=} t'$ *for every valuation* $v : X \to \underline{A}$, i.e.:

$\underline{A} \models t \stackrel{e}{=} t'$ if and only if $\underline{A} \models t \stackrel{e}{=} t'[v]$ for every valuation $v : X \to \underline{A}$.

Remarks and further notation

(i) The above semantics shows that $\underline{A} \models t \stackrel{e}{=} t'$ if and only if $(t, t') \in char_X\{\underline{A}\}$.

(ii) If t and t' are identical terms, then $t \stackrel{e}{=} t$ will be called a *term existence statement* (briefly *TE-statement*), since we have for $v : X \to \underline{A}$: $\underline{A} \models t \stackrel{e}{=} t[v]$ if and only if the interpretation of t w.r.t. v *exists*, if and only if $v \in dom\, t^{\underline{A}}$ for the term operation induced by t. This shows that in our semantics of E-equations the "diagonal" of $F(X, TAlg(\tau))^2$ gains importance, while for total algebras $\underline{A}$ the statement $\underline{A} \models t \stackrel{e}{=} t$ is trivially true. Thus this gives additional expressive power to our E-equations.

(iii) If an existence equation $t \stackrel{e}{=} t'$ is valid in a partial algebra $\underline{A}$, then this is equivalent to the fact that t and t' induce on $\underline{A}$ *total and identical* term operations:
$\underline{A} \models t \stackrel{e}{=} t'$, if and only if $t^{\underline{A}} = t'^{\underline{A}}$, and $t^{\underline{A}}$ and $t'^{\underline{A}}$ are total term operations on $\underline{A}$.

The fact that validity of an E-equation implies totalness of the induced term operations has long prevented the acceptance of existence equality as a suitable concept of equality for partial algebras — besides the fact that existence equality does only satisfy a very restricted kind of reflexivity, as we shall see below. However, the concept of existence equations gives rise to a new kind of quasi-equations which we will call "existentionally conditioned existence equations", in which the premise only consists of a conjunction of TE-statements, while the conclusion is just one E-equation. In connection with total algebras these ECE-equations just become equations, thus showing that the concept of E-equations opens the possibility of even more expressive power of the object language in the generalization of equations from total to partial algebras. We shall discuss this in some more detail in the next subsection. Let us here just reinterpret our results from section 2.6.

In what follows, let $\tau : \Omega \to N_0$ be an arbitrary similarity type and all partial algebras under consideration are assumed to be of that type. Moreover, let X be any countably infinite set of variables (with $X \cap \Omega = \emptyset$).

Definition For any sets $G \subseteq F(X, TAlg(\tau))^2$ and $G_0 \subseteq F(\emptyset, TAlg(\tau))^2$ as well as for any class $\mathfrak{K}$ of partial algebras define $\mathfrak{K} \models (G, G_0)$ if and only if for every $\underline{K} \in \mathfrak{K}$, for every $(t, t') \in G$ and for every $(t_0, t_0') \in G_0$ one has $\underline{K} \models t \stackrel{e}{=} t'$ and $\underline{K} \models t_0 \stackrel{e}{=}_{\emptyset} t_0'$, where $\underline{K} \models t_0 \stackrel{e}{=}_{\emptyset} t_0'$ means that for the (only) valuation $v_0 : \emptyset \to K$ of the empty set of variables one has $\underline{K} \models t_0 \stackrel{e}{=} t_0'[v_0]$ — we shall sometimes indicate this by writing $\underline{K} \models_{\emptyset} t_0 \stackrel{e}{=} t_0'[v_0]$.

If one of the sets or classes involved has only one element, then we will omit the "set brackets".

Finally we introduce the operators

$$\begin{aligned}
Eeq_X\,\mathfrak{K} &:= \{(t,t') \in F(X,TAlg(\tau))^2 \mid \mathfrak{K} \models t \stackrel{e}{=} t'\},\\
Eeq_\emptyset\,\mathfrak{K} &:= \{(t_0,t_0') \in F(\emptyset,TAlg(\tau))^2 \mid \mathfrak{K} \models t_0 \stackrel{e}{=}_\emptyset t_0'\},\\
Eeq\,\mathfrak{K} &:= (Eeq_X\,\mathfrak{K},\ Eeq_\emptyset\,\mathfrak{K}),\\
Mod\,G &:= \{\underline{A} \in PAlg(\tau) \mid \underline{A} \models G\},\\
Mod\,G_0 &:= \{\underline{A} \in PAlg(\tau) \mid \underline{A} \models_\emptyset G_0\}, \text{ and}\\
Mod\,(G,G_0) &:= \{\underline{A} \in PAlg(\tau) \mid \underline{A} \models t \stackrel{e}{=} t' \text{ for all } (t,t') \in G \text{ and}\\
&\qquad\qquad \underline{A} \models t_0 \stackrel{e}{=}_\emptyset t_0' \text{ for all } (t_0,t_0') \in G_0\}.
\end{aligned}$$

Observe that

$$\begin{aligned}
Eeq_X\,\mathfrak{K} &= char_X\,\mathfrak{K},\\
Eeq_\emptyset\,\mathfrak{K} &= char_\emptyset\,\mathfrak{K},\\
Eeq\,\mathfrak{K} &= (char_X\,\mathfrak{K}, char_\emptyset\,\mathfrak{K}), \text{ and}\\
Mod\,(G,G_0) &= Mod\,G \cap Mod\,G_0.
\end{aligned}$$

Moreover, Theorem 2.10 can now be reformulated as

Theorem 3.1 (i) *Let $\mathfrak{K}$ be an arbitrary class of partial algebras. Then*

$$\begin{aligned}
Mod\,Eeq_X\,\mathfrak{K} &= \mathcal{HSP}\,\mathfrak{K} \cup \{\underline{\emptyset}\},\\
Mod\,Eeq\,\mathfrak{K} &= \mathcal{HSP}\,\mathfrak{K}.
\end{aligned}$$

(ii) *For any set $M \in \{\emptyset, X\}$ and for any set $G \subseteq Eeq_M$ the following statements are equivalent:*

(A) $G = Eeq_M\,Mod\,G$.

(B) *G is a closed and fully invariant congruence relation on the relative subalgebra $\underline{P_G} := \underline{dom}\,G$ of $\underline{F}(M,TAlg(\tau))$ such that $\underline{P_G}$ is freely generated by X.*

(iii) *For any sets $G \subseteq Eeq_X$ and $G_0 \subseteq Eeq_\emptyset$ the following statements are equivalent:*

(a) $(G,G_0) = Eeq\,Mod\,(G,G_0)$.

(b) *$G = Eeq_X\,Mod\,G$, $G_0 = Eeq_\emptyset\,Mod\,G_0$, and $G_0 = \emptyset$ or $G_0 = G \cap F(\emptyset,TAlg(\tau))^2$ (if one considers $\underline{F}(\emptyset,TAlg(\tau))$ as a subalgebra of $\underline{F}(X,TAlg(\tau))$)*[5].

[5] In the many-sorted case one has to be more specific (for the concepts and notation see the appendix): Let $v_S : \underline{F}(X,TAlg(\tau) \downarrow \{\underline{S}\}) \to$ be the sort homomorphism, and assume that for each $s \in S$ $v_S{}^{-1}(\{s\})$ is countably infinite. Moreover, for any subset $U \subseteq S$, let $X_U := v_S{}^{-1}(U)$, $\underline{F}_U := \underline{F}(X_U, TAlg(\tau) \downarrow \{\underline{S}\})$ as a subalgebra of $\underline{F}_S$, and $v_U : \underline{F}_U \longrightarrow \underline{S}$ the corresponding sort homomorphism. Then one has to consider as possible E-equational theories families $(G_U)_{U \subseteq S}$ with $G_U \subseteq F_U{}^2$, and such a family is an E-equational theory (i.e. for each $U \subseteq S$ one has $G_U = Eeq_{X_U}(\bigcap_{W \subseteq S} Mod\,G_W))$ if and only if, for each $U, W \subseteq S$,

- G_U is a closed and fully invariant congruence relation on a relative subalgebra of $\underline{F}_U$ freely generated by X_U, and
- for $U \subseteq W$ one has $G_U \subseteq G_W$, and in case of $v_W(X_W) \subseteq v_U(F_U)$ one has $G_U = G_W \cap F_U{}^2$.

Thus the primitive classes are exactly the existence equationally definable classes. However, if all existence equations are related to the infinite set X of variables, then always $\underline{\emptyset}$ has to belong in addition to the existence equationally definable (i.e. primitive) classes providing in them the so-called initial object. Since in theoretical computer science one wants in general to specify a non-empty initial object, one has either to choose the a little bit more complex description of primitive classes by pairs of sets of existence equations or one has to forbid the empty partial algebra as a model (what may only cause trouble in general, if in a partial algebra no constant exist, since then the empty set will generate no subalgebra).

3.2 About the first order language for partial algebras based on existence equations

Based on existence equations as atomic formulas one can now build the syntax of a usual first order language.

Definition Let X be any set of variables, $\tau : \Omega \to \mathbb{N}_0$ a given similarity type such that $X \cap \Omega = \emptyset$. A first order language $\mathcal{L}(X, \tau)$ is now defined in the usual way:

(F_X1) $t \stackrel{e}{=} t'$ is an atomic formula and hence a formula of $\mathcal{L}(X, \tau)$ for any $t, t' \in F(X, TAlg(\tau))$.

(F_X2) If F and F' are formulas of $\mathcal{L}(X, \tau)$ then $\neg F$, $(F \wedge F')$, $(F \vee F')$, $(F \Rightarrow F')$, and $(F \Leftrightarrow F')$ are formulas of $\mathcal{L}(X, \tau)$.

(F_X3) If F is a formula of $\mathcal{L}(X, \tau)$, and if $x \in X$ is a variable, then $(\forall x)F$ and $(\exists x)F$ are formulas of $\mathcal{L}(X, \tau)$.

Before we can also define $\mathcal{L}(\emptyset, \tau)$, let us define the functions

$$fvar : \mathcal{L}(X, \tau) \to \mathfrak{P}(X)$$

assigning to each formula the set of its *free variables*, and

$$var : F(X, TAlg(\tau)) \to \mathfrak{P}(X),$$

assigning to each term the set of variables on which it really depends:

- $var(x) := \{x\}$ for every variable $x \in X$,
- $var(\omega t_1 \dots t_{\tau(\omega)}) := \bigcup_{i=1}^{\tau(\omega)} var(t_i)$ for all $\omega \in \Omega$ and $t_1, \dots, t_{\tau(\omega)} \in F(X, TAlg(\tau))$, for which var has already been defined, e.g. $var(\omega) := \emptyset$, if $\tau(\omega) = 0$.

Once we know the function var, we have

- $fvar(t_1 \stackrel{e}{=} t_2) := var(t_1) \cup var(t_2)$ for all $t_1, t_2 \in F(X, TAlg(\tau))$.

– If $fvar$ is defined for F and F' of $\mathcal{L}(X,\tau)$, then we have

$$fvar(\neg F) := fvar(F),$$
$$fvar((f \wedge F')) := fvar((F \vee F')) :=$$
$$fvar((F \Rightarrow F')) := fvar((F \Leftrightarrow F')) := fvar(F) \cup fvar(F').$$

– If $fvar$ is defined for F, and if $x \in X$, then

$$fvar((\forall x)F) := fvar((\exists x)F)) := fvar(F) \setminus \{x\}.$$

In the same way one can define the set $ovar(F)$ of variables occurring in F with the only difference that

$$ovar((\forall x)F) := ovar((\exists x)F) := ovar(F) \cup \{x\}.$$

We then have $\mathcal{L}(\emptyset,\tau)$ defined by the rules (F_X1) and (F_X2) for $X = \emptyset$ and

$(F_\emptyset 3^*)$ If $F \in \mathcal{L}(X,\tau)$ and $fvar(F) = \emptyset$, i.e. if F is a *sentence* of $\mathcal{L}(X,\tau)$ then $F_\emptyset \in \mathcal{L}(\emptyset,\tau)$, where $F_\emptyset$ is the same formula as F, where only the index indicates that it is now considered as a formula of $\mathcal{L}(\emptyset,\tau)$.

We set $\mathcal{L}(\tau) := \mathcal{L}(X,\tau) \cup \mathcal{L}(\emptyset,\tau)$. For $M \in \{\emptyset, X\}$ we define the semantics of existence equations as before, and for $F, F' \in \mathcal{L}(M,\tau)$, $\underline{A}$ a partial algebra of type τ and $v : M \to A$ we define:

$\underline{A} \models (F \wedge F')[v]$	iff	$\underline{A} \models F[v]$ *and* $\underline{A} \models F'[v]$,
$\underline{A} \models (F \vee F')[v]$	iff	$\underline{A} \models F[v]$ *or* $\underline{A} \models F'[v]$,
$\underline{A} \models (F \Rightarrow F')[v]$	iff	(*if* $\underline{A} \models F[v]$ *then* $\underline{A} \models F'[v]$),
$\underline{A} \models (F \Leftrightarrow F')[v]$	iff	$\underline{A} \models ((F[v] \Rightarrow F') \wedge (F' \Rightarrow F))[v]$,
$\underline{A} \models (\neg F)[v]$	iff	it is *not* true that $\underline{A} \models F[v]$.

If x is any variable, and if $F \in \mathcal{L}(M,\tau)$, then

$\underline{A} \models (\forall x)F[v]$ iff *for all* $a \in A$ and *for all* valuations $v^x_a : M \cup \{x\} \to A$ with

$$v^x_a(y) := \begin{cases} v(y), & \text{if } y \in M \setminus \{x\} \\ a, & \text{if } y = x \end{cases}$$

one has $\underline{A} \models F[v^x_a]$,

$\underline{A} \models (\exists x)F[v]$ iff there *exists* an element $a \in A$ and a valuation $v^x_a : M \cup \{x\} \to A$ (like above) such that $\underline{A} \models F[v^x_a]$.

Finally, we say that a formula $F \in \mathcal{L}(M,\tau)$ ***holds*** in a partial algebra, if and only if $\underline{A} \models F[v]$ *for all* valuations $v : M \to \underline{A}$.

A complete and correct system of rules w.r.t. satisfaction of formulas from $\mathcal{L}(X,\tau)$ in partial algebras has been given in [B82]. Since in what follows we shall restrict our considerations to (mainly positive) universal Horn formulas, we do not go here into more details about the general language. Let us only observe again what we already have mentioned earlier: If we do not forbid the empty partial algebra, the first order language only gains its full expressive power, if it also refers to valuations of the empty set of variables. Moreover, the category theoretical translation below of elementary implications also supports the above approach.

3.3 Elementary implications and their translation into epimorphisms

Definition For a given similarity type τ we define an *elementary implication* ι to be a formula (we omit brackets in the usual way)

$$\iota \equiv (\bigwedge_{i\in I} t_i \overset{e}{=} t'_i \Rightarrow \bigwedge_{j\in J} t^*_j \overset{e}{=} t^{*\prime}_j).$$

"$\bigwedge_{i\in I} t_i \overset{e}{=} t'_i$" stands for the formation of an "arbitrarily long" conjunction in generalization of "$\wedge$". Observe that we have

$$\underline{A} \models \iota \text{ iff } \{(t_i, t'_i) \mid i \in I\} \subseteq char_{fvar(\iota)}\{\underline{A}\} \text{ implies } \{(t^*_j, t^{*\prime}_j) \mid j \in J\} \subseteq char_{fvar(\iota)}\{\underline{A}\}).$$

Particular elementary implications are

- *ECE-equations* (i.e. *existentially conditioned existence equations*)

$$\bigwedge_{i\in I} t_i \overset{e}{=} t_i \Rightarrow t \overset{e}{=} t',$$

- *QE-equations* (i.e. *quasi existence equations*)

$$\bigwedge_{i\in I} t_i \overset{e}{=} t'_i \Rightarrow t \overset{e}{=} t',$$

where the conclusion consists of one existence equation only.

In principle in what follows I and J may be arbitrary sets, while w.r.t. $\mathcal{L}(\tau)$ they have to be finite.

Special elementary implications occur in connection with three further equational concepts, two of which are also frequently used as axioms for the description of (classes of) partial algebras, while the third one has been used by T.Evans in [Ev51] as mentioned in the "Motivation":

- *Weak equations* $t \overset{w}{=} t'$ (for $t, t' \in F(X, TAlg(\tau))$) are in our approach special ECE-equations[6]:

$$t \overset{w}{=} t' := (t \overset{e}{=} t \wedge t' \overset{e}{=} t' \Rightarrow t \overset{e}{=} t').$$

- *Strong equations* (or *Kleene equations*) $t \overset{s}{=} t'$ (for $t, t' \in F(X, TAlg(\tau))$) are conjunctions of special ECE-equations:

$$t \overset{s}{=} t' := ((t \overset{e}{=} t \Rightarrow t \overset{e}{=} t') \wedge (t' \overset{e}{=} t' \Rightarrow t \overset{e}{=} t')).$$

[6]We want to mention in this connection that P.Kosiuczenko — see [Kos90] — has recently used a combination of E-equations and weak equations in order to characterize axiomatic classes of partial algebras, in which each partial algebra has a "permutable" respectively "distributive lattice of *closed* congruence relations.

- *Evans equations* $t \overset{E}{=} t'$ are even more complicated:

$$t \overset{E}{=} t' := (((t \overset{e}{=} t \wedge \bigwedge_{t'' \in \downarrow t', t'' \neq t'} t'' \overset{e}{=} t'') \Rightarrow t \overset{e}{=} t') \wedge$$

$$((t' \overset{e}{=} t' \wedge \bigwedge_{t'' \in \downarrow t, t'' \neq t} t'' \overset{e}{=} t'') \Rightarrow t \overset{e}{=} t')).$$

Their very special implicational form makes it understandable that their theory is not so easily describable as the one of the existence equations (see e.g. S.C.Kleene [Kl52], R.Kerkhoff [Ke70], H.Höft [Ho70] and [Ho73], R.John [J75] and [J78], L.Rudak [Ru83], and W.Craig [Cr89]).

For ECE-equations and QE-equations in general we still have "nice" Birkhoff type theorems, when we denote for an arbitrary class $\mathfrak{K}$ of partial algebras of type τ and for a set $M \in \{\emptyset, X\}$ of variables by

- $ECEeq_M(\mathfrak{K})$ the set of all ECE-equations in $\mathcal{L}(M, \tau)$, which are valid in all $\mathfrak{K}$-algebras.
- $QEeq_M(\mathfrak{K})$ the set of all QE-equations in $\mathcal{L}(M, \tau)$, which are valid in all $\mathfrak{K}$-algebras.

Theorem 3.2 *Let $\mathfrak{K}$ be any class of partial algebras of type τ. In each instance below the statements* (a) *and* (b) *are equivalent:*

(i) (a) $\mathfrak{K} = Mod\, ECEeq_X(\mathfrak{K})$,
(b) $\mathfrak{K} = \mathcal{H}_c\mathcal{S}\mathcal{P}_r(\mathfrak{K}) \cup \{\underline{\emptyset}\}$.

(ii) (a) $\mathfrak{K} = Mod\, (ECEeq_X(\mathfrak{K}) \cup ECEeq_\emptyset(\mathfrak{K}))$,
(b) $\mathfrak{K} = \mathcal{H}_c\mathcal{S}\mathcal{P}_r(\mathfrak{K})$.

(iii) (a) $\mathfrak{K} = Mod\, QEeq_X(\mathfrak{K})$,
(b) $\mathfrak{K} = \mathcal{I}\mathcal{S}\mathcal{P}_r(\mathfrak{K}) \cup \{\underline{\emptyset}\}$.

(iv) (a) $\mathfrak{K} = Mod\, (QEeq_X(\mathfrak{K}) \cup QEeq_\emptyset(\mathfrak{K}))$,
(b) $\mathfrak{K} = \mathcal{I}\mathcal{S}\mathcal{P}_r(\mathfrak{K})$.

As an **example** for ECE-varieties let us list the axioms for the class of all small categories considered as homogeneous partial algebras of type

$$(\Omega, \tau) = (\{D, C, \circ\}, \{(D,1), (C,1), (\circ, 2)\})$$

satisfying the axioms (cf. subsection 1.1 for the properties (C 1) through (C 4) formulated there).

1. $\circ x D x \overset{e}{=} x$ (this implies $Dx \overset{e}{=} Dx$, i.e. D has to be total),
2. $\circ C x x \overset{e}{=} x$ (this implies $Cx \overset{e}{=} Cx$, i.e. C has to be total),

3. $oyx \stackrel{e}{=} oyx \Rightarrow Cx \stackrel{e}{=} Dy \wedge C \circ yx \stackrel{e}{=} Cy \wedge D \circ yx \stackrel{e}{=} Dx$,
4. $oyx \stackrel{e}{=} oyx \wedge ozy \stackrel{e}{=} ozy \Rightarrow oz \circ yx \stackrel{e}{=} \circ \circ zyx$.

One might have expected in addition the QE-equation

5. $Dy \stackrel{e}{=} Cx \Rightarrow oyx \stackrel{e}{=} oyx$,

but this can be proved from the other axioms, showing that the class of all small categories is really an ECE-variety. We briefly sketch the proof of 5 :

Let $\underline{K}$ be any small category, $f, g \in K$ such that $D^{\underline{K}}g = C^{\underline{K}}f$. By axioms 1 and 2 we have the existence of $g \circ^{\underline{K}} D^{\underline{K}}g (= g)$ and $C^{\underline{K}}f \circ^{\underline{K}} f (= f)$. Because of the assumption $D^{\underline{K}}g = C^{\underline{K}}f$ the premise of axiom 4 is satisfied: $g \circ^{\underline{K}} C^{\underline{K}}f$ and $C^{\underline{K}}f \circ^{\underline{K}} f$ exist, and therefore $(g \circ^{\underline{K}} C^{\underline{K}}f) \circ^{\underline{K}} f$ and $g \circ^{\underline{K}} (C^{\underline{K}}f \circ^{\underline{K}} f)$ exist and are equal. But $C^{\underline{K}}f \circ^{\underline{K}} f = f$, i.e. $g \circ^{\underline{K}} f$ exists, and this was to be proved. □

Another **example** may be the specification of an interval $\mathbf{Z}_{lk} := [-l, k]$ of integers $(l, k \in \mathbf{N})$. We choose the similarity type

$$(\Omega, \tau) = (\{0, s, p\}, \{(0,0), (s,1), (p,1)\}).$$

Observe that the algebra $(\mathbf{Z}; 0^{\underline{\mathbf{Z}}}, s^{\underline{\mathbf{Z}}}, p^{\underline{\mathbf{Z}}})$ of all integers can be specified as the initial object $\underline{F}(\emptyset, \mathfrak{K})$ of the model class $\mathfrak{K}$ of the axioms

$$0 \stackrel{e}{=} 0, \; psx \stackrel{e}{=} x, \; spx \stackrel{e}{=} x.$$

$\underline{\mathbf{Z}}_{lk}$ can be specified as the initial object $\underline{F}(\emptyset, \mathfrak{K}_{lk})$ of the model class $\mathfrak{K}_{lk}$ of the axioms ($s^n(x)$ stands as abbreviation for $\underbrace{s \ldots s}_{n \text{ times}} x$) :

$$s^k(0) \stackrel{e}{=} s^k(0),$$
$$p^l(0) \stackrel{e}{=} p^l(0),$$
$$sx \stackrel{e}{=} sx \Rightarrow psx \stackrel{e}{=} x,$$
$$px \stackrel{e}{=} px \Rightarrow spx \stackrel{e}{=} x.$$

The reader is asked (as an exercise) to find a specification of $\underline{\mathbf{Z}}_{lk}$ as $\underline{F}(\emptyset, \mathfrak{K}')$ of an E-variety $\mathfrak{K}'$. He will realize that the above ECE-equational one is much simpler. Recall the implementation of integers by TURBO PASCAL already mentioned in the "Motivation", where an interval $[-32768, 32767]$ of $\mathbf{Z}$ is implemented in such a way that $s(32767) = -32768$, while the value is not defined in the (initial) data type belonging to our specification.

Elementary implications are of special interest, since the classes defined by them have still a fairly simple description as seen above and they have free algebras and in more generality universal solutions of any partial algebra of the same type.

Moreover, they allow a relatively simple translation into a category theoretical language[7]:

[7] See however [AN79], where such a translation has been carried through for all first order formulas.

Definition Let again

$$\iota \equiv (\bigwedge_{i\in I} t_i \stackrel{e}{=} t_i' \Rightarrow \bigwedge_{j\in J} t_j^* \stackrel{e}{=} t_j^{*\prime}).$$

be an elementary implication. Then we assign to ι a homomorphism $e_\iota : P_\iota \to C_\iota$ — which is indeed an epimorphism — as follows: Let

$$P_0 := var(\iota) \cup \downarrow \{t_i, t_i' \mid i \in I\},$$

where for any set T of terms in $F(X, TAlg(\tau))$, $\downarrow T$ designates the set of all subterms of terms occurring in T, and

$$C_0 := P_0 \cup \downarrow \{t_j^*, (t_j')^* \mid j \in J\}.$$

Moreover, let $\underline{P}_0$ and $\underline{C}_0$ be the relative subalgebras of $\underline{F}(X, TAlg(\tau))$ with carrier sets P_0 and C_0, respectively. Let

$$\begin{aligned} \theta_{\underline{P}_0} &:= Con_{\underline{P}_0}\{(t_i, t_i') \mid i \in I\}, \\ \underline{P}_\iota &:= \underline{P}_0/\theta_{\underline{P}_0}, \\ \theta_{\underline{C}_0} &:= Con_{\underline{C}_0}(\{(t_i, t_i') \mid i \in I\} \cup \{(t_j^*, (t_j')^*) \mid j \in J\}), \text{ and} \\ \underline{C}_\iota &:= \underline{C}_0/\theta_{\underline{C}_0}. \end{aligned}$$

Then $e_\iota : \underline{P}_\iota \to \underline{C}_\iota$ is the homomorphism induced by the inclusion mapping from $\underline{P}_0$ into $\underline{C}_0$ (see Figure 11).

$$\begin{tikzcd} \underline{var}(\iota)_{disc.} \arrow[r, tail, "id_{var(\iota)P_0}"] & \underline{P}_0 \arrow[r, tail, "id_{P_0C_0}"] \arrow[d, "nat_{\theta_{\underline{P}_0}}"'] & \underline{C}_0 \arrow[d, "nat_{\theta_{\underline{C}_0}}"] \\ & \underline{P}_\iota := \underline{P}_0/\theta_{\underline{P}_0} \arrow[r, "e_\iota \in Epi"'] & \underline{C}_\iota := \underline{C}_0/\theta_{\underline{C}_0} \end{tikzcd}$$

Figure 11: e_ι is the epimorphism encoding the elementary implication ι.

Lemma 3.1 (i) *For a given elementary implication ι the encoding homomorphism $e_\iota : \underline{P}_\iota \to \underline{C}_\iota$ is an epimorphism. Moreover, ι holds in a partial algebra $\underline{A}$ iff for every homomorphism $f : \underline{P}_\iota \to \underline{A}$ there exists a (unique) homomorphism $g : \underline{C}_\iota \to \underline{A}$ such that $g \circ e_\iota = f$, i.e. iff e_ι is an $\{\underline{A}\}$-extendable epimorphism, iff $\underline{A}$ is injective w.r.t. e_ι.*

(ii) *Every epimorphism $e : \underline{P} \to \underline{C}$ encodes a — possibly infinitary — implication; namely, if X is a generating subset of $\underline{P}$ and $\beta : X \to P$ the inclusion mapping, R a generating subset of ker $\beta^\sim$, and S a generating subset of ker $(e \circ \beta)^\sim$, then*

$$\iota_e := (\bigwedge_{(t,t')\in R} t \stackrel{e}{=} t' \Rightarrow \bigwedge_{(s,s')\in S} s \stackrel{e}{=} s')$$

is an elementary implication encoded by e. If the sets X, R and S can be chosen to be finite, then $\iota_e \in \mathcal{L}(\tau)$.

Definition If $\mathcal{E}$ is a class of epimorphisms, then we define

$$Inj(\mathcal{E}) := \{\underline{A} \in Alg(\tau) \mid \text{ each } e \in \mathcal{E} \text{ is } \{\underline{A}\}\text{-extendable }\}.$$

Thus $Inj(\mathcal{E}) = Mod\{\iota_e \mid e \in \mathcal{E}\}$, when we extend the concept of models also to infinitary elementary implications.

Observe that existence equations $\iota \equiv t \stackrel{e}{=} t'$ are special kinds of elementary implications, where the premise is empty; however for the encoding epimorphism $\iota_e : \underline{P}_\iota \to \underline{C}_\iota$ one has $\underline{P}_\iota := fvar(\iota)$, which is a discrete partial algebra. $\underline{P}_\iota$ only allows a homomorphism into the empty partial algebra, if it is empty itself; else ι trivially holds in $\underline{\emptyset}$ and $\underline{\emptyset} \in Inj(\{e_\iota\})$ is also true.

For the description of closed sets of ECE- and QE-equations another representation is useful, by which one can also include the one of E-equations:

Definition Let

$$\iota := (\bigwedge_{i=1}^{n} t_i \stackrel{e}{=} t_i' \Rightarrow t \stackrel{e}{=} t')$$

be any QE-equation. Then ι may be set theoretically represented by an ordered pair

$$(\{(t_i, t_i') \mid 1 \leq i \leq n\}, (t, t')) \in \mathfrak{P}_{fin}(F \times F) \times (F \times F),$$

where $F := F(X, TAlg(\tau))$ and for any set S, $\mathfrak{P}_{fin}(S)$ designates the set of all finite subsets of S. If ι is an ECE-equation, then the corresponding pair belongs to $\mathfrak{P}_{fin}(\{(t,t) \mid t \in F\}) \times (F \times F)$, and if ι is an E-equation, then the corresponding pair belongs to $\mathfrak{P}_{fin}(\{(x,x) \mid x \in X\}) \times (F \times F)$. Thus we define

$$Prem_E := \mathfrak{P}_{fin}(\{(x,x) \mid x \in X\}),$$

$$Prem_{ECE} := \mathfrak{P}_{fin}(\{(t,t) \mid t \in F\}), \text{ and}$$

$$Prem_{QE} := \mathfrak{P}_{fin}(\{(t,t') \mid t, t' \in F\}).$$

Let now $Prem \in \{Prem_E, Prem_{ECE}, Prem_{QE}\}$, and consider $Q \subseteq Prem \times (F \times F)$ to be any set encoding elementary implications of the corresponding type. For $P \in Prem$ define

$$Q(P) := \{(t,t') \mid (P, (t,t')) \in Q\}.$$

For any class $\mathfrak{K}$ of partial algebras define

$$Imp_{Prem}(\mathfrak{K}) := \{(P, (t,t')) \mid P \in Prem, t, t' \in F, \mathfrak{K} \models (\bigwedge_{(p,p') \in P} p \stackrel{e}{=} p' \Rightarrow t \stackrel{e}{=} t')\}$$

and set $\underline{\downarrow} E$ to be the relative subalgebra of $\underline{F} = \underline{F}(X, TAlg(\tau))$ consisting of all subterms of terms occurring in $E \subseteq F \times F$.

With the above notation one has the following description of closed sets of elementary implications of one of the three kinds of *Prem*:

Theorem 3.3 *Let* $Prem \in \{Prem_E, Prem_{ECE}, Prem_{QE}\}$*, and let* $Q \subseteq Prem \times (F \times F)$ *(with* $\underline{F} = \underline{F}(X, TAlg(\tau))$*) be any set representing elementary implications connected with* $Prem$*.*

(a) *Then the following statements are equivalent:*

(i) $Q = Imp_{Prem}(Mod(Q))$.

(ii) *Q has the following properties* (I1) *through* (I4) *for any* $P, P' \in Prem$*:*

(I1) $\underline{\downarrow} Q(P)$ *is an X-generated relative subalgebra of* $\underline{F}(X, TAlg(\tau))$.

(I2) $Q(P)$ *is a closed congruence relation on* $\underline{\downarrow} Q(P)$.

(I3) $P \subseteq Q(P)$.

(I4) *For every homomorphism* $f : X \cup \underline{\downarrow} P \to \underline{\downarrow} Q(P')$ *which satisfies* $(f \times f)(P) \subseteq Q(P')$*, there exists a homomorphic extension* $f_{PP'} : \underline{\downarrow} Q(P) \to \underline{\downarrow} Q(P')$*, which satisfies* $(f_{PP'} \times f_{PP'})(Q(P)) \subseteq Q(P')$.

(b) *If* $Q = Imp_{Prem}(Mod(Q))$*, and* $P \in Prem$*, then*

$$Q(P) = \bigcap \{ker\ f^{\sim} \mid f : \underline{\downarrow} P \to \underline{A},\ \underline{A} \in Mod(Q) \text{ and } P \subseteq kerf^{\sim}\}.$$

3.4 Preservation and reflection of formulas

One application of our formulas is the classification of many important properties of homomorphisms between partial algebras by reflection and preservation of formulas.

Definition Let $\underline{A}, \underline{B}$ be partial algebras of type τ, $F \in \mathcal{L}(\tau)$ a formula w.r.t. the set M of variables ($M \in \{\emptyset, X\}$) and $f : A \to B$ any mapping. We say that

(i) f *preserves* the formula F, if and only if for every valuation $v : M \to A$ one has that $\underline{A} \models F[v]$ implies $\underline{B} \models F[f \circ v]$;

(ii) f *reflects* the formula F, if and only if for every valuation $v : M \to A$ one has that $\underline{B} \models F[f \circ v]$ implies $\underline{A} \models F[v]$.

This definition can also be applied to infinitary elementary implications.

This well known model theoretic concept shows that our notion of homomorphism is a basic model theoretic one and closely related to existence equations, since one has the

Proposition 3.2 *Let $\underline{A}, \underline{B}$ be partial algebras of type τ, and let $f : A \to B$ be any mapping. Then the following statements are equivalent:*

(i) *f is a homomorphism* $f : \underline{A} \to \underline{B}$*;*

(ii) *f preserves the existence equations*

$$\omega x_1 \dots x_{\tau(\omega)} \stackrel{e}{=} y,$$

for all $\omega \in \Omega$ and for all $x_1, \ldots, x_{\tau(\omega)}, y \in X$ (where one may assume that these variables are pairwise distinct).

(iii) *f preserves all existence equations in $L(\tau)$.*

The relationship between preservation and reflection of formulas is very close, since one has

Lemma 3.3 *Let $\underline{A}, \underline{B}$ be partial algebras of type τ, $f : A \to B$ any mapping and F any formula of $\mathcal{L}(\tau)$ or an arbitrary elementary implication. Then the following statements are equivalent:*

(i) *f preserves F;*

(ii) *f reflects $\neg F$.*

Let us recall from category theory that in the category $\underline{Alg(\tau)}$ a homomorphism $f : \underline{A} \to \underline{B}$ is called *initial*, if for every partial algebra $\underline{C}$ and for every mapping $g : C \to A$, g is a homomorphism from $\underline{C}$ into $\underline{A}$ if and only if $f \circ g$ is a homomorphism from $\underline{C}$ into $\underline{B}$. The dual concept is called a *final* homomorphism. Thus we get the following examples:

Proposition 3.4 *In Table 1 some properties of homomorphisms are listed together with the sets of formulas, the reflection of which characterizes them. Different variables are assumed to be distinct. If an operation symbol $\varphi \in \Omega$ occurs, then the reflection of this kind of formulas for all $\varphi \in \Omega$ is meant. If some kind of TE-statement $t \overset{e}{=} t$ or E-equation $t \overset{e}{=} t'$ is mentioned, this means reflection of all such axioms. Observe that "injective and initial" is equivalent to "injective and full".*

Property of homomorphisms	Kind of reflected formulas
injective	$x \overset{e}{=} y$
closed	$\varphi x_1 \ldots x_{\tau(\varphi)} \overset{e}{=} \varphi x_1 \ldots x_{\tau(\varphi)}$
closed	$t \overset{e}{=} t$
initial	$\varphi x_1 \ldots x_{\tau(\varphi)} \overset{e}{=} y$
injective and closed	$x \overset{e}{=} y, \varphi x_1 \ldots x_{\tau(\varphi)} \overset{e}{=} \varphi x_1 \ldots x_{\tau(\varphi)}$
injective and closed	$t \overset{e}{=} t'$
injective and initial	$x \overset{e}{=} y, \varphi x_1 \ldots x_{\tau(\varphi)} \overset{e}{=} z$

Table 1: Properties of homomorphisms defined by reflection of formulas (see Prop. 3.4)

It may attract attention that such important properties like epimorphic, surjective, full and surjective (quotient), etc. do not occur in this table. In order to get their characterization we need the concept of a factorization system, which is discussed in the next subsection 3.5. However, let us first add some more facts about reflection and preservation of formulas.

Proposition 3.5 *Let $\underline{A}, \underline{B}, \underline{C}$ be any partial algebras, let $f : A \to B$ and $g : B \to C$ be any mappings, and let F and F_i $(i \in I)$ be any formulas from $\mathcal{L}(\tau)$ or arbitrary elementary implications. Then one has:*

(i) *If both f and g reflect (preserve) F, then so does $g \circ f$. In particular the class of all homomorphisms reflecting some set of formulas or elementary implications constitutes a subcategory of $\underline{Alg(\tau)}$.*

(ii) *If $g \circ f$ reflects F, and if g preserves F, then f reflects F.*

(iii) *If $g \circ f$ reflects F, and if f is surjective and preserves F, then g reflects F.*

(iv) *If f reflects F_i for each $i \in I$, then f reflects $\bigwedge_{i \in I} F_i$ and $\bigvee_{i \in I} F_i$.*

More properties can be found in [B86], Lemma 9.1.7. Observe that full homomorphisms cannot be defined by reflection of formulas, since their composition need not to be full. One could define some form of "weak reflection" in order to characterize full homomorphisms (see [B86]), Observation 9.2.16), but main applications of fullness are in connection with injectivity, where it is equivalent to "initial and injective" (as we have seen), or in connection with surjectivity, where they are isomorphic to quotient homomorphisms, which are characterized in subsection 3.5 below.

3.5 Factorization systems

Before we introduce the concept of a factorization system let us investigate the category theoretical interpretation of the reflection of elementary implications ι encoded by the epimorphism $e_\iota : \underline{P}_\iota \to \underline{C}_\iota$. Thus, let $f : \underline{A} \to \underline{B}$ be any homomorphism. If $\underline{A}$ does not satisfy the premise of ι, then f trivially reflects ι. Hence we should assume that $\underline{A}$ satisfies the premise of ι with respect to some valuation $v : fvar(\iota) \to A$. But this is equivalent to the fact that v induces a homomorphism, say $p : \underline{P}_\iota \to \underline{A}$, while the fact "$\underline{B} \models \iota[f \circ v]$" is equivalent to the existence of a homomorphism $q : \underline{C}_\iota \to \underline{B}$ such that $f \circ p = q \circ e_\iota$. Then the reflection of ι by f is expressed by the existence of a homomorphism $d : \underline{C}_\iota \to \underline{A}$ such that $d \circ e_\iota = p$ (see Figure 12).

$$\begin{tikzcd} \underline{P}_\iota \arrow[r, "e_\iota \in Epi"] \arrow[d, "p"'] & \underline{C}_\iota \arrow[d, "q"] \\ \underline{A} \arrow[r, "f"'] & \underline{B} \end{tikzcd} \quad \mapsto \quad \begin{tikzcd} \underline{P}_\iota \arrow[r, "e_\iota \in Epi"] \arrow[d, "p"'] & \underline{C}_\iota \arrow[d, "q"] \arrow[dl, "d"'] \\ \underline{A} \arrow[r, "f"'] & \underline{B} \end{tikzcd}$$

Figure 12: Reflection by f of the el. implication ι encoded by the epi. e_ι

Observe that d is unique and also satisfies $f \circ d = q$, since e_ι is an epimorphism. This motivates the following

Definition A pair (e, m) of homomorphisms $e : \underline{P} \to \underline{C}$, $m : \underline{A} \to \underline{B}$ has the *unique-*

diagonal-fill-in-property, i.e. it satisfies $Difip(e, m)$, if and only if for all homomorphisms $p : \underline{P} \to \underline{A}$ and $q : \underline{C} \to \underline{B}$ satisfying $q \circ e = m \circ p$ there exists a unique homomorphism $d : \underline{C} \to \underline{A}$ such that $d \circ e = p$ and $m \circ d = q$ (see Figure 13).

$$\begin{tikzcd} \underline{P} \arrow[r, "e"] \arrow[d, "p"'] & \underline{C} \arrow[d, "q"] \\ \underline{A} \arrow[r, "m"'] & \underline{B} \end{tikzcd} \quad \mapsto \quad \begin{tikzcd} \underline{P} \arrow[r, "e"] \arrow[d, "p"'] & \underline{C} \arrow[d, "q"] \arrow[dl, "d"'] \\ \underline{A} \arrow[r, "m"'] & \underline{B} \end{tikzcd}$$

Figure 13: Diagonal-fill-in-property

Observe that here e need not to be an epimorphism.

As usual such a relation induces two operators Λ and Λ^{op} of a Galois correspondence applicable here to classes of homomorphisms:

Definition Let $\mathcal{E}, \mathcal{M} \subseteq Hom$ be two classes of homomorphisms. Then one defines

$$\Lambda(\mathcal{E}) := \{m \in Hom \mid Difip(e, m) \text{ for all } e \in \mathcal{E}\},$$
$$\Lambda^{op}(\mathcal{M}) := \{e \in Hom \mid Difip(e, m) \text{ for all } m \in \mathcal{M}\}.$$

Proposition 3.6 *One always has for Λ (and similarly for Λ^{op}) and $\mathcal{E}, \mathcal{E}' \subseteq Hom$:*

(i) $Iso \subseteq \Lambda(\mathcal{E})$,

(ii) $\Lambda(\mathcal{E}) \circ \Lambda(\mathcal{E}) \subseteq \Lambda(\mathcal{E})$,

(iii) $\Lambda(\mathcal{E}) \cap \mathcal{E} \subseteq Iso$,

(iv) $\mathcal{E} \subseteq \mathcal{E}'$ *implies* $\Lambda(\mathcal{E}) \supseteq \Lambda(\mathcal{E}')$, $\Lambda\Lambda^{op}\Lambda(\mathcal{E}) = \Lambda(\mathcal{E})$ *and* $\mathcal{E} \subseteq \Lambda^{op}\Lambda(\mathcal{E})$,

(v) $g \circ f \in \Lambda(\mathcal{E})$ *and* $g \in Mono$ *imply* $f \in \Lambda(\mathcal{E})$;
$g \circ f \in \Lambda^{op}(\mathcal{E})$ *and* $f \in Epi$ *imply* $g \in \Lambda^{op}(\mathcal{E})$.

For those who know a little bit more about category theory we mention that Λ is preserved by multiple pullbacks, products and induced product morphisms, while Λ^{op} is preserved by multiple pushouts, coproducts and induced coproduct morphisms (for more details see G.E.Strecker [S72], or see [B86], section 10).

Of special interest for us are the situations where the related classes $\Lambda(\mathcal{E})$ and $\Lambda\Lambda^{op}(\mathcal{E})$ satisfy some additional properties:

Definition Let $\mathcal{E}, \mathcal{M} \subseteq Hom$ be arbitrary classes of homomorphisms of partial algebras of type τ. The pair $(\mathcal{E}, \mathcal{M})$ is said to form a *factorization system* in $Alg(\tau)$ (and $\mathcal{E}$ is sometimes called the left factor (or *epi*factor) and $\mathcal{M}$ is sometimes called the right factor (or *mono*factor[8])), if the following conditions are satisfied:

[8] Observe, however, that there are interesting examples (see below or [Pa79]), where $\mathcal{E}$ does not consist of epimorphisms or $\mathcal{M}$ does not consist of monomorphisms.

(FS1) $\mathcal{M} \circ \mathcal{E} = Hom$,

(FS2) $\mathcal{M} \circ \mathcal{M} = \mathcal{M}$, $\mathcal{E} \circ \mathcal{E} = \mathcal{E}$,

(FS3) $Iso \subseteq \mathcal{M} \cap \mathcal{E}$,

(FS4) $\mathcal{E} \times \mathcal{M} \subseteq Difip$.

Lemma 3.7 *$\mathcal{E}, \mathcal{M} \subseteq Hom$ form a factorization system, if and only if* (FS1), (FS2) *and* (FS3) *are satisfied together with*

(FS4') *The factorization in* (FS1) *is unique up to isomorphism, i.e. if $f = m \circ e = m' \circ e'$ with $e, e' \in \mathcal{E}$, $m, m' \in \mathcal{M}$, then there is a unique isomorphism i from codom e onto codom e',*[9] *such that $e' = i \circ e$ and $m = m' \circ i$ (see Figure* 14*).*

$$
\begin{tikzcd}
\bullet \arrow[r, "e \in \mathcal{E}"] \arrow[d, "\mathcal{E} \ni e'"'] & \bullet \arrow[d, "m \in \mathcal{M}"] \\
\bullet \arrow[r, "m' \in \mathcal{M}"'] & \bullet
\end{tikzcd}
\quad \mapsto \quad
\begin{tikzcd}
\bullet \arrow[r, "e \in \mathcal{E}"] \arrow[d, "\mathcal{E} \ni e'"'] & \bullet \arrow[d, "m \in \mathcal{M}"] \arrow[dl, "(\cong)\, i"'] \\
\bullet \arrow[r, "m' \in \mathcal{M}"'] & \bullet
\end{tikzcd}
$$

Figure 14: Uniqueness of $(\mathcal{E}, \mathcal{M})$-factorizations

Factorization systems are abundant in $\underline{Alg(\tau)}$, since we have the

Theorem 3.4 *Let $\mathcal{E} \subseteq Epi$ and $\mathcal{M} \subseteq Mono$ in $\underline{Alg(\tau)}$, then*

$$(\Lambda^{op}\Lambda(\mathcal{E}), \Lambda(\mathcal{E})) \text{ as well as } (\Lambda^{op}(\mathcal{M}), \Lambda\Lambda^{op}(\mathcal{M}))$$

are factorization systems in $\underline{Alg(\tau)}$.

Particular examples are described in

Proposition 3.8 *In $Alg(\tau)$ we have among others the factorization systems $(\mathcal{E}, \mathcal{M})$ shown in Table* 2*, where $e : \underline{P} \to \underline{C}$ in $\mathcal{E}$ and $m : \underline{A} \to \underline{B}$ in $\mathcal{M}$.*

The first factorization system in Table 2 has already been considered in Corollary 2.5, while the third one has been considered in Lemma 1.14.

Observe that final homomorphisms are full homomorphisms which totally induce the structure on their image but need not be epimorphisms (since outside of the image the structure is just discrete). On the other hand closed homomorphisms as well as initial homomorphisms are in general not injective, i.e. no monomorphisms.

[9] "*codom e*" designates here the target algebra of the homomorphism e.

$\mathcal{E}$ is the class of all homomorphisms e which are	$\mathcal{M}$ is the class of all homomorphisms m which are
$TAlg(\tau)$-extendable epimorphisms	closed
epimorphisms	closed and injective
full and surjective (= quotients)	injective
surjective	initial and injective (=full and injective)
surjective, and $c \in C \setminus \bigcup_{\varphi\in\Omega} \varphi^{\underline{C}}(C^{\tau(\varphi)})$ implies $\#e^{-1}(\{c\}) = 1$	initial
final	bijective

Table 2: Some interesting factorization systems (see Prop. 3.8)

Proposition 3.8 shows us that the most interesting properties of homomorphisms that have shown up so far are either definable by the reflection of existence equations or are their partners in a factorization system (representing all the reflected epimorphisms). The only exception here from this "rule" are final, bijective and full homomorphisms, respectively, where at least final homomorphisms and bijective homomorphisms are "partners" in a factorization system, too.

In particular, the fact that the classes *Ext* of all $TAlg(\tau)$-extendable epimorphisms and $\mathcal{M}_c$ of all closed homomorphisms form a factorization system shows that our closed initial homomorphic extensions are definable within the category $\underline{Alg(\tau)}$ without using partial mappings between partial algebras: for a homomorphism $f : \underline{A} \to \underline{B}$ the pair $(id_{A\ dom\ f^\sim}, f^\sim)$ is just its $(Ext, \mathcal{M}_c)$-factorization (up to isomorphism).

3.6 A Meta Birkhoff Theorem by Andréka, Németi and Sain

Now we have almost all the tools available, which are needed for the formulation (and the proof) of a (still quite restricted version of a) result by H.Andréka, I.Németi and I.Sain (see [AN82] and [NSa82]) which yields many Birkhoff-type theorems for partial algebras w.r.t. very different kinds of implications. However, we still have to generalize our category theoretical description of formulas:

Definition

(i) A family $\mathsf{c} := (e_i : \underline{P} \to \underline{C_i})_{i\in I}$ of epimorphisms will be called a *cone* in what follows.

(ii) We say that a cone c *holds* in a partial algebra $\underline{A}$, or that $\underline{A}$ *is injective* w.r.t. c, in symbols $\underline{A} \models \mathsf{c}$ or $\underline{A} \in Inj(\{\mathsf{c}\})$, if and only if for all homomorphisms $f : \underline{P} \to \underline{A}$ there are $k \in I$ and a homomorphism $g : \underline{C_k} \to \underline{A}$ such that $f = g \circ e_k$ (see Figure 15). If $I = \emptyset$, then $\mathsf{c} = (\underline{P})$, and we have $\underline{A} \models (\underline{P})$ if and only if *there is no* homomorphism $f : \underline{P} \to \underline{A}$.

(iii) If K is a class of cones, we say that $\underline{A} \in InjK$ if and only if $\underline{A} \models \mathsf{c}$ for all $\mathsf{c} \in K$, i.e.

$$InjK := \{\underline{A} \in Alg(\tau) \mid \underline{A} \models \mathsf{c} \text{ for all } \mathsf{c} \in K\}.$$

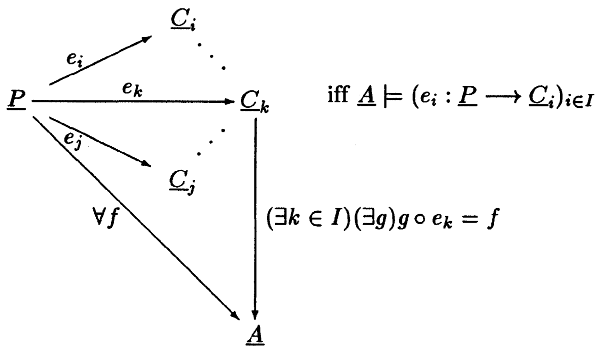


Figure 15: Validity of cones

Observe that in a model theoretic interpretation a cone c represents an infinitary implication (for $I \neq \emptyset$) of the form

$$\iota_{\mathsf{c}} \equiv (\bigwedge_{j \in J} t_j \overset{e}{=} t'_j \Rightarrow \bigvee_{i \in I} \bigwedge_{k \in K_i} t_{ik} \overset{e}{=} t'_{ik})$$

and the injectivity of $\underline{A}$ w.r.t. c just means that ι_{c} holds in $\underline{A}$. The cone $(\underline{P})$ for $I = \emptyset$ corresponds to the formula

$$\neg \bigwedge_{j \in J} t_j \overset{e}{=} t'_j.$$

In what follows we shall use the letters $\mathcal{H}$ and $\mathcal{S}$ both for classes of homomorphisms and for special operators induced by them; $\mathcal{H}$ for "$\mathcal{H}$-homomorphic images" and $\mathcal{S}$ for "$\mathcal{S}$-subobjects", i.e. for $\mathfrak{K} \subseteq Alg(\tau)$ we define

$$\mathcal{H}(\mathfrak{K}) := \{\underline{B} \in Alg(\tau) \mid \text{ there are } \underline{A} \in \mathfrak{K} \text{ and } f : \underline{A} \to \underline{B} \text{ in } \mathcal{H}\},$$

$$\mathcal{S}(\mathfrak{K}) := \{\underline{B} \in Alg(\tau) \mid \text{ there are } \underline{B} \in \mathfrak{K} \text{ and } f : \underline{A} \to \underline{B} \text{ in } \mathcal{S}\}.$$

Recall that one has the following concept dual to injectivity:

Definition Let $\mathcal{H}$ be a class of homomorphisms. A partial algebra $\underline{P}$ is called *$\mathcal{H}$-projective* if and only if for every homomorphism $h : \underline{A} \to \underline{B}$ from $\mathcal{H}$ and for every homomorphism $p : \underline{P} \to \underline{B}$ there is a homomorphism $f : \underline{P} \to \underline{A}$ such that $h \circ f = p$ (see Figure 16).

As the last preparation of the following theorem we have to specify different kinds of cones.

Definition Let $\mathsf{c} = (e_i : \underline{P} \to \underline{C}_i)_{i \in I}$ be a cone (of epimorphisms) then we say that

(i) c is an $\mathcal{H}$-cone (for $\mathcal{H}$-images), if and only if $\underline{P}$ is $\mathcal{H}$-projective;

$$\begin{tikzcd} & \underline{P} \arrow[d, "p"] \\ \underline{A} \arrow[r, "h \in \mathcal{H}"'] & \underline{B} \end{tikzcd} \quad \mapsto \quad \begin{tikzcd} & \underline{P} \arrow[d, "p"] \arrow[dl, "f"'] \\ \underline{A} \arrow[r, "h \in \mathcal{H}"'] & \underline{B} \end{tikzcd}$$

Figure 16: $\underline{P}$ is $\mathcal{H}$-projective

(ii) c is an $\mathcal{S}$-cone (for $\mathcal{S}$-subobjects), if and only if $e_i \in \bigwedge^{op}(\mathcal{S})$ for all $i \in I$;

(iii) c is a $\mathcal{P}$-cone (for products), if and only if $\#I = 1$;

(iv) c is a $\mathcal{P}_+$-cone (for products with non-empty index sets), if and only if $\#I \leq 1$;

(v) c is a $\mathcal{P}_r$-cone (for reduced products), if and only if $\#I = 1$ and $\underline{P}$ is totally finite;

(vi) c is a $\mathcal{P}_{r+}$-cone (for reduced products with non-empty index sets), if and only if $\#I \leq 1$ and $\underline{P}$ is totally finite;

(vii) c is an e-cone (for the empty product), if and only if $\#I \geq 1$;

(viii) c is a $\mathcal{P}_u$-cone (for ultraproducts), if and only if I is finite and $\underline{P}$ and all $\underline{C}_i$ $(i \in I)$ are totally finite.

From the results of H.Andréka, I.Németi and I.Sain in [AN82] and [NSa82] one can extract the following

Theorem 3.5 (Meta Birkhoff Theorem) *Let $(\mathcal{O}, \mathcal{M})$ be a category of partial algebras with $\mathcal{O} \subseteq Alg(\tau)$ and the class $\mathcal{M} \subseteq Hom$ of homomorphisms such that*

- *$(\mathcal{O}, \mathcal{M})$ has products and direct limits,*
- *every $\underline{A} \in \mathcal{O}$ is the direct limit (in $(\mathcal{O}, \mathcal{M})$) of totally finite partial algebras belonging to $\mathcal{O}$.*

Moreover, let $\mathcal{H}, \mathcal{S} \subseteq \mathcal{M}$ be classes of morphisms such that:

(1) *Each $\underline{A} \in \mathcal{O}$ is the $\mathcal{H}$-image of an $\mathcal{H}$-projective $\underline{P} \in \mathcal{O}$.*

(2) *Every $\mathcal{H}$-projective object $\underline{P} \in \mathcal{O}$ is the direct limit of totally finite $\mathcal{H}$-projective partial algebras from $\mathcal{O}$.*

(3) *$(\bigwedge^{op}(\mathcal{S}), \mathcal{S})$ is a factorization system in $(\mathcal{O}, \mathcal{M})$.*

(4) *If $g \circ f \in \mathcal{H}$ and $f \in \bigwedge^{op}(\mathcal{S})$, then $g \in \mathcal{H}$.*

(5) *From each $\underline{A} \in \mathcal{O}$ there starts up to isomorphism only a* set *of $\bigwedge^{op}(\mathcal{S})$-morphisms.*

(6) *$\mathcal{S} = \bigwedge\{e : \underline{P} \to \underline{C} \mid e \in \bigwedge^{op}(\mathcal{S})$ and $\underline{P}$ and $\underline{C}$ are totally finite$\}$.*

Let $\mathfrak{K} \subseteq \mathcal{O}$ be any subclass of partial algebras and let $\mathcal{F}$ be one of the operators $\mathcal{P}$, $\mathcal{P}_+$, $\mathcal{P}_r$, $\mathcal{P}_{r+}$, e or $\mathcal{P}_u$. Then

$$InjK_{\mathcal{HSF}}(\mathfrak{K}) = \mathcal{HSF}(\mathfrak{K}),$$

where $K_{\mathcal{HSF}}$ designates the class of all $\mathcal{HSF}$-cones, which hold in $\mathfrak{K}$.

For applications of this theorem we only consider the category $\underline{Alg(\tau)}$ of all partial algebras of type τ and of all homomorphisms between them. However, some $\overline{ECE}$-varieties might also do. Moreover, let in this category

$\mathcal{I}$ be the class of all isomorphisms,
$\mathcal{S}_w$ be the class of all injective homomorphisms,
$\mathcal{H}_w$ be the class of all surjective homomorphisms,
$\mathcal{M}_c$ be the class of all closed homomorphisms,
$\mathcal{M}_i$ be the class of all initial homomorphisms,
$\mathcal{H}_f$ be the class of all full and surjective homomorphisms,
$\mathcal{H}_i := \mathcal{M}_i \cap \mathcal{H}_w$ be the class of all initial and surjective homomorphisms,
$\mathcal{H}_b := \mathcal{S}_w \cap \mathcal{H}_w$ be the class of all bijective homomorphisms,
$\mathcal{H}_c := \mathcal{M}_c \cap \mathcal{H}_w$ be the class of all closed and surjective homomorphisms,
$\mathcal{S}_i := \mathcal{M}_i \cap \mathcal{S}_w$ be the class of all initial and injective homomorphisms,
$\mathcal{S}_c := \mathcal{M}_c \cap \mathcal{S}_w$ be the class of all closed and injective homomorphisms.

Then we get the

Proposition 3.9 *In Table 3 it is indicated by + in a row for a class of homomorphisms chosen for $\mathcal{H}$ and in the column for a class of homomorphisms chosen for $\mathcal{S}$, when it is known that this pair $(\mathcal{H}, \mathcal{S})$ satisfies the assumptions of Theorem 3.5. A missing entry means that the corresponding pair has not yet been investigated*[10].

$\mathcal{H} \setminus \mathcal{S}$	$\mathcal{S}_i$	$\mathcal{S}_w$	$\mathcal{M}_i$	$\mathcal{S}_c$	$\mathcal{M}_c$
$\mathcal{I}$	+	+	+	+	+
$\mathcal{H}_c$	+	+	+	+	+
$\mathcal{H}_f$	+	+	+	+	+
$\mathcal{H}_w$	+	+	+	+	+
$\mathcal{H}_i$	+	+	+		
$\mathcal{H}_b$	+	+	+		

Table 3: Compatible pairs $(\mathcal{H}, \mathcal{S})$ (see Proposition 3.9)

This yields already 156 different Birkhoff type theorems, since we have 6 possibilities for the operator $\mathcal{F}$ in Theorem 3.5. For $\mathcal{F} = \mathcal{P}_r$, $\mathcal{S} = \mathcal{S}_c$, and for $\mathcal{H}$ being one of the classes $\mathcal{I}, \mathcal{H}_c$ or $\mathcal{H}_w$, we get the three results from Theorem 3.2, when we observe in addition the influence of the conditions (1) through (6) from Theorem 3.5 on the implications under

[10] It seems that a student of W.Bartol at Warsaw has worked on them recently, and that two entries are positive, two are negative.

consideration[11]. Some descriptions of premises and conclusions — derived from these conditions — for special operators are collected in Table 4 (equality "=" here really means that the terms have to be identical, while X is the set of (free) variables under consideration).

Restrictions on the *premise* $\bigwedge_{i\in I} t_i \stackrel{e}{=} t'_i$ in case of $\mathcal{H}$ as:	
$\mathcal{I}$	no restrictions
$\mathcal{H}_w$	$t_i = t'_i$ is a variable $(i \in I)$
$\mathcal{H}_f$	$t_i = t'_i = \varphi_i x_{1i} \ldots x_{\tau(\varphi_i)i}$ $(i \in I, \varphi_i \in \Omega)$, and for $(k,i) \neq (k',i')$ the variables x_{ki} and $x_{k'i'}$ are distinct
$\mathcal{H}_c$	$t_i = t'_i$, arbitrary term $(i \in I)$
Restrictions on the *conclusion* $t \stackrel{e}{=} t'$ with respect to the premise $\bigwedge_{i\in I} t_i \stackrel{e}{=} t'_i$ for $\mathcal{S}$ as:	
$\mathcal{S}_w$	$t, t' \in\downarrow \{t_i, t'_i \mid i \in I\} \cup X$
$\mathcal{S}_i$	t arbitrary term, $t' \in\downarrow \{t_i, t'_i \mid i \in M\} \cup X$
$\mathcal{S}_c$	t, t' arbitrary terms
$\mathcal{M}_i$	t arbitrary term, $t' \in\downarrow \{t_i, t'_i \mid i \in M\} \cup X$, and not both of t, t' are variables
$\mathcal{M}_c$	$t = t'$ arbitrary term

Table 4: Premises and conclusions for some special operators

It should be observed that one consequence of Theorem 3.5 is the fact that the quasi-primitive classes $\mathfrak{K} = \mathcal{I}\mathcal{S}_c\mathcal{P}(\mathfrak{K})$ are exactly the classes definable by elementary implications with no restrictions on the lengths of premise or conclusion.

Appendix: Some remarks on many-sorted partial algebras

Since computer scientists need rather many-sorted than only one-sorted (partial) algebras we want to conclude these notes with some observations concerning many-sorted (partial) algebras. Above all we want to indicate how many-sorted partial and total algebras can be treated within the category of partial algebras of the corresponding similarity type. From this it will not be difficult to conclude that the basic category theoretical constructions work for heterogeneous partial algebras quite analogously to the homogeneous ones. However, we also want to point out some differences, e.g. concerning the model theoretic properties. More about this can be found in [B86].

Let us first recall that a *many-sorted similarity type* or *signature* $\Sigma := (S, \tau^*)$ is usually specified by

– a set, say S, the elements of which are called *sorts*,

– a set, say Ω, of operation symbols,

[11] Observe that in the first case the operators $\mathcal{H}\mathcal{S}_c\mathcal{P}_r$ and $\mathcal{H}\mathcal{S}_c\mathcal{P}$ are identical, since reduced products are weak homomorphic images of direct products.

– a mapping τ^* from Ω into the set $S^* \times S$ consisting of all pairs $\tau^*(\varphi) = (s_1^\varphi \dots s_{\tau(\varphi)}^\varphi, s^\varphi)$ (for $\varphi \in \Omega$), of which the first component is a finite word $w = s_1^\varphi \dots s_{\tau(\varphi)}^\varphi$ of elements from S, the length $\tau(\varphi)$ of which is just the arity of the operation symbol φ, while for $1 \leq i \leq \tau(\varphi)$ the i-th "letter" s_i^φ indicates that the i-th argument of each realization of φ has always to be of sort s_i^φ; the second component, s^φ, of $\tau^*(\varphi)$ designates the sort of the value of any realization of φ, whenever this value exists.

A *many-sorted (partial) algebra* $\underline{A}$ of signature Σ is then defined as $((A_s)_{s\in S}, (\varphi^{\underline{A}})_{\varphi\in\Omega})$, where $(A_s)_{s\in S}$ is a family of sets, A_s being called the *phylum* or *carrier set* of sort s, and if $\tau^*(\varphi) = (s_1^\varphi \dots s_{\tau(\varphi)}^\varphi, s^\varphi)$, then $\varphi^{\underline{A}}$ is a (possibly partial) mapping from (or out of) $A_{s_1^\varphi} \times \dots \times A_{s_{\tau(\varphi)}^\varphi}$ into A_{s^φ}. A *homomorphism*, say $h : \underline{A} \to \underline{B}$, from a many-sorted (partial) algebra $\underline{A}$ into a many-sorted (partial) algebra $\underline{B}$ is then usually defined as being a sequence $(h_s : A_s \to B_s)_{s\in S}$ of mappings h_s between corresponding phyla, such that $\varphi^{\underline{A}}(a_1, \dots, a_{\tau(\varphi)}) = a$ $(a_i \in A_{s_i^\varphi},\ a \in A_s)$ implies $\varphi^{\underline{B}}(h_{s_1^\varphi}(a_1), \dots, h_{s_{\tau(\varphi)}^\varphi}(a_{\tau(\varphi)})) = h_{s^\varphi}(a)$, or in the total case just

$$h_{s^\varphi}(\varphi^{\underline{A}}(a_1, \dots, a_{\tau(\varphi)})) = \varphi^{\underline{B}}(h_{s_1^\varphi}(a_1), \dots, h_{s_{\tau(\varphi)}^\varphi}(a_{\tau(\varphi)})).$$

These definitions suggest that:

– Different phyla of a many-sorted partial algebra of signature Σ may, without loss of generality, be assumed to be disjoint. (Inclusions should be specified by appropriate unary operations.)

– The specification of the signature Σ can be considered as the description of a particular homogeneous partial algebraic structure $(\varphi^{\underline{S}})_{\varphi\in\Omega}$ of the homogeneous similarity type $\tau = (\tau(\varphi))_{\varphi\in\Omega}$ on the set S of sorts, where for each $\varphi \in \Omega$, $\varphi^{\underline{S}}$ is defined only on the sequence $(s_1^\varphi, \dots, s_{\tau(\varphi)}^\varphi)$, and for this it takes the value s^φ. We shall call this partial algebra $\underline{S}$ the *sort algebra* (or *specification algebra*) for the signature under consideration.

– One can replace now the family $(A_s)_{s\in S}$ of phyla of a many-sorted partial algebra by its disjoint union, say $A^* := \dot{\bigcup} A_s$, in which case each many-sorted (partial) operation $\varphi^{\underline{A}}$ becomes a partial operation on A^*, which we shall designate again by $\varphi^{\underline{A}}$.

– The original partition of A into phyla can be replaced by a mapping, say $v_{A^*} : A^* \to S$ such that $v_{A^*}(a) = s$ if and only if $a \in A_s$ $(a \in A^*,\ s \in S)$. v_{A^*} will be called the *canonical sort mapping* for $\underline{A}^*$. One then has the following

Theorem 3.6 *With the concepts and notations introduced above for any many-sorted partial algebra $\underline{A}$ the canonical sort mapping v_{A^*} is always a homomorphism*[12]

$$v_{A^*} : (\dot{\bigcup}_{s\in S} A_s, (\varphi^{\underline{A}})_{\varphi\in\Omega}) \to \underline{S};$$

and this homomorphism is closed if and only if $\underline{A}$ is a total many-sorted algebra.

Conversely, if $\underline{A}^*$ is any partial algebra and $v_{A^*} : \underline{A}^* \to \underline{S}$ is a homomorphism, then $\underline{A}^*$ can be considered as a many-sorted partial algebra $((\varphi^{-1}(\{s\}))_{s\in S}, (\varphi^{\underline{A}})_{\varphi\in\Omega})$ which is total if and only if v_{A^*} is closed.

[12]Remember that we assume in particular the phyla to be pairwise disjoint!

Example Let us consider the signature for stack automata of integers (in the way computer scientists usually represent it):

sorts:	stack, integer
operations:	0: $\rightarrow$ integer
	succ, pred: integer $\rightarrow$ integer
	empty: $\rightarrow$ stack
	push: stack $\times$ integer $\rightarrow$ stack
	pop: stack $\rightarrow$ stack
	top: stack $\rightarrow$ integer.

This means that we have two nullary, four unary and one binary (possibly partial) operations, e.g. "push: stack $\times$ integer $\rightarrow$ stack" means that push is a binary operation with first argument of sort "stack", second argument of sort "integer" and value of sort "stack".

A closed homomorphism v_A from a partial algebra $\underline{A}$ into $\underline{S}$ means that there exist fundamental constants $0^{\underline{A}} \in v_A^{-1}(\{\text{integer}\})$ and $\text{empty}^{\underline{A}} \in v_A^{-1}(\{\text{stack}\})$ and that e.g. the binary partial operation $\text{push}^{\underline{A}}$ maps each pair (and only such) (c, d) with $c \in v_A^{-1}(\{\text{stack}\})$ and $d \in v_A^{-1}(\{\text{integer}\})$ onto some element of $v_A^{-1}(\{\text{stack}\})$, etc., but that is just how a total two-sorted algebra

$$\underline{A} : (v_A^{-1}(\{\text{stack}\}), v_A^{-1}(\{\text{integer}\}); 0^{\underline{A}}, \text{succ}^{\underline{A}}, \text{pred}^{\underline{A}}, \text{empty}^{\underline{A}}, \text{push}^{\underline{A}}, \text{pop}^{\underline{A}}, \text{top}^{\underline{A}})$$

is described.

Thus, we can consider a many-sorted partial algebra of signature Σ and sort algebra $\underline{S}$ as a pair $(\underline{A}^*, v_{A^*} : \underline{A}^* \rightarrow \underline{S})$, where $\underline{A}^*$ is a partial algebra of similarity type $\tau = (\tau(\varphi))_{\varphi\in\Omega}$[13], and v_{A^*} is a homomorphism (which is closed if and only if the many-sorted algebra is total).

A homomorphism $h : \underline{A} \rightarrow \underline{B}$ between many-sorted partial algebras then corresponds to a homomorphism $h^* : \underline{A}^* \rightarrow \underline{B}^*$ such that $v_{B^*} \circ h^* = v_{A^*}$, where the graph of h^* is the disjoint union of the graphs of the mappings $h_s : A_s \rightarrow B_s$ $(s \in S)$; and vice versa: if $h^* : \underline{A}^* \rightarrow \underline{B}^*$ is a homomorphism between partial algebras of type τ provided with homomorphisms $v_{A^*} : \underline{A}^* \rightarrow \underline{S}$ and $v_{B^*} : \underline{B}^* \rightarrow \underline{S}$ such that $v_{B^*} \circ h^* = v_{A^*}$, then one has that $(h^*|_{A_s} : A_s \rightarrow B_s)_{s\in S}$ is a homomorphism between the corresponding many-sorted partial algebras $\underline{A}$ and $\underline{B}$. If $\underline{A}$ and $\underline{B}$ are total, then h^* is closed.

In category theory, for any category $\mathfrak{C}$ and any object S the category $\mathfrak{C} \downarrow \{S\}$, the objects of which are pairs $(C, v : C \rightarrow S)$, where C is an object and v is a morphism of $\mathfrak{C}$, and where morphisms $h : (C, v) \rightarrow (D, w)$ are morphisms $h : C \rightarrow D$ of $\mathfrak{C}$ which satisfy $w \circ h = v$, is called a *comma category.*

Thus we have established an equivalence between the category $\underline{Alg(\Sigma)}$ of all partial many-sorted algebras of signature $\Sigma = (S, \Omega, \tau^*)$ and the comma category $\underline{Alg(\tau)} \downarrow \{\underline{S}\}$ of the category $\underline{Alg(\tau)}$ of all partial algebras of the corresponding homogeneous type τ with respect to the sort algebra $\underline{S}$.

[13] Here $\tau(\varphi)$ is the length of the word in the first component of $\tau^*(\varphi)$.

In order to form *coproducts* in $\underline{Alg(\tau)} \downarrow \{\underline{S}\}$ for a family $((\underline{A}_i, v_{A_i}))_{i\in I}$, one forms in $\underline{Alg(\tau)}$ the coproduct $((\iota_i : \underline{A}_i \to \coprod_{i\in I}\underline{A}_i)_{i\in I}, \coprod_{i\in I}\underline{A}_i) =: ((\iota_i)_{i\in I}, \underline{A})$; and the sort homomorphism $v_A : \underline{A} \to \underline{S}$ is then the homomorphism in the category $\underline{Alg(\tau)}$ induced by the family $(v_{A_i})_{i\in I}$ of sort homomorphisms. Since $\coprod_{i\in I}\underline{A}_i$ carries the weakest structure such that all ι_i $(i \in I)$ are still homomorphisms, it is not difficult to realize that in the case of closed v_{A_i} $(i \in I)$ also v_A and all ι_i $(i \in I)$ are closed.

If one wants to construct in $\underline{Alg(\tau)} \downarrow \{\underline{S}\}$ a *product* of a family $((\underline{A}_i, v_{A_i}))_{i\in I}$, it is easy to realize that this corresponds to the construction of a pullback $(\underline{P}, (p_i : \underline{P} \to \underline{A}_i)_{i\in I})$ with respect to the given sink $((\underline{A}_i, v_{A_i}))_{i\in I}$. Since the class of all closed homomorphisms is equal to $\Lambda(Ext)$, where Ext is the class of all $TAlg(\tau)$-extendable epimorphisms, it is easy to realize that all induced homomorphisms p_i (and hence $v_p = v_{A_i} \circ p_i$ $(i \in I)$) become closed, if all v_i $(i \in I)$ are closed.

These observations extend to all other category theoretical constructions.

There is no problem with *subalgebras*. With respect to *congruence relations* one has to observe that $ker\ v_{A^*} := \{(a, a')|v_{A^*}(a) = v_{A^*}(a')\}$ becomes the largest admissible congruence (and equivalence) relation of the pair $(\underline{A}^*, v_{A^*})$, and that the compatibility condition $v_{A^*} = v_{B^*} \circ h^*$ for homomorphisms $h^* : (\underline{A}^*, v_{A^*}) \to (\underline{B}^*, v_{B^*})$ implies that always $ker\ h^* \subseteq ker\ v_{A^*}$. Observe that $ker\ v_{A^*} = \bigcup_{s\in S} A_s \times A_s$ is just the union of the largest equivalence relations of the phyla.

One advantage of this approach may be the observation that it easily also allows to handle *overloading*: one just has to drop the requirement that in the sort algebra each partial operation is to be defined on exactly one sequence. Everything which has been observed so far does not depend on this requirement. Because of the fact that one uses variables of different sorts (or the corresponding equivalent representation with a sort mapping) in order to formulate axioms, different instances of the same operation can satisfy different axioms.

Example Let us consider integers, sequences of integers of length n (for some fixed natural number n) and $n \times n$ square matrices of integers with addition and multiplication as operations:

sorts: int, seq_n, mat_n

operations: $+$: int $\times$ int $\to$ int
$\mathrm{seq}_n \times \mathrm{seq}_n \to \mathrm{seq}_n$
$\mathrm{mat}_n \times \mathrm{mat}_n \to \mathrm{mat}_n$
$\cdot$: int $\times$ int $\to$ int
int $\times\ \mathrm{seq}_n \to \mathrm{seq}_n$
$\mathrm{mat}_n \times \mathrm{seq}_n \to \mathrm{seq}_n$
$\mathrm{mat}_n \times \mathrm{mat}_n \to \mathrm{mat}_n$
int $\times\ \mathrm{mat}_n \to \mathrm{mat}_n$

Among the axioms one may formulate

$$(x, y : \mathrm{int}; \cdot xy \overset{e}{=} \cdot yx)$$

$$(x, y, z : \mathrm{int}; \cdot \cdot xyz \stackrel{e}{=} \cdot x \cdot yz)$$

$$(x, y : \mathrm{int}, z : \mathrm{seq}_n; \cdot \cdot xyz \stackrel{e}{=} \cdot x \cdot yz)$$

etc., while there is no commutativity law for matrix multiplication.

One thing we have already used here, which is of much more importance in the many-sorted than in the one-sorted case, is the fact that with each axiom one has also to specify the variables to which it refers — there may be more variables and more sorts specified than actually needed, and in connection with additional sorts or e.g. for existence-equations without free variables this really will have meaning, if empty phyla are allowed. Namely, if a partial algebra has an empty phylum of sort s, then each existence-equation which refers to a variable of this sort in its specification is trivially valid in this partial algebra, e.g. in our specification above the axiom

$$(x, y : \mathrm{int}, z : \mathrm{mat}_n; \cdot xy \stackrel{e}{=} x)$$

is trivially satisfied in every algebra of this signature, in which the phylum of sort mat_n is empty.

Concerning its model theoretic meaning an axiom of the form

$$(x, y : \mathrm{int}, z : \mathrm{seq}_n; \cdot \cdot xyz \stackrel{e}{=} \cdot x \cdot yz)$$

can be replaced by the implication

$$(x \stackrel{e}{=} x \wedge z \stackrel{e}{=} z \rightarrow \cdot \cdot xyz \stackrel{e}{=} \cdot x \cdot yz)$$

($y \stackrel{e}{=} y$ is superfluous in the premise), which means: if the variable x (of sort int) and the variable z (of sort mat_n) are interpretable, i.e. if the corresponding phyla are non-empty, then the conclusion has to hold true.

This observation explains, why many-sorted (existence-) equations (also in the total case) behave already much like implications rather than like equations. Namely, it may no longer be true that a variety of many-sorted algebras is always generated (w.r.t. the operators $\mathcal{H}$, $\mathcal{S}$ and $\mathcal{P}$) by a single algebra.

In addition, in the case of many-sorted partial algebras which allow empty phyla one has to distinguish, whether the set S of sorts is finite or infinite. However, we have to admit that our example below is rather artificial and pathological, and we do not see right now one which is more likely to be realized in computer science:

<u>sorts:</u> $\underline{n}$ for each natural number n
<u>operations:</u> $\varphi_{\underline{n}}$, $\varphi'_{\underline{n}} :\rightarrow \underline{n}$ for each n.

Consider a set $\{\underline{A}_n | n \in \mathrm{N}\}$ of partial algebras, where — using the usual many-sorted terminology —

$$A_{n,m} := \begin{cases} \{\varphi_{\underline{m}}, \varphi'_{\underline{m}}\}, & \text{if } m \neq n, \\ \emptyset, & \text{if } m = n, \end{cases}$$

$$\varphi_{\underline{n}}^{\underline{A}_m} := \varphi_{\underline{n}},\ \varphi'^{\underline{A}_m}_{\underline{n}} := \varphi'_{\underline{n}},\ \text{if } m \neq n,$$

and if $m = n$, then $\varphi_{\underline{n}}^{\underline{A}_m}$ and $\varphi'^{\underline{A}_m}_{\underline{n}}$ are undefined. Let us consider the class

$$\mathfrak{K} := \mathcal{HSP}\{\underline{A}_n | n \in \mathrm{N}\}.$$

Then the only partial algebras in $\mathfrak{K}$, which have non-empty phyla for all $n \in \mathrm{N}$, are isomorphic to the sort algebra for this signature, i.e. $\mathfrak{K}$ satisfies the existence-equations

$$(*) \quad (x_n, y_n : \underline{n}(n \in \mathrm{N}); x_m \stackrel{e}{=} y_m)$$

for each natural number m, while the category theoretical reduced product of the family $(\underline{A}_n)_{n \in \mathrm{N}}$ with respect to the Fréchet filter $\mathcal{F}$ of cofinite subsets of N:

$$\mathcal{F} = \{E | E \subseteq \mathrm{N} \text{ and } \mathrm{N} \setminus E \text{ is finite}\}$$

(i.e. the direct limit of the directed system

$$(\prod_{n \in E} \underline{A}_n, pr_{E,E'} : \prod_{n \in E} \underline{A}_n \to \prod_{n \in E'} \underline{A}_n | E, E' \in \mathcal{F}, E \supseteq E')$$

where $pr_{E,E'}(a_n | n \in E) := (a_n | n \in E'))$ is containing the algebra $\underline{B}$ as subalgebra, where $B_n = \{\varphi_{\underline{n}}, \varphi'_{\underline{n}}\}$, and $\varphi_{\underline{n}}^{\underline{B}} = \varphi_{\underline{n}}$, $\varphi'^{\underline{B}}_{\underline{n}} = \varphi'_{\underline{n}}$ for each $n \in \mathrm{N}$. This shows that $\underline{B} \notin \mathfrak{K}$, i.e. $\mathfrak{K}$ is not closed w.r.t. direct limits, since these do not preserve existence-equations of the form $(*)$ which are actually equivalent to implications with an infinite conjunction of existence-equations in the premise (according to our earlier translation).

In [B86] the approach to many-sorted (partial) algebras via the comma category $\underline{Alg(\tau)} \downarrow \{\underline{S}\}$ has been carried through in parallel to the development of the theory of partial algebras.

In the introduction ("Motivation") we have already indicated the fact that relations can very easily be treated as partial operations. This observation again indicates that structures containing total operations, partial operations and/or relations can be treated in a unified way within the theory of partial algebras. Namely, in order to specify that an n-ary operation symbol, say φ, has to represent a total operation one just has to require the term-existence statement

$$\varphi x_1 \ldots x_n \stackrel{e}{=} \varphi x_1 \ldots x_n.$$

And that an n-ary operation symbol, say ψ, has to represent an n-ary relation may be expressed by the ECE-equation, say,

$$\psi x_1 \ldots x_n \stackrel{e}{=} \psi x_1 \ldots x_n \to \psi x_1 \ldots x_n \stackrel{e}{=} x_1,$$

if the context does not yield any "better" axioms.

This shows that the unified representation needs additional axioms, while a distinction within the type specification usually needs the discussion of several cases in the presentation or proof of statements.

Since "one has to pay for everything" it is just a matter of taste which approach one is preferring — the same naturally also applies to the representation of many-sorted partial algebras.

References

[AN79] Andrka, H., Nmeti, I., Injectivity in categories to represent all first order formulas. I., *Demonstratio Math.* **12** (1979), 717–732.

[AN82] Andréka, H., Németi, I., A general axiomatizability theorem formulated in terms of cone-injective subcategories, in: *Universal Algebra* (Proc. Coll. Esztergom 1977), Colloq. Math. Soc. János Bolyai **29**, North-Holland, Amsterdam, 1982, 13–35.

[AN83] Andréka, H., Németi, I., Generalization of the concept of variety and quasivariety to partial algebras through category theory, *Dissertationes Math.* (*Rozprawy Mat.*) **204**, Warszawa, 1983.

[B82] Burmeister, P., Partial algebras - survey of a unifying approach towards a two-valued model theory for partial algebras, *Algebra Universalis* **15** (1982), 306–358.

[B86] Burmeister, P., *A Model Theoretic Oriented Approach to Partial Algebras. Introduction to Theory and Application of Partial Algebras – Part I*, Mathematical Research Vol. 32, Akademie-Verlag, Berlin, 1986. (Part II: [Re84]).

[Cr89] Craig, W., Near-equational and equational systems of logic for partial functions. I and II, *J. Symbolic Logic* **54** (1989), 759–827, and 1181–1215.

[Ev51] Evans, T., Embeddability and the word problem, *J. London Math. Soc.* **26** (1951), 64–71.

[GoTcWa78] Goguen, J.A., Thatcher, J.W., Wagner, E.G., An initial algebra approach to the specification, correctness, and implementation of abstract data types (IBM Research Report RC-6487, 1976) in: *Current Trends in Programming Methodology, IV: Data Structuring* (R. Yeh, ed.), Prentice Hall, 1978, 80–144.

[GSchm63] Grätzer, G., Schmidt, E.T., Characterization of congruence lattices of abstract algebras, *Acta Sci. Math.* (*Szeged*) **24** (1963), 34–59.

[HS73] Herrlich, H., Strecker, G.E., *Caterogy Theory – An Introduction*, Allyn and Bacon, 1973 (2nd ed.: Heldermann-Verlag).

[Ho70] Höft, H., *Equations in Partial Algebras*, Ph.D. Thesis, University of Houston, 1970.

[Ho73] Höft, H., Weak and strong equations in partial algebras, *Algebra Universalis* **3** (1973), 203–215.

[J75] John, R., *Gültigkeitsbegriffe für Gleichungen in partiellen Algebren*, Dissertation, TH Darmstadt, 1975.

[J78] John, R., Gültigkeitsbegriffe für Gleichungen in partiellen Algebren, *Math. Z.* **159** (1978), 25–35.

[Ke70] Kerkhoff, R., Gleichungsdefinierbare Klassen partieller Algebren, *Math. Ann.* **185** (1970), 112–133.

[Kl52] Kleene, S.C., *Introduction to Metamathematics*, North-Holland, Amsterdam, 1952.

[KoSp68] Kochen, S., Specker, E.P., The problem of hidden variables in Quantum Mechanics, *J. Math. Mech.* **17** (1968), 59–87.

[Kos90] Kosiuczenko, P., Mal'cev type conditions for partial algebras, Manuscript, Politechnika Warszawska 1990, submitted to *Algebra Universalis.*

[La69] Lampe, W.A., *On Related Structures of a Universal Algebra*, Ph.D. Thesis, Pennsylvania State University, 1969.

[Mar58] Marczewski, E., A general scheme of independence in mathematics, *Bull. Acad. Polon. Sci. Sér. Math. Astr. Phys.* **6** (1958), 731–736.

[NSa82] Németi, I., Sain, I., Cone-implicational subcategories and some Birkhoff-type theorems, in: *Universal Algebra* (Proc. Coll. Esztergom 1977), Colloq. Math. Soc. János Bolyai **29**, North-Holland, Amsterdam, 1982, 535–578.

[Pa79] Pasztor, A., *Faktorisierungssysteme in der Kategorie der partiellen Algebren – Kennzeichnung von (Homo-)Morphismen*, Dissertation, TH Darmstadt, 1979; appeared in: Hochschulverlag, Freiburg, 1979.

[Re84] Reichel, H., *Structural Induction on Partial Algebras. Introduction to Theory and Applications of Partial Algegras – Part II*, Mathematical Research, Vol. 18, Akademie-Verlag, Berlin, 1984. (Part I: [B86]).

[Ru83] Rudak, L., A completeness theorem for weak equational logic, *Algebra Universalis* **16** (1983), 331–337.

[Sch62] Schmidt, J., Die Charakteristik einer Allgemeinen Algebra I., *Arch. Math.* **13** (1962), 457–470.

[Sch64] Schmidt, J., Die Charakteristik einer Allgemeinen Algebra II., *Arch. Math.* **15** (1964), 286–301.

[Sch70] Schmidt, J., A homomorphism theorem for partial algebras, *Colloq. Math.* **21** (1970), 5–21.

[Sl64] Słomiński, J., A theory of extensions of quasi-algebras to algebras, *Rozprawy Mat.* **40** (1964).

[S72] Strecker, G.E., Epireflection operators vs perfect morphisms and closed classes of epimorphisms, *Bull. Austral. Math. Soc.* **7** (1972), 359–366.

Duality Theory on Ten Dollars a Day*

Brian A. DAVEY
Department of Mathematics
La Trobe University
Bundoora, Victoria 3083
Australia

Abstract

Duality theory grew out of two classical special cases—Pontryagin's duality for abelian groups and Stone's duality for Boolean algebras. In the late 1960s and early 1970s, it was further fertilized by Priestley's duality for distributive lattices and the Hofmann-Mislove-Stralka duality for semilattices. Until the early 1970s, general approaches to duality theory were firmly rooted in category theory. The study of duality theory within general algebra began in the mid-1970s but blossomed in the 1980s. This paper presents an overview of duality theory from 1980 up to early 1992 as seen through the eyes of an algebraist. The presentation is in the style of a travel guide and is aimed at beginning graduate students. A minimum of general algebra and topology is assumed and category theory is completely avoided.

Many young travellers in the realm of general algebra find that the signposts along the road to duality theory point in directions which they would not, of their own accord, choose to travel: to the limits of category theory, to topology's tortuous terrain, to the myriad byways of unfamiliar examples. For them, and perhaps for a few of the not-so-young, we offer this traveller's guide. Here they will find a low cost yet comprehensive tour of the field, avoiding category theory and keeping excursions into topology to a minimum. Our tour is aimed at beginning graduate students who have already completed a first course in topology (up to compactness) and a first course in general algebra (up to Birkhoff's theorems on free algebras, varieties and subdirect representations). Those who would prefer a more comprehensive guide book, including the category-theoretic requisites as well as examples of dualities in action, are referred to the forthcoming monograph Clark and Davey [4] of which this tour guide is a zeroth draft.

To ensure that we are in shape for the longer journey into general algebra, we begin with three day trips into more familiar territory: abelian groups, Boolean algebras and distributive lattices. The reader, and especially the first-time traveller, is warned that the commentary during the guided tour of abelian groups will contain a lot of important chit-chat which will not be repeated during the other two trips.

Abelian groups Denote the class of abelian groups by $\mathcal{A}$. The *circle group* is the subgroup $T := \{ z \in \mathbb{C} : |z| = 1 \}$ of the group of nonzero complex numbers under multipli-

*Dedicated to the memory of Evelyn Nelson and Alan Day

I. G. Rosenberg and G. Sabidussi (eds.), Algebras and Orders, 71–111.

cation. For each abelian group A we denote the set of all homomorphisms $x : A \to T$ by $\mathcal{A}(A,T)$. As will soon become apparent, such *homsets* play a crucial role in duality theory.

There is a natural map

$$e : A \to T^{\mathcal{A}(A,T)}, \text{ given by } (e(a))(x) := x(a)$$

for all $a \in A$ and all $x \in \mathcal{A}(A,T)$. We say that the map e is *given by evaluation* since, for each $a \in A$, the map $e(a) : \mathcal{A}(A,T) \to T$ is given by the rule *"evaluate at a"*. It is easily seen that e is a homomorphism. Indeed, since each $x \in \mathcal{A}(A,T)$ is a homomorphism and since the operation on a power of T is pointwise, we have

$$(e(a \cdot b))(x) = x(a \cdot b) = x(a) \cdot x(b) = (e(a))(x) \cdot (e(b))(x) = (e(a) \cdot e(b))(x)$$

for each $x \in \mathcal{A}(A,T)$ and hence $e(a \cdot b) = e(a) \cdot e(b)$ for all $a, b \in A$. It is a fundamental fact about abelian groups that, if $A \in \mathcal{A}$ and $a, b \in A$ with $a \neq b$, then there exists a homomorphism $x : A \to T$ with $x(a) \neq x(b)$. In other words, if $a \neq b$ in A, then there exists $x \in \mathcal{A}(A,T)$ such that $(e(a))(x) \neq (e(b))(x)$ and thus $e(a) \neq e(b)$. Hence e is an embedding. Consequently, every abelian group is isomorphic to a subgroup of a power of T. Using the usual class operators, $\mathbb{I}$ (all isomorphic copies of), $\mathbb{S}$ (all subgroups of) and $\mathbb{P}$ (all products of), we have $\mathcal{A} = \mathbb{ISP}(T)$.

Thus we have represented each abelian group A as a group of functions: the group A is isomorphic to the group

$$\{ e(a) : \mathcal{A}(A,T) \to T \mid a \in A \}$$

of evaluations maps. This representation would be greatly strengthened if we had some intrinsic description of the evaluation maps. We wish to find some property (expressed in terms of the sets $\mathcal{A}(A,T)$ and T with no reference to the elements of A) which will distinguish the evaluation maps in the set of all maps $\varphi : \mathcal{A}(A,T) \to T$. This search is at the heart of duality theory.

First note that T inherits a topology $\mathcal{T}$ from $\mathbb{C}$. In fact, $\langle T; \cdot, ^{-1}, 1, \mathcal{T} \rangle$ is a compact topological group. We impose the product topology on the power T^A: sets of the form

$$[a : V] := \{ u : A \to T \mid u(a) \in V \},$$

where $a \in A$ and V is open in T, form a subbase for the product topology on T^A. By Tychonoff's Theorem (a product of compact spaces is compact), T^A is compact. It is an easy exercise to see that the set $\mathcal{A}(A,T)$ of homomorphisms is a closed subspace of T^A. So let's do it!

Let $u : A \to T$ and assume that u is not a homomorphism. Thus there exist $a, b \in A$ such that $u(a \cdot b) \neq u(a) \cdot u(b)$. Since the topology on T is Hausdorff, there exist open sets U and V in T such that $u(a \cdot b) \in U$, $u(a) \cdot u(b) \in V$ and $U \cap V = \emptyset$. Since multiplication on T is continuous, there exist open sets V_a and V_b in T such that $u(a) \in V_a$, $u(b) \in V_b$ and $V_a \cdot V_b \subseteq V$. Thus

$$W := [ab : U] \cap [a : V_a] \cap [b : V_b]$$

is an open set in T^A which contains u. Moreover, if $v \in W$, then $v(a \cdot b) \in U$ and $v(a) \cdot v(b) \in V_a \cdot V_b \subseteq V$, whence $v(a \cdot b) \neq v(a) \cdot v(b)$ as $U \cap V = \emptyset$. Thus W is disjoint from $\mathcal{A}(A,T)$ and hence $\mathcal{A}(A,T)$ is closed in T^A.

Since $\mathcal{A}(A,T)$ is a closed subspace of the compact space T^A, it follows immediately that $\mathcal{A}(A,T)$ is also compact. It is a triviality that the evaluation maps are continuous with respect to this topology: if U is open in T and $a \in A$, then

$$\begin{aligned} e(a)^{-1}(U) &= \{\, x \in \mathcal{A}(A,T) \mid (e(a))(x) \in U \,\} \\ &= \{\, x \in \mathcal{A}(A,T) \mid x(a) \in U \,\} \\ &= \mathcal{A}(A,T) \cap [a : U] \end{aligned}$$

which is open in $\mathcal{A}(A,T)$. Nevertheless, the evaluation maps are not the only continuous maps from $\mathcal{A}(A,T)$ to T.

In order to distinguish the evaluation maps we must impose further structure on $\mathcal{A}(A,T)$. Note that $\mathcal{A}(A,T)$ is closed under the pointwise multiplication: if $x, y \in \mathcal{A}(A,T)$, then $x \cdot y \in \mathcal{A}(A,T)$ since, for all $a, b \in A$,

$$\begin{aligned} (x \cdot y)(a \cdot b) &= x(a \cdot b) \cdot y(a \cdot b) \\ &= (x(a) \cdot x(b)) \cdot (y(a) \cdot y(b)) \\ &= (x(a) \cdot y(a)) \cdot (x(b) \cdot y(b)) \\ &= (x \cdot y)(a) \cdot (x \cdot y)(b) \end{aligned}$$

which says that $x \cdot y$ is a homomorphism. The crucial identity in this calculation is

$$(s \cdot t) \cdot (u \cdot v) = (s \cdot u) \cdot (t \cdot v)$$

which says precisely that multiplication on T, regarded as a map from T^2 to T, is a homomorphism. The set $\mathcal{A}(A,T)$ contains the identity element of T^A, namely the constant map onto $\{1\}$, because $\{1\}$ is a one-element subgroup of T. Finally, $\mathcal{A}(A,T)$ is closed under forming inverses: if $x : A \to T$ is a homomorphism, then $x^{-1} : A \to T$ (defined pointwise) is also a homomorphism since, for all $a, b \in A$,

$$x^{-1}(a \cdot b) = (x(a \cdot b))^{-1} = (x(a) \cdot x(b))^{-1} = x(a)^{-1} \cdot x(b)^{-1} = x^{-1}(a) \cdot x^{-1}(b).$$

Again, the crucial identity, namely $(s \cdot t)^{-1} = s^{-1} \cdot t^{-1}$, says precisely that ${}^{-1} : T \to T$ is a homomorphism. Thus $\mathcal{A}(A,T)$ is a subgroup of T^A, and so we may add this natural pointwise group structure to the topology on $\mathcal{A}(A,T)$. Once more it is trivial that the evaluation maps preserve the additional structure. The evaluation $e(a) : \mathcal{A}(A,T) \to T$ is a homomorphism for each $a \in A$ since

$$(e(a))(x \cdot y) = (x \cdot y)(a) = x(a) \cdot y(a) = (e(a))(x) \cdot (e(a))(y)$$

for all $x, y \in \mathcal{A}(A,T)$.

To summarize, $\mathcal{A}(A,T)$ is a closed subgroup of T^A (and therefore is a compact topological group) and, for each $a \in A$, the evaluation map $e(a) : \mathcal{A}(A,T) \to T$ is a continuous homomorphism. It is a surprising and highly nontrivial result that the evaluation maps are the only continuous homomorphisms from $\mathcal{A}(A,T)$ to T. This is part of the Pontryagin duality for locally compact abelian groups [32]. Hence, in the case of abelian groups, we have found a natural intrinsic structure on $\mathcal{A}(A,T)$ and T—both are (compact) topological abelian groups—which distinguishes the evaluation maps. Thus every abelian group is isomorphic to the group of all continuous homomorphisms from some compact topological abelian group into the circle group T. The compact topological abelian group $\mathcal{A}(A,T)$ is called the *dual of* A. We shall denote it by $D(A)$.

At this point it is important to draw the distinction between a *representation theory* for a class $\mathcal{A}$ of algebras and a *duality theory* for $\mathcal{A}$. What we have described so far is a representation theory for the class $\mathcal{A}$ of abelian groups. To lift this up to a duality theory for $\mathcal{A}$ we must show that the representation respects homomorphisms (while turning them on their heads). If $A, B \in \mathcal{A}$ and $u : A \to B$ is a homomorphism, then the *dual of* u is the natural map $D(u) : D(B) \to D(A)$ defined by *"compose on the right with* u*"*, i.e.

$$(D(u))(x) := x \circ u \in D(A) = \mathcal{A}(A,T) \text{ for all } x \in D(B) = \mathcal{A}(B,T).$$

The map $D(u)$ is continuous since, if V is open in T and $a \in A$, then

$$D(u)^{-1}([a : V] \cap \mathcal{A}(A,T)) = [u(a) : V] \cap \mathcal{A}(B,T).$$

Moreover, $D(u)$ is a homomorphism since, for all $x, y \in \mathcal{A}(B,T)$ and all $a \in A$,

$$\begin{aligned} ((D(u))(x \cdot y))(a) &= ((x \cdot y) \circ u)(a) = (x \cdot y)(u(a)) = x(u(a)) \cdot y(u(a)) \\ &= (x \circ u)(a) \cdot (y \circ u)(a) = ((x \circ u) \cdot (y \circ u))(a) \\ &= ((D(u))(x) \cdot (D(u))(y))(a) \\ \text{whence} \quad (D(u))(x \cdot y) &= (D(u))(x) \cdot (D(u))(y). \end{aligned}$$

The picture we have painted so far during this brief excursion into Pontryagin duality has been intentionally one-sided. We commenced the trip with the cultural mind set of an algebraist for which we make no apology. Nevertheless, since the total picture is highly symmetrical, the other side warrants fuller description.

The duals $D(A)$ for $A \in \mathcal{A}$ need a home. Since, by construction, each $D(A) = \mathcal{A}(A,T)$ is a closed subgroup of a power of T (regarded as a topological group), a natural choice for their home is the class $\mathbb{IS}_c\mathbb{P}(T)$ of all isomorphic copies of closed subgroups of powers of T. (A map will be an isomorphism in this context if it is simultaneously a group isomorphism and a topological homeomorphism.) Another natural choice would be the class $\mathcal{X}$ of all compact topological abelian groups. In fact, $\mathcal{X} = \mathbb{IS}_c\mathbb{P}(T)$ as we shall see once some further notation is established.

For each $X \in \mathcal{X}$, the homset $\mathcal{X}(X,T)$, consisting of the continuous homomorphisms from X to T, is a subgroup of T^X. The proof is identical to the proof given above that $\mathcal{A}(A,T)$ is a subgroup of T^A except that we must now observe that (a) if $\alpha, \beta : X \to T$ are continuous then $\alpha \cdot \beta : X \to T$ is continuous (since $\cdot : T^2 \to T$ is continuous), (b) if $\alpha : X \to T$ is continuous then $\alpha^{-1} : X \to T$ is continuous (since ${}^{-1} : T \to T$ is continuous), and (c) the constant map from A onto $\{1\} \subseteq T$ is continuous. Thus $\mathcal{X}(X,T) \in \mathcal{A}$. We refer to $\mathcal{X}(X,T)$ as the *dual* of X and denote it by $E(X)$. Just as the map $D : \mathcal{A} \to \mathcal{X}$ respects homomorphisms, it is very easily seen that the map $E : \mathcal{X} \to \mathcal{A}$ respects continuous homomorphisms (modulo turning them on their heads). If $X, Y \in \mathcal{X}$ and $\varphi : X \to Y$ is a continuous homomorphism, then the *dual of* φ is the natural map $E(\varphi) : E(Y) \to E(X)$ defined by *"compose on the right with* φ*"*, i.e.

$$(E(\varphi))(\alpha) := \alpha \circ \varphi \in E(X) = \mathcal{X}(X,T) \text{ for all } \alpha \in E(Y) = \mathcal{X}(Y,T).$$

We now have two natural maps given by evaluation: for all $A \in \mathcal{A}$ and $X \in \mathcal{X}$,

$$e : A \to ED(A) = \mathcal{X}(D(A), \underset{\sim}{T}) = \mathcal{X}(\mathcal{A}(A, \underline{T}), \underset{\sim}{T})$$

defined by $(e(a))(x) := x(a)$ for all $a \in A$ and $x \in \mathcal{A}(A, \underline{T})$, and

$$\varepsilon : X \to DE(X) = \mathcal{A}(E(X), \underline{T}) = \mathcal{A}(\mathcal{X}(X, \underset{\sim}{T}), \underline{T})$$

defined by $(\varepsilon(x))(\alpha) := \alpha(x)$ for all $x \in X$ and $\alpha \in \mathcal{X}(X, \underset{\sim}{T})$. While it is clear that $\mathbb{IS}_c\mathbb{P}(\underset{\sim}{T}) \subseteq \mathcal{X}$, the reverse inclusion is far from clear. The vital (and difficult) fact is that if X is a compact topological abelian group then there are enough continuous homomorphisms from X into T to separate the points of X, i.e. if $x \neq y$ in X, then there exists a continuous homomorphism $\alpha : X \to T$ such that $\alpha(x) \neq \alpha(y)$. From this it is easily seen that the map $\varepsilon : X \to DE(X)$ is an isomorphism of X onto a closed subgroup of a power of T. Thus $\mathbb{IS}_c\mathbb{P}(\underset{\sim}{T})$ is the class of all compact topological abelian groups.

As was discussed earlier, the map $e : A \to ED(A)$ is an isomorphism for all $A \in \mathcal{A}$. This is what we mean when we say that *we have a duality between* $\mathcal{A}$ *and* $\mathcal{X}$. In many applications this is all that is needed: each $A \in \mathcal{A}$ has a representation as $E(X)$ for some $X \in \mathcal{X}$, but X need not be unique up to isomorphism. If, in addition, the map $\varepsilon : X \to DE(X)$ is an isomorphism for all $X \in \mathcal{X}$, then we say that *the duality between* $\mathcal{A}$ *and* $\mathcal{X}$ *is full.* The Poltryagin duality between the class $\mathcal{A}$ of abelian groups and the class $\mathcal{X}$ of compact topological abelian groups is a full duality and hence every abelian group A can be represented as the group of continuous homomorphisms from a unique-up-to-isomorphism compact topological abelian group into the circle group.

The circle group has a split personality. It lives in $\mathcal{A}$ as the abelian group $\underline{T} = \langle T; \cdot, ^{-1}, 1\rangle$ and in $\mathcal{X}$ as the compact topological group $\underset{\sim}{T} = \langle T; \cdot, ^{-1}, 1, \mathcal{T}\rangle$. As we shall see, this schizophrenic behaviour is completely typical within duality theory. In general, our choice of notation will make it clear which class an object belongs to: A, B, C for groups in $\mathcal{A}$ and X, Y, Z for topological groups in $\mathcal{X}$. But, to make it clear which role the circle group is playing, we shall henceforth use the $\underline{T}$ versus $\underset{\sim}{T}$ notation.

(A professional compositor would probably cringe at the use of the "underscore" and "twiddle" rather than some change of font. In our defense we make three points: (a) it is essential in this context to distinguish the two roles of the circle group and the $\underline{T}$ versus $\underset{\sim}{T}$ usage is far clearer than some subtle change of font, and moreover can be reproduced on a blackboard, (b) the underscore–twiddle notation is already well established in the literature, (c) since this is being produced in TeX, the author and the compositor are one and the same so no argument arises!)

Boolean algebras The dual of a Boolean algebra $\langle A; \vee, \wedge, ', 0, 1\rangle$ is usually defined to be the set $\mathcal{U}(A)$ of ultrafilters of A endowed with the topology generated by the sets of the form

$$\mathcal{U}_a := \{\, F \in \mathcal{U}(A) \mid a \in F \,\}$$

for $a \in A$. Note that $\mathcal{U}_a$ is clopen (i.e. both closed and open) in this topology since

$$X \backslash \mathcal{U}_a = \{\, F \in \mathcal{U}(A) \mid a \notin F \,\} = \{\, F \in \mathcal{U}(A) \mid a' \in F \,\} = \mathcal{U}_{a'}$$

which is a basic open set. Stone's duality for Boolean algebras [33] (or see [14]) asserts, in part, that the map $e : a \mapsto \mathcal{U}_a$ is an isomorphism of A onto the Boolean algebra of clopen subsets of $\mathcal{U}(A)$. Our task now is to see that this can be expressed naturally in terms of homsets in a manner strictly analogous to what we observed during our day trip into the Pontryagin duality for abelian groups. The role of the circle group $\underline{T}$ will now be played by the two-element Boolean algebra $\underline{2} = \langle\{0,1\}; \vee, \wedge, ', 0, 1\rangle$ while the topological group $\underset{\sim}{T}$ will be replaced by a much simpler object, namely $\underset{\sim}{2} = \langle\{0,1\}; \mathcal{T}\rangle$ where $\mathcal{T}$ is the discrete topology.

Let $\mathcal{B}$ denote the class of all Boolean algebras. For any subset F of A we define a map $\chi_F : A \to \{0,1\}$, the *characteristic function* of F, by

$$\chi_F(a) := \begin{cases} 1 & \text{if } a \in F, \\ 0 & \text{if } a \notin F. \end{cases}$$

It is easily seen that F is a prime filter of the Boolean algebra A if and only if χ_F is a lattice homomorphism onto $\underline{2}$. But a filter of a Boolean algebra A is prime if and only if it is an ultrafilter, and a lattice homomorphism from A onto $\underline{2}$ is automatically a Boolean algebra homomorphism. Thus $\varphi : F \mapsto \chi_F$ is a bijection between the set $\mathcal{U}(A)$ of ultrafilters of A and the set $\mathcal{B}(A, \underline{2})$ of all Boolean algebra homomorphisms $x : A \to \underline{2}$. A simple modification of the proof for the circle group shows that the natural map

$$e : A \to \underline{2}^{\mathcal{B}(A,\underline{2})}, \text{ given by } (e(a))(x) := x(a)$$

for all $a \in A$ and all $x \in \mathcal{B}(A, \underline{2})$, is a homomorphism. The Boolean Ultrafilter Theorem says precisely that if $a \neq b$ in A, then there exists an ultrafilter F of A which contains exactly one of a and b. Thus, taking $x = \chi_F$, we have

$$(e(a))(x) = x(a) = \chi_F(a) \neq \chi_F(b) = x(b) = (e(b))(x)$$

and consequently e is an embedding. Hence $\mathcal{B} = \mathbb{ISP}(\underline{2})$.

By mimicing the proof for the circle group, it is easily seen that $\mathcal{B}(A, \underline{2})$ is a closed subspace of the product space $\underset{\sim}{2}^A$. (All that is needed is that the topology on $\underset{\sim}{2}$ is Hausdorff and that the Boolean algebra operations on $\underline{2}$ are continuous with respect to the topology on $\underset{\sim}{2}$: both are trivially true since the topology on $\underset{\sim}{2}$ is discrete.) Recall that if $a \in A$ and $V \subseteq 2 = \{0,1\}$, then

$$[a : V] := \{\, u : A \to T \mid u(a) \in V \,\}.$$

Since

$$\varphi(\mathcal{U}_a) = [a : \{1\}] \cap \mathcal{B}(A, \underline{2}) \text{ and } \varphi(\mathcal{U}_{a'}) = [a : \{0\}] \cap \mathcal{B}(A, \underline{2}),$$

for all $a \in A$, the map $\varphi : \mathcal{U}(A) \to \mathcal{B}(A, \underline{2})$ is a homeomorphism. Thus we may define the *dual*, $D(A)$, to be the compact topological space $\mathcal{B}(A, \underline{2})$ with its topology inherited as a subspace of the power $\underset{\sim}{2}^A$.

As a home for the dual spaces $D(A)$ for $A \in \mathcal{B}$ we take the class $\mathcal{X} := \mathbb{IS}_c\mathbb{P}(\underset{\sim}{2})$ of all isomorphic (i.e. homeomorphic) copies of closed subspaces of powers of the two-element discrete space $\underset{\sim}{2}$. For each $X \in \mathcal{X}$, the homset $\mathcal{X}(X, \underset{\sim}{2})$ of all continuous maps from X

into $\underset{\sim}{2}$ is a subalgebra of $\underline{2}^X$ and hence $\mathcal{X}(X, \underset{\sim}{2}) \in \mathcal{B}$. Thus we define the *dual* of X to be $E(X) := \mathcal{X}(X, \underset{\sim}{2})$, a subalgebra of $\underline{2}^X$. Note that a subset U of X is clopen if and only if its characteristic function χ_U is continuous and hence $E(X)$ is isomorphic to the Boolean algebra of clopen subsets of X.

We leave it to the reader to define the dual $D(u) : D(B) \to D(A)$ of a homomorphism $u : A \to B$ and the dual $E(\varphi) : E(Y) \to E(X)$ of a continuous map $\varphi : X \to Y$. (Just replace $\mathcal{A}$ by $\mathcal{B}$ and T by either $\underline{2}$ or $\underset{\sim}{2}$ in the definition given in the abelian group case.)

As in the abelian group case, we have two natural maps given by evaluation: for all $A \in \mathcal{B}$ and all $X \in \mathcal{X}$,

$$e : A \to ED(A) = \mathcal{X}(D(A), \underset{\sim}{2}) = \mathcal{X}(\mathcal{B}(A, \underline{2}), \underset{\sim}{2}),$$

defined by $(e(a))(x) := x(a)$ for all $a \in A$ and $x \in \mathcal{B}(A, \underline{2})$, and

$$\varepsilon : X \to DE(X) = \mathcal{B}(E(X), \underline{2}) = \mathcal{B}(\mathcal{X}(X, \underset{\sim}{2}), \underline{2}),$$

defined by $(\varepsilon(x))(\alpha) := \alpha(x)$ for all $x \in X$ and $\alpha \in \mathcal{X}(X, \underset{\sim}{2})$. The fact that A is isomorphic to the Boolean algebra of clopen subsets of $\mathcal{U}(A)$ translates into the statement that the map $e : A \to ED(A)$ is an isomorphism. In this case, all that is needed to distinguish the evaluation maps is the topology on $\mathcal{B}(A, \underline{2})$ and on $\underset{\sim}{2}$: a map $u : \mathcal{B}(A, \underline{2}) \to \underset{\sim}{2}$ is an evaluation map $e(a)$ for some $a \in A$ if and only if it is continuous. Thus we have a *duality* between $\mathcal{B}$ and $\mathcal{X}$. In fact the duality is *full*, i.e. $\varepsilon : X \to DE(X)$ is also an isomorphism for each $X \in \mathcal{X}$.

It is natural to ask for an axiomatization of the class $\mathcal{X}$. While not all applications of a duality require an axiomatization of the dual structures, the utility of the duality is greatly increased if we have such an axiomatization. It is a very easy exercise to see that $X \in \mathcal{X} = \mathbb{IS}_c\mathbb{P}(\underset{\sim}{2})$ if and only if X is a compact Hausdorff space which is *totally disconnected* (i.e. has a basis of clopen sets). Such spaces are referred to as *Stone spaces* or *Boolean spaces*.

That completes our second day trip. Much of what we have seen so far has an air of general algebra about it. We can already take a step back and survey the scene at a higher level. We need a class of algebras of the form $\mathcal{A} = \mathbb{ISP}(\underline{P})$ for some algebra $\underline{P} = \langle P; F\rangle$. The algebra $\underline{P}$ should have a compact topology $\mathcal{T}$ with respect to which each operation $f \in F$ is continuous. We will define $\mathcal{X}$ to be $\mathbb{IS}_c\mathbb{P}(\underset{\sim}{P})$ and define the dual of $A \in \mathcal{A}$ to be $D(A) := \mathcal{A}(A, \underline{P})$ regarded as a substructure of $\underset{\sim}{P}^A$ where $\underset{\sim}{P} = \langle P; ???, \mathcal{T}\rangle$. Unfortunately, with only two examples under our belts, it is not yet clear what structure will be appropriate on $\underset{\sim}{P}$. We need an example where the general framework is the same but where the second personality, $\underset{\sim}{P}$, of the algebra $\underline{P}$ has a character quite different from the strongly algebraic nature of $\underset{\sim}{T} = \langle T; \cdot, ^{-1}, 1, \mathcal{T}\rangle$ and the purely topological nature of $\underset{\sim}{2} = \langle\{0,1\}; \mathcal{T}\rangle$. Hence we commence our third and final day trip.

Distributive lattices Just as the dual of a Boolean algebra is usually defined in terms of ultrafilters, the dual of a bounded distributive lattice $\langle A; \vee, \wedge, 0, 1\rangle$ is usually defined in terms of prime filters. We may define the dual of A to be the set $\mathcal{F}(A)$ of prime filters of A. As in the Boolean case, we endow $\mathcal{F}(A)$ with a topology $\mathcal{T}$: take the sets

$$\mathcal{F}_a := \{\, F \in \mathcal{F}(A) \mid a \in F \,\},$$

where $a \in A$, and their complements as a subbase for $\mathcal{T}$. We also order $\mathcal{F}(A)$ by set inclusion. Thus the dual of A is the ordered topological space $\langle \mathcal{F}(A); \subseteq, \mathcal{T}\rangle$. According to Priestley's duality for the class $\mathcal{D}$ of bounded distributive lattices [28, 29] (or see [14]), $e : a \mapsto \mathcal{F}_a$ is an isomorphism of A onto the lattice of clopen increasing subsets of $\mathcal{F}(A)$. As in the Boolean case, this translates easily into a statement about homsets and evaluation maps.

We are now in fairly familiar territory. Let $\underline{2} = \langle\{0,1\}; \vee, \wedge, 0, 1\rangle$ be the two-element bounded distributive lattice. Once again, a very simple calculation shows that the natural map

$$e : A \to \underline{2}^{\mathcal{D}(A,\underline{2})}, \text{ given by } (e(a))(x) := x(a)$$

for all $a \in A$ and all $x \in \mathcal{D}(A,\underline{2})$, is a homomorphism. The Distributive Prime Ideal Theorem guarantees that if $a \neq b$ in A, then there exists a prime filter F which contains exactly one of a and b. Thus the characteristic function $\chi_F : A \to \{0,1\}$ separates a and b. The argument given for Boolean algebras applies without change, whence e is an embedding. Consequently, $\mathcal{D} = \mathbb{ISP}(\underline{2})$.

Let $\underset{\sim}{2} = \langle\{0,1\}; \leqslant, \mathcal{T}\rangle$ be the two-element chain with the discrete topology. Since $\mathcal{D}(A,\underline{2})$ is a closed subspace of $\underset{\sim}{2}^A$, it inherits both a compact topology and an order from the power $\underset{\sim}{2}^A$. It is a very easy exercise to show that $\varphi : F \mapsto \chi_F$ is a homeomorphism and an order-isomorphism between $\mathcal{F}(A)$ and $\mathcal{D}(A,\underline{2})$. Thus we define the *dual* of A to be the ordered Boolean space $\mathcal{D}(A,\underline{2})$. The algebraic half of Priestley duality can now be reformulated in terms of homsets as: *for every bounded distributive lattice A, the evaluation maps $e(a)$ for $a \in A$ are the only continuous, order-preserving maps from $\mathcal{D}(A,\underline{2})$ into $\underset{\sim}{2}$.*

The natural home for the duals $D(A)$ for $A \in \mathcal{D}$ is the class $\mathcal{X} := \mathbb{IS}_c\mathbb{P}(\underset{\sim}{2})$ of all isomorphic (i.e. simultaneously homeomorphic and order-isomorphic) copies of closed subsets of powers of $\underset{\sim}{2} = \langle\{0,1\}; \leqslant, \mathcal{T}\rangle$. For each $X \in \mathcal{X}$, the homset $\mathcal{X}(X, \underset{\sim}{2})$ of all continuous, order-preserving maps from X into $\underset{\sim}{2}$ is a $\{0,1\}$-sublattice of $\underline{2}^X$ and thus $\mathcal{X}(X, \underset{\sim}{2}) \in \mathcal{D}$. Although the proof is easy, it is essential for our further travels that we gauge the general-algebraic import of this observation.

Let $\alpha, \beta \in \mathcal{X}(X, \underset{\sim}{2})$. Then $\alpha \vee \beta : X \to \underset{\sim}{2}$ and $\alpha \wedge \beta : X \to \underset{\sim}{2}$ are continuous since $\vee : 2^2 \to 2$ and $\wedge : 2^2 \to 2$ are continuous. Moreover, $\alpha \vee \beta$ and $\alpha \wedge \beta$ are order-preserving since, for all $x, y \in X$,

$$\begin{aligned}
x \leqslant y \quad &\Longrightarrow \quad \alpha(x) \leqslant \alpha(y) \;\;\&\;\; \beta(x) \leqslant \beta(y) \;\text{ as }\; \alpha, \beta \in \mathcal{X}(X, \underset{\sim}{2}) \\
&\Longrightarrow \quad \alpha(x) \vee \beta(x) \leqslant \alpha(y) \vee \beta(y) \;\;\&\;\; \alpha(x) \wedge \beta(x) \leqslant \alpha(y) \wedge \beta(y) \\
&\Longrightarrow \quad (\alpha \vee \beta)(x) \leqslant (\alpha \vee \beta)(y) \;\;\&\;\; (\alpha \wedge \beta)(x) \leqslant (\alpha \wedge \beta)(y).
\end{aligned}$$

This calculation depends upon the fact that $\underline{2}$ satisfies

$$u \leqslant v \ \& \ s \leqslant t \implies u \vee s \leqslant v \vee t \ \& \ u \wedge s \leqslant v \wedge t$$

or, equivalently, (recalling that $\leqslant$ is a subset of 2^2),

$$(u,v) \in \leqslant \ \& \ (s,t) \in \leqslant \implies (u,v) \vee (s,t) \in \leqslant \ \& \ (u,v) \wedge (s,t) \in \leqslant .$$

This says precisely that $\leqslant$ is a sublattice of $\underline{2}^2$. The constant maps are in $\mathcal{X}(X, \underset{\sim}{2})$ since $(0,0) \in \leqslant$ and $(1,1) \in \leqslant$. Hence we have used the fact that $\leqslant$ is a $\{0,1\}$-sublattice of $\underline{2}^2$, i.e. that $\leqslant$ is a $\mathcal{D}$-subalgebra of $\underline{2}^2$.

We define the dual of X to be $E(X) := \mathcal{X}(X, \underset{\sim}{2})$, a subalgebra of $\underline{2}^X$. A subset U of X is clopen and increasing if and only if its characteristic function $\chi_U : X \to \underset{\sim}{2}$ is both continuous and order-preserving. Thus $E(X)$ is isomorphic to the lattice of clopen increasing subsets of X.

The maps $D : \mathcal{D} \to \mathcal{X}$ and $E : \mathcal{X} \to \mathcal{D}$ can be defined on morphisms via composition exactly as in the Boolean case. Of course, we once again have the two natural maps given by evaluation: for all $A \in \mathcal{D}$ and all $X \in \mathcal{X}$,

$$e : A \to ED(A) = \mathcal{X}(D(A), \underset{\sim}{2}) = \mathcal{X}(\mathcal{D}(A, \underline{2}), \underset{\sim}{2}),$$

defined by $(e(a))(x) := x(a)$ for all $a \in A$ and $x \in \mathcal{D}(A, \underline{2})$, and

$$\varepsilon : X \to DE(X) = \mathcal{D}(E(X), \underline{2}) = \mathcal{D}(\mathcal{X}(X, \underset{\sim}{2}), \underline{2}),$$

defined by $(\varepsilon(x))(\alpha) := \alpha(x)$ for all $x \in X$ and $\alpha \in \mathcal{X}(X, \underset{\sim}{2})$. Priestley duality tells us that we have a *full duality* between $\mathcal{D}$ and $\mathcal{X}$, i.e. $e : A \to ED(A)$ and $\varepsilon : X \to DE(X)$ are isomorphisms for all $A \in \mathcal{D}$ and $X \in \mathcal{X}$.

An ordered topological space X is call *totally order-disconnected* if, for all $x, y \in X$ with $x \nleqslant y$, there exists a clopen increasing subset U of X such that $x \in U$ but $y \notin U$. This is precisely the notion needed to axiomatize $\mathcal{X}$: an ordered topological space X is in $\mathcal{X} = \mathbb{IS}_c\mathbb{P}(\underset{\sim}{2})$ if and only if X is compact and totally order-disconnected. Such ordered topological spaces are often called *TODC spaces* or *Priestley spaces.*

Applications of Priestley's duality for $\mathcal{D}$ abound—see, for example, the survey articles Davey and Duffus [11] and Priestley [30].

Having completed our three day-trips, we are now ready to commence our guided tour of general duality theory. But before we do, we should address a fundamental question: *"Why bother?"* There are many reasons for developing a duality (of the type described in this guide) for your favourite class $\mathcal{A}$ of algebras. Here are a few.

- Once we have a duality for $\mathcal{A}$ we have a uniform way of representing each algebra $A \in \mathcal{A}$ as an algebra of continuous functions.

- If we have a full duality and have axiomatized the class $\mathcal{X}$, we can find examples of particular algebras in $\mathcal{A}$ by constructing objects in $\mathcal{X}$, which often turns out to be easier.

- Algebraic questions in $\mathcal{A}$ can be answered by translating them into (often simpler) questions in $\mathcal{X}$. For example,
    1. free algebras in $\mathcal{A}$ are easily described via their duals in $\mathcal{X}$,
    2. while a free product $A * B$ is often difficult to describe in $\mathcal{A}$, the dual, $D(A * B)$, is simply the cartesian product $D(A) \times D(B)$,
    3. congruence lattices in $\mathcal{A}$ may be studied by looking at lattices of closed substructures in $\mathcal{X}$,
    4. injective algebras in $\mathcal{A}$ may be characterized by first studying projective structures in $\mathcal{X}$
    5. algebraically closed and existentially closed algebras may be described via their duals.
- Some dualities have the particularly powerful property of being "logarithmic" in that they turn products into sums. For example, for both Boolean algebras and bounded distributive lattices we have $D(A \times B) \simeq D(A) \,\dot{\cup}\, D(B)$.

1 Setting the scene

Since we wish to cover as much ground as possible and to reach the more interesting destinations as quickly as possible, most of what follows will be presented without proof. Nevertheless, the odd proof will be given in order that the reader should taste some of the local flavour. In lieu of proofs, signposts to the literature will appear.

We begin by listing the general assumptions which will pertain for the remainder of our trip.

1. $\mathcal{A} := \mathbb{ISP}(\underline{P})$ is the quasivariety (= prevariety) generated by the non-trivial finite algebra $\underline{P} = \langle P; F\rangle$.
2. $\underset{\sim}{P} = \langle P; G, H, R, \mathcal{T}\rangle$ where
    (a) G is a set of (total) operations on P such that if $g \in G$ is nullary then $\{g\}$ is a subalgebra of $\underline{P}$ and if g is n-ary for $n \geqslant 1$ then $g : \underline{P}^n \to \underline{P}$ is a homomorphism;
    (b) H is a set of partial operations on P (of arity at least 1) such that if $h \in H$ is n-ary then the domain, $\mathrm{dom}(h)$, of h is a (non-empty) subalgebra of $\underline{P}^n$ and $h : \mathrm{dom}(h) \to \underline{P}$ is a homomorphism;
    (c) R is a set of finitary relations on P (of arity at least 1) such that if $r \in R$ is n-ary then r is a subalgebra of $\underline{P}^n$;
    (d) $\mathcal{T}$ is the discrete topology on P.
3. $\mathcal{X} := \mathbb{IS}_c\mathbb{P}(\underset{\sim}{P})$ is the class of all topological structures of the same type as $\underset{\sim}{P}$ which are isomorphic (i.e. simultaneously isomorphic and homeomorphic) to a closed substructure of a power of $\underset{\sim}{P}$.

We regard $\underset{\sim}{P}$ as an *alter ego* for the algebra $\underline{P}$. Whenever the conditions given in 2 above hold, we say that *the structure on* $\underset{\sim}{P}$ *is algebraic over* $\underline{P}$.

A couple of passing remarks are called for. If you wish to include examples where $\underline{P}$ is infinite, then you must assume that $\mathcal{T}$ is a compact Hausdorff topology and that each operation in F is continuous with respect to $\mathcal{T}$. This then provides a framework within which each of our three day-trip examples sits comfortably. For abelian groups, we let $\underline{P}$ be the circle group and take $G = \{\cdot, ^{-1}, 1\}$ and $H = R = \emptyset$. In the case of Boolean algebras, we let $\underline{P}$ be the two-element Boolean algebra and take $G = H = R = \emptyset$. To obtain the duality for bounded distributive lattices, we now let $\underline{P}$ be the two-element bounded distributive lattice and take $R = \{\leqslant\}$ and $G = H = \emptyset$. Fortunately, none of these examples required the use of partial operations. As we shall see, there are natural examples where partial operations are essential in order to obtain a full duality.

To avoid empty-headed nit picking, we shall henceforth insist that all products have a nonempty index set and that subalgebras and substructures are nonempty. While this has the advantage of eliminating a number of "ı"s which would otherwise have to be dotted, it has the disadvantage of excluding the one-element algebra from $\mathcal{A}$ whenever $\underline{P}$ has no one-element subalgebras. Whenever this happens, a very simple patch will rectify the situation: add all one-element algebras (of the same type as $\underline{P}$) to the class $\mathcal{A}$ and simultaneously add the empty structure (of the same type as $\underset{\sim}{P}$) to $\mathcal{X}$. For example, Priestley's duality for bounded distributive lattices applies to the one-element lattice—its dual is the empty ordered topological space. Finally, we should make quite clear what we intend by a substructure and by a power $\underset{\sim}{P}^S$. Let $(X; G_X, H_X, R_X, \mathcal{T}_X)$ and $(Y; G_Y, H_Y, R_Y, \mathcal{T}_Y)$ be topological structures of the same type as $\underset{\sim}{P}$; then Y is a *substructure* of X if (a) $\emptyset \neq Y \subseteq X$, (b) if $g_X \in G_X$ is n-ary and g_Y is the corresponding operation in G_Y, then $g_X(y_1, \ldots, y_n) = g_Y(y_1, \ldots, y_n)$ for all $y_1, \ldots, y_n \in Y$ (in particular, $g_X = g_Y$ if the operation is nullary), (c) if $h_X \in H_X$ is n-ary and h_Y is the corresponding partial operation in H_Y, then $\mathrm{dom}(h_Y) = \mathrm{dom}(h_X) \cap Y^n$ and $h_X(y_1, \ldots, y_n) = h_Y(y_1, \ldots, y_n)$ for all $(y_1, \ldots, y_n) \in \mathrm{dom}(h_Y)$, (d) if $r_X \in R_X$ is n-ary and r_Y is the corresponding relation in R_Y, then $r_Y = r_X \cap Y^n$, and (e) $\mathcal{T}_Y$ is the subspace topology induced on Y by $\mathcal{T}_X$. Note that it is possible for a partial operation in H_Y to have an empty domain even though the corresponding partial operation in H_X has a non-empty domain, and, likewise, a relation in R_Y may be empty while the corresponding relation in R_X is nonempty. Let $S \neq \emptyset$, then the operations, partial operations, relations and topology on the *power* $\underset{\sim}{P}^S$ are all defined in the obvious pointwise manner. In particular, the domain of an n-ary partial operation h on P^S is

$$\{\, (x_1, \ldots, x_n) \in (P^S)^n \mid (\forall s \in S)(x_1(s), \ldots, x_n(s)) \in \mathrm{dom}(h) \,\}$$

and, since $\mathcal{T}$ is the discrete topology, the sets

$$[s : a] := \{\, x \in P^S \mid x(s) = a \,\} \text{ with } s \in S \text{ and } a \in P$$

form a subbase for the topology on $\underset{\sim}{P}^S$.

The proofs of all results stated in the remainder of this section can be found in [19]. In fact, all proofs are of the straightforward, follow-your-nose variety.

If $X, Y \in \mathcal{X}$, then a continuous map $\varphi : X \to Y$ which preserves the operations, partial operations and relations will be called a *morphism*. (The map φ preserves an n-ary relation r if $(\forall a_1, \ldots, a_n \in X)\ (a_1, \ldots, a_n) \in r_X \implies (\varphi(a_1), \ldots, \varphi(a_n)) \in r_Y$. Note that we do not insist on the reverse implication.) The homset $\mathcal{X}(X, Y)$ consists of all morphisms from X to Y. Similarly, if $A, B \in \mathcal{A}$, then the homset $\mathcal{A}(A, B)$ consists of all homomorphisms from A to B. We shall denote the identity map on a set S by id_S.

We now show that the assumption that the structure on $\underset{\sim}{P}$ is algebraic over $\underline{P}$ allows us to mimic the homset approach to duality theory illustrated in our earlier discussion of abelian groups, Boolean algebras and bounded distributive lattices.

Lemma 1.1 *Assume that the structure on $\underset{\sim}{P}$ is algebraic over $\underline{P}$. For all $A \in \mathcal{A}$, define $D(A) := \mathcal{A}(A, \underline{P})$ and for all $A, B \in \mathcal{A}$ and each homomorphism $u : A \to B$ define a map $D(u) : D(B) \to D(A)$ by $(D(u))(x) := x \circ u$ for all $x \in D(B) = \mathcal{A}(B, \underline{P})$.*

1. *$D(A)$ is a closed substructure of $\underset{\sim}{P}^A$ and hence $D(A) \in \mathcal{X}$.*

2. *$D(u) : D(B) \to D(A)$ is a morphism in $\mathcal{X}$.*

3. *$D(\mathrm{id}_A) = \mathrm{id}_{D(A)}$ and if $u : A \to B$ and $v : B \to C$ are homomorphisms, then $D(v \circ u) = D(u) \circ D(v)$.*

Proof See Lemma 1.3 on page 116 of [19]. The fact that $\mathcal{A}(A, \underline{P})$ is closed in $\underset{\sim}{P}^A$ is a simple extension of the proof given earlier in the abelian group case and uses the fact that the operations on $\underline{P}$ are continuous with respect to the topology $\mathcal{T}$ on $\underset{\sim}{P}$. The proof that $\mathcal{A}(A, \underline{P})$ is closed under the (partial) operations in $G \cup H$ uses the assumption that the (partial) operations in $G \cup H$ are homomorphisms in $\mathcal{A}$. □

Thus the map $D = \mathcal{A}(-, \underline{P}) : \mathcal{A} \to \mathcal{X}$ is well defined on both algebras and homomorphisms. In proving that the map $E = \mathcal{X}(-, \underset{\sim}{P}) : \mathcal{X} \to \mathcal{A}$ is well defined we cannot avoid the assumption that the structure on $\underset{\sim}{P}$ is algebraic over $\underline{P}$.

Lemma 1.2 *The following conditions are equivalent:*

1. *for each $X \in \mathcal{X}$, the set $\mathcal{X}(X, \underset{\sim}{P})$ is a subalgebra of $\underline{P}^X$;*

2. *for each $n \in \mathrm{N}$, the set $\mathcal{X}(\underset{\sim}{P}^n, \underset{\sim}{P})$ is a subalgebra of $\underline{P}^{P^n}$;*

3. *for each $n \in \mathrm{N}$, every n-ary fundamental operation, and more generally every n-ary term function, f on $\underline{P}$, is a morphism $f : \underset{\sim}{P}^n \to \underset{\sim}{P}$ in $\mathcal{X}$;*

4. *the structure on $\underset{\sim}{P}$ is algebraic over $\underline{P}$.*

Proof See Lemma 1.1 on page 112 of [19]. □

Lemma 1.3 *Assume that the structure on $\underset{\sim}{P}$ is algebraic over $\underline{P}$. For all $X \in \mathcal{X}$, define $E(X) := \mathcal{X}(X, \underset{\sim}{P})$ and for all $X, Y \in \mathcal{X}$ and each morphism $\varphi : X \to Y$ define a map $E(\varphi) : E(Y) \to E(X)$ by $(E(\varphi))(\alpha) := \alpha \circ \varphi$ for all $\alpha \in E(Y) = \mathcal{X}(Y, \underset{\sim}{P})$.*

1. *$E(X)$ is a subalgebra of $\underline{P}^X$ and hence $E(X) \in \mathcal{A}$.*

2. *$E(\varphi): E(Y) \to E(X)$ is a homomorphism in $\mathcal{A}$.*

3. *$E(\mathrm{id}_X) = \mathrm{id}_{E(X)}$ and if $\varphi: X \to Y$ and $\psi: Y \to Z$ are morphisms, then $E(\psi \circ \varphi) = E(\varphi) \circ E(\psi)$.*

Proof See Lemma 1.4 on page 117 of [19]. □

It should come as no surprise that we can define the evaluation maps $e: A \to ED(A)$ and $\varepsilon: X \to DE(X)$, for all $A \in \mathcal{A}$ and all $X \in \mathcal{X}$, and that these maps are always embeddings. In $\mathcal{A}$, an embedding is simply a one-to-one homomorphism, but in $\mathcal{X}$ we require more. Let $X, Y \in \mathcal{X}$. Then a map $\varphi: X \to Y$ is called an *embedding* if it is an isomorphism of X onto a closed substructure of Y. Specifically, this means that φ is continuous, one-to-one, preserves the operations in G, satisfies

$$(x_1, \ldots, x_n) \in \mathrm{dom}(h_X) \iff (\varphi(x_1), \ldots, \varphi(x_n)) \in \mathrm{dom}(h_Y)$$

and

$$\varphi(h_X(x_1, \ldots, x_n)) = h_Y(\varphi(x_1), \ldots, \varphi(x_n)) \text{ for all } (x_1, \ldots, x_n) \in \mathrm{dom}(h_X)$$

for each partial operation h, and satisfies

$$(x_1, \ldots, x_n) \in r_X \iff (\varphi(x_1), \ldots, \varphi(x_n)) \in r_Y$$

for each relation r. Since X is compact, Y is Hausdorff and φ is continuous and one-to-one, it follows that φ is a homeomorphism onto a closed subspace.

Lemma 1.4 *Assume that the structure on $\underset{\sim}{P}$ is algebraic over $\underline{P}$. For all $A \in \mathcal{A}$ and $X \in \mathcal{X}$, the natural maps given by evaluation,*

$$e_A: A \to ED(A) = E(\mathcal{A}(A, \underline{P})) = \mathcal{X}(\mathcal{A}(A, \underline{P}), \underset{\sim}{P})$$

defined by $(e_A(a))(x) := x(a)$ for all $a \in A$ and $x \in \mathcal{A}(A, \underline{P})$, and

$$\varepsilon_X: X \to DE(X) = D(\mathcal{X}(X, \underset{\sim}{P})) = \mathcal{A}(\mathcal{X}(X, \underset{\sim}{P}), \underline{P})$$

defined by $(\varepsilon_X(x))(\alpha) := \alpha(x)$ for all $x \in X$ and $\alpha \in \mathcal{X}(X, \underset{\sim}{P})$, are embeddings. Moreover,

1. *for all $u \in \mathcal{A}(A, B)$ and $\varphi \in \mathcal{X}(X, Y)$ we have $ED(u) \circ e_A = e_B \circ u$ and $DE(\varphi) \circ \varepsilon_X = \varepsilon_Y \circ \varphi$, i.e. the diagrams in Figure 1 commute;*

2. *for all $u \in \mathcal{A}(A, E(X))$ there is a unique $\varphi \in \mathcal{X}(X, D(A))$, namely $\varphi = D(u) \circ \varepsilon_X$, such that $u = E(\varphi) \circ e_A$ (see Figure 2);*

3. *for all $\varphi \in \mathcal{X}(X, D(A))$ there is a unique $u \in \mathcal{A}(A, E(X))$, namely $u = E(\varphi) \circ e_A$, such that $\varphi = D(u) \circ \varepsilon_X$ (see Figure 2).*

$$\begin{tikzcd} A \arrow[r, "u"] \arrow[d, "e_A"'] & B \arrow[d, "e_B"] \\ ED(A) \arrow[r, "ED(u)"'] & ED(B) \end{tikzcd} \qquad \begin{tikzcd} X \arrow[r, "\varphi"] \arrow[d, "\varepsilon_X"'] & Y \arrow[d, "\varepsilon_Y"] \\ DE(X) \arrow[r, "DE(\varphi)"'] & DE(Y) \end{tikzcd}$$

Figure 1: $ED(u) \circ e_A = e_B \circ u$ and $DE(\varphi) \circ \varepsilon_X = \varepsilon_Y \circ \varphi$

$$\begin{tikzcd} A \arrow[r, "e_A"] \arrow[dr, "u"'] & ED(A) \arrow[d, "E(\varphi)"] \\ & E(X) \end{tikzcd} \qquad \begin{tikzcd} X \arrow[r, "\varepsilon_X"] \arrow[dr, "\varphi"'] & DE(X) \arrow[d, "D(u)"] \\ & D(A) \end{tikzcd}$$

Figure 2: $u = E(D(u) \circ \varepsilon_X) \circ e_A$ and $\varphi = D(E(\varphi) \circ e_A) \circ \varepsilon_X$

Proof See Lemma 1.5 on page 118 of [19]. □

Important notice Throughout the remaining sectors of this guided tour it will be taken for granted that the structure on $\underset{\sim}{P}$ is algebraic over $\underline{P}$ and this will no longer be stated explicitly.

For experienced travellers only! The categorical cognoscenti will have noted that Lemmas 1.1 and 1.3 say simply that $D : \mathcal{A} \to \mathcal{X}$ and $E : \mathcal{X} \to \mathcal{A}$ are contravariant functors, while Lemma 1.4 says that D and E are adjoint to each other with e and ε as the natural transformations which act as the units of the adjunction. (See [24] for the requisite categorical concepts.)

2 Duality theorems

Having set the scene, our guided tour may begin in earnest. We begin, where our algebraic hearts belong, in the class $\mathcal{A}$. Our first aim in developing a duality for the class $\mathcal{A} = \mathbb{ISP}(\underline{P})$ is to find some choice of G, H and R such that, when we choose $\underset{\sim}{P} = \langle P; G, H, R, \mathcal{T}\rangle$, the natural map given by evaluation,

$$e_A : A \to ED(A) = \mathcal{X}(D(A), \underset{\sim}{P}) = \mathcal{X}(\mathcal{A}(A, \underline{P}), \underset{\sim}{P}),$$

is an isomorphism for all $A \in \mathcal{A}$. Since, by Lemma 1.4, the map e_A is always an embedding, this says precisely that for each $A \in \mathcal{A}$ the evaluation maps are the only morphisms from

$\mathcal{A}(A, \underline{P})$ to $\underset{\sim}{P}$. If e_A is an isomorphism, we say that $\underset{\sim}{P}$ (or $G \cup H \cup R$) ***yields a duality on** A*. When $\underset{\sim}{P}$ yields a duality on each $A \in \mathcal{A}$, we say that $\underset{\sim}{P}$ (or $G \cup H \cup R$) ***yields a duality on** $\mathcal{A}$*. This view of duality theory puts the emphasis on the algebra $\underline{P}$ and the structure $\underset{\sim}{P}$, which in the view of your guide is precisely where the emphasis should be. Nevertheless, it would be reasonable for someone on this tour to pull their favourite class $\mathcal{C}$ of algebras from their pocket and ask the tour guide, *"Is there a duality for the class $\mathcal{C}$?"* We shall say that there is a ***natural duality*** for a class $\mathcal{C}$ of algebras if $\mathcal{C} = \mathbb{ISP}(\underline{P})$ for some finite algebra $\underline{P}$ and there is some structure $\underset{\sim}{P}$ which is algebraic over $\underline{P}$ and which yields a duality on $\mathcal{C}$. Sometimes the answer to the tourist's question is *"Yes there is a duality for $\mathcal{C}$, but not a natural one!"* For example, a category which is dual to the variety of Heyting algebras may be obtained by restricting Priestley duality. But there can be no natural duality for the simple reason that there is a proper class of non-isomorphic subdirectly irreducible Heyting algebras, whereas, if $\underline{P}$ is finite, then $\mathbb{ISP}(\underline{P})$ contains only finitely many non-isomorphic subdirectly irreducible algebras. (See Section 5 for a further discussion of restricted versus natural dualities.)

Our first lemma shows that, when trying to prove that $\underset{\sim}{P}$ yields a duality on $\mathcal{A}$, we may delete an operation from G or a partial operation from H provided we add its graph to the set R of relations. If $\mathrm{dom}(h) \subseteq P^n$ and $h : \mathrm{dom}(h) \to P$, then the *graph* of h is

$$\mathrm{graph}(h) := \{ (a_1, \ldots, a_n, h(a_1, \ldots, a_n)) \mid (a_1, \ldots, a_n) \in \mathrm{dom}(h) \} \subseteq P^{n+1}.$$

Note that $\mathrm{graph}(h)$ is a subalgebra of $\underline{P}^{n+1}$ if and only if $\mathrm{dom}(h)$ is a subalgebra of $\underline{P}^n$ and $h : \mathrm{dom}(h) \to \underline{P}$ is a homomorphism. Thus it makes sense to delete h from H and to add $\mathrm{graph}(h)$ to R. Although the proof of the following lemma is of the follow-your-nose variety, we present some of the calculations in order to illustrate the notational manipulations involved when working with maps defined on subsets of powers of P.

Lemma 2.1 *Let $\mathrm{dom}(h) \subseteq P^n$ and let $h : \mathrm{dom}(h) \to P$ for some $n \in \mathbb{N}$. Let S be a nonempty set, extend h pointwise to a partial operation on P^S and let X be a nonempty subset of P^S which is closed under h. Then a map $\alpha : X \to P$ preserves the partial operation h if and only if α preserves the graph of h.*

Proof Let $\mathrm{dom}(h)$, h, S and X be as described above and assume that $\alpha : X \to P$ preserves h. We shall show that α preserves the relation $r := \mathrm{graph}(h)$. The converse we leave as an easy exercise. Denote by r^S the pointwise extension of r to P^S. Let $x_1, \ldots, x_n, x_{n+1} \in X$; then

$$\begin{aligned}
& (x_1, \ldots, x_n, x_{n+1}) \in r^S \\
\Longrightarrow\quad & (\forall s \in S)\ (x_1(s), \ldots, x_n(s), x_{n+1}(s)) \in r = \mathrm{graph}(h) \\
\Longrightarrow\quad & (\forall s \in S)\ (x_1(s), \ldots, x_n(s)) \in \mathrm{dom}(h)\ \ \&\ \ h(x_1(s), \ldots, x_n(s)) = x_{n+1}(s) \\
\Longrightarrow\quad & (x_1, \ldots, x_n) \in \mathrm{dom}_X(h)\ \ \&\ \ h(x_1, \ldots, x_n) = x_{n+1} \\
\Longrightarrow\quad & (\alpha(x_1), \ldots, \alpha(x_n)) \in \mathrm{dom}(h)\ \ \&\ \ h(\alpha(x_1), \ldots, \alpha(x_n)) = \alpha(x_{n+1}) \\
\Longrightarrow\quad & (\alpha(x_1), \ldots, \alpha(x_n), \alpha(x_{n+1})) \in \mathrm{graph}(h) = r.
\end{aligned}$$

$\square$

Of course, this lemma applies equally well to total maps $g : P^n \to P$. We can now argue in the following way. If there is some structure $\underset{\sim}{P} = \langle P; G, H, R, \mathcal{T}\rangle$ which yields a duality on $\mathcal{A}$, then there is a relational structure $\underset{\sim}{P} = \langle P; R, \mathcal{T}\rangle$ which yields a duality on $\mathcal{A}$: simply replace the maps in $G \cup H$ by their graphs, which, by the lemma, preserves duality since $D(A) = \mathcal{A}(A, \underline{P}) \subseteq P^A$ is closed under the operations in G and the partial operations in H, by Lemma 1.1, so that the same maps are morphisms from $\mathcal{A}(A, \underline{P})$ to $\underset{\sim}{P}$. Now if some set of algebraic relations yields a duality on $\mathcal{A}$, then any larger set of algebraic relations will also yield a duality. In particular, $\underset{\sim}{P} = \langle P; B, \mathcal{T}\rangle$ will yield a duality on $\mathcal{A}$ where

$$B = \bigcup\{\,\mathbb{S}(\underline{P}^n) \mid n \geqslant 1\,\}$$

is the set of all finitary algebraic relations on $\underline{P}$. We refer to this as the *brute force* construction. Thus the issue of the existence of a duality for $\mathcal{A}$ may, on one level, be summed up as in the theorem below.

Theorem 2.2 *The following are equivalent:*

1. *there is some structure $\underset{\sim}{P} = \langle P; G, H, R, \mathcal{T}\rangle$ which yields a duality on $\mathcal{A}$;*

2. *brute force yields a duality on $\mathcal{A}$ (i.e. $\underset{\sim}{P} = \langle P; B, \mathcal{T}\rangle$ yields a duality on $\mathcal{A}$);*

3. *for all $A \in \mathcal{A}$, the evaluation maps $e(a) : \mathcal{A}(A, \underline{P}) \to P$, where $a \in A$, are the only maps from $\mathcal{A}(A, P)$ to P which preserve every finitary algebraic relation on $\underline{P}$.*

While the brute force construction is important at a theoretical level, in practice we try to make the structure on $\underset{\sim}{P}$ as simple as possible. Part of the beauty of Priestley's duality for bounded distributive lattices is that it is given by a single, particularly simple relation. No-one in their right mind would use the brute force duality for the class $\mathcal{D}$ of bounded distributive lattices. Nevertheless, brute force does yield a (full) duality on $\mathcal{D}$.

We return now to the situation where $\underset{\sim}{P} = \langle P; G, H, R, \mathcal{T}\rangle$ and we seek user-friendly conditions which will guarantee that $\underset{\sim}{P}$ yields a duality on $\mathcal{A}$. The following two lemmas are easily proved (see Lemmas 1.6 and 1.7 on page 120 of [19])—the first uses the fact that the structure on $\underset{\sim}{P}$ is algebraic over $\underline{P}$.

Lemma 2.3 *Let S be a nonempty set, let $F\mathcal{A}(S)$ be the free S-generated algebra in $\mathcal{A}$ and let*

$$\rho : D(F\mathcal{A}(S)) = \mathcal{A}(F\mathcal{A}(S), \underline{P}) \to \underset{\sim}{P}^S$$

be the map which restricts each homomorphism $x : F\mathcal{A}(S) \to P$ to the set S. Then ρ is an isomorphism in $\mathcal{X}$.

Lemma 2.4 *Let $A, B \in \mathcal{A}$ and let $u : A \to B$ be a surjective homomorphism. Then $D(u) : D(B) \to D(A)$ is an embedding in $\mathcal{X}$. Similarly, if $X, Y \in \mathcal{X}$ and $\psi : X \to Y$ is a surjective morphism, then $E(\psi) : E(Y) \to E(X)$ is an embedding in $\mathcal{A}$.*

We can now give necessary and sufficient conditions for $\underset{\sim}{P}$ to yield a duality on $\mathcal{A}$. (See Theorem 1.8 on page 121 of [19].)

Theorem 2.5 (The First Duality Theorem) *The following are equivalent:*

1. $\underset{\sim}{P}$ *yields a duality on* $\mathcal{A}$*;*
2. *for all* $A \in \mathcal{A}$*, every morphism* $\alpha : D(A) \to \underset{\sim}{P}$ *extends to an* A*-ary term function* $t : P^A \to P$*, i.e. for all* $x \in D(A)$*, we have* $\alpha(x) = t(x)$*;*
3. *the following two conditions hold—*

   (D1) $\underset{\sim}{P}$ *is injective with respect to embeddings in* $\mathcal{X}$ *of the form* $D(u) : D(B) \to D(A)$ *where* $u : A \to B$ *is a surjective homomorphism, i.e. for each morphism* $\alpha : D(B) \to \underset{\sim}{P}$ *there exists a morphism* $\beta : D(A) \to \underset{\sim}{P}$ *such that* $\beta \circ D(u) = \alpha$*,*

   (D2) *for each* $n \in \mathbb{N}$*, every morphism* $t : \underset{\sim}{P}^n \to \underset{\sim}{P}$ *is an* n*-ary term function on* $\underline{P}$*.*

In practice, rather than prove **(D1)** we show that $\underset{\sim}{P}$ is *injective in* $\mathcal{X}$, i.e. if $Y \in \mathcal{X}$ and X is a closed substructure of Y, then every morphism $\alpha : X \to \underset{\sim}{P}$ extends to a morphism $\beta : Y \to \underset{\sim}{P}$ satisfying $\beta{\restriction}X = \alpha$.

Since each n-ary term function $t : P^n \to P$ of the algebra $\underline{P}$ preserves every algebraic relation on $\underline{P}$, it follows that **(D2)** says precisely that $\mathcal{X}(\underset{\sim}{P}^n, \underset{\sim}{P})$ is the set of all n-ary term functions on $\underline{P}$ for each $n \in \mathbb{N}$. In fact, it is easily seen that **(D2)** implies that $\mathcal{X}(\underset{\sim}{P}^S, \underset{\sim}{P})$ is the set of all S-ary term functions on $\underline{P}$ for every nonempty set S. (See page 123 of [19].) Thus $G \cup H \cup R$ determines the clone of term functions on $\underline{P}$. It is important to note that the converse is false. There is a set R of relations which determines the clone of the three-element Kleene algebra but does not yield a duality. Nevertheless, increasing R by one further relation does yield a duality. (See the discussion of Kleene algebras in Section 4 below or see page 176 of [19].)

Conditions **(D1)** and **(D2)** combine to yield a natural *interpolation condition*:

(IC) *for each* $n \in \mathbb{N}$ *and each substructure* X *of* $\underset{\sim}{P}^n$*, every morphism* $\alpha : X \to \underset{\sim}{P}$ *extends to a term function* $t : P^n \to P$ *of the algebra* $\underline{P}$*.*

Denote the class of all finite algebras in $\mathcal{A}$ by $\mathcal{A}_{fin}$. If **(IC)** holds, then **(D1)** holds for all $A, B \in \mathcal{A}_{fin}$ and **(D2)** holds—just take $X = \underset{\sim}{P}^n$. Thus **(IC)** implies that $\underset{\sim}{P}$ yields a duality on $\mathcal{A}_{fin}$ and that $\underset{\sim}{P}$ is injective in $\mathcal{X}_{fin}$ (since every finite structure in $\mathcal{X}$ can be embedded in a finite power of $\underset{\sim}{P}$). We would like to lift the duality, provided on $\mathcal{A}_{fin}$ by **(IC)**, up to a duality on the whole of $\mathcal{A}$. The following result (see Theorem 1.16 on page 136 of [19]) shows that this is achievable provided $H = \emptyset$ and R is finite.

Theorem 2.6 (The Second Duality Theorem) *Assume that* $\underset{\sim}{P} = \langle P; G, R, \mathcal{T}\rangle$ *(i.e. the structure on* $\underset{\sim}{P}$ *includes no partial operations) and assume that* R *is finite. If* **(IC)** *holds, then* $\underset{\sim}{P}$ *yields a duality on* $\mathcal{A}$ *and* $\underset{\sim}{P}$ *is injective in* $\mathcal{X}$*.*

This result is rather surprising. It yields a topological representation of every algebra in $\mathcal{A}$ yet requires us to do no topology! A natural question now is, *"How can we force* **(IC)** *to hold?"* The answer is, *"Use brute force!"* (See 1.15 on page 135 of [19].)

Theorem 2.7 (The Brute-Force Duality Theorem) *Brute force yields a duality on $\mathcal{A}_{fin}$. Indeed, if $\underset{\sim}{P} = \langle P; B, \mathcal{T}\rangle$, where $B = \bigcup\{\, \mathbb{S}(\underline{P}^n) \mid n \in \mathbb{N}\,\}$, then* **(IC)** *holds.*

The Second Duality Theorem and the Brute-Force Duality Theorem are in a tug-of-war—the former asks that the structure on $\underset{\sim}{P}$ be small while the latter wants the structure on $\underset{\sim}{P}$ to be as big as possible. Fortunately, there is one very important instance when the two can resolve their differences.

A $(k+1)$-ary term $n(x_1, \dots, x_{k+1})$ is called a *near-unanimity term* on $\underline{P}$ if $\underline{P}$ satisfies the identities

$$n(x, \dots, x, y) \approx n(x, \dots, x, y, x) \approx \dots \approx n(y, x, \dots, x) \approx x.$$

A 3-ary near-unanimity term on $\underline{P}$ is usually called a *majority term.* For example, on any algebra with an underlying lattice structure, the median,

$$m(x, y, z) := (x \wedge y) \vee (y \wedge z) \vee (z \wedge x)$$

is a majority term since it satisfies the identities

$$m(x, x, y) \approx m(x, y, x) \approx m(y, x, x) \approx x.$$

The theorem below is extremely useful. We shall see in Section 7 that it has a very strong converse.

Theorem 2.8 (The NU-Duality Theorem) *Let $k \geqslant 2$ and assume that $\underline{P}$ has a $(k+1)$-ary near-unanimity term. Then $\underset{\sim}{P} = \langle P; \mathbb{S}(\underline{P}^k), \mathcal{T}\rangle$ yields a duality on $\mathcal{A}$ and $\underset{\sim}{P}$ is injective in $\mathcal{X}$.*

Proof Since $\mathbb{S}(\underline{P}^k)$ is a finite set of relations, by the Second Duality Theorem it is enough show that **(IC)** holds with respect to $\underset{\sim}{P} = \langle P; \mathbb{S}(\underline{P}^k), \mathcal{T}\rangle$. Let X be a subset of P^n and let $\alpha : X \to P$ preserve every subalgebra of $\underline{P}^k$. Since $\underline{P}$ has a $(k+1)$-ary near-unanimity term, it follows that α preserves every subalgebra of every finite power of $\underline{P}$. (This is implicit in [1] and a direct proof is easy—see Lemma 1.17 on page 138 of [19]). Hence, by the Brute-Force Duality Theorem, α extends to an n-ary term function on $\underline{P}$. Thus **(IC)** holds with respect to $\underset{\sim}{P}$, as required. □

It follows from the NU-Duality Theorem that if $\underline{P}$ has an underlying lattice structure, then $R = \mathbb{S}(\underline{P}^2)$ yields a duality on $\mathcal{A} = \mathbb{ISP}(\underline{P})$. Unfortunately, even when $\underline{P}$ is small, $\mathbb{S}(\underline{P}^2)$ can be monstrous. In practice, we

- replace any relation which is the graph of a (partial) operation by the corresponding (partial) operation—this decreases the size of R, increases the size of $G \cup H$, but has no effect on the size of $G \cup H \cup R$,

- remove any relations which can be "generated" by the remaining ones.

While, strictly speaking, the first process is irrelevant as far as duality is concerned, it is more natural to work with a function rather than its graph and, moreover, this will be essential when we come to the issue of full dualities. We now make clear what we mean by the second process.

Let $\underset{\sim}{P} = \langle P; G, H, R, \mathcal{T}\rangle$. We say that $G \cup H \cup R$ *generates* a relation r on an algebra $A \in \mathcal{A}$ if whenever a continuous map $\varphi : \mathcal{A}(A, \underline{P}) \to P$ preserves the (partial) operations and relations in $G \cup H \cup R$ it also preserves r. If $G \cup H \cup R$ generates r on every algebra in $\mathcal{A}$, we say simply that $G \cup H \cup R$ *generates* r. So, for example, if R yields a duality on $\mathcal{A}$ and $R\backslash\{r\}$ generates r, then the smaller set, $R\backslash\{r\}$, also yields a duality on $\mathcal{A}$. When proving that $G \cup H \cup R$ generates r it is common to prove the following stronger statement: *for all nonempty sets S, if X is a closed substructure of $\underset{\sim}{P}^S$ and $\varphi : X \to \underset{\sim}{P}$ is a morphism, then φ preserves r.* Below is a (non-exhaustive) list of useful constructs for finding relations generated by $G \cup H \cup R$.

Trivial relations The relations $\Delta = \{ (x, x) \mid x \in P \}$ and $\nabla = P \times P$ are preserved by any map $\varphi : X \to P$ where $X \subseteq P^S$ and hence are generated by any set $G \cup H \cup R$.

Projections If D is a subalgebra of $\underline{P}^n$ and $h : D \to \underline{P}$ is a projection (restricted to D), then again any map will preserve h and hence h is generated by any set $G \cup H \cup R$.

Symmetry If r if a subalgebra of $\underline{P}^2$, then $\{r\}$ generates $r^* := \{ (b, a) \mid (a, b) \in r \}$.

Intersection $\{r, s\}$ generates $r \cap s$ (provided $r \cap s \neq \emptyset$).

Domain $G \cup H \cup R$ generates $\mathrm{dom}(h)$ for each $h \in H$.

Composition $G \cup H \cup R$ generates the "clone" of partial functions generated by $G \cup H$.

Equalizer Let $g, h \in G \cup H$ both be n-ary. Then $G \cup H \cup R$ generates

$$\mathrm{eq}(g, h) := \{ a \in P^n \mid a \in \mathrm{dom}(g) \cap \mathrm{dom}(h) \ \ \& \ \ g(a) = h(a) \}.$$

Fixpoints Let $h \in G \cup H$ be unary. Then $G \cup H \cup R$ generates

$$\mathrm{fix}(h) := \{ a \in P \mid a \in \mathrm{dom}(h) \ \ \& \ \ h(a) = a \}.$$

Kernels Let $h \in G \cup H$ be n-ary. Then $G \cup H \cup R$ generates

$$\ker(h) := \{ (a, b) \in (P^n)^2 \mid a, b \in \mathrm{dom}(h) \ \ \& \ \ h(a) = h(b) \}.$$

Relational product of a map and a relation. Let $h \in G \cup H$ be unary and let $r \in R$ be binary. Then $G \cup H \cup R$ generates

$$h \circ r := \{ (a, b) \in P^2 \mid a \in \mathrm{dom}(h) \ \ \& \ \ (h(a), b) \in r \}.$$

Some of these constructs can be generalized slightly: see pages 140–142 of [19]. Note that if $\varphi : P^S \to P$ preserves relations r and s, then φ preserves their relational product $s \circ r$. This is no longer true when we consider $\varphi : X \to P$ with $X \subseteq P^S$ and consequently $\{r, s\}$ does not necessarily generate $s \circ r$. This explains why a set R of algebraic relations on $\underline{P}$ which determines the clone of term functions on $\underline{P}$ will not necessarily yield a duality on $\mathcal{A}$.

Stone and Priestley revisited The NU-Duality Theorem and the "trivial relations" and "symmetry" constructs listed above yield very short proofs of the algebraic halves of the Stone and Priestley dualities (modulo the fact that $\mathcal{B} = \mathbb{ISP}(\underline{2})$ and $\mathcal{D} = \mathbb{ISP}(\underline{2})$). In the terminology of this section, the algebraic half of Stone duality for Boolean Algebras says that $\underset{\sim}{2} = \langle 2; \emptyset, \mathcal{T} \rangle$ yields a duality on $\mathcal{B}$ while the algebraic half of Priestley duality for bounded distributive lattices says that $\underset{\sim}{2} = \langle 2; \leqslant, \mathcal{T} \rangle$ yields a duality on $\mathcal{D}$. Since both Boolean algebras and distributive lattices have an underlying lattice structure the NU-Duality Theorem applies with $k = 2$. When $\underline{2}$ is the two-element Boolean algebra, the only subalgebras of $\underline{2}^2$ are the trivial relations Δ and ∇; hence $R = \emptyset$ generates the subalgebras of $\underline{2}^2$ and consequently $\underset{\sim}{2} = \langle 2; \emptyset, \mathcal{T} \rangle$ yields a duality on $\mathcal{B}$. When $\underline{2}$ is the two-element bounded distributive lattice, there are two further subalgebras of $\underline{2}^2$, namely

$$\leqslant = \{(0,0), (0,1), (1,1)\} \text{ and } \geqslant = \{(0,0), (1,0), (1,1)\}.$$

Since $\geqslant$ is just the converse of $\leqslant$, it follows that $\leqslant$ generates all subalgebras of $\underline{2}^2$ and hence $\underset{\sim}{2} = \langle 2; \leqslant, \mathcal{T} \rangle$ yields a duality on $\mathcal{D}$.

We close this guided tour of our duality theorems with an open problem.

Problem 1 Find a family of constructs, like those listed above, such that if $G \cup H \cup R$ generates r, then r can be obtained by a finite number of applications of the constructs—or show that no such family exists.

Note that if we wish to generate r on the powers of $\underset{\sim}{P}$, rather than on the substructures of the form $\mathcal{A}(A, \underline{P})$, then a complete set of constructs is available from the theory of clones.

3 Full-duality and strong-duality theorems

The time has come to cross the border between $\mathcal{A}$ and $\mathcal{X}$. Now the operations and partial operations in the structure on $\underset{\sim}{P}$ come into their own. We know from Lemma 1.4 that, regardless of the choice of G, H and R, the map $\varepsilon_X : X \to DE(X)$ is an embedding for all $X \in \mathcal{X}$. But if the structure on $\underset{\sim}{P}$ is too weak there may be structures X in $\mathcal{X}$ for which ε_X is not an isomorphism. If $\underset{\sim}{P}$ yields a duality on $\mathcal{A}$ and moreover ε_X is an isomorphism for all $X \in \mathcal{X}$, then we say that $\underset{\sim}{P}$ yields a *full duality* on $\mathcal{A}$. Once we have a structure $\underset{\sim}{P}$ which yields a duality on $\mathcal{A}$, we would like to modify $\underset{\sim}{P}$ so that a full duality is obtained. As we saw in the previous section, strengthening the structure on $\underset{\sim}{P}$ cannot destroy the duality and may have the effect of eliminating from $\mathcal{X}$ some of the structures X for which ε_X is not an isomorphism. Thus we now feel free to remove from R any relation which is the graph of a (partial) operation while adding the corresponding map to $G \cup H$. We may also add new (partial) operations, which are algebraic over $\underline{P}$, to $G \cup H$.

Much of the territory covered below has only recently been opened up for tourists. The results stated here are a common refinement and extension due to David Clark and the author of the full-duality theorems of Clark and Krauss [5] and Davey and Werner [19] and were obtained after the Montréal lectures were given. The proofs will eventually appear in Clark and Davey [3].

Full duality, strong duality and the injectivity of $\underset{\sim}{P}$ It is easily seen that if $\underset{\sim}{P}$ yields a duality on $\mathcal{A}$ then this duality is full if and only if every $X \in \mathcal{X}$ is isomorphic to $D(A)$ for some $A \in \mathcal{A}$; in symbols, $\mathcal{X} = \mathbb{I}D(\mathcal{A})$. Hence our first aim is to describe $\mathbb{I}D(\mathcal{A})$.

Let $I \neq \emptyset$, let $\mathrm{dom}(h)$ be a subalgebra of $\underline{P}^I$ and let $h : \mathrm{dom}(h) \to \underline{P}$ be a homomorphism, i.e. h is an algebraic I-ary partial operation on $\underline{P}$. Note that I may be infinite. Just as we did in the finitary case, we may extend h pointwise to an I-ary partial operation on P^S. Let $\pi_s : P^S \to P$ denote the s-th projection for each $s \in S$. Thus

$$\mathrm{dom}_{P^S}(h) := \{\, x \in (P^S)^I \mid \pi_s \circ x \in \mathrm{dom}(h) \text{ for all } s \in S \,\}$$

and $h : \mathrm{dom}_{P^S}(h) \to P^S$ is defined by $(h(x))(s) := h(\pi_s \circ x)$. A subset X of P^S is closed under h provided $h(x) \in X$ whenever $x \in X \cap \mathrm{dom}_{P^S}(h)$. We shall say that X is *hom-closed* (*in* P^S) if, for each nonempty set I, the set X is closed under every algebraic I-ary partial operation on $\underline{P}$.

We say that a subset X of P^S is *term-closed* (*in* P^S) if for all $y \in P^S \backslash X$ there exist S-ary term functions $\sigma, \tau : P^S \to P$ on $\underline{P}$ such that $\sigma{\restriction}X = \tau{\restriction}X$ and $\sigma(y) \neq \tau(y)$. Thus X is term-closed in P^S provided it is an intersection of the equalizer sets of pairs of S-ary term functions on $\underline{P}$.

Theorem 3.1 *Let $A \in \mathcal{A}$, let $S \neq \emptyset$ and let X be a nonempty subset of P^S.*

1. *$D(A)$ is both term-closed and hom-closed in P^A.*
2. *X is term-closed in P^S if and only if X is hom-closed in P^S.*
3. *If X is term-closed in P^S, then X is a closed substructure of $\underset{\sim}{P}^S$ and hence $X \in \mathcal{X}$.*

Proof 1 is an easy calculation, 2 is a tricky calculation (see Lemma 2.15 of [5]) and 3 follows easily from 2. □

The following two results constitute Theorem 2.26 of [5].

Theorem 3.2 *Let $X \in \mathcal{X}$. Then $X \in \mathbb{I}D(\mathcal{A})$ if and only if X is isomorphic to a term-closed (= hom-closed) subset of P^S for some nonempty set S.*

Theorem 3.3 (The Full-Duality Theorem) *Assume that the structure $\underset{\sim}{P}$ yields a duality on $\mathcal{A}$. Then the following are equivalent:*

1. *$\underset{\sim}{P}$ yields a full duality on $\mathcal{A}$;*
2. *$\mathcal{X} = \mathbb{I}D(\mathcal{A})$;*
3. *if X is a closed substructure of $\underset{\sim}{P}^S$ for some non-empty sets S, then there is a non-empty set T and a term-closed (= hom-closed) subset Y of P^T such that X is isomorphic to Y.*

Unless we can choose $Y = X$, Condition 3 of this theorem is rather awkward. If $\underset{\sim}{P}$ yields a duality on $\mathcal{A}$ and every closed substructure of $\underset{\sim}{P}^S$ is term-closed (= hom-closed) in P^S, then we say that $\underset{\sim}{P}$ *yields a strong duality on* $\mathcal{A}$. In practice, we always establish this stronger condition. In fact, every known full duality is strong. This more natural condition is intimately related to the more natural version of condition **(D1)** of the First Duality Theorem: $\underset{\sim}{P}$ *is injective* in $\mathcal{X}$. In every known example of a full duality, $\underset{\sim}{P}$ is injective in $\mathcal{X}$.

Assume that $\underset{\sim}{P}$ yields a duality on $\mathcal{A}_{fin}$. Then we say that $\underset{\sim}{P}$ *yields a full duality on* $\mathcal{A}_{fin}$ if $\varepsilon_X : X \to DE(X)$ is an isomorphism for every finite structure in $\mathcal{X}$, and we say that $\underset{\sim}{P}$ *yields a strong duality on* $\mathcal{A}_{fin}$ if every (closed) substructure of a finite power of $\underset{\sim}{P}$ is term-closed (= hom-closed).

Theorem 3.4 (The First Strong-Duality Theorem) *If $\underset{\sim}{P}$ yields a strong duality on $\mathcal{A}$, then $\underset{\sim}{P}$ is injective in $\mathcal{X}$ and yields a full duality on $\mathcal{A}$. The converse holds at the finite level, i.e. $\underset{\sim}{P}$ yields a strong duality on $\mathcal{A}_{fin}$ if and only if $\underset{\sim}{P}$ is injective in $\mathcal{X}_{fin}$ and yields a full duality on $\mathcal{A}_{fin}$.*

If $\underset{\sim}{P} = \langle P; G, R, \mathcal{T}\rangle$, i.e. the structure on $\underset{\sim}{P}$ includes no partial operations, then we call $\underset{\sim}{P}$ a *total structure.* In this case, we have necessary and sufficient conditions for a strong duality which are devoid of topology. The following result is based on Lemma 2.33 of [5]. Note that Condition 3 in the theorem says simply that every (closed) substructure of a finite power of $\underset{\sim}{P}$ is term-closed.

Theorem 3.5 (The Second Strong-Duality Theorem) *Assume that $\underset{\sim}{P}$ is a total structure which yields a duality on $\mathcal{A}$. Then the following conditions are equivalent:*

1. *$\underset{\sim}{P}$ yields a strong duality on $\mathcal{A}$;*
2. *$\underset{\sim}{P}$ is injective in $\mathcal{X}$ and yields a full duality on $\mathcal{A}$;*
3. *$\underset{\sim}{P}$ satisfies the* Finite Term Closure *condition—*

   (FTC) *if X is a substructure of $\underset{\sim}{P}^n$ for some $n \in \mathbb{N}$ and $y \in P^n \backslash X$, then there exist morphisms $\sigma, \tau : \underset{\sim}{P}^n \to \underset{\sim}{P}$, i.e. n-ary term functions on $\underline{P}$, such that $\sigma{\restriction}X = \tau{\restriction}X$ and $\sigma(y) \neq \tau(y)$.*

The Second Duality Theorem and the Second Strong-Duality Theorem combine to give purely finite conditions for the existence of a strong duality.

Theorem 3.6 (The Third Strong-Duality Theorem) *Assume that $\underset{\sim}{P} = \langle P; G, R, \mathcal{T}\rangle$ is a total structure and that R is finite. Then the following are equivalent:*

1. *$\underset{\sim}{P}$ yields a strong duality on $\mathcal{A}$;*
2. *$\underset{\sim}{P}$ yields a strong duality on $\mathcal{A}_{fin}$;*
3. **(IC)** *and* **(FTC)** *hold.*

Note that the injectivity of $\underset{\sim}{P}$ says precisely that E maps embeddings to surjections. In this setting we get more. The proof of the following lemma is a purely set-theoretic argument based on Lemma 2.4 and Figure 1 from Lemma 1.4.

Lemma 3.7 *Assume that $\underset{\sim}{P}$ yields a full duality on $\mathcal{A}$. Then the following are equivalent:*

1. *$\underset{\sim}{P}$ is injective in $\mathcal{X}$;*
2. *$\varphi : X \to Y$ is an embedding in $\mathcal{X}$ (if and) only if $E(\varphi) : E(Y) \to E(X)$ is surjective;*
3. *$u : A \to B$ is surjective in $\mathcal{A}$ if (and only if) $D(u) : D(B) \to D(A)$ is an embedding in $\mathcal{X}$.*

We note without further ado that Lemma 3.7 remains valid when rewritten for $\underline{P}$ rather than $\underset{\sim}{P}$. It may come as a surprise that the injectivity of $\underline{P}$ in $\mathcal{A}$ and the injectivity of $\underset{\sim}{P}$ in $\mathcal{X}$ are closely linked. (See Proposition 1.11 on page 128 of [19].)

Lemma 3.8 *If $\underset{\sim}{P}$ yields a full duality on $\mathcal{A}$ and $\underline{P}$ is injective in $\mathcal{A}$, then $\underset{\sim}{P}$ is injective in $\mathcal{X}$.*

Total structures and the injectivity of $\underline{P}$ While we have no counterexample to the conjecture that $\underset{\sim}{P}$ is injective in $\mathcal{X}$ whenever we have a full duality, the corresponding statement about $\underline{P}$ is false! For example, let $\underline{P}$ be the n-element chain regarded as a Heyting algebra and define $\mathrm{End}(\underline{P})$ to be the set of all endomorphisms of $\underline{P}$ and H to be the set of all homomorphisms $h : \mathrm{dom}(h) \to \underline{P}$ where $\mathrm{dom}(h)$ is a proper subalgebra of $\underline{P}$. We shall see in Section 5 that $\underset{\sim}{P} = \langle P; \mathrm{End}(\underline{P}), H, \mathcal{T}\rangle$ yields a strong duality on $\mathcal{A} = \mathbb{ISP}(\underline{P})$. Nevertheless, $\underline{P}$ is not injective in $\mathcal{A}$ for $n \geqslant 4$. As the discussion below shows, the injectivity of $\underline{P}$ is intimately connected to the need for partial operations in the structure on $\underset{\sim}{P}$.

We turn now to the problem of eliminating partial operations from the structure on $\underset{\sim}{P}$ without destroying a strong duality—when is it possible and how do we go about it?

Let $\underset{\sim}{P} = \langle P; G, H, R, \mathcal{T}\rangle$ and $\underset{\sim}{P}' = \langle P; G', H', R', \mathcal{T}\rangle$. We say that $\underset{\sim}{P}'$ *dominates* $\underset{\sim}{P}$ (or that $\underset{\sim}{P}$ *is dominated by* $\underset{\sim}{P}'$) if, for every nonempty set S, a closed subset X of P^S is a substructure with respect to $\underset{\sim}{P}$ whenever it is a substructure with respect to $\underset{\sim}{P}'$, and, moreover, for each closed substructure X of $\underset{\sim}{P}^S$ a continuous map $\varphi : X \to P$ is a morphism with respect to $\underset{\sim}{P}$ whenever it is a morphism with respect to $\underset{\sim}{P}'$. We say that the structures $\underset{\sim}{P}$ and $\underset{\sim}{P}'$ are *structurally equivalent* if each dominates the other.

The next two lemmas indicate the relationship between domination, structural equivalence and strong duality. The first says simply that we cannot kill off a (strong) duality by strengthening the structure on $\underset{\sim}{P}$.

Lemma 3.9 *Assume that $\underset{\sim}{P}'$ dominates $\underset{\sim}{P}$. If $\underset{\sim}{P}$ yields a (strong) duality on $\mathcal{A}$, then so does $\underset{\sim}{P}'$.*

Lemma 3.10 *Let $\underset{\sim}{P}$ and $\underset{\sim}{P}'$ be structures and assume that $\underset{\sim}{P}$ yields a strong duality on $\mathcal{A}$. Then $\underset{\sim}{P}'$ yields a strong duality on $\mathcal{A}$ if and only if $\underset{\sim}{P}'$ is structurally equivalent to $\underset{\sim}{P}$.*

The following lemma is a follow-your-nose exercise.

Lemma 3.11 *Let $\underset{\sim}{P} = \langle P; G, H, R, \mathcal{T}\rangle$ and let $h : \mathrm{dom}(h) \to \underline{P}$ be an element of H with $\mathrm{dom}(h)$ a subalgebra of $\underline{P}^n$. Assume that h extends to a homomorphism $g : \underline{P}^n \to \underline{P}$ and define $G' = G \cup \{g\}$, $H' = H\backslash\{h\}$ and $R' = R \cup \{\mathrm{dom}(h)\}$. Then $\underset{\sim}{P}$ is dominated by $\underset{\sim}{P}' := \langle P; G', H', R', \mathcal{T}\rangle$.*

Thus we have one instance where it is possible to purge $\underset{\sim}{P}$ of its partial operations. Assume that $\underset{\sim}{P}$ yields a strong duality on $\mathcal{A}$. By the previous three lemmas, if every n-ary partial operation $h \in H$ extends to a homomorphism $g : \underline{P}^n \to \underline{P}$, then $\underset{\sim}{P}$ is structurally equivalent to a total structure. Rather surprisingly, as Theorem 3.13 below shows, the converse is also true! Our next lemma is a nice example of schizophrenia at work: an algebraic partial operation $h : \mathrm{dom}(h) \to \underline{P}$ with $\mathrm{dom}(h)$ a subalgebra of $\underline{P}^n$ can be viewed as part of the structure on $\underset{\sim}{P}$ but can also be viewed as a collection of algebras and homomorphisms in $\mathcal{A}$ to which the map D may be applied.

Lemma 3.12 *Let A be a subalgebra of $\underline{P}^n$, let $i : A \to \underline{P}^n$ be the inclusion map and let $h : A \to \underline{P}$ be a homomorphism. Then h extends to a homomorphism $g : \underline{P}^n \to \underline{P}$ if and only if the set*

$$\{\, x{\restriction}A : A \to P \mid x : \underline{P}^n \to \underline{P} \textit{ is a homomorphism } \} \subseteq P^A$$

is closed under h, i.e. the image of the map $D(i) : D(\underline{P}^n) \to D(A)$ is closed under h.

Those unaccustomed to structures with partial operations find a most insidious feature of these objects to be the fact that the images of morphisms are not in general substructures. Hard experience has shown that many lovely but invalid theorems can be proved by overlooking just this fact! So a shopping list of *nice* properties of $\underline{P}$, $\underset{\sim}{P}$ and $\mathcal{X}$ would include: (a) $\underline{P}$ is injective in $\mathcal{A}$, (b) $\underset{\sim}{P}$ is structurally equivalent to a total structure, and (c) the image of every morphism in $\mathcal{X}$ is a substructure.

The theorem below which ties all these threads together says, in essence, that if anything is nice then everything is! The proof is obtained by tweaking slightly the proof of Proposition 1.11 on page 128 of [19].

Theorem 3.13 (The Total-Structure Theorem) *Assume that $\underset{\sim}{P}$ yields a strong duality on $\mathcal{A}$. Then the following are equivalent:*

1. *there is some total structure $\underset{\sim}{P}$ which yields a strong duality on $\mathcal{A}$;*
2. *$\underset{\sim}{P}$ is structurally equivalent to a total structure;*
3. *the image of every morphism in $\mathcal{X}$ is a closed substructure;*
4. *$\underline{P}$ is injective in $\mathcal{A}$;*
5. *for all $n \in \mathbb{N}$, every n-ary partial operation $h \in H$ in the structure on $\underset{\sim}{P}$ extends to a homomorphism $g : \underline{P}^n \to \underline{P}$.*

This theorem provides us with a simple algorithm, based on the generating algebra $\underline{P}$, to determine if and how partial operations can be eliminated from a given strong duality for $\mathcal{A}$.

Producing strong dualities All known full dualities are strong and can be obtained by applying one of the theorems stated below.

If $\underset{\sim}{P} = \langle P; G, \mathcal{T}\rangle$, i.e. the structure on $\underset{\sim}{P}$ includes no partial operations and no relations, then we call $\underset{\sim}{P}$ a *total algebra.* The Second Duality Theorem and the Third Strong-Duality Theorem combine to show that we get two strong dualities for the price of one whenever $\underset{\sim}{P}$ is a total algebra.

Theorem 3.14 (The Two-for-One Strong-Duality Theorem) *Let $\underline{P} = \langle P; F\rangle$ and let $\underset{\sim}{P} = \langle P; G, \mathcal{T}\rangle$ be a total algebra. Define $\underline{P}' := \langle P; G\rangle$ and $\mathcal{A}' := \mathbb{ISP}(\underline{P}')$, and define $\underset{\sim}{P}' := \langle P; F, \mathcal{T}\rangle$ and $\mathcal{X}' := \mathbb{IS}_c\mathbb{P}(\underset{\sim}{P}')$. Then the following are equivalent:*

1. **(IC)** *and* **(FTC)** *hold with respect to $\underset{\sim}{P}$;*
2. **(IC)** *and* **(FTC)** *hold with respect to $\underset{\sim}{P}'$;*
3. *the following symmetric conditions hold:*
   (a) *$\underline{P}$ is injective in $\mathcal{A}_{fin}$,*
   (b) *$\underline{P}'$ is injective in $\mathcal{A}'_{fin}$,*
   (c) *every homomorphism $u : (\underline{P}')^n \to \underline{P}'$ is an n-ary term function on $\underline{P}$,*
   (d) *every homomorphism $v : \underline{P}^n \to \underline{P}$ is an n-ary term function on $\underline{P}'$;*
4. **(IC)** *holds with respect to both $\underset{\sim}{P}$ and $\underset{\sim}{P}'$;*
5. *the algebras $\underline{P}$ and $\underline{P}'$ are injective in $\mathcal{A}$ and $\mathcal{A}'$ respectively and the structures $\underset{\sim}{P}$ and $\underset{\sim}{P}'$ yield strong dualities on $\mathcal{A}$ and $\mathcal{A}'$ respectively.*

Given $\underline{P}$, where do we find the total algebra $\underset{\sim}{P}$? Many applications come from the following corollary which tells us when we can use the first candidate which comes to mind.

Theorem 3.15 (The Strong Self-Duality Theorem) *Let $\underset{\sim}{P}$ be obtained by augmenting $\underline{P}$ with the discrete topology. Then $\underset{\sim}{P}$ yields a strong duality on $\mathcal{A}$ and $\underline{P}$ is injective in $\mathcal{A}$ if and only if* **(IC)** *holds.*

Our experience indicates that if $\underset{\sim}{P}$ yields a full duality on $\mathcal{A}$, then $\underset{\sim}{P}$ will be injective in $\mathcal{X}$. Consequently, if $\underline{P}$ is not injective in $\mathcal{A}$ and we wish to obtain a full duality, then Theorem 3.13 tells us that we must expect to use some partial operations in the structure on $\underset{\sim}{P}$. In such cases, we cannot use **(FTC)** as a means of proving that the duality is strong and therefore full. Fortunately, all is not lost, as we shall now see. The development to this point has concentrated on establishing a strong duality by showing that every closed substructure of a power of $\underset{\sim}{P}$ is term-closed. But it also suffices to show that every closed substructure of a power of $\underset{\sim}{P}$ is hom-closed. We now turn and head in that direction.

Every congruence on a finite algebra Q is a meet of meet-irreducible congruences on Q. Let $m(Q)$ be the least n such that the zero congruence on Q is a meet of n meet-irreducible congruences, and let

$$M(\underline{P}) := \max\{\, m(Q) \mid Q \text{ is a subalgebra of } \underline{P}\,\}.$$

The following lemma is a substantial improvement on Theorem 2.37 of Clark and Krauss [5]. Its proof is obtained by giving Clark and Krauss' proof a rather substantial tweak—in particular the use of filtrality is replaced by a direct application of Jónsson's Lemma. By the *clone of* $\underset{\sim}{P}$ we mean the clone of partial functions generated by $G \cup H$.

Lemma 3.16 *Assume that $\underline{P}$ generates a congruence-distributive variety and that the clone of $\underset{\sim}{P}$ includes all n-ary algebraic partial operations on $\underline{P}$ for $n \leqslant M(\underline{P})$. Then every closed substructure of a power of $\underset{\sim}{P}$ is hom-closed.*

Thus if $\underline{P}$ generates a congruence-distributive variety and $\underset{\sim}{P} = \langle P; G, H, R, \mathcal{T}\rangle$ yields a duality on $\mathcal{A}$, we can always obtain a strong duality by adding finitely many (partial) operations to $G \cup H$. Indeed, it suffices to add to G and H the sets G^+ and H^+, respectively, where

$$\begin{aligned} G^+ &= \bigcup\{\, \mathcal{A}(\underline{P}^n, \underline{P}) \mid n \leqslant M(\underline{P})\,\}, \text{ and} \\ H^+ &= \bigcup\{\, \mathcal{A}(A, \underline{P}) \mid A \text{ is a proper subalgebra of } \underline{P}^n \ \&\ n \leqslant M(\underline{P})\,\}. \end{aligned}$$

Since any algebra which has a near-unanimity term generates a congruence-distributive variety, we can combine this lemma with the NU-Duality Theorem to obtain our final strong-duality theorem.

Theorem 3.17 (The NU-Strong-Duality Theorem) *Let $k \geqslant 2$ and assume that $\underline{P}$ has a $(k+1)$-ary near-unanimity term. If the structure on $\underset{\sim}{P}$ generates all subalgebras of $\underline{P}^k$ and the clone of $\underset{\sim}{P}$ includes all n-ary algebraic partial operations on $\underline{P}$ for $n \leqslant M(\underline{P})$, then $\underset{\sim}{P}$ yields a strong duality on $\mathcal{A}$.*

An immediate consequence of this result is that if $\underline{P}$ has a near-unanimity term, then there is a structure $\underset{\sim}{P}$ of finite type (i.e. $G \cup H \cup R$ is finite) which yields a strong duality on $\mathcal{A}$.

Stone and Priestley yet again In the previous section we gave a very short proof of the algebraic halves of the Stone and Priestley dualities. The NU-Strong-Duality Theorem gives even shorter proofs that these dualities are strong and therefore full. In both cases, $M(\underline{2}) = 1$ and the only algebraic unary partial operation on $\underline{2}$ is the identity map which, of course, can be omitted. Thus, by the NU-Strong-Duality Theorem, both dualities are strong. Since neither of these dualities involves partial operations, the Second Strong-Duality Theorem is also available. In both cases, **(FTC)** is extremely easy to prove.

So ends our guided tour of the available full-duality and strong-duality theorems. Although we now have a much better appreciation of the role of the injectivity of $\underset{\sim}{P}$ and $\underline{P}$, the problem posed in the lectures remains.

Problem 2 (a) Find an example of a finite algebra $\underline{P}$ and a structure $\underset{\sim}{P}$ which is algebraic over $\underline{P}$ such that $\underset{\sim}{P}$ is not injective in $\mathcal{X}$ yet yields a full duality on $\mathcal{A}$. (b) Find an example of a natural duality which is full but not strong.

By Lemma 3.8, the algebra $\underline{P}$ in (a) will not be injective in $\mathcal{A}$ and hence, by the Total-Structure Theorem, the structure on $\underset{\sim}{P}$ will include proper partial operations. The main result of Section 7 (Theorem 7.2) shows that the algebra $\underline{P}$ cannot generate a congruence-distributive variety. Of course, a solution to (a) would also solve (b).

4 Examples

At this stage our tour becomes more local as we see our theorems at work on the ground. Given the scope and utility of our duality and strong-duality theorems, we usually find that most of the pain has been taken out of proving that $\underset{\sim}{P}$ yields a strong duality on $\mathcal{A}$. The real work is in finding axioms for the dual class $\mathcal{X} = \mathbb{IS}_c\mathbb{P}(\underset{\sim}{P})$. Here we have no general theorems. The proofs are example-specific and necessarily involve topological arguments. The steps involved in obtaining an axiomatization of $\mathcal{X}$ are set out below.

- Give a set Σ of "axioms" (involving G, H, R and the topology) such that:

  1. $\underset{\sim}{P}$ satisfies Σ,
  2. if X satisfies Σ and Y is isomorphic to X, then Y satisfies Σ,
  3. if X satisfies Σ and Y is a closed substructure of X then Y satisfies Σ,
  4. if X_i satisfies Σ for all $i \in I$, then $\prod\{\, x_i \mid i \in I \,\}$ satisfies Σ.

  These conditions give $\mathcal{X} \subseteq \mathrm{Mod}(\Sigma)$.

- Show that each $X \in \text{Mod}(\Sigma)$ satisfies the following separation conditions:
  1. if $x, y \in X$ with $x \neq y$, then there exists a morphism $\alpha : X \to \underset{\sim}{P}$ such that $\alpha(x) \neq \alpha(y)$,
  2. if h is an n-ary partial operation in H and $(x_1, \ldots, x_n) \in X^n$ is not in the domain of h on X, then there exists a morphism $\alpha : X \to \underset{\sim}{P}$ such that $(\alpha(x_1), \ldots, \alpha(x_n))$ is not in the domain of h on $\underset{\sim}{P}$,
  3. if r is an n-ary relation in R and $(x_1, \ldots, x_n) \in X^n$ with $(x_1, \ldots, x_n) \notin r$, then there exists a morphism $\alpha : X \to \underset{\sim}{P}$ such that $(\alpha(x_1), \ldots, \alpha(x_n)) \notin r$.

  These separation conditions guarantee that each $X \in \text{Mod}(\Sigma)$ is isomorphic to a closed substructure of a power of $\underset{\sim}{P}$, whence $\text{Mod}(\Sigma) \subseteq \mathcal{X}$.

While an informal use of the word "axiom" will be fully adequate for the present tour, note that Clark and Krauss [5] give an explicit formal language in which $\mathcal{X}$ may be axiomatized for any choice of $\underset{\sim}{P}$.

In each of the examples below, we describe $\underline{P}$ and $\underset{\sim}{P}$, state the axiomatization of the dual class $\mathcal{X}$ and indicate which of our theorems should be applied when proving that $\underset{\sim}{P}$ yields a (strong) duality on $\mathcal{A}$. Any missing details may be found in Section 2 of Davey and Werner [19].

Abelian groups of exponent at most m The variety $\mathcal{A}_m$ of abelian groups of exponent at most m is generated by the cyclic group $\underline{\mathbb{Z}}_m = \langle \mathbb{Z}_m; +, -, 0 \rangle$, in fact $\mathcal{A}_m = \mathbb{ISP}(\underline{\mathbb{Z}}_m)$. By choosing $\underset{\sim}{\mathbb{Z}}_m = \langle \mathbb{Z}_m; +, -, 0, \mathcal{T} \rangle$ we obtain Pontryagin's duality restricted to $\mathcal{A}_m$. It is easy to see that $\underline{\mathbb{Z}}_m$ is injective in $(\mathcal{A}_m)_{fin}$ and that every homomorphism from $\underline{\mathbb{Z}}_m^n$ to $\underline{\mathbb{Z}}_m$ is a term function on $\underline{\mathbb{Z}}_m$. Hence, by the Strong Self-Duality Theorem, $\underset{\sim}{\mathbb{Z}}_m$ yields a strong duality on $\mathcal{A}_m$. The class $\mathcal{X} = \mathbb{IS}_c\mathbb{P}(\underset{\sim}{\mathbb{Z}}_m)$ is the class of all compact topological abelian groups of exponent at most m whose topology is Boolean.

Vector spaces over a finite field The variety $\mathcal{V}_K$ of vector spaces over a finite field K is generated by the one-dimensional vector space $\underline{K}$ and $\mathcal{V}_K = \mathbb{ISP}(\underline{K})$. A very easy application of the Strong Self-Duality Theorem shows that choosing $\underset{\sim}{K}$ to be $\underline{K}$ with the discrete topology yields a strong duality on $\mathcal{V}_K$. The class $\mathcal{X} = \mathbb{IS}_c\mathbb{P}(\underset{\sim}{K})$ is the class of all compact topological vector spaces over K whose topology is Boolean.

Semilattices The variety $\mathcal{H}_1$ of all meet-semilattice with one is generated by the two-element semilattice $\underline{S}_1 = \langle \{0, 1\}; \wedge, 1 \rangle$ and $\mathcal{H}_1$ equals $\mathbb{ISP}(\underline{S}_1)$. With $\underset{\sim}{S}_1 := \langle \{0, 1\}; \wedge, 1, \mathcal{T} \rangle$ we obtain the (strong) duality due to Hofmann, Mislove and Stralka [22] between $\mathcal{H}_1$ and the class $\mathcal{X}$ of all compact topological meet-semilattices with 1 which carry a Boolean topology. Once again, the Strong Self-Duality Theorem does the work for us.

To obtain a strong duality for the class $\mathcal{H}_{01}$ of bounded meet-semilattices we simply replace $\underline{S}_1$ by $\underline{S}_{01} = \langle \{0, 1\}; \wedge, 0, 1 \rangle$ and $\underset{\sim}{S}_1$ by $\underset{\sim}{S}_{01} = \langle \{0, 1\}; \wedge \rangle$. Note that $\mathbb{ISP}(\underline{S}_{01})$

is the class of non-trivial bounded meet-semilattices and $\mathbb{IS}_c\mathbb{P}(\underset{\sim}{S}_{01})$ is the class of non-empty compact topological meet-semilattices which carry a Boolean topology. To extend the duality to the variety $\mathcal{H}_{01}$ we must add the empty topological meet-semilattice to $\mathcal{X}$ in order to provide a dual for the one-element semilattice.

By the Two-for-One Strong-Duality Theorem we get a second strong duality for free. Thus $\underset{\sim}{S} = \langle \{0,1\}; \wedge, 0, 1\rangle$ yields a strong duality on $\mathcal{A} = \mathbb{ISP}(\underline{S})$ where $\underline{S} = \langle \{0,1\}; \wedge\rangle$. In this case, $\mathcal{A}$ is the class of all (non-empty) meet-semilattices while $\mathcal{X} = \mathbb{ISP}(\underset{\sim}{S})$ is the class of all *non-trivial* compact topological bounded meet-semilattices which carry a Boolean topology.

Sets Let $\underline{S}$ be the set $\{0,1\}$ with an empty set of operations. Then $\mathcal{S} := \mathbb{ISP}(\underline{S})$ is the variety of all (non-empty) sets. We have already seen that $\underset{\sim}{2} = \langle \{0,1\}; \mathcal{T}\rangle$ yields a strong duality on the class $\mathcal{B}$ of Boolean algebras. Since $\underset{\sim}{2}$ is a total algebra, the Two-for-One Strong-Duality Theorem implies that $\underset{\sim}{S} := \langle \{0,1\}; \vee, \wedge, ', 0, 1, \mathcal{T}\rangle$ yields a strong duality on $\mathcal{S}$. The class $\mathcal{X} := \mathbb{IS}_c\mathbb{P}(\underset{\sim}{S})$ turn out to be the class of all compact topological Boolean algebras which carry a Boolean topology. This duality was first proved by Banaschewski [2].

Median algebras The simplest example of an algebra with a near-unanimity term is $\underline{M} = \langle \{0,1\}; m\rangle$ where $m : \{0,1\}^3 \to \{0,1\}$ is the median, i.e. $m(x,y,z) = w$ if and only if w is the (unique) repeated value in the triple (x,y,z). The class $\mathbb{ISP}(\underline{M})$ is the variety $\mathcal{M}$ of median algebras. The subalgebras of $\underline{M}^2$ are the trivial relations Δ and ∇, the products of the subalgebras $\{0\}, \{1\}$ and $\{0,1\}$ of $\underline{M}$, the orders $\leqslant$ and $\geqslant$, the relations $M^2\backslash\{(0,0)\}$ and $M^2\backslash\{(1,1)\}$, and the graph of the automorphism $' : \underline{M}_2 \to \underline{M}_2$ given by $0' = 1$ and $1' = 0$. In fact, every subset of $\underline{M}^2$ is a subalgebra. (Note that in the discussion of median algebras on page 170 of [19], the relations $M^2\backslash\{(0,0)\}$ and $M^2\backslash\{(1,1)\}$ were not listed.) Let $\underset{\sim}{M} = \langle \{0,1\}; ', 0, \leqslant, \mathcal{T}\rangle$. It is easily seen that $'$, 0 and $\leqslant$ generate all subalgebras of $\underline{M}^2$; in particular, note that

$$M^2\backslash\{(0,0)\} = {'}\circ\leqslant \quad\text{and}\quad M^2\backslash\{(1,1)\} = {'}\circ\geqslant .$$

Every unary partial operation on $\underline{M}$ extends to either the constant map onto 0, the constant map onto 1, the identity map or the map $'$. Thus no proper partial maps are required to obtain a strong duality. Since we have $'$ and 0 in the structure on $\underset{\sim}{M}$, the two constant maps are not required. The identity map is never required. Since $M(\underline{M}) = 1$, the NU-Strong-Duality Theorem implies that $\underset{\sim}{M}$ yields a strong duality on $\mathcal{M}$. The class $\mathcal{X} = \mathbb{IS}_c\mathbb{P}(\underset{\sim}{M})$ is the class of all Priestley spaces with a least element 0 and a homeomorphism $'$ satisfying

$$x'' \approx x \;\;\&\;\; (x \leqslant y \iff y' \leqslant x') \;\;\&\;\; (x \leqslant x' \iff x = 0).$$

This duality is due to Werner [35].

Kleene algebras The variety $\mathcal{K}$ of Kleene algebras is $\mathcal{K} = \mathbb{ISP}(\underline{K})$, with the one-element algebras adjoined, where

$$\underline{K} = \langle \{0, d, 1\}; \vee, \wedge, ', 0, 1\rangle.$$

The underlying lattice is the three-element chain $0 < d < 1$ and $'$ is the negation $0' = 1$, $1' = 0$ and $d' = d$. This is a very important 3-valued logic: 0 and 1 correspond to the usual

Boolean truth values of *"false"* and *"true"* while d corresponds to *"don't know"*. The only proper subalgebra of $\underline{K}$ is $K_0 = \{0, 1\}$. (Think of K_0 as *"being in the state of knowledge"*—if we are in the state of knowledge, then we know whether a given statement is true of false.) The *"uncertainty order"*,

$$\preccurlyeq = \{(0,0), (d,d), (1,1), (0,d), (1,d)\},$$

in which 0 and 1 have minimal uncertainty and d has the most uncertainty, is a subalgebra of $\underline{K}^2$. The unary algebraic (partial) operations on $\underline{K}$ are id_{K_0} and id_K. The subalgebras of $\underline{K}^2$ which are not graphs of (partial) maps are the products $K_0 \times K_0$, $K_0 \times K$, $K \times K_0$, $K \times K$, the order, $\preccurlyeq$, and its converse, $\succcurlyeq$, along with

$$\preccurlyeq \cap (K_0 \times K), \ \succcurlyeq \cap (K \times K_0), \ \text{and} \ \succcurlyeq \circ \preccurlyeq \, = \, \preccurlyeq \circ \succcurlyeq .$$

Denote the relation $\succcurlyeq \circ \preccurlyeq = K^2 \backslash \{(0,1), (1,0)\}$ by $-$. Thus $\{\preccurlyeq, K_0, -\}$ generates all subalgebras of $\underline{K}^2$. Since id_{K_0} extends to id_K, we require no (partial) operations in order to obtain duality provided K_0 is in the structure on $\underset{\sim}{K}$. Since $M(\underline{K}) = 1$, the NU-Strong-Duality Theorem again shows that $\underset{\sim}{K} = \langle \{0, d, 1\}; \preccurlyeq, -, K_0, \mathcal{T} \rangle$ yields a strong duality on $\mathbb{ISP}(\underline{K})$. The class $\mathcal{X} = \mathbb{IS}_c\mathbb{P}(\underset{\sim}{K})$ is the class of all nonempty Priestley spaces X with a distinguished closed subspace X_0 (possibly empty) and a closed binary relation, $-$, satisfying

1. $(\forall x \in X)\ x - x$
2. $(\forall x, y \in X)\ (x - y \ \ \& \ \ x \in X_0) \implies x \preccurlyeq y$
3. $(\forall x, y, z \in X)\ (x - y \ \ \& \ \ y \preccurlyeq z) \implies z - x.$

In order to extend the duality to the one-element algebras and therefore to the whole variety $\mathcal{K}$, we must add the empty structure to $\mathcal{X}$.

If we wish only to describe the clone of term functions on $\underline{K}$, then, since $-$ is just $\succcurlyeq \circ \preccurlyeq$, we can drop $-$. Thus a map $t : K^n \to K$ is a term function on $\underline{K}$ if and only if it preserves the unary relation of *"being in the state of knowledge"* and the binary relation of *"uncertainty"*. This duality is due to Davey and Werner [19].

A number of other examples may be found in Davey and Werner [19]. These include varieties generated by quasi-primal algebras (see also [5]), varieties of weakly associative lattices, de Morgan algebras, Stone algebras and double Stone algebras. In each case, the duality was proved to be full by establishing **(FTC)**. We can now give an easier proof via the NU-Strong-Duality Theorem since, in each case, we have $M(\underline{P}) = 1$.

5 Piggyback dualities

We have already travelled a considerable distance on foot. The time has come to hitch a ride and let someone else do a lot of the work for us. We shall now see how to obtain a natural duality for a class $\mathcal{A} = \mathbb{ISP}(\underline{P})$ by riding piggyback on an existing duality. The theory of piggyback dualities has its roots in the author's thesis [6] and the papers [7, 9] which followed from it, but the general- algebraic framework was developed somewhat later in Davey and Werner [20, 21] and Davey and Priestley [13]. The discussion below is the special case of the theory developed in [20, 21, 13] which occurs when we ride piggyback on Priestley's duality for bounded distributive lattices.

For this section only, denote the two-element bounded distributive lattice by $\underline{\underline{2}}$ and denote the two-element Priestley space by $\underset{\approx}{2}$. Thus

$$\underline{\underline{2}} = \langle\{0,1\}; \vee, \wedge, 0, 1\rangle \text{ and } \underset{\approx}{2} = \langle\{0,1\}; \leqslant, \mathcal{T}\rangle,$$

and hence $\mathcal{D} := \mathbb{ISP}(\underline{\underline{2}})$ is the variety of bounded distributive lattices and $\mathcal{P} := \mathbb{IS}_c\mathbb{P}(\underset{\approx}{2})$ is the class of Priestley spaces (modulo one-element lattices and empty spaces).

Let $\underline{P} = \langle P; F\rangle$ be a finite algebra which has a term-definable bounded-distributive-lattice structure, i.e. there are binary terms $\vee, \wedge$ and nullary or constant-unary terms $0, 1$ on $\underline{P}$ such that $\underline{\underline{P}} := \langle P; \vee, \wedge, 0, 1\rangle$ is a bounded distributive lattice. Thus each algebra $A \in \mathcal{A} := \mathbb{ISP}(\underline{P})$ has a term-definable bounded-distributive-lattice structure (via the terms $\vee, \wedge, 0, 1$) and every homomorphism $u : A \to B$ with $A, B \in \mathcal{A}$ is a $\{0,1\}$-lattice homomorphism between the underlying lattices.

Thus we may view $\mathcal{A}$ as a subclass of $\mathcal{D}$. By restricting the maps $\mathcal{D}(-, \underline{\underline{2}}) : \mathcal{D} \to \mathcal{P}$ and $\mathcal{P}(-, \underset{\approx}{2}) : \mathcal{P} \to \mathcal{D}$ to the class $\mathcal{A}$ we obtain a duality (i.e. a dual category equivalence) between $\mathcal{A}$ and a subclass of $\mathcal{P}$. This is the *restricted Priestley duality* for $\mathcal{A}$. While this hand-me-down duality can be very useful for the study of the class $\mathcal{A}$, it is only rarely a natural duality. Our aim is to use the Priestley duality to read off a natural duality for $\mathcal{A}$. We can then use the restricted Priestley duality and the natural duality in tandem to study the class $\mathcal{A}$. Note that $\underline{P}$ has a 3-ary majority term function (since it has a term-definable lattice structure) and hence $\mathcal{A}$ does have a natural duality by the NU-Duality Theorem. For a discussion of natural dualities with a particular emphasis on algebras with a term-definable bounded-distributive-lattice structure, we highly recommend H.A. Priestley's survey [31]. The following theorem is a special case of Theorem 2.2 from [13] which has been slightly sharpened owing to the fact that the natural duality is riding piggyback on Priestley duality.

Theorem 5.1 (The Piggyback-Duality Theorem) *Let $\underline{P}$ be a nontrivial finite algebra which has a term-definable bounded-distributive-lattice structure $\underline{\underline{P}}$. Let Ω be a set of $\mathcal{D}$-homomorphisms from $\underline{\underline{P}}$ into $\underline{\underline{2}}$ and let G be a set of $\mathcal{A}$-endomorphisms of $\underline{P}$ such that*

$$\{\, \alpha \circ g : P \to 2 \mid \alpha \in \Omega,\ g \in G \,\}$$

separates the points of P. Define R to be the set of all $\mathcal{A}$-subalgebras of $\underline{P}^2$ which are maximal in

$$(\alpha, \beta)^{-1}(\leqslant) := \{\, (a,b) \in P^2 \mid \alpha(a) \leqslant \beta(b) \,\}$$

for some $\alpha, \beta \in \Omega$. Then $\underset{\sim}{P} := \langle P; G, R, \mathcal{T}\rangle$ yields a duality on $\mathcal{A} = \mathbb{ISP}(\underline{P})$.

There is always at least one choice of Ω and G as required by the theorem. Indeed, since $\underline{\underline{P}} \in \mathcal{D} = \mathbb{ISP}(\underline{\underline{2}})$, the choice $\Omega = \mathcal{D}(\underline{\underline{P}}, \underline{\underline{2}})$ and $G = \emptyset$ will suffice. In practice we try to minimize the size of Ω at the expense of increasing the size of G as this will reduce the size of R. We shall give two examples of the Piggyback-Duality Theorem in action.

Kleene algebras revisited Since the three-element Kleene algebra $\underline{K}$ has no non-identity endomorphisms, there is only one possible choice for Ω, namely $\Omega = \{\alpha, \beta\}$ where

$$\alpha(0) = 0,\ \alpha(d) = \alpha(1) = 1 \text{ and } \beta(0) = \beta(d) = 0,\ \beta(1) = 1.$$

We must now find all Kleene subalgebras of $\underline{K}^2$ which are maximal in the following sublattices of $\underline{\underline{K}}^2$:

$$\begin{aligned}(\alpha,\alpha)^{-1}(\leqslant) &= \{(0,0),(d,d),(1,1),(1,d),(d,1),(0,d),(0,1)\},\\ (\alpha,\beta)^{-1}(\leqslant) &= \{(0,0),(1,1),(d,1),(0,d),(0,1)\},\\ (\beta,\beta)^{-1}(\leqslant) = (\beta,\alpha)^{-1}(\leqslant) &= \{(0,0),(d,d),(1,1),(1,d),(d,0),(d,1),(0,d),(0,1)\}.\end{aligned}$$

A simple argument (see Lemma 3.5 of [13]) shows that each of these has a unique maximal Kleene subalgebra. First note that $\mathrm{id}_{K_0} = \{(0,0),(1,1)\}$ is the largest Kleene subalgebra of $(\alpha,\beta)^{-1}(\leqslant)$. Hence we may replace this binary relation by the unary relation K_0. The largest Kleene subalgebra of $(\alpha,\alpha)^{-1}(\leqslant)$ is the order

$$\preccurlyeq = \{(0,0),(d,d),(1,1),(0,d),(1,d)\}$$

and the largest Kleene subalgebra of $(\beta,\alpha)^{-1}(\leqslant) = (\beta,\beta)^{-1}(\leqslant)$ is the relation

$$- = K^2 \backslash \{(0,1),(1,0)\}.$$

Hence, by the Piggyback-Duality Theorem, $\underset{\sim}{K} = \langle K; \preccurlyeq, -, K_0, \mathcal{T}\rangle$ yields a duality on the class $\mathcal{K}$ of Kleene algebras. The distinct advantage that piggybacking has over the bare-hands approach of the previous section is that we do not have to list all subalgebras of $\underline{K}^2$ and then see how to generate them from $\preccurlyeq$, $-$ and K_0.

Relative-Stone Heyting algebras Let $\underline{C}_n$ be the n-element chain regarded as a Heyting algebra. Thus $\underline{C}_n = \langle C_n; \vee, \wedge, *, 0, 1\rangle$, where 0 and 1 are the bounds of the chain and $*$ is relative pseudocomplementation given by

$$a * b = \begin{cases} 1 & \text{if } a \leqslant b,\\ b & \text{if } a > b.\end{cases}$$

The algebra $\underline{C}_n$ is an important n-valued logic (see Horn [23]). The class $\mathbb{ISP}(\underline{C}_n)$ is the variety, $\mathcal{L}_n$, of Heyting algebras generated by $\underline{C}_n$ (give or take the one-element algebras). It is easily seen that a map $e : C_n \to C_n$ is an endomorphism of $\underline{C}_n$ if and only if e is order-preserving, $e(0) = 0$, and there exists $a \in C_n$ such that $e(b) = 1 \iff b \geqslant a$ and e is one-to-one on $\{c \in C_n \mid c < a\}$. From this it follows that $G = \mathrm{End}(\underline{C}_n)$ and $\Omega = \{\alpha\}$, where $\alpha(1) = 1$ and $\alpha(a) = 0$ if $a \neq 1$, satisfy the conditions of the Piggyback-Duality Theorem.

Let r be a subalgebra of $\underline{C}_n^2$ which is contained in

$$(\alpha,\alpha)^{-1}(\leqslant) = (C_n \backslash \{1\} \times C_n) \cup \{(1,1)\}.$$

Then $(1,a) \in r$ if and only if $a = 1$. Hence

$$(a,b),(a,c) \in r \implies (1, b*c) = (a*a, b*c) = (a,b)*(a,c) \in r,$$

whence $b * c = 1$ and so $b \leqslant c$. Thus, by summetry,

$$(a, b), (a, c) \in r \implies b = c,$$

i.e. r is the graph of a unary (partial) operation on $\underline{C}_n$. Let H be the set of all proper unary partial operations which are algebraic over $\underline{C}_n$. By the Piggyback-Duality Theorem, the structure

$$\underset{\sim}{C}_n := \langle C_n; \mathrm{End}(\underline{C}_n), H, \mathcal{T}\rangle$$

yields a duality on $\mathcal{L}_n$ which, by the NU-Strong-Duality Theorem is strong. (We again have $M(\underline{C}_n) = 1$.) As usual, we allow the empty structure in $\mathcal{X}$ in order to provide a dual for the one-element algebra. The fact that this duality is strong is a new result—it was proved in the aeroplane on the way to Montréal when your tour guide first realised that the $M(\underline{P}) = 1$ case of the NU-Strong-Duality Theorem held.

In fact, it is proved in Davey [7] that the structure

$$\underset{\sim}{C}'_n := \langle C_n; \mathrm{End}(\underline{C}_n), \mathcal{T}\rangle$$

yields a duality on $\mathcal{L}_n$. While this duality need not be full, the class $\mathcal{X}' := \mathbb{I}\mathbb{S}_c\mathbb{P}(\underset{\sim}{C}'_n)$ is simpler than the class $\mathcal{X} := \mathbb{I}\mathbb{S}_c\mathbb{P}(\underset{\sim}{C}_n)$. In [7] the duality via $\underset{\sim}{C}'_n$ is used in order to describe free algebras and injective algebras in $\mathcal{L}_n$. At this stage, it is far from clear that we can drop H and still have a duality. Rather *ad hoc* proofs of this fact are given in [19] and [20]. A new and extremely simple proof will be given during the next stage of our tour.

6 Optimal dualities—Schizophrenia Strikes Again!

Once again we change tack and sail into recently charted waters. Having applied the results of the earlier sections to find a natural duality for a class $\mathcal{A} = \mathbb{ISP}(\underline{P})$ of algebras, we now wish to know how to find a structure $\underset{\sim}{P}$ which is minimal with respect to yielding a duality on $\mathcal{A}$. More specifically, we ask: *If $\underset{\sim}{P}$ yields a duality on $\mathcal{A}$ and $\underset{\sim}{P}'$ is obtained by deleting one relation or (partial) operation from the structure on $\underset{\sim}{P}$, when will $\underset{\sim}{P}'$ yield a duality on $\mathcal{A}$?*

Since we have to check that $\underset{\sim}{P}'$ yields a duality on each algebra $A \in \mathcal{A}$, there is no obvious reason why this should be a finite problem—yet it is! Since we are concerned here with duality rather than full or strong duality, we shall replace any (partial) operations by their graphs (see Lemma 2.1) and assume that $\underset{\sim}{P} = \langle P; R, \mathcal{T}\rangle$ is a relational structure which yields a duality on $\mathcal{A}$. We shall now take advantage of an extension of the schizophrenia inherent in the $\underline{P}$ versus $\underset{\sim}{P}$ personality split. Each relation $r \in R$ lives a second life as a subalgebra of some $\underline{P}^n$; we shall denote this algebra by $\underline{r}$. Thus $\underline{r} \in \mathcal{A}$. It is a tantalizing fact that in order to prove that $\underset{\sim}{P}' := \langle P; R\backslash\{r\}, \mathcal{T}\rangle$ still yields a duality on $\mathcal{A}$, it suffices to check that $\underset{\sim}{P}'$ yields a duality on a single, finite algebra—namely the *test algebra* $\underline{r} \in \mathcal{A}$. The main result of this section is surprisingly easy to prove—see Section 2 of Davey and Priestley [15].

Theorem 6.1 (The Test-Algebra Theorem) *Let R be a finite set of finitary algebraic relations on $\underline{P}$ and assume that $\underset{\sim}{P} = \langle P; R, \mathcal{T}\rangle$ yields a duality on $\mathcal{A}$. Let $R' \subseteq R$ and define $\underset{\sim}{P}' := \langle P; R', \mathcal{T}\rangle$. The following are equivalent:*

1. *$\underset{\sim}{P}'$ yields a duality on $\mathcal{A}$;*

2. *$\underset{\sim}{P}'$ yields a duality on the test algebra $\underline{r}$ for each $r \in R\backslash R'$;*

3. *$\underset{\sim}{P}'$ generates the relation r on the test algebra $\underline{r}$ for each $r \in R\backslash R'$.*

If $\underline{P}$ has a majority term, in particular if $\underline{P}$ has a term-definable lattice structure, then, by the NU-Duality Theorem, there is a set R of binary algebraic relations on $\underline{P}$ such that $\underset{\sim}{P} = \langle P; R, \mathcal{T}\rangle$ yields a duality on $\mathcal{A} = \mathbb{ISP}(\underline{P})$. For any set S of binary relations on P define $S^* := \{\, s^* \mid s \in S\,\}$, where s^* is the converse of s. We say that a set S of binary algebraic relations on $\underline{P}$ is *unavoidable (amongst binary relations)* if any set R of binary algebraic relations on $\underline{P}$, such that $\underset{\sim}{P} = \langle P; R, \mathcal{T}\rangle$ yields a duality on $\mathcal{A}$, intersects $S \cup S^*$. A single relation s is *unavoidable* if $\{s\}$ is. In this sense, the relation $\leqslant$ on the two-element bounded distributive lattice is unavoidable. We have a very strong optimality for Priestley duality: $\leqslant$ yields a duality on $\mathcal{D}$ and if R is a set of binary algebraic relations which yields a duality on $\mathcal{D}$, then R contains either $\leqslant$ or $\geqslant$. Davey and Priestley [15] is devoted to finding unavoidable sets of relations and optimal sets of relations which yield natural dualities on the varieties $\mathcal{B}_n$ of pseudocomplemented distributive lattices. In [15], the piggyback philosophy is extended and the interplay between the natural duality and the restricted Priestley duality plays a crucial role.

Relative-Stone Heyting algebras revisited We close this section by applying the Test-Algebra Theorem to prove that $\underset{\sim}{C}'_n := \langle C_n; \mathrm{End}(\underline{C}_n), \mathcal{T}\rangle$ yields a duality on the variety $\mathcal{L}_n$ generated by the Heyting algebra $\underline{C}_n$, as claimed at the end of the previous section. This proof comes from Davey and Priestley [16], a paper on optimal dualities for varieties of Heyting algebras which is still in a state of ferment.

Recall that H is the set of all proper unary partial operations which are algebraic over $\underline{C}_n$. Since graph(h), *qua* algebra, is isomorphic to dom(h) for each $h \in H$ we may take dom(h) as the test algebra for the partial operation h. We proved in Section 5 that $\underset{\sim}{P} = \langle C_n; \mathrm{End}(\underline{C}_n), H, \mathcal{T}\rangle$ yields a duality on $\mathcal{L}_n$. Thus the Test-Algebra Theorem tells us that in order to prove that H can be deleted without destroying the duality it suffices to show that $\mathrm{End}(\underline{C}_n)$ yields a duality on dom(h) for all $h \in H$, i.e. that $\mathrm{End}(\underline{C}_n)$ yields a duality on every proper subalgebra of $\underline{C}_n$. This will follow from the next two lemmas which apply to arbitrary finite algebras, not just to $\underline{C}_n$. Note the schizophrenia inherent in the proof below: $\mathrm{End}(\underline{P})$ is the structure on $\underset{\sim}{P}$ and simultaneously is the dual of $\underline{P}$ since $D(\underline{P}) = \mathcal{A}(\underline{P}, \underline{P}) = \mathrm{End}(\underline{P})$.

Lemma 6.2 $\mathrm{End}(\underline{P})$ *yields a duality on* $\underline{P}$.

Proof Let $\underset{\sim}{P} = \langle P; \mathrm{End}(\underline{P}), \mathcal{T}\rangle$. We must show that the map

$$e_{\underline{P}} : \underline{P} \to \mathcal{X}(\mathcal{A}(\underline{P}, \underline{P}), \underset{\sim}{P})$$

is surjective. To this end, let $\alpha : \mathcal{A}(\underline{P}, \underline{P}) \to \underset{\sim}{P}$ be a morphism and define $a := \alpha(\mathrm{id}_{\underline{P}})$. Let $g \in \mathcal{A}(\underline{P}, \underline{P})$; then, since α preserves g, we have

$$(e_{\underline{P}}(a))(g) = g(a) = g(\alpha(\mathrm{id}_{\underline{P}})) = \alpha(g(\mathrm{id}_{\underline{P}})) = \alpha(g \circ \mathrm{id}_{\underline{P}}) = \alpha(g).$$

Thus $e_{\underline{P}}(a) = \alpha$ and consequently $e_{\underline{P}}$ is an isomorphism. □

A simple modification of this proof shows that if A is a subalgebra of $\underline{P}^n$, then the set $\mathcal{A}(A, \underline{P})$ of partial operations with domain A yields a duality on A.

Lemma 6.3 *If $\underset{\sim}{P}$ yields a duality on an algebra $A \in \mathcal{A}$, then $\underset{\sim}{P}$ yields a duality on every retract B of A.*

Proof Recall that B is a retract of A if there is an embedding $u : B \to A$ and a surjective homomorphism $v : A \to B$ such that $v \circ u = \mathrm{id}_B$. Thus, by Lemma 1.1, we have $D(u) \circ D(v) = \mathrm{id}_{D(B)}$ and hence $D(B)$ is a retract of $D(A)$. It is now a follow-your-nose argument to see that if e_A is an isomorphism, then e_B is an isomorphism also. □

We wish to prove that $\mathrm{End}(\underline{C}_n)$ yields a duality on every proper subalgebra of $\underline{C}_n$. By the two lemmas above, it would suffice to show that every proper subalgebra of $\underline{C}_n$ is a retract—unfortunately this is false! But a structure $\underset{\sim}{P}$ yields a duality on A provided it yields a duality on any isomorphic copy of A. Hence it suffices to prove that every proper subalgebra of $\underline{C}_n$ is isomorphic to a retract of $\underline{C}_n$, and this is true! If we let

$$C_n = \{c_1, c_2, \ldots, c_n\} \text{ with } 0 = c_1 < c_2 < \ldots < c_n = 1,$$

then, for $2 \leqslant k \leqslant n$, the map $v : C_n \to \{c_1, c_2, \ldots, c_{k-1}, 1\} \subseteq C_n$, defined by

$$v(c_i) = \begin{cases} c_i & \text{if } 1 \leqslant i < k, \\ 1 & \text{if } k \leqslant i \leqslant n, \end{cases}$$

is a retraction of $\underline{C}_n$ onto a k-element subalgebra. Hence $\underset{\sim}{C}'_n = \langle C_n; \mathrm{End}(\underline{C}_n), \mathcal{T}\rangle$ yields a duality on $\mathcal{L}_n$. This duality is strong if and only if $n = 2$ or $n = 3$ as $\underline{C}_n$ is not injective in $\mathcal{L}_n$ for $n \geqslant 4$.

7 Algebras which admit a duality

As our tour draws to a close, we bring the focus of our attention back to where it all began, with the algebra $\underline{P}$. Up to now, we have concentrated more on the structure $\underset{\sim}{P}$. How do we find $\underset{\sim}{P}$? Once found, how do we refine it? How do we show that a putative $\underset{\sim}{P}$ will, in fact, yield a (full or strong) duality? Here we consider a more fundamental question.

We shall say that a finite algebra $\underline{P}$ *admits a duality* (or, when we are being more colloquial, is *dualizable*) if there is some structure $\underset{\sim}{P}$ (algebraic over $\underline{P}$) which yields a duality on $\mathcal{A} := \mathbb{ISP}(\underline{P})$. By Theorem 2.2, this is equivalent to saying that brute force yields a duality on $\mathcal{A}$.

Problem 3 Which finite algebras admit a duality?

We present some refinements of this problem later in the section. The two theorems below are from Davey, Heindorf and McKenzie [12].

Assume that $\underline{P}$ has a majority term. Then by the NU-Duality Theorem, $\underline{P}$ admits a duality. Moreover, $\underline{P}$ admits a particularly well-behaved duality since $\underset{\sim}{P}$ can be chosen such that

1. $\underset{\sim}{P}$ is of finite type, i.e. $G \cup H \cup R$ is finite,
2. the operations in G and the partial operations in H are at most unary and the relations in R are at most binary, and
3. $\underset{\sim}{P}$ is injective in $\mathcal{X}$.

Our first theorem provides a converse.

Theorem 7.1 *Assume that $\underset{\sim}{P}$ is of finite type and that the (partial) operations in $G \cup H$ are at most unary. Let m be the maximum of the arities of the relations in R. If $\underset{\sim}{P}$ yields a duality on $\mathcal{A}$ with $\underset{\sim}{P}$ injective in $\mathcal{X}_{fin}$, then $\underline{P}$ has a $(k+1)$-ary near-unanimity term where $k = \max\{2, m\}$.*

Proof Let $\underset{\sim}{P}$ and k be as described in the statement of the theorem. For $a, b \in P$ define

$$X_{ab} := \{(a, \ldots, a, b), (a, \ldots, b, a), \ldots, (b, a, \ldots, a)\} \subseteq P^{k+1}$$

and define $X := \bigcup\{\, X_{ab} \mid a, b \in P\,\}$. Since the (partial) operations in $G \cup H$ are at most unary, X is a substructure of $\underset{\sim}{P}^{k+1}$. Define $\alpha : X \to P$ by $\alpha(x) = a$ if $x \in X_{ab}$. On any k-or-fewer-element subset of X, the map α is a projection. Hence α preserves all the relations in R since they are at most k-ary, and clearly α preserves the (partial) operations in $G \cup H$ as they are at most unary. Thus α is a morphism. Since $\underset{\sim}{P}$ is injective in $\mathcal{X}_{fin}$ there is a morphism $t : \underset{\sim}{P}^{k+1} \to \underset{\sim}{P}$ which extends α. As $\underset{\sim}{P}$ yields a duality on $\mathcal{A}$, the map t is a $(k+1)$-ary term function on $\underline{P}$ and $t{\restriction}X = \alpha$ says exactly that t is a near-unanimity function. □

This is our only result where the size of $G \cup H \cup R$ has played a role. In fact, in every known duality, $\underset{\sim}{P}$ can be chosen to be of finite type.

Problem 4 Is it true that if $\underline{P}$ admits a duality, then it admits a duality of finite type?

Recall that a lattice is join-semi-distributive if it satisfies the quasi-identity

$$x \vee y = x \vee z \implies x \vee (y \wedge z) = x \vee y.$$

Clearly distributive lattices are join-semi-distributive. Associated with an algebra $A \in \mathcal{A}$ are two natural lattices of congruences. The first is simply the lattice $\mathrm{Con}(A)$ of all congruences. The second is the lattice $\mathrm{Con}_{\mathcal{A}}(A)$ consisting of $\nabla = A^2$ along with all relative congruences, i.e. those congruences θ on A such that $A/\theta \in \mathcal{A}$. Since $\mathcal{A}$ need not be a variety (and therefore closed under homomorphic images), $\mathrm{Con}_{\mathcal{A}}(A)$ is, in general, a proper subset of $\mathrm{Con}(A)$.

Note that $\mathrm{Con}_{\mathcal{A}}(A)$ is closed under arbitrary meets in $\mathrm{Con}(A)$ but need not be a sublattice. A class $\mathcal{A}$ is called *congruence distributive* if $\mathrm{Con}A$ is distributive for all $A \in \mathcal{A}$ and is called *relatively congruence distributive* if $\mathrm{Con}_{\mathcal{A}}A$ is distributive for all $A \in \mathcal{A}$. *Congruence join-semi-distributivity* and *relative congruence join-semi-distributivity* are defined similarly. It is well known and if $\underline{P}$ has a near-unanimity term, then the variety generated by $\underline{P}$, and hence the class $\mathbb{ISP}(\underline{P})$, is congruence distributive.

Theorem 7.2 *Let $\underline{P}$ be a finite algebra and let $\mathcal{A} := \mathbb{ISP}(\underline{P})$. The following are equivalent:*

1. *$\underline{P}$ has a near-unanimity term;*
2. *$\underline{P}$ generates a congruence-distributive variety and $\underline{P}$ admits a duality;*
3. *every finite algebra in $\mathcal{A}$ is congruence join-semi-distributive and $\underline{P}$ admits a duality.*

Furthermore, if every finite algebra in $\mathcal{A}$ is relatively congruence join-semi-distributive and $\underline{P}$ admits a duality, then $\underline{P}$ has a near-unanimity term.

This theorem has an immediate application to order-primal algebras. An algebra $\underline{P} = \langle P; F\rangle$ is *order-primal* if there is some order on P such that for all $n \in \mathbb{N}$ a map $t : P^n \to P$ is a term function on $\underline{P}$ just when t preserves this order. For example, the two-element bounded distributive lattice is order-primal with respect to its underlying order. It is proved in Davey, Quackenbush and Schweigert [17] (see also McKenzie [25]) that if $\underline{P}$ is order-primal, then $\mathcal{A} := \mathbb{ISP}(\underline{P})$ is relatively congruence distributive. Thus we conclude that *an order-primal algebra $\underline{P}$ admits a duality if and only if $\underline{P}$ has a near-unanimity term.* This improves Theorem 2.3 of [17].

Theorem 7.2 makes it very easy to find algebras which do not admit a duality. For example, if $\langle P; \leqslant\rangle$ is a crown with more than four elements, and $\underline{P}$ is an order-primal algebra with respect to $\langle P; \leqslant\rangle$, then $\underline{P}$ has no near-unanimity term (see [10]) and hence $\underline{P}$ does not admit a duality. Let $\underline{P} = \langle\{0,1\}; \to\rangle$ be the two-element implication algebra—thus $x \to y := x' \vee y$. Then, by [26], $\underline{P}$ generates a congruence-distributive variety but $\underline{P}$ has no near-unanimity term. Thus $\underline{P}$ does not admit a duality. (This was proved directly in Section 2.1 on pages 148–151 of [19].)

We now step out of the congruence-distributive realm into the wider congruence-modular realm in order to pose two questions.

Problem 5 Is every finite group $\underline{G}$ which admits a duality necessarily abelian?

Problem 6 Prove or disprove and refine: if $\mathcal{A} = \mathbb{ISP}(\underline{P})$ is congruence modular (congruence permutable) and $\underline{P}$ admits a duality, then either $\underline{P}$ has a near-unanimity term or $\underline{P}$ is abelian.

There are several questions concerning algebras which admit a duality which are most naturally expressed in terms of clones. If $\mathrm{Clo}(\underline{P})$ is the clone of all term functions on $\underline{P}$, then we obtain a new algebra $\underline{P}^* := \langle P; \mathrm{Clo}(\underline{P})\rangle$. In almost all respects the algebras $\underline{P}$ and $\underline{P}^*$ are equivalent since the operations of each are definable in terms of the operations of the other. Of course, $\underline{P}$ admits a duality if and only if $\underline{P}^*$ admits a duality. If C is a clone on the set P, then we say that C *admits a duality* (or is *dualizable*) if the algebra $\underline{P}_C = \langle P; C \rangle$ admits a duality. If R is a set of finitary relations on a set P, then

$$C_R := \bigcup_{n \in \mathbb{N}} \{\, \varphi : P^n \to P \mid \varphi \text{ preserves the relations in } R \,\}$$

is a clone on P. If P is a finite set, then every clone on P is of the form C_R for some set R of relations on P. (Indeed, the Brute-Force Duality theorem implies that if C is a clone on P, then $C = C_R$ where R is the set of all subalgebras of finite powers of the algebra $\underline{P}_C$.) We shall say that the clone C is *determined* by R. While every clone on the finite set P is determined by some set R of relations, not every clone is determined by a finite set R of relations. Since there are finite order-primal algebras which do not admit a duality, not every clone determined by a finite set of relations admits a duality. To date, we have no counter-example to the converse.

Problem 7 Is every dualizable clone determined by a finite set of relations?

If the answer to Problem 4 is 'Yes', then the answer to Problem 7 is 'Yes' also. A positive answer to Problem 7 would also provide a positive answer to our next question.

Problem 8 Let P be a finite set. Is there only a countable number of clones on P which admit a duality?

Note that for $n \geqslant 3$ there is an uncountable number of clones on an n-element set—L. Heindorf has recently shown that an uncountable number of these do not admit a duality. The lattice of clones on a two-element set is countable and has been completely described—see Post [27] or Szendrei [34]. Nineteen of the clones on $\{0,1\}$ include the median operation and hence admit a duality. There are eight infinite descending chains of clones on $\{0,1\}$ of the form

$$C_1 \supset C_2 \supset \ldots \supset C_n \supset \ldots,$$

where C_n contains an $(n+3)$-ary near-unanimity term but contains no $(n+2)$-ary near-unanimity term: each of these clones admits a duality. We saw in Section 4 that the four clones on $\{0,1\}$ generated by $\{\wedge\}$, $\{\wedge, 0\}$, $\{\wedge, 1\}$ and $\{\wedge, 0, 1\}$ admit dualities, and by symmetry the clones generated by $\{\vee\}$, $\{\vee, 0\}$, $\{\vee, 1\}$ and $\{\vee, 0, 1\}$ also admit dualities. The clone generated by $\{+\}$ is the clone of $\underline{\mathbb{Z}}_2 = \langle \{0,1\}; +, -, 0 \rangle$ and so admits a duality. By symmetry the clone of $\{\mp\}$, where $x \mp y := x + y + 1 \pmod 2$, admits a duality. Straightforward applications of the Two-for-One Strong-Duality Theorem show that $\{+, '\}$ dualizes $\{p\}$ and *vice versa* (where $p(x,y,z) = x + y + z$ is the Mal′cev function) and that $\{p, '\}$ is self dual. The remaining six clones on $\{0,1\}$ which admit a duality are generated by $\{', 0, 1\}$, $\{'\}$, $\{0\}$, $\{1\}$, and $\emptyset$. Since each of these clones has the median as a homomorphism of the algebra $\underline{P}_C$ we may apply the NU-Strong-Duality Theorem and the Two-for-One Strong-Duality Theorem to show that each of these clones is dual to one of the clones which contain the median. This leaves just eight clones on $\{0,1\}$, the meets of the eight infinite descending chains. None of these admits a duality. Four of them form the interval between the clone C_1 generated by d, where $d(x,y,z) := x \vee (y \wedge z)$, and the clone C_2 generated by $\{\to\}$. Since $\underline{P}_{C_1}$ generates a congruence-distributive variety and since C_2 contains no near-unanimity function, it follows from Theorem 7.2 that these four clones do not admit dualities. The other four clones form the interval between the clone generated by

$\{d'\}$, where $d'(x,y,z) := x \wedge (y \vee z)$, and the clone generated by $\{\leftarrow\}$, where $x \leftarrow y := x \wedge y'$. By symmetry, these clones do not admit dualities. Thus there are exactly 8 clones on $\{0,1\}$ which do not admit a duality. In relation to Problems 4 and 7, it should be noted that the dualities obtained above are of finite type and that the eight clones on $\{0,1\}$ which do not admit a duality are precisely the clones which are not determined by a finite set of relations. The full details of these dualities will appear in Davey and Rosenberg [18].

So ends our guided tour of duality theory. The many loose ends and unsolved problems are sure to drive the topic forward for some time yet. Thus we break the journey, not at some final destination but somewhere *en route.*

Your tour guide leaves you with *his* final verse for the conference song.

If I should wander from the Path, (Now Alouette knows it all.)
Mark my tomb with this epitaph. (Now Alouette knows it all.)
"Like Galois, the romantic fool
Died fighting in an $\mathcal{A}$-class duel."

References

[1] Baker, K.A., and Pixley, A.F., Polynomial interpolation and the Chinese remainder theorem for algebraic systems, *Math. Z.* **143** (1975), 165–174.

[2] Banaschewski, B., Projective covers in categories of topological spaces and topological algebras, in *General Topology and its Relation to Modern Analysis and Algebra* (*Kanpur 1968*), Academia, Prague 1971, 63–91.

[3] Clark, D.M., and Davey, B.A., The quest for full dualities, (in preparation).

[4] Clark, D.M., and Davey, B.A., *Dualities for the Working Algebraist*, (in preparation).

[5] Clark, D.M., and Krauss, P.H., Topological quasi varieties, *Acta. Sci. Math.* (Szeged) **47** (1984), 3–39.

[6] Davey, B.A., *Duality Theory for Quasi-Varieties of Universal Algebras*, Ph.D. thesis, University of Manitoba, Winnipeg, Manitoba, Canada, 1974.

[7] Davey, B.A., Dualities for equational classes of Brouwerian algebras and Heyting algebras, *Trans. Amer. Math. Soc.* **221** (1976), 119–146.

[8] Davey, B.A., Topological dualities for prevarieties of universal algebras, in G.-C. Rota (ed.), *Studies in Foundations and Combinatorics*, Advances in Mathematics Supplementary Studies, **1**, 1978, 61–99.

[9] Davey, B.A., Dualities for Stone algebras, double Stone algebras, and relative Stone algebras, *Colloq. Math.* **46** (1982), 1–14.

[10] Davey, B.A., Monotone clones and congruence modularity, *Order* **6** (1990), 389–400.

[11] Davey, B.A., and Duffus, D., Exponentiation and duality, in I. Rival (ed.), *Ordered Sets*, NATO ASI Series C 83, Reidel, Dordrecht 1982, 43–95.

[12] Davey, B.A., Heindorf, L., and McKenzie, R., Near unanimity: an obstacle to general duality theory, to appear in *Algebra Universalis.*

[13] Davey, B.A., and Priestley, H.A., Generalized piggyback dualities and applications to Ockham algebras, *Houston Math. J.* **13** (1987), 151–197.

[14] Davey, B.A., and Priestley, H.A., *Introduction to Lattices and Order*, Cambridge University Press, Cambridge, 1990.

[15] Davey, B.A., and Priestley, H.A., Optimal natural dualities, to appear in *Trans. Amer. Math. Soc.*

[16] Davey, B.A., and Priestley, H.A., Optimal dualities for varieties of Heyting algebras, (in preparation).

[17] Davey, B.A., Quackenbush, R.W., and Schweigert, D., Monotone clones and the varieties they determine, *Order* **7** (1990), 145–167.

[18] Davey, D.A., and Rosenberg, I., Clones and duality theory, (in preparation).

[19] Davey, B.A., and Werner, H., Dualities and equivalences for varieties of algebras, in A.P. Huhn and E.T. Schmidt (eds.), *Contributions to Lattice Theory*, Colloq. Math. Soc. János Bolyai **33**, North-Holland, Amsterdam, 1983, 101–275.

[20] Davey, B.A., and Werner, H., Piggyback-dualitäten, *Bull. Austral. Math. Soc.* **32** (1985), 1–32.

[21] Davey, B.A., and Werner, H., Piggyback dualities, in L. Szabó and A. Szendrei (eds.), *Lectures in Universal Algebra*, Colloq. Math. Soc. János Bolyai **43**, North-Holland, Amsterdam, 1986, 61–83.

[22] Hofmann, K.H., Mislove, M., and Stralka, A., *The Pontryagin Duality of Compact 0-dimensional Semilattices and its Applications*, Lecture Notes in Mathematics **396**, Springer, Berlin–Heidelberg–New York, 1974.

[23] Horn, A., Logic with truth values in a linearly ordered Heyting algebra, *J. Symbolic Logic* **34** (1969), 395–408.

[24] MacLane, S., *Categories for the Working Mathematician*, Graduate Texts in Mathematics **5**, Springer, New York–Heidelberg–Berlin, 1971.

[25] McKenzie, R., Monotone clones, residual smallness and congruence distributivity, *Bull. Austral. Math. Soc.* **41** (1990), 283–300.

[26] Mitschke, A., Implication algebras are 3-permutable and 3-distributive, *Algebra Universalis* **1** (1971), 182–186.

[27] Post, E.L., *The Two-Valued Iterative Systems of Mathematical Logic*, Annals of Math. Studies No. **5**, Princeton University Press, Princeton, N.J., 1941.

[28] Priestley, H.A., Representation of distributive lattices by means of ordered Stone spaces, *Bull. London Math. Soc.* **2** (1970), 186–190.

[29] Priestley, H.A., Ordered topological spaces and the representation of distributive lattices, *Proc. London Math. Soc.* (3) **2** (1972), 507–530.

[30] Priestley, H.A., Ordered sets and duality for distributive lattices, *Annals of Discrete Math.* **23** (1984), 39–60.

[31] Priestley, H.A., Natural dualities, Proceedings of the Birkhoff Symposium, Darmstadt, 1991.

[32] Pontryagin, L.S., *Topological Groups*, 2nd edition, Gordon and Breach, New York, 1966.

[33] Stone, M.H., The theory of representations for Boolean algebras, *Trans. Amer. Math. Soc.* **4** (1936), 37–111.

[34] Szendrei, A., *Clones in Universal Algebra*, Séminaire de Mathématiques Supérieures, Vol. **99**, Les Presses de l'Université de Montréal, 1986.

[35] Werner, H., A duality for weakly associative lattices, in B. Csákány and I. Rosenberg (eds), *Finite Algebra and Multi-Valued Logic*, Colloq. Math. Soc. János Bolyai **28**, North-Holland, Amsterdam, 1981, 781–808.

Algebraic Ordered Sets and Their Generalizations

Marcel ERNÉ
Institut für Mathematik
Universität Hannover
Welfengarten 1
DW-3000 Hannover
Germany

Abstract

We study order-theoretical, algebraic and topological aspects of compact generation in ordered sets. Today, algebraic ordered sets (a natural generalization of algebraic lattices) have their place not only in classical mathematical disciplines like algebra and topology, but also in theoretical computer sciences. Some of the main statements are formulated in the language of category theory, because the manifold facets of algebraic ordered sets become more transparent when expressed in terms of equivalences between suitable categories. In the second part, collections of directed subsets are replaced with arbitrary selections of subsets $\mathcal{Z}$. Many results on compactness remain true for the notion of $\mathcal{Z}$-compactness, and the theory is now general enough to provide a broad spectrum of seemingly unrelated applications. Among other representation theorems, we present a duality theorem encompassing diverse specializations such as the Stone duality, the Lawson duality, and the duality between sober spaces and spatial frames.

Acknowledgement

The TEXnical version of this paper has been written with skill, care and patience by Jürgen Reinhold, who has also contributed a lot of corrections and improvements. Needless to say, the remaining errors have to be attributed to the author alone.

I. G. Rosenberg and G. Sabidussi (eds.), Algebras and Orders, 113–192.

1 Chains and directed sets

In many mathematical developments, and in particular, in the theory of computation, various types of approximations and limit processes for "ideal objects" carry a natural order structure (see e.g. [Pl], [Scl], [Sm], [Sty], [We]). Most convenient for practical use is the approximation by sequences or "ω-chains"; but sometimes these are not sufficient, so that one has to work with certain well-ordered or linearly ordered collections of approximations, or even with directed sets or nets — a well-known phenomenon in topology and functional analysis.

In the order-theoretical framework we are concerned with, the "ideal objects" are usually represented by certain *joins*, i.e. least upper bounds of their approximations. In many problems of lattice theory and its applications, we are confronted with situations where it is necessary to form joins and meets not only of linearly ordered or directed sets, but also of finite, countable or even arbitrary sets. It is therefore desirable to bring all these situations under a common umbrella.

The natural tool for such a uniform approach are so-called *subset selections*; these are (class-theoretical) functions $\mathcal{Z}$ assigning to each (partially) ordered set or, more generally, to each quasiordered set Q a collection $\mathcal{Z}Q$ of subsets. As usual, we mean by a *quasiordered set* a set equipped with a *quasiorder*, i.e. a transitive and reflexive relation $\leq$, and in case of antisymmetric quasiorders, we speak of *orders* and of *ordered sets.*

We shall use certain fixed capital letters for the most frequently used subset selections, listed below:

$\mathcal{Z}$	*members of* $\mathcal{Z}Q$	*description*
$\mathcal{A}$	*arbitrary lower sets*	subsets A of Q such that $x \leq y \in A$ implies $x \in A$
$\mathcal{B}$	*binary subsets*	subsets with one or two elements
$\mathcal{C}$	*chains*	nonempty linearly ordered subsets
$\mathcal{D}$	*directed subsets*	subsets D of Q such that each finite subset of D has an upper bound in D (in particular, $D \neq \emptyset$)
$\mathcal{E}$	*singletons*	one-element subsets
$\mathcal{F}$	*finite subsets*	subsets equipotent to a natural number
$\mathcal{M}$	*principal ideals*	subsets of the form $\downarrow y = \downarrow_Q y = \{x \in Q \mid x \leq y\}$
$\mathcal{P}$	*members of the power set*	arbitrary subsets
$\mathcal{W}$	*well-ordered chains*	chains whose nonempty subsets have least elements

By a *system on a set* S we mean a subset $\mathcal{X}$ of the power set $\mathcal{P}S$. If $\mathcal{X}$ is closed under arbitrary intersections (resp. unions) then we speak of a *closure system* (resp. *kernel system*). More generally, a *point closure system* on S is a collection $\mathcal{X}$ of subsets of S

containing each of the *point closures*

$$\downarrow y = \bigcap\{Y \in \mathcal{X} \mid y \in Y\} \qquad (y \in S).$$

For any system $\mathcal{X}$, the *specialization* (*quasiorder*) $\leq_{\mathcal{X}}$ is defined by

$$x \leq_{\mathcal{X}} y \text{ iff for all } Y \in \mathcal{X}, \quad y \in Y \text{ implies } x \in Y.$$

Hence the point closures of $\mathcal{X}$ are nothing but the principal ideals with respect to the specialization quasiorder, justifying the common notation $\downarrow y$ for them.

The *Alexandroff completion* or *additive completion* $\mathcal{A}Q$ is the largest (point) closure system whose specialization is the given quasiorder on Q, while the *minimal extension* $\mathcal{M}Q$ is the smallest point closure system with this property. Moreover, the quasiorder of Q is the specialization of a point closure system $\mathcal{X}$ iff $\mathcal{M}Q \subseteq \mathcal{X} \subseteq \mathcal{A}Q$. In this case we call $\mathcal{X}$ a *standard extension* of Q, and a *standard completion* if, in addition, $\mathcal{X}$ is a closure system (cf. [Ba2], [EW], [Fl], [Sch3]). In the order-theoretical literature, the members of $\mathcal{A}Q$ are also referred to as *lower ends, decreasing sets, down-sets, order ideals,* (*initial*) *segments* etc. (in German: *Abschnitte*). Obviously, $\mathcal{A}Q$ is both a closure and a kernel system, hence an *Alexandroff-discrete topology* (*A-topology* for short), and $\mathcal{M}Q$ is its unique minimal basis (cf [Al]).

Any subset selection $\mathcal{Z}$ has two "companions" $\mathcal{Z}_\perp$ and $\mathcal{Z}_0$, defined by

$$\mathcal{Z}_\perp Q = \mathcal{Z}Q \cup \{\emptyset\},$$

$$\mathcal{Z}_0 Q = \mathcal{Z}Q \setminus \{\emptyset\}.$$

Almost every subset selection $\mathcal{Z}$ occurring in practice agrees either with $\mathcal{Z}_\perp$ or with $\mathcal{Z}_0$. Thus, for example, we have $\mathcal{Z} = \mathcal{Z}_\perp$ for $\mathcal{Z} \in \{\mathcal{A}, \mathcal{F}, \mathcal{P}\}$ and $\mathcal{Z} = \mathcal{Z}_0$ for $\mathcal{Z} \in \{\mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}, \mathcal{W}\}$, while $\mathcal{F}_0 Q$ is the collection of all nonempty finite subsets of Q.

Notational remark. Of course, the choice of the symbols for various fixed subset selections is not stringent and may change a bit from time to time (although it goes without saying that the chosen letters are used consistently within one article). For example, in [E8], we have denoted by $\mathcal{C}Q$ the collection of *all* linearly ordered subsets (including the empty set), while in [E14], $\mathcal{C}Q$ denoted the system of all lower sets generated by (nonempty) chains, and similarly for $\mathcal{W}$.

An ordered set Q is called *$\mathcal{Z}$-join complete* or, for short, *$\mathcal{Z}$-complete*, if each $Z \in \mathcal{Z}Q$ has a join, i.e. a least upper bound in Q, denoted by $\bigvee Z$ (or, if the reference to the underlying ordered set has to be stressed, by $\bigvee_Q Z$). Notice that an ordered set is $\mathcal{Z}_\perp$-complete iff it is $\mathcal{Z}$-complete and has a least element (frequently denoted by the symbol $\perp$).

A $\mathcal{W}$-complete ordered set is also called *up-complete* and, mainly in the more recent literature on ordered sets in computer science, a $\mathcal{W}_\perp$-complete ordered set is referred to as a *complete* (*partially*) *ordered set*, abbreviated *CPO* (cf. [DP]). This definition is in accordance with that of a complete lattice, since in a lattice which is a CPO, *every* subset has a join and a meet (see 1.6). Moreover, we shall see later on that up-complete ordered sets are already $\mathcal{C}$- and $\mathcal{D}$-complete (see 1.12). A $\mathcal{B}$-complete ordered set is a *join-semilattice*, and a $\mathcal{B}_\perp$-complete ordered set is a join-semilattice with least element ($\perp$-$\vee$-*semilattice* for short). Finally, a $\mathcal{P}$- (or $\mathcal{A}$-)complete ordered set is just a *complete lattice.*

Turning to the dual completeness concepts, we call an ordered set Q *$\mathcal{Z}$-meet complete* or *dually $\mathcal{Z}$-complete* if the dually ordered set Q^* (obtained by inverting the order relation) is $\mathcal{Z}$-complete. For subset selections $\mathcal{Z}$ with $\mathcal{Z}Q = \mathcal{Z}Q^*$ like $\mathcal{B}, \mathcal{C}, \mathcal{E}, \mathcal{F}$ or $\mathcal{P}$, dual $\mathcal{Z}$-completeness means that each $Z \in \mathcal{Z}Q$ has a meet (greatest lower bound) in Q, denoted by $\bigwedge Z$ or $\bigwedge_Q Z$. But observe that, for example, an ordered set Q is dually $\mathcal{D}$-complete iff every *down-directed* (*filtered*) subset of Q (that is, every directed subset of Q^*) has a meet in Q. Such ordered sets are called *down-complete.* An up- and down-complete ordered set is called *chain-complete.* By definition, a $\mathcal{B}$-meet-complete ordered set is a *meet-semilattice.* Instead of *$\mathcal{P}_0$-complete* we write *$\bigvee$-complete*, and *$\bigwedge$-complete* has the dual meaning, viz. that each nonempty subset has a meet, or equivalently, each upper bounded subset has a join.

Since every collection $\mathcal{X}$ of sets is ordered by the inclusion order $\subseteq$, it makes sense to form the system $\mathcal{Z}\mathcal{X}$ (more precisely, $\mathcal{Z}(\mathcal{X}, \subseteq)$) and to call $\mathcal{X}$ *$\mathcal{Z}$-union complete* (*$\mathcal{Z}$-$\bigcup$-complete*) if for all $\mathcal{Y} \in \mathcal{Z}\mathcal{X}$, the union $\bigcup \mathcal{Y}$ belongs to $\mathcal{X}$. Dually, we say $\mathcal{X}$ is *$\mathcal{Z}$-intersection complete* (*$\mathcal{Z}$-$\bigcap$-complete*) if $\bigcap \mathcal{Y} \in \mathcal{X}$ for all $\mathcal{Y} \in \mathcal{Z}(\mathcal{X}, \supseteq)$. Observe that $\mathcal{Z}$-$\bigcup$-completeness implies $\mathcal{Z}$-completeness, but not conversely. In particular, every kernel system and every closure system is a complete lattice, but of course, a system which is a complete lattice with respect to the containment order need neither be a kernel system nor a closure system. The $\mathcal{P}_0$-$\bigcap$-complete set systems are the *$\bigcap$-structures*, and the closure systems are the *topped $\bigcap$-structures* in the sense of [DP].

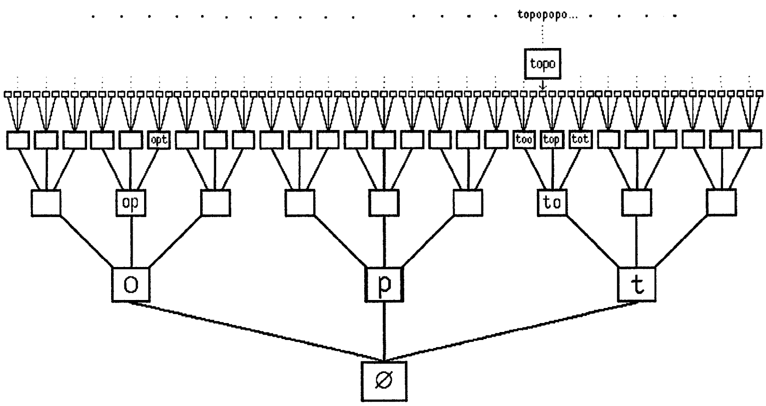
A $\mathcal{W}$-$\bigcup$-complete system is also called *inductive* (again, we shall see soon that such systems are already $\mathcal{C}$- and $\mathcal{D}$-$\bigcup$-complete); however, in other contexts, the word "*inductive*" may have a different meaning (see e.g. Section 7). Notice that every finite ordered set is up-complete and every finite system is inductive. Two tables of completeness properties are to be found at the end of this section.

A further general definition involving subset selections will occur in due course: a map φ between ordered sets P and P' *preserves $\mathcal{Z}$-joins* if for all $Z \in \mathcal{Z}P$, $x = \bigvee Z$ implies $\varphi x = \bigvee \varphi[Z]$. Notice that no completeness assumptions are necessary for this definition, and that the formula "$x = \bigvee Z$" is simply an abbreviation for the statement "*x is the join, i.e. the least upper bound of Z*". Similarly, a map Φ on a system $\mathcal{X}$ *preserves $\mathcal{Z}$-unions* if $\Phi(\bigcup \mathcal{Y}) = \bigcup \Phi[\mathcal{Y}]$ for all $\mathcal{Y} \in \mathcal{Z}\mathcal{X}$ with $\bigcup \mathcal{Y} \in \mathcal{X}$. Passing to the dually ordered sets P^* and P'^*, we say a map φ between ordered sets P and P' *preserves $\mathcal{Z}$-meets* iff φ preserves $\mathcal{Z}$-joins as a map between P^* and P'^*.

A $\bigwedge$-complete CPO is sometimes referred to as a *complete meet-semilattice* (see e.g. [DP]). Thus a complete meet-semilattice is characterized by the existence of joins for all directed and all upper bounded subsets. By a *tree*, we mean a nonempty ordered set in which any two incomparable elements have a lower but no upper bound, and by a *$\mathcal{W}$-tree* an ordered set whose well-ordered subsets are precisely those which have a join. Alternately, a tree may be described as a connected ordered set whose principal ideals are chains, and a

$\mathcal{W}$-tree as a complete meet-semilattice whose principal ideals are well-ordered chains (the proof of these statements is an easy exercise.) A *forest* is a disjoint union of trees, in other words, an ordered set whose principal ideals are chains.

1.1 Example A typical $\mathcal{W}$-tree is formed by the (finite or infinite) words over an alphabet (e.g. $\{o, p, t\}$), ordered by the prefix relation.



Another example of fundamental order- and set-theoretical interest is the $\mathcal{W}$-tree $\mathcal{Y}S$ of all well-orderings R whose carrier

$$X_R = \{x \mid (x, x) \in R\}$$

is contained in a fixed set S, where $\mathcal{Y}S$ is ordered by *propagation*:

$$R \trianglelefteq R' \iff R = R' \cap (X_{R'} \times X_R).$$

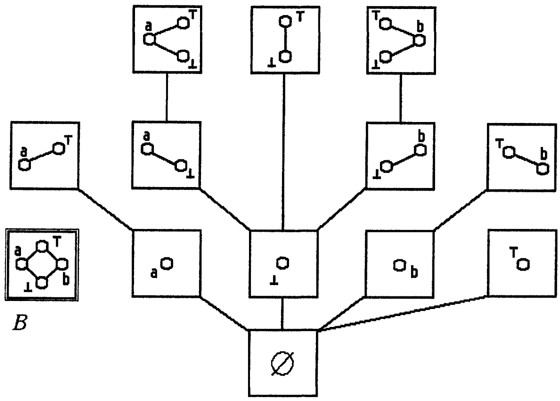
By definition, $R \trianglelefteq R'$ means that X_R is a lower set of the well-ordered set $(X_{R'}, R')$ and the order R is induced by R'. Since the system of all lower sets of a well-ordered set is again well-ordered, we see that each principal ideal of $\mathcal{Y}X$ is well-ordered (by propagation); conversely, if $\mathcal{X} \subseteq \mathcal{Y}S$ is well-ordered by propagation then the union $\bigcup \mathcal{X}$ is easily seen to be an upper bound (in fact, the join) of $\mathcal{X}$, so that $\mathcal{Y}S$ is actually a $\mathcal{W}$-tree.

The previous example may be modified as follows: the set $\mathcal{W}_\perp Q$ of all well-ordered subsets of a given ordered set $Q = (S, \leq)$ is again ordered by propagation:

$$W \leq V \iff \leq|_W \trianglelefteq \leq|_V \iff W \text{ is a lower set of } V \text{ (with respect to } \leq \text{)}.$$

Endowed with this propagation order, $\mathcal{W}_\perp Q$ becomes a $\mathcal{W}$-tree.

1.2 Example The $\mathcal{W}$-tree $\mathcal{W}_\perp B$ of a four-element Boolean lattice B.



In connection with various maximal principles and fixpoint theorems, the following lemma on $\mathcal{W}$-trees is quite helpful ($x \prec y$ means that x is covered by y, i.e. $x < y$ and there is no z with $x < z < y$).

1.3 Lemma *If φ is an arbitrary selfmap of a $\mathcal{W}$-tree T then the set*

$$T_\varphi = \{y \in T \mid x \prec \varphi x \leq y \text{ for all } x < y\}$$

is a principal ideal, hence well-ordered. No y with $y \prec \varphi y$ is the greatest element of T_φ.

Proof First, we observe that T_φ is a well-ordered chain: if for some nonempty $Z \subseteq T_\varphi$, the meet $x = \bigwedge Z$ would not belong to Z then we would have $x \prec \varphi x \leq y$ for all $y \in Z$, which leads to the contradiction $\varphi x \leq x$. Since T is up-complete, T_φ has a join y_φ, and as $\downarrow y_\varphi$ is a chain, $x < y_\varphi$ implies $x < y$ and $x \prec \varphi x \leq y \leq y_\varphi$ for some $y \in T_\varphi$. Thus y_φ is the greatest element of the lower set T_φ. Finally, $y \in T_\varphi$ and $y \prec \varphi y$ imply $\varphi y \in T_\varphi$ since $x < \varphi y$ entails $x \leq y$ and then $x \prec \varphi x \leq \varphi y$. $\square$

The Axiom of Choice (AC) ensures that for any $\mathcal{W}$-tree T there exists a map $\varphi : T \to T$ such that φx covers x for any non-maximal $x \in T$. Hence, by 1.3, AC implies the Maximal Principle for Trees:

(MT) *Every $\mathcal{W}$-tree has a maximal element.*

On the other hand, applying MT to $\mathcal{W}$-trees of the form $\mathcal{W}_{\perp}Q$, one obtains the Maximal Principle (alias Zorn's Lemma) in its full generality:

(MP) *Every ordered set with upper bounded well-ordered subsets has a maximal element.*

Indeed, if Y is a maximal member of $\mathcal{W}_{\perp}Q$ (with respect to propagation) then any upper bound of Y must be maximal in Q.

Similarly, an application of MT to the $\mathcal{W}$-tree $\mathcal{Y}S$ of all well-orderings on subsets of a fixed set S yields a maximal member of $\mathcal{Y}S$, and this must be a well-ordering on the whole set S. In all, this establishes, in a short and elementary way, the equivalence of the Axiom of Choice with several maximal principles of progressive strength and with the Well-Ordering Principle (WP) (cf. [Ba1]).

Moreover, MT can be used to prove an interesting strengthening of Bourbaki's Fixpoint Lemma [Bo], stating that *every extensive selfmap of a nonempty up-complete ordered set Q has a fixpoint* (where $\varphi : Q \to Q$ is *extensive* if $x \leq \varphi x$ for all $x \in Q$). To this aim, consider a selfmap φ of an arbitrary ordered set Q and call a subset X of Q ***φ-stable*** if it is ***φ-invariant*** (i.e. $\varphi[X] \subseteq X$) and *$\bigvee$-closed* (i.e. whenever a subset of X has a join x in Q then $x \in X$). Since the φ-stable subsets form a closure system, there exists a smallest φ-stable set S_φ. The crucial step in proving the Fixpoint Lemma is to show that in case of an *extensive* selfmap φ of an up-complete ordered set, the least φ-stable subset S_φ is *well-ordered* and, consequently, has a greatest element which must be a fixpoint of φ. While Bourbaki's method [Bo] may be regarded as a refinement of Zermelo's ingenious *second* proof of the Well-Ordering Principle [Ze2], the subsequent arguments are strongly influenced by Zermelo's *first* proof [Ze1] (which is not less ingenious) and its refinements due to Witt [Wi] and Banaschewski [Ba1].

By a ***φ-chain***, we mean a subset W of Q with the following three properties:

(W1) W is well-ordered (by the order induced from Q).
(W2) If a subset of W has a join in W, then this is also the join in Q.
(W3) If x is a non-maximal member of W then φx covers x in W.

1.4 Theorem *Let φ be a selfmap of an arbitrary ordered set Q. Then:*
(1) *Every φ-chain is contained in every φ-stable set.*
(2) *If φ is extensive then the smallest φ-stable set S_φ is the greatest φ-chain.*
(3) *If, in addition, Q is up-complete then S_φ has a greatest element, and this is a fixpoint of φ.*

Proof (1) Assume X is a φ-stable set and W is a φ-chain not contained in X. Then, by (W1), $W \setminus X$ has a least element y, and $V = \{w \in W \mid w < y\}$ is a subset of X. Being upper bounded by y, the subset V has a join x in W; by (W2), x is also the join of V in Q and, therefore, in the $\bigvee$-closed set X. But then x would be covered by y in W, and (W3) would imply $y = \varphi x \in X$, a contradiction.

(2) Again, we consider the $\mathcal{W}$-tree $T = \mathcal{W}_{\perp}Q$ of all well-ordered subsets of Q, ordered by $W \leq V$ iff W is a lower set ("initial segment".) of V. For $W \in T$, define

$$\Phi W = \begin{cases} W \cup \{x\} & \text{if } x = \bigvee W \text{ but } x \notin W \\ W \cup \{\varphi x\} & \text{if } x \text{ is the greatest element of W} \\ W & \text{otherwise} \end{cases}$$

With respect to the propagation order, ΦW covers or equals W. In particular, we have an extensive map $\Phi : T \to T$, and Lemma 1.3 ensures the existence of a greatest element V of T such that $W < V$ implies $W \prec \Phi W \leq V$, and this greatest element V must be a fixpoint of Φ (otherwise $V \prec \Phi V$). Hence

$$W \leq V \text{ implies } \Phi W \leq V, \text{ in particular, } \Phi W \subseteq V.$$

Using these implications, we show that V is φ-stable. For $y \in V$, we have $W = V \cap \downarrow y \leq V$ and therefore $\varphi y \in \Phi W \subseteq V$. If y is not maximal in V then $W < \Phi W = W \cup \{\varphi y\}$, and so φy covers y in V. If a subset Z of V has a join x in Q with $x \notin Z$ then $W = V \cap \downarrow Z \leq V$ and $x = \bigvee W \in \Phi W \subseteq V$.

By (1), it remains to show that whenever x is the join of a set Z in V then also in Q: otherwise, $W = V \cap \downarrow Z$ were an initial segment of V without join in Q, whence $W = \Phi W$, while $x = \bigvee W \notin W$ implies $W < V$ and therefore $W \prec \Phi W$. Hence V is a φ-stable φ-chain and coincides, by (1), with S_{φ}.

(3) Since S_{φ} is well-ordered, it has a join x. By φ-stability of S_{φ}, x and φx must be members of S_{φ}, whence $x = \varphi x$ is the greatest element of S_{φ}. □

It is a frequent experience of the mathematician working with ordered sets that certain statements or properties remain unchanged if directed sets are replaced with (well-ordered) chains. Assuming AC, the passage from chains to well-ordered chains is easy, on account of the following well-known Cofinality Principle:

> (CP) *Every chain C contains a cofinal well-ordered subchain, that is, a well-ordered subchain W that generates the same lower set as C. In particular, C and W have the same upper bounds.*

Indeed, any well-ordered subchain of C maximal with respect to propagation must be cofinal with C. By CP, it is clear that $\mathcal{W}$-(union) completeness and $\mathcal{C}$-(union) completeness are equivalent properties. Much harder to prove is the equivalence of $\mathcal{C}$- and $\mathcal{D}$-(union) completeness. Since the usual proofs of these basic but non-trivial equivalences involve rather complicated transfinite tools, we feel the need to give here an elementary proof, borrowed from [E11] and avoiding completely the machinery of ordinal numbers. However, we shall see that a weak version of the Axiom of Choice is indispensable. Besides that, we shall make use of Bourbaki's Fixpoint Lemma (which is trivial in the light of AC; but the point is that it has been proved in Zermelo-Fraenkel set theory without AC).

Our arguments will rely on a certain set-theoretical induction principle which, despite of its manifold applications, occurs only sporadically in the literature and is usually shown by transfinite induction via ordinal numbers (see e.g. [Sch1]). Instead, we shall use the Fixpoint Lemma in combination with the so-called Axiom of Multiple Choice which is effectively weaker than the full AC, as was shown by A. Levy (see [Le]):

(AMC) *For any family $(X_i : i \in I)$ of nonempty sets, there exists a function assigning to each $i \in I$ a nonempty finite subset of X_i.*

1.5 Set Induction Principle *If an inductive set system $\mathcal{X}$ contains all finite subsets of a certain set S, then S must also be a member of $\mathcal{X}$.*

Proof Forming the intersection of all inductive set systems containing all finite subsets of S, we may assume without loss of generality that $\mathcal{X}$ is the *smallest* inductive system with this property, whence $\mathcal{X} \subseteq \mathcal{P}S$. But the system

$$\mathcal{X}' = \{X \mid X \cup F \in \mathcal{X} \text{ for all } F \in \mathcal{F}S\}$$

is still inductive and contains all finite subsets of S, whence $\mathcal{X}' = \mathcal{X}$. Any maximal principle equivalent to AC immediately gives a maximal member of $\mathcal{X}$, and this must be the set S. Alternatively, under the assumption $S \notin \mathcal{X}$, AMC provides a choice function Φ assigning to each $X \in \mathcal{X}$ a nonempty finite subset of $S \setminus X$. But then the map $X \mapsto X \cup \Phi(X)$ would be a fixpoint-free extensive selfmap of the inductive (hence up-complete) system $\mathcal{X}$, contradicting the Fixpoint Lemma. □

Concerning the position of the Set Induction Principle (SIP) within the framework of set theory, it is of interest that SIP together with the Axiom of Choice for Families of Finite Sets (ACF) implies the following axiom of set theory (cf. [Mo]):

(CS) *Every infinite set contains a countable infinite subset.*

To see this, consider the collection $\mathcal{F}S$ of all finite subsets of an infinite set S. By SIP, $\mathcal{F}S$ cannot be inductive, so there exists a well-ordered subset $\mathcal{Y}$ of $\mathcal{F}S$ such that $\bigcup \mathcal{Y}$ is infinite. Hence there is a strictly increasing sequence $(F_n : n \in \omega)$ in $\mathcal{Y}$, and by ACF we obtain a function selecting one element from each of the difference sets $F_{n+1} \setminus F_n$. The image of this function is then a countable subset of S.

While the Well-Ordering Principle is equivalent to AC, the Ordering Principle (OP), postulating the existence of a linear ordering on each set, is known to be strictly weaker than AC in ZF set theory. In fact, in the classical Mostowski model [Ms], OP holds but CS fails.

On the other hand, OP clearly implies ACF (but not conversely; cf. [Mo]). Summarizing these observations, we see that SIP is neither a consequence of OP (otherwise, we would have the implications OP $\Longrightarrow$ ACF + SIP $\Longrightarrow$ CS), nor does SIP imply ACF, because AMC + ACF is obviously equivalent to AC. Hence we have the following diagram of implications:

$$\begin{tikzcd}
& \mathrm{WP} \approx \mathrm{AC} \arrow[dl] \arrow[d] \arrow[dr] & \\
\mathrm{OP} \arrow[d] & \mathrm{ACF} + \mathrm{SIP} \arrow[dl] \arrow[d] \arrow[dr] & \mathrm{AMC} \arrow[d] \\
\mathrm{ACF} & \mathrm{CS} & \mathrm{SIP}
\end{tikzcd}$$

The basic ideas for these considerations on the strength of SIP are due to N. Weaver, whose cooperation on this subject is gratefully acknowledged.

Now let us turn to a few straightforward applications of 1.5. Applying SIP to the system of all subsets of an ordered set which have a join, we obtain immediately

1.6 Corollary *A join-semilattice in which every well-ordered subset has a join is already a complete lattice.*

Another immediate consequence of SIP is the following:

1.7 Corollary *A set system $\mathcal{X}$ is inductive and decreasing (i.e. $X \subseteq Y \in \mathcal{X}$ implies $X \in \mathcal{X}$) iff it is of finite character, i.e.*

$$Y \in \mathcal{X} \Longleftrightarrow \mathcal{F}Y \subseteq \mathcal{X}.$$

Now it is not difficult to derive J. Schmidt's "Hauptsatz über induktive Hüllensysteme" [Sch2] which plays a fundamental role in universal algebra (cf. [MMT]):

1.8 Theorem *The following statements on a map $\Gamma : \mathcal{P}S \to \mathcal{P}S$ are equivalent:*
(a) Γ *is finitary, i.e.* $\Gamma Y = \bigcup\{\Gamma F \mid F \in \mathcal{F}Y\}$ *for all* $Y \subseteq S$.
(b), (c), (d) Γ *preserves* $\mathcal{W}$-($\mathcal{C}$-,$\mathcal{D}$-)*unions.*
Each of these conditions implies that the closure system $\{Y \subseteq S \mid \Gamma Y \subseteq Y\}$ *is inductive.*
If Γ *is a closure operator then the above statements are also equivalent to*
(b′), (c′), (d′) *The closure system* $\Gamma[\mathcal{P}S]$ *is* $\mathcal{W}$-($\mathcal{C}$-,$\mathcal{D}$-)$\bigcup$*-complete.*

Proof The implications (a) $\Longrightarrow$ (d) $\Longrightarrow$ (c) $\Longrightarrow$ (b) are obvious, and the equivalence of (a) and (b) follows from 1.7 since Γ is finitary iff for each $x \in S$, the system $\{Y \subseteq S \mid x \notin \Gamma Y\}$ is of finite character, i.e. inductive and decreasing. Thus, for well-ordered $\mathcal{Y} \subseteq \mathcal{P}S$, $x \in \Gamma(\bigcup \mathcal{Y})$ is equivalent to $x \in \Gamma Y$ for some $Y \in \mathcal{Y}$.

The last two statements in 1.8 follow from a more general result on so-called **O**-*invariant* subset selections $\mathcal{Z}$ like $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}, \mathcal{F}, \mathcal{P}$ or $\mathcal{W}$ (see Section 6): *If a map $\Gamma : \mathcal{P}S \to \mathcal{P}S$ preserves $\mathcal{Z}$-unions then the corresponding closure system $\{Y \subseteq S \mid \Gamma Y \subseteq Y\}$ is $\mathcal{Z}$-$\bigcup$-complete; the converse holds whenever Γ is a closure operator.*

The proof of this claim is an easy exercise.

The previous considerations already establish the equivalence of $\mathcal{W}$-, $\mathcal{C}$- and $\mathcal{D}$-$\bigcup$-completeness in case of closure systems. However, the general case of arbitrary systems requires two auxiliary lemmas. The most general situation needed in mathematical practice is covered by the following definition: given an ordered set Q and a subset X of Q, denote by $\mathcal{J}_X Q$ the system of all subsets Y of X such that the join $\bigvee_Q Y$ exists in Q and belongs to X (whence $\bigvee_Q Y$ is also the join of Y in X).

1.9 Lemma *Suppose $D \subseteq X \subseteq Q$, and let be given a finitary operator $\Gamma : \mathcal{P}D \to \mathcal{P}D$. Then*

$$\mathcal{W}X \cup \Gamma[\mathcal{F}_0 D] \subseteq \mathcal{J}_X Q \text{ implies } \Gamma[\mathcal{P}_0 D] \subseteq \mathcal{J}_X Q.$$

Proof By hypothesis, the system

$$\mathcal{X} = \{Y \subseteq D \mid \Gamma Y \in \mathcal{J}_X Q \text{ or } Y = \emptyset\}$$

contains $\mathcal{F}D$. If a subsystem $\mathcal{Y}$ of $\mathcal{X}$ is well-ordered by inclusion and $\mathcal{Y} \nsubseteq \{\emptyset\}$ then $W = \{\bigvee \Gamma Y \mid Y \in \mathcal{Y} \setminus \{\emptyset\}\}$ is a well-ordered chain in X, whence $\bigvee W$ exists and belongs to X. Thus $\bigvee \Gamma(\bigcup \mathcal{Y}) = \bigvee \bigcup \Gamma[\mathcal{Y}] = \bigvee W \in X$, i.e. $\bigcup \mathcal{Y} \in \mathcal{X}$; hence, by 1.5, $\mathcal{P}D \subseteq \mathcal{X}$. □

The next lemma enables us to "blow up" the subsets of a directed set to directed subsets in such a manner that containment is preserved and finite sets are mapped to finite sets with greatest elements (cf. [Kr]).

1.10 Lemma *For any directed set D, there is an order-preserving map $\psi : \mathcal{F}_0 D \to D$ such that $\psi\{x\} = x$ for $x \in D$ and ψF is an upper bound of F for each $F \in \mathcal{F}_0 D$. Hence the operator*

$$\Gamma : \mathcal{P}D \longrightarrow \mathcal{P}D, \quad Y \longmapsto \psi[\mathcal{F}_0 Y]$$

is extensive, finitary, and sends finite nonempty sets to finite sets with greatest elements. Moreover, the range of Γ consists of directed subsets of D.

Proof Using AC, we find a map φ assigning to each nonempty finite subset of D an upper bound in D. Then we may define $\psi : \mathcal{F}_0 D \to D$ recursively by

$$\psi\{x\} = x \quad \text{for} \quad x \in D,$$

$$\psi F = \varphi\{\psi(F \setminus \{x\}) \mid x \in F\} \quad \text{for} \quad F \in \mathcal{F}_0 D \setminus \mathcal{E}D.$$

It is then straightforward to verify the claimed properties for ψ and Γ. □

Now we are ready for the main result of this introductory section.

1.11 Theorem *Let X be a subset of an ordered set Q and $\mathcal{J} = \mathcal{J}_X Q$. Then the following statements are equivalent:*

(a) *For each nonempty subset Y of X, $\mathcal{B}Y \subseteq \mathcal{J}$ implies $Y \in \mathcal{J}$.*
(b) $\mathcal{W}X \subseteq \mathcal{J}$. (c) $\mathcal{C}X \subseteq \mathcal{J}$. (d) $\mathcal{D}X \subseteq \mathcal{J}$.
(b′) *$\mathcal{J}$ is $\mathcal{W}$-$\bigcup$-complete.* (c′) *$\mathcal{J}$ is $\mathcal{C}$-$\bigcup$-complete.* (d′) *$\mathcal{J}$ is $\mathcal{D}$-$\bigcup$-complete.*

Proof The implications (a) $\Longrightarrow$ (c) $\Longrightarrow$ (b) and (d′) $\Longrightarrow$ (c′) $\Longrightarrow$ (b′) are clear.
(b) $\Longrightarrow$ (d): Let $D \in \mathcal{D}X$ and choose Γ as in 1.10. Then $\mathcal{W}X \subseteq \mathcal{J}$ and $\Gamma[\mathcal{F}_0 D] \subseteq \mathcal{J}$, so that by 1.9, we may conclude $\Gamma[\mathcal{P}_0 D] \subseteq \mathcal{J}$, and in particular, $D = \Gamma D \in \mathcal{J}$.
(d) $\Longrightarrow$ (d′): If $\mathcal{Y} \in \mathcal{D}\mathcal{J}$ then $D = \{\bigvee Y \mid Y \in \mathcal{Y}\} \in \mathcal{D}X$ and therefore $\bigvee \bigcup \mathcal{Y} = \bigvee D \in X$.
(b′) $\Longrightarrow$ (a): By induction, $\mathcal{B}Y \subseteq \mathcal{J}$ implies $\mathcal{F}_0 Y \subseteq \mathcal{J}$, and then 1.5 yields $Y \in \mathcal{J}$. □

The case $X = P$ amounts to various characterizations of up-completeness (cf. Cohn [Co] and Isbell [Is1]):

1.12 Corollary *For an ordered set Q, the following conditions are equivalent:*
(a) *If every binary subset of a nonempty subset Y of Q has a join then Y has a join, too.*
(b) *Every well-ordered chain of Q has a join, i.e. Q is up-complete.*
(c) *Every chain of Q has a join.*
(d) *Every directed subset of Q has a join.*
(e) *The system of all subsets of Q possessing a join is inductive.*

As another consequence of 1.11, we obtain the aforementioned equivalence of various types of union-completeness, by taking for Q power set lattices (cf. [MS]).

1.13 Corollary *$\mathcal{W}$-, $\mathcal{C}$- and $\mathcal{D}$-$\bigcup$-completeness are equivalent properties.*

Notice that SIP has been used for the proof of this equivalence but is in turn a trivial consequence of it, because the collection of all finite subsets of a set is directed, so that the only $\mathcal{D}$-$\bigcup$-complete subset of $\mathcal{P}S$ containing $\mathcal{F}S$ is $\mathcal{P}S$ itself.

A certain intrinsic topology for ordered sets will play a central role in our later considerations, namely the so-called *Scott topology* (cf. [Com])

$$\sigma Q = \{U \subseteq Q \mid \text{For all } D \in \mathcal{D}Q \text{ possessing a join, } \bigvee D \in U \text{ iff } D \text{ intersects } U\}.$$

By definition, a subset X of Q is *Scott closed* iff if is a lower set such that for each directed subset D of X possessing a join in Q, this join belongs to X. The corresponding closure system is denoted by $\mathcal{D}^{\vee}Q$ (for a more general construction, see Section 6). By Theorem 1.11, directed sets may be replaced with (well-ordered) chains in the definition of Scott-open or Scott-closed subsets of up-complete ordered sets, respectively. Moreover, in a complete lattice L, a subset X is Scott closed iff for all $Y \subseteq L$,

$$\bigvee Y \in X \Longleftrightarrow \bigvee F \in X \text{ for all } F \in \mathcal{F}Y,$$

in other words, iff the system $\mathcal{J}_X L$ is of finite character. Conversely, a system is of finite character iff it is a Scott-closed subset of some power set lattice.

A map between ordered sets is known to be Scott continuous, i.e. continuous with respect to the Scott topologies, iff it preserves $\mathcal{D}$-joins (see e.g. [Com]), and by the previous remarks, it suffices to postulate that it preserves $\mathcal{C}$- or at least $\mathcal{W}$-joins. Applying this remark to the unary meet-operations

$$\wedge_x : S \longrightarrow S, \quad y \longmapsto x \wedge y$$

of an up-complete meet-semilattice S, we see that the distributive law

$$x \wedge \bigvee Y = \bigvee \{x \wedge y \mid y \in Y\}$$

holds for all directed subsets Y iff this is true at least for all well-ordered chains Y iff the maps $\wedge_x$ are continuous with respect to the Scott topologies, justifying the name *meet-continuous* (cf.[Bi]) or *upper continuous* (cf. [CD]) for this property of semilattices.

Since the main subject of this volume are *ordered sets and algebras*, we would like to include in this introductory section an application of the Set Induction Principle to universal algebra. First, let us mention the following purely set-theoretical consequence of SIP:

1.14 Corollary *Every infinite set Y is the union of a well-ordered system of subsets having a smaller cardinality than Y.*

In this context, "well-ordered" refers to set inclusion, and a subset "has a smaller cardinality than Y" iff it is not equipotent to Y, i.e., there is no bijection between Y and this subset. Hence no ordinal or cardinal numbers are needed, neither for the formulation of 1.14 nor for its proof: consider the system $\mathcal{X}$ of all subsets X of Y which are either finite or may be represented as a well-ordered union of subsets with smaller cardinality. If a subsystem $\mathcal{Y}$ of $\mathcal{X}$ is well-ordered by inclusion then its union is either equipotent to some member of $\mathcal{Y}$, or all members of $\mathcal{Y}$ have a smaller cardinality than $\bigcup \mathcal{Y}$; in any case, $\bigcup \mathcal{Y} \in \mathcal{X}$. Hence, by SIP, we obtain $Y \in \mathcal{X}$, as desired. Now we can prove:

1.15 Corollary *Every uncountable algebra A with at most countably many (finitary) operations may be represented as the union of a well-ordered system of subalgebras having a smaller cardinality than A.*

For the proof, we need only one basic fact from cardinal number theory, namely, that for any infinite set Y, the set $\mathcal{F}Y$ of all finite subsets is equipotent to Y. From this it is clear that for any subset Y of an algebra A with countably many operations, the subalgebra ΓY generated by Y is either countable or equipotent to Y. By 1.14, A is the union of a well-ordered collection $\mathcal{Y}$ of subsets of smaller cardinality, and if A is uncountable, then the set $\{\Gamma Y \mid Y \in \mathcal{Y}\}$ is a well-ordered system of subalgebras each of which has a smaller cardinality than A.

Two concluding remarks are in order. First, we observe that 1.15 extends neither to countable algebras (counterexample: the additive group of integers) nor to algebras with uncountably many operations (counterexamples: real vector spaces of finite dimension). Second, from 1.15, we may derive a famous lemma on directed sets due to Iwamura (cf. [Iw], [Ma]):

1.16 Corollary *Every infinite directed set is the union of a well-ordered system of directed subsets of smaller cardinality.*

For countable directed sets, this is verified easily by induction, and for uncountable directed sets D, consider the operations $\psi_n(x_1,\ldots,x_n) = \psi\{x_1,\ldots,x_n\}$ where ψ assigns to every finite subset of D an upper bound in D. Every subalgebra with respect to these countably many operations is directed, so 1.15 applies.

Originally Iwamura's Lemma was proved by tools of ordinal number theory, and it was used for the proof of various results concerning the exchange of directed sets with chains (see, e.g., [Ma], [MS]). An easy proof of Markowsky's strenghtened version of this lemma [My] is obtained with the help of 1.10: given a directed set D of cardinality γ, choose a bijection $\delta : \gamma \to D$ and an operator $\Gamma : \mathcal{P}D \to \mathcal{P}D$ according to 1.10. Then the sets

$$D_\alpha = \Gamma\delta[\alpha] \qquad (\alpha < \gamma)$$

are directed and have the following properties:

(1) If α is finite then so is D_α, while if α is infinite then $|D_\alpha| = \alpha < \gamma$.

(2) If $\alpha < \beta < \gamma$ then $D_\alpha \subset D_\beta$.

(3) $D = \bigcup\{D_\alpha \mid \alpha < \gamma\}$.

Finally, we would like to point out that, in spite of many positive results, the exchange between chains and directed sets does not work throughout. For example, 1.12 becomes wrong if "join" is replaced by "upper bound":

1.17 Example In the set $\mathcal{X}$ of all countable subsets of the reals, ordered by

$$X \leq Y \text{ iff } X = Y \text{ or } X \text{ is a finite subset of } Y,$$

every chain must be countable and has therefore an upper bound (its union; notice that this need not be the join of the chain!). However, the system of all finite sets of real numbers is directed but has no upper bounds in $\mathcal{X}$.

1.18 Table *$\mathcal{Z}$-complete and dually $\mathcal{Z}'$-complete ordered sets*

$\mathcal{Z}$ \ $\mathcal{Z}'$	$\mathcal{A}, \mathcal{P}$	$\mathcal{B}, \mathcal{F}_0$	$\mathcal{C}, \mathcal{D}, \mathcal{W}$	$\mathcal{E}$
$\mathcal{A}, \mathcal{P}$	complete lattice	complete lattice	complete lattice	complete lattice
$\mathcal{B}, \mathcal{F}_0$	complete lattice	lattice	down-complete join-semilattice	join-semilattice
$\mathcal{B}_\perp, \mathcal{F}$	complete lattice	lattice with $\perp$	down-complete $\perp$-$\vee$-semilattice	$\perp$-$\vee$-semilattice
$\mathcal{C}, \mathcal{D}, \mathcal{W}$	complete lattice	up-complete meet-semilattice	chain-complete ordered set	up-complete ordered set
$\mathcal{E}$	complete lattice	meet-semilattice	down-complete ordered set	ordered set

1.19 Table *$\mathcal{Z}$-$\bigcup$-complete and $\mathcal{Z}'$-$\bigcap$-complete systems*

$\mathcal{Z}$ \ $\mathcal{Z}'$	$\mathcal{A}, \mathcal{P}$	$\mathcal{B}, \mathcal{F}_0$	$\mathcal{C}, \mathcal{D}, \mathcal{W}$	$\mathcal{E}$
$\mathcal{A}, \mathcal{P}$	A-topology	topology	$\mathcal{C}$-$\bigcap$-complete kernel system	kernel system
$\mathcal{B}, \mathcal{F}_0$	$\mathcal{B}$-$\bigcup$-complete closure system	set lattice	$\mathcal{B}$-$\bigcup$-complete, $\mathcal{C}$-$\bigcap$-complete	$\mathcal{B}$-$\bigcup$-complete
$\mathcal{B}_\perp, \mathcal{F}$	topological closure system	set lattice with $\perp$	$\mathcal{F}$-$\bigcup$-complete, $\mathcal{C}$-$\bigcap$-complete	$\mathcal{F}$-$\bigcup$-complete
$\mathcal{C}, \mathcal{D}, \mathcal{W}$	algebraic closure system	inductive and $\mathcal{B}$-$\bigcap$-complete	inductive and $\mathcal{C}$-$\bigcap$-complete	inductive
$\mathcal{E}$	closure system	$\mathcal{B}$-$\bigcap$-complete	$\mathcal{C}$-$\bigcap$-complete	set system

2 Compact generation in ordered sets

Algebraic or *compactly generated* complete lattices are a familiar tool of universal algebra whose importance and applicability to various fields of classical and modern mathematics need no particular emphasis. For a comprehensive treatment of the underlying theory, the reader may consult, among other sources like G. Birkhoff's classical monograph *Lattice Theory* [Bi], two excellent books on this topic: *Algebraic Theory of Lattices* by P. Crawley and R.P. Dilworth [CD], and *Algebras, Lattices and Varieties* by R. N. McKenzie, G. F. McNulty and W. F. Taylor [MMT].

Recent trends and developments in this area are characterized by a more and more increasing interest in non-complete versions of the concept of compact generation; new impulses arose from the viewpoint of logic, computer science and the theory of computation (see, for example, the well-written *Introduction to Order and Lattices* by B. Davey and H. Priestley [DP]. For more specific topics, see e.g. [Gu1-2], [Pl], [Sc1-2], [Sm], [SP], [Sty], [We]).

A frequently encountered situation in the theory of computation is that at least suprema of certain chains exist (or are adjoined as new "ideal objects"), while binary joins occur rather rarely. Indeed, it is not clear *a priori* what kind of suprema are most adequate for the intended theory, but the consideration of directed suprema turned out to be one of the most reasonable and convenient settings; fortunately, we know from the first section that in most cases, it suffices to work with well-ordered chains instead of directed sets, while the restriction to sequences or to countable directed sets sometimes requires a slightly modified approach. In the present section, we shall discuss the basic facts for the case of directed sets and chains, while generalizations to other types of suprema via a "uniform approach" (as suggested by Wright, Wagner and Thatcher [WWT]) are reserved to the second part.

Before we are dealing with the main facts about compactly generated ordered sets, let us supply the basic definitions and some instructive examples.

Given a subset Y of a quasiordered set Q, we denote by $Y^{\uparrow}$ and $Y^{\downarrow}$ the set of all upper and lower bounds of Y, respectively. These sets should not be confused with the *upper set generated by* Y,

$$\uparrow Y = \uparrow_Q Y = \{x \in Q \mid y \leq x \text{ for some } y \in Y\},$$

and the *lower set generated by* Y,

$$\downarrow Y = \downarrow_Q Y = \{x \in Q \mid x \leq y \text{ for some } y \in Y\}.$$

The set

$$\Delta Y = Y^{\uparrow\downarrow}$$

is usually referred to as the (*lower*) *cut generated by* Y; the map

$$\Delta : \mathcal{P}Q \to \mathcal{P}Q, \quad Y \mapsto Y^{\uparrow\downarrow}$$

is a closure operator, and the corresponding closure system

$$\mathcal{N}Q = \{Y \subseteq Q \mid Y = \Delta Y\}$$

is known as the *Dedekind-MacNeille completion* or *completion by cuts* of Q. Every cut is a lower set, but not conversely. Important for us is the fact that whenever Y has a join (= least upper bound) $\bigvee Y$ then

$$x \in \Delta Y \iff Y^{\uparrow} \subseteq \uparrow x \iff x \leq \bigvee Y.$$

Any isomorphic copy of $\mathcal{N}Q$ is called a (or "the") *normal completion* of Q. Normal completions are characterized by the property of being complete lattices in which the original ordered set is $\bigvee$- and $\bigwedge$-densely embedded (see, for example, [Ba2]). In fact, the principal ideal map

$$\eta_Q^{\mathcal{N}} : Q \longrightarrow \mathcal{N}Q, \quad x \longmapsto \downarrow x$$

is such a $\bigvee$- and $\bigwedge$-dense embedding. Thus, if convenient, an ordered set Q may always be regarded as a subset of its normal completion.

Now, generalizing the classical definition of compactness from topology and lattice theory, we consider an arbitrary subset selection $\mathcal{Z}$ and call an element x of a (quasi-)ordered set Q *$\mathcal{Z}$-compact* or *$\mathcal{Z}$-prime* (cf. [WWT], [E15-19]) if for all $Z \in \mathcal{Z}Q$, $x \in \Delta Z$ implies $x \in \downarrow Z$. Later on, we shall study this general concept in greater detail, but for the moment we are only interested in the special cases $\mathcal{Z} = \mathcal{W}, \mathcal{C}, \mathcal{D}$, and $\mathcal{I}$, where

$$\mathcal{I}Q = \{Y \subseteq Q \mid \Delta F \subseteq Y \text{ for all } F \in \mathcal{F}Y\}$$

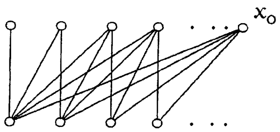
is the closure system of all *ideals* in the sense of Frink [Fr]. Since for arbitrary $Y \subseteq Q$, $\mathcal{I}Y = \bigcup\{\Delta F \mid F \in \mathcal{F}Y\}$ is the ideal generated by Y, we see that an element x is $\mathcal{I}$-compact iff $x \in \Delta Y$ implies $x \in \Delta F$ for some finite $F \subseteq Y$. In particular, for elements of *complete lattices*, the usual notion of compactness ($x \leq \bigvee Y$ implies $x \leq \bigvee F$ for some $F \in \mathcal{F}Y$) is equivalent to $\mathcal{I}$-compactness, but also, as we shall see in 2.2, to $\mathcal{W}$-, $\mathcal{C}$- and $\mathcal{D}$-compactness. However, in the absence of certain joins, the situation is more complicated. While the implications

$$\mathcal{I}\textit{-compact} \implies \mathcal{D}\textit{-compact} \implies \mathcal{C}\textit{-compact} \iff \mathcal{W}\textit{-compact}$$

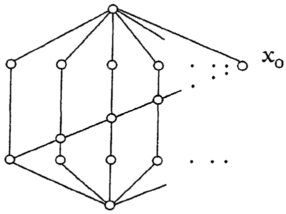
are rather evident (for the last equivalence, use the Cofinality Principle), neither of the first two arrows can be inverted in general, as the following examples show (cf. [E2], [E5], [E6]):

2.1 Example Order the integers by

$$x \sqsubseteq y \iff x = y \text{ or } x < 0 = y \text{ or } -y \leq x < 0.$$



P



$\mathcal{N}P$

In this ordered set P of height 2, each element is $\mathcal{D}$-compact, but the element $x_0 = 0$ is not $\mathcal{I}$-compact, and it is not compact in the normal completion $\mathcal{N}P$ (but see 2.7!)

An example where not all $\mathcal{C}$-compact elements are $\mathcal{D}$-compact has been given in 1.17: the system $\mathcal{X}$ of all countable subsets of the reals, ordered by $X \leq Y$ iff X equals Y or is a finite subset of Y, consists of $\mathcal{C}$-compact members only, since any chain $\mathcal{Y}$ in $\mathcal{X}$ must be countable, and consequently, $\Delta\mathcal{Y}$ is the lower set generated by $\mathcal{Y}$ (notice that any two infinite members of $\mathcal{X}$ are incomparable!) But only the finite members of $\mathcal{X}$ are $\mathcal{D}$-compact.

Nevertheless, we have:

2.2 Lemma (1) *For elements of an up-complete ordered set, $\mathcal{W}$-, $\mathcal{C}$- and $\mathcal{D}$-compactness are equivalent.*
(2) *An element of a join-semilattice is $\mathcal{I}$-compact iff it is $\mathcal{D}$-compact. Hence in complete lattices, $\mathcal{W}$-, $\mathcal{C}$-, $\mathcal{D}$- and $\mathcal{I}$-compactness are equivalent properties.*

Proof (1) x is $\mathcal{W}$-($\mathcal{C}$-,$\mathcal{D}$-)compact iff the set X of all elements y with $x \not\leq y$ satisfies the equivalent conditions (b), (c), (d) in Theorem 1.11.

(2) In a join-semilattice, the nonempty ideals are precisely the directed lower sets, while the empty set is an ideal iff the join-semilattice has no least element. □

For convenience, $\mathcal{D}$-compact elements will simply be called *compact.* Thus an element x of an up-complete ordered set is compact iff for all directed subsets (or at least all well-ordered chains) Z, $x \leq \bigvee Z$ implies $x \leq y$ for some $y \in Z$. The set of all compact elements of an ordered set Q is denoted by $\mathcal{K}Q$, and Q is called *compactly generated* if Q is up-complete and each element of Q is a join of compact elements. More important for certain theories and applications is the following strengthening: An up-complete ordered set Q is called *algebraic* if for each $y \in Q$, there is a *directed* set D of compact elements with join y. In this case, it is easy to see that the set

$$\kappa y = \mathcal{K}Q \cap \downarrow y$$

is directed with join y. For complete lattices, "*compactly generated*" is tantamount to "*alge-*

braic", because the sets κy are automatically directed, being join-subsemilattices. Similarly, a compactly generated complete meet-semilattice is already algebraic, because each of its principal ideals $\downarrow x$ is a complete lattice and, consequently, the sets κx are directed. However, as we shall see in 2.3, a compactly generated CPO need not be algebraic, even if it is a meet-semilattice.

Adjoining a top element to a complete meet-semilattice yields a complete lattice, and conversely, deleting a compact top element of a complete lattice leaves us with a complete meet-semilattice. Under these mutually inverse processes, algebraic complete meet-semilattices correspond to algebraic complete lattices with compact top elements. Therefore, from the purely algebraic point of view, we may restrict our attention to complete lattices rather than complete meet-semilattices, although the "non-topped" versions are of major significance in computer science (cf. [DP]).

Clearly every ordered set satisfying the Ascending Chain Condition (ACC) is an algebraic poset in which *every* element is compact, and conversely, any up-complete ordered set consisting of compact elements only must satisfy the ACC. In particular, all finite ordered sets are algebraic. But notice that the chain ω of all natural numbers is not up-complete (hence not algebraic), although each of its elements is compact. In contrast to this example, *no* element of $\omega \times \omega$, except the least one, is compact.

2.3 Examples (1) Let S be an uncountable set and $\mathcal{X}$ the system of all subsets which have either at most one element or a countable complement. Then $\mathcal{X}$ is an up-complete meet-semilattice whose compact elements are the empty set and the singletons, so that $\mathcal{X}$ is compactly generated, but the sets κy are not directed unless y is compact ($y \in \mathcal{X}$).

(2) A similar but countable example of a compactly generated meet-semilattice which fails to be algebraic is this: take all subsets of $\omega \times \omega$ that have at most one element or are of the form

$$M(F) = \{(x,y) \mid \text{ if } (x,z) \in F \text{ then } y \leq z\}$$

for some finite $F \subseteq \omega \times \omega$. The reader is encouraged to carry out the details.

(3) The tree of all (finite or infinite) words over an alphabet is an algebraic complete meet-semilattice. Here the compact elements are the finite words. More generally, the $\mathcal{W}$-trees are precisely those algebraic complete meet-semilattices in which every principal ideal is well-ordered.

(4) An interesting example af a non-topped inductive $\bigcap$-structure is the set of all (partial) order relations on a fixed set S with more than one element. The compact elements are here the order relations with only a finite number of related pairs outside of the diagonal

$$1_S = \{(x,x) \mid x \in S\}.$$

The maximal elements of this algebraic semilattice are the linear orderings on S.

(5) Other important examples of inductive $\bigcap$-structures are the systems of finite character, in particular, the system

- of all linearly ordered subsets of an ordered set
- of all linearly independent subsets of a vector space
- of all complete subgraphs of a graph
- of all partial maps between two sets
- of all subsets of a complete lattice whose joins belong to a fixed Scott-closed set.

In a system of finite character, the compact elements are precisely the finite sets belonging to this system.

With regard to the previous examples, the compact elements of an up-complete ordered set are also referred to as *finite elements.* Though not explicitly related to our theory, the so-called *finite element method* of modern numerical analysis is a typical instance of approximation by certain "finite" objects, so the common nomenclature is perhaps not entirely casual. The remark that the finite elements of a power set $\mathcal{P}S$ are just the finite subsets of S is generalized by the observation that the compact elements of any inductive closure system $\mathcal{X}$ are precisely the *finitely generated* members of $\mathcal{X}$, i.e. the sets of the form

$$\Gamma F = \bigcap\{Y \in \mathcal{X} \mid F \subseteq Y\}$$

for finite F, where Γ denotes the corresponding closure operator. Since every closed set $Y \in \mathcal{X}$ is the join (moreover, the union) of the point closures $\downarrow x = \Gamma\{x\}$ with $x \in Y$, it is clear that every inductive closure system is an algebraic complete lattice; such closure systems are also called *algebraic*, because they are precisely the subalgebra systems of universal algebras. Conversely, any algebraic complete lattice L is isomorphic to an algebraic closure system, namely the ideal lattice of its join-semilattice of compact elements. Thus we have the following set representation for algebraic lattices (cf. [BF], [Bü]; for generalizations to non-complete ordered sets, see 2.16, 2.17 and 4.5):

2.4 Theorem *The algebraic (= compactly generated) complete lattices (resp. meet-semilattices) are, up to isomorphism, the inductive closure systems (resp. $\bigcap$-structures).*

But observe that a closure system which is an algebraic lattice need not be inductive. For example, the principal ideals of any algebraic lattice L form a closure system $\mathcal{M}L$ isomorphic to L, while $\mathcal{M}L$ is inductive only if L satisfies the ACC. It is therefore of some interest to characterize those closure systems $\mathcal{X}$ on a set S which are algebraic lattices (but not necessarily algebraic closure systems) by means of their closure operator

$$\Gamma : \mathcal{P}S \to \mathcal{P}S, \quad X \mapsto \bigcap\{Y \in \mathcal{X} \mid X \subseteq Y\}.$$

To this aim, we proceed as follows. Generalizing the notion of $\mathcal{I}$-compactness, we consider an arbitrary closure operator $\Gamma : \mathcal{P}S \to \mathcal{P}S$ and call a subset C of S Γ-*compact* if for all $X \subseteq S$ with $C \subseteq \Gamma X$, there exists a finite $F \subseteq X$ with $C \subseteq \Gamma F$; and an element $x \in S$ is said to be Γ-*compact* if so is the singleton $\{x\}$; in particular, for the cut operator Δ, "Δ-*compact*" means "$\mathcal{I}$-*compact*" (cf. [E12]).

A straightforward verification shows that *a set C is Γ-compact iff its closure ΓC is a compact member of the corresponding closure system $\mathcal{X}$ (considered as a complete lattice)*. Notice also that every compact member of $\mathcal{X}$ is of the form ΓC where C is a *finite* Γ-compact subset of S. Furthermore, it is easy to see that the closure operator Γ is finitary (i.e. $\mathcal{X}$ is an inductive closure system) iff all finite subsets of S are Γ-compact. Finally, setting

$$K_\Gamma x = \bigcup\{C \subseteq {\downarrow}x \mid C \text{ is finite and } \Gamma\text{-compact}\},$$

$$\kappa_\Gamma x = \{c \in {\downarrow}x \mid c \text{ is } \Gamma\text{-compact}\},$$

we have:

2.5 Proposition *A closure system $\mathcal{X}$ on a set S is an algebraic lattice iff $x \in \Gamma(K_\Gamma x)$ for all $x \in S$.*

Proof Suppose $\mathcal{X}$ is an algebraic, i.e. compactly generated lattice. Then, for $x \in S$, we get

$$x \in \Gamma\{x\} = \bigvee\{\Gamma C \mid C \subseteq {\downarrow}x, C \text{ is finite and } \Gamma\text{-compact}\} = \Gamma(K_\Gamma x).$$

Conversely, if $x \in \Gamma(K_\Gamma x)$ for all $x \in S$ then we compute for $Y \in \mathcal{X}$:

$$Y = \Gamma Y = \Gamma(\bigcup\{\Gamma(K_\Gamma x) \mid x \in Y\})$$

$$= \bigvee\{\Gamma C \mid C \subseteq {\downarrow}x \text{ for some } \Gamma\text{-compact finite } C \text{ and some } x \in Y\}.$$

Hence Y is a join of compact members of $\mathcal{X}$. □

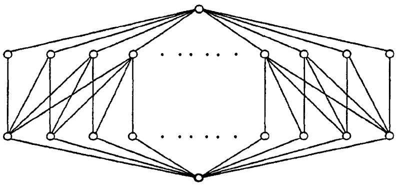
2.6 Corollary *If Γ is a closure operator with $x \in \Gamma(\kappa_\Gamma x)$ for all $x \in S$ then the corresponding closure system is an algebraic lattice. The converse holds whenever each nonempty finitely generated Γ-closed set is a point closure.*

For most of our purposes, the notion of $\mathcal{D}$-compactness is adequate; however, the definition of $\mathcal{I}$-compactness has the advantage to be *completion-invariant* in the following sense (cf. [E2], [E17]):

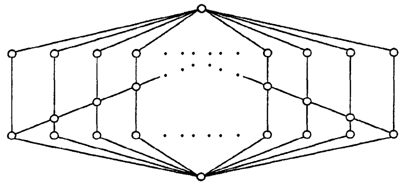
2.7 Corollary *An element x of an ordered set Q is $\mathcal{I}$-compact iff x (more precisely, the principal ideal generated by x) is ($\mathcal{I}$-)compact in the normal completion $\mathcal{N}Q$. Thus, if each element of Q is a join of $\mathcal{I}$-compact elements then the normal completion of Q is an algebraic lattice, and the converse implication holds for join-semilattices.*

These assertions are easy consequences of the more general results 2.5 and 2.6, applied to the cut operator Δ instead of an arbitrary closure operator Γ. In contrast to the above result, a slight modification of our introductory Example 2.1 shows that the normal completion of an algebraic CPO need not be algebraic (cf. [E5]):

2.8 Example



Q



$\mathcal{N}Q$

Although this ordered set Q satisfies the ACC and is therefore certainly an algebraic CPO, the normal completion $\mathcal{N}Q$ is extremely non-algebraic: the only compact element of $\mathcal{N}Q$ is the bottom element.

On the other hand, in contrast to the situation with *join*-semilattices, an up-complete *meet*-semilattice with algebraic normal completion may fail to be algebraic, as Example 2.3(1) shows: here the power set $\mathcal{P}S$ (which is certainly an algebraic lattice) is a normal completion of the non-algebraic semilattice $\mathcal{X}$. Moreover, it may happen that the $\mathcal{I}$-compact elements of an ordered set with ACC are not $\bigvee$-dense (cf. 2.14), although the normal completion is an algebraic lattice:

2.9 Example Employ the ordered set $P = P_0$ from 2.1 as a "molecule" for an inductive construction of a "fractal" ordered set P_∞: assuming that ordered sets $P_0, ..., P_n$ have been constructed, attach to each minimal element x of P_n a new copy of P_0 by identifying x with the element x_0 (the unique maximal element of P_0 generating an infinite principal ideal), and denote the resulting ordered set by P_{n+1}. Then the limit (= union) P_∞ of these ordered sets P_n satisfies the ACC, so each of its elements is compact. However, the principal ideal $\downarrow x_0$ of P_∞ does not contain any $\mathcal{I}$-compact element: for each $x \leq x_0$, the set $Y = \downarrow x \setminus \{x\}$ is an (undirected) Frink ideal with $x = \bigvee Y$. Nevertheless, the normal completion $\mathcal{N}P_\infty$ is an algebraic lattice violating the ACC.

Recall that the Scott-closed subsets of power set lattices are precisely the systems of finite character. The fact that any such system is an algebraic ordered set may be generalized as follows (cf. [E16], [Com]):

2.10 Proposition *Every Scott-open and every Scott-closed subset of an algebraic ordered set is algebraic.*

Proof If Q is algebraic and U is Scott open then the compact elements of U are the compact elements of Q contained in U: if x is compact in U and $D \in \mathcal{D}Q$ satisfies $x \leq \bigvee D$ then $D \cap U$ is a (nonempty!) directed subset of U with $\bigvee(D \cap U) = \bigvee D$ (because $D \cap U$ is cofinal with D). Hence $x \leq y$ for some $y \in D \cap U$. Similarly, if z is an arbitrary element of U

then $z = \bigvee \kappa z = \bigvee(\kappa z \cap U)$ because κz is directed. Hence U is algebraic (up-completeness is clear). For the Scott-closed set $A = Q \setminus U$, we know that A is up-complete (being closed under directed joins in Q), that for $z \in A$, κz is a directed subset of A (because A is a lower set), and that the elements of κz are also compact in A. □

In contrast to the situation with complete lattices, an interval of an algebraic ordered set need not be algebraic. A counterexample has been given in [E16].

As was observed by Birkhoff and Frink [BF], the following important order-theoretical property is shared by all algebraic lattices: an ordered set Q is called *weakly atomic* if each of its intervals

$$[a, b] = \{x \in Q \mid a \leq x \leq b\}$$

with at least two elements contains a covering pair, that is, a two-element subinterval. Although there exist examples of algebraic ordered sets which fail to be weakly atomic, one can prove a positive result for up- and down-complete ordered sets (see [E16]):

2.11 Proposition *Every compactly generated up- and down-complete ordered set is weakly atomic.*

With respect to the usual order-theoretical constructions like sums and products, compactly generated and algebraic ordered sets behave quite well, as was shown in [E16]:

2.12 Proposition (1) *Cardinal (disjoint) sums of algebraic ordered sets are again algebraic.*

(2) *A product of nonempty ordered sets is algebraic iff each factor is algebraic and all but a finite number of the factors are componentwise minimized.*

Similar statements hold for "compactly generated" instead of "algebraic".

By a *componentwise minimized* ordered set we mean one in which every connected component has a minimum; in other words, each element dominates a unique minimal element. As a consequence of 2.12, we obtain the remarkable fact that infinite powers of the algebraic chain ω^* (the dual of ω) are not algebraic, although ω^* satisfies the ACC.

Assertion 2.12(1) may be essentially generalized as follows. The *ordinal sum* of a family $(P_i \mid i \in I)$ of ordered sets over an ordered index set I is the set-theoretical sum

$$P = \textstyle\sum_{i \in I} P_i = \{(i, p) \mid i \in I, p \in P_i\},$$

ordered "lexicographically" by

$$(i, p) \leq (j, q) \text{ iff } i < j \text{ in } I \text{ or } (i = j \text{ and } p \leq q \text{ in } P_i = P_j).$$

2.13 Proposition *The ordinal sum of a family $(P_i \mid i \in I)$ of ordered sets over an ordered index set I is an algebraic CPO iff I and each P_i is an algebraic CPO.*

We omit the straightforward but somewhat tedious proof of this fact. Note, however, that the ordinal sum of two-element antichains over the algebraic index chain $\omega + 1$ is not even up-complete, so the existence of least elements is essential in 2.13.

A particularly important property of algebraic lattices carries over to compactly generated ordered sets, viz. the existence of $\bigwedge$-decompositions into $\bigwedge$-irreducible elements. In case of ordered sets which are not complete lattices, the definition of irreducibility requires a bit more care than in the complete case: we call an element x *completely join-irreducible*, written $\bigvee$*-irreducible* (respectively, *completely meet-irreducible*, written $\bigwedge$*-irreducible*) if it is not the join (respectively, meet) of any set Y unless $x \in Y$. By this definition, a maximal element of an ordered set is $\bigwedge$-irreducible iff it is not the greatest element. Recall that a subset Z of an ordered set Q is said to be $\bigvee$*-dense* or a $\bigvee$*-generator* (respectively, $\bigwedge$*-dense* or a $\bigwedge$*-generator*) if each element of Q is a join (respectively, meet) of elements from Z. The following equivalent description of $\bigvee$- and $\bigwedge$-density by a certain "separation property" is helpful for practical purposes; its proof is left as an easy exercise.

2.14 Lemma *A subset G of an ordered set Q is $\bigvee$-dense iff for all $x, y \in Q$ with $x \not\leq y$, there is some $z \in G$ with $z \leq x$ but $z \not\leq y$. Dually, G is $\bigwedge$-dense iff for all $x, y \in Q$ with $x \not\leq y$, there is some $z \in G$ with $y \leq z$ but $x \not\leq z$.*

Using this lemma, the proof for the Birkhoff-Frink decomposition theorem on complete lattices (see [BF] or [CD]) is easily translated to up-complete ordered sets, and one obtains:

2.15 Proposition *In a compactly generated ordered set, the set of $\bigwedge$-irreducible elements is $\bigwedge$-dense.*

Proof By hypothesis, the set of all compact elements is $\bigvee$-dense. Hence, for $x \not\leq y$, we find a compact z with $z \leq x$ and $z \not\leq y$. As the set of all q with $z \not\leq q$ is closed under directed joins, we find (by Zorn's Lemma) a maximal element $q \geq y$ with $z \not\leq q$. This element must be $\bigwedge$-irreducible, because z is a lower bound for all elements greater than q, and consequently, q cannot be the greatest lower bound of these. As $y \leq q$ and $x \not\leq q$, the $\bigwedge$-irreducible elements form a $\bigwedge$-dense subset of Q. □

Here are some applications of 2.15:

(1) In the algebraic meet-semilattice of all order relations on a fixed set (see 2.3 (4)), only the maximal members, i.e. the linear orders are $\bigwedge$-irreducible. Thus Szpilrajn's Theorem [Sz] on the representation of arbitrary orders as intersections of linear orders is a special instance of 2.15.

(2) Since the linearly ordered subsets of a fixed ordered set Q form a system of finite character, we know from 2.15 that each chain of Q is an intersection of $\bigwedge$-irreducible ones. If Q itself is a chain then the $\bigwedge$-irreducible subchains are precisely the coatoms of the power set $\mathcal{P}Q$. In a non-linearly ordered set Q, a chain C is $\bigwedge$-irreducible iff it is either maximal, or there exists a unique element $x \in Q \setminus C$ such that $C \cup \{x\}$ is a maximal chain.

(3) Let V be a vector space with more than two elements. The linearly independent subsets of V form a system of finite character. Since the $\bigwedge$-irreducible members of this system are the maximal ones, i.e. the bases, we conclude that every linearly independent subset of V is an intersection of bases.

(4) For any ordered set Q, the $\mathcal{W}$-tree $\mathcal{W}_\perp Q$ of all well-ordered subsets of Q, ordered by propagation, is an algebraic ordered set and closed under nonempty intersections. Hence each well-ordered subset of Q is an intersection of $\bigwedge$-irreducible ones (with respect to the propagation order), and it is easy to see that W is $\bigwedge$-irreducible in $\mathcal{W}_\perp Q$ iff $W \neq Q$ has no upper bound, or W has a join that is $\bigwedge$-irreducible in Q or does not belong to W.

As mentioned earlier, a similar set representation as for algebraic lattices (see 2.4) holds for any compactly generated ordered set Q: the system

$$\mathcal{X}Q = \{\mathcal{K}Q \cap \downarrow y \mid y \in Q\}$$

is inductive, but it is not a closure system unless Q is a complete lattice; however, it is always a point closure system, the point closures being the principal ideals of the ordered set $\mathcal{K}Q$ of all compact elements. The map

$$\varepsilon_Q : \mathcal{X}Q \longrightarrow Q, \quad Y \longmapsto \bigvee Y$$

turns out to be an isomorphism between the inductive point closure system $\mathcal{X}Q$ and the compactly generated ordered set Q. On the other hand, every inductive point closure system is easily seen to be a compactly generated ordered set (because the point closures are compact), and we arrive at the following *Representation Theorem for Compactly Generated Ordered Sets*:

2.16 Theorem *An ordered set is compactly generated iff it is isomorphic to an inductive point closure system.*

Next, let us establish similar set representations for algebraic ordered sets. Thus we are looking for suitable representatives where the order relation is set inclusion and, moreover, directed joins agree with directed unions. Without saying, such a representation theory has diverse advantages. For example, in order to prove certain statements on compactly generated or algebraic ordered sets, it suffices to restrict the attention to the selected representatives which are often better visualizable. On the other hand, general results on abstract algebraic ordered sets provide us with concrete applications to the set-theoretical representatives. Later on, in Sections 4-6, we shall supply the corresponding "$\mathcal{Z}$-generalizations" and the necessary proofs, while in the present "classical" situation we shall only touch upon the main ideas and omit the proof details. For more background on this subject, the reader may consult [BH], [Com], [E16] and [We].

Roughly speaking, our first representation theorem for algebraic ordered sets will state that they are, up to isomorphism, certain ideal systems of ordered sets (a fact that is commonly known at least in the more restricted case of algebraic lattices). Given an arbitrary ordered set P, let us denote by $\mathcal{D}^\wedge P$ the collection of all directed lower sets of P. Although this is not always a closure system, it is certainly a point closure system: the point closures are the principal ideals, and these are trivially directed. In some sources (e.g. in [Com]), the members of $\mathcal{D}^\wedge P$ are referred to as the *ideals* of P, but it should be observed that an ideal in the sense of Frink, that is, a member of the ideal completion $\mathcal{I}P$, need not be directed. However, in case of a join-semilattice with least element, $\mathcal{D}^\wedge P$ coincides with $\mathcal{I}P$, and the ideals are precisely those lower sets which are closed under finite joins. While Frink's ideal completion $\mathcal{I}P$ is the *smallest inductive closure system* containing all principal ideals of P, $\mathcal{D}^\wedge P$ is the *smallest inductive (point closure) system* with this property. In particular, $\mathcal{D}^\wedge P$ is always an up-complete ordered set containing a $\bigvee$-dense isomorphic copy of P (viz. $\mathcal{M}P$); therefore, $\mathcal{D}^\wedge P$ will be referred to as the *up-completion* of P. It is evident that the principal ideals are precisely the compact members of the ordered set $\mathcal{D}^\wedge P$. Hence the map

$$\eta_P : P \longrightarrow \mathcal{K}\mathcal{D}^\wedge P, \quad x \longmapsto \downarrow x$$

is an isomorphism. On the other hand, a straightforward verification shows that for any algebraic ordered set Q, the map

$$\varepsilon_Q : \mathcal{D}^\wedge \mathcal{K} Q \longrightarrow Q, \quad Y \longmapsto \bigvee Y$$

is an isomorphism. These remarks suffice to provide our *First Representation Theorem for Algebraic Ordered Sets* (cf. [Hf1], [HM1], [E16]):

2.17 Theorem *An ordered set is algebraic iff it is isomorphic to the up-completion of some ordered set. Conversely, every ordered set P is the subposet of all compact elements of some algebraic ordered set which is isomorphic to the up-completion of P.*

The second representation theorem will be of more topological nature. In particular, we need the notion of *soberness*. A topological space is said to be *sober* iff it is T_0 and the only irreducible (i.e. $\vee$-prime) closed sets are the point closures. In modern topology and its lattice-theoretical aspects, sober spaces play an important role, because they are completely determined, up to isomorphism, by the lattice structure of their topologies (see, for example, [Com], [BH] or [Jo1]).

In a recent paper, entitled *"The ABC of Order and Topology"* [E19], we have discussed thoroughly three specific classes of topological spaces which play a prominent role in the interplay between ordered and topological structures:

A-spaces (*Alexandroff-discrete spaces*) are characterized by the property that arbitrary intersections of open sets are open, or equivalently, that each point has a smallest neighborhood (cf. [Al]).

B-spaces (or *monotope spaces*; cf. [Bt], [HM1]) are topological spaces with a minimal basis; it is easy to see that such a minimal basis must already be the *smallest basis* and consist of all *open cores* (= *monogenerated open sets*), where by the *core* of a point x, we mean the intersection of all neighborhoods of this point (this is the principal filter generated by x with respect to the specialization quasiorder). Furthermore, one can show that a topological space is a B-space iff its topology is *isomorphic* to some A-topology.

C-spaces are topological spaces in which every point has a neighborhood basis of (not necessarily open) cores. Alternatively, these spaces may be characterized by the property that their topology is completely distributive (cf. [Ba3], [Hf3-4], [EW]). A third possibility to describe such spaces is the following: a topological space is a C-space iff it is locally compact (without Hausdorff separation) and "has a dual" (cf. [E6], [Hf5]), i.e. its topology is dually isomorphic to some other topology. It is easy to see that every A-space is a B-space, and every B-space is a C-space, but not conversely.

Recall that a subset of an ordered set is closed in the Scott topology iff it is a lower set and closed under directed joins. Now the *Second Representation Theorem for Algebraic Ordered Sets* reads as follows (cf. [Bt], [Hf1], [HM1]):

2.18 Theorem *The algebraic ordered sets, equipped with their Scott topology, are precisely the sober B-spaces (considered as ordered sets with respect to specialization).*

The main ingredients for the proof of this theorem are the following two observations on *compatible* topologies on ordered sets, i.e. topologies such that the corresponding closure system has the given order as specialization:

2.19 Lemma *Every compatible completely distributive topology on an ordered set contains the Scott topology, which in turn contains every compatible sober topology. Hence every ordered set carries at most one compatible sober and completely distributive topology, namely the Scott topology.*

Proof Let U be some subset of an ordered set Q which is not open in a given compatible C-space topology. Then we find a point $y \in U$ such that for no $x \in U$, y is an inner point of the core $\uparrow x$. Hence the set

$$D = \{x \mid y \in (\uparrow x)^\circ\}$$

is a subset of $Q \setminus U$. The crucial observation is now that D is directed. Indeed, for finite $F \subseteq D$, we obtain $y \in \bigcap\{(\uparrow x)^\circ \mid x \in F\}$, so we find a w such that $y \in (\uparrow w)^\circ \subseteq \uparrow w \subseteq \bigcap\{(\uparrow x)^\circ \mid x \in F\} \subseteq F^\uparrow$, i.e., w is an upper bound of F in D. Moreover, y is the join of D, as $\uparrow y = \bigcap\{\uparrow x \mid x \in D\}$. But y belongs to U, so this set cannot be Scott open.

For the second claim, let $(S, \mathcal{X})$ be any sober closure space and $Q = (S, \leq)$ the associated ordered set. If D is a directed subset of some closed set $X \in \mathcal{X}$, then the closure ΓD is a $\vee$-prime member of $\mathcal{X}$ contained in X; by soberness, ΓD must be a point closure $\downarrow x$, and x must be the join of D. Thus Q is an up-complete ordered set, and X is a Scott-closed subset of Q (cf. [Com] and 6.11). □

The question of which ordered sets actually admit such a compatible sober C-space topology has been answered in [Hf3-4] and [La1]: these are precisely the *continuous ordered sets*, i.e. up-complete ordered sets such that for every element there exists a smallest directed ideal whose join dominates this element. As every algebraic ordered set is continuous, the Scott topology of such an ordered set is indeed a sober C-space topology, and moreover, it has a smallest basis, consisting of all principal dual ideals (= cores) generated by compact elements.

The third representation theorem is merely a consequence of the second. By an $\mathcal{A}$-*lattice* or *superalgebraic lattice* we mean a complete lattice in which every element is a join of *supercompact* (= *completely join-prime*, $\bigvee$*-prime*) elements. It is well known that these lattices are, up to isomorphism, just the A-topologies, so that their duals are again $\mathcal{A}$-lattices. Combining this observation with the above remark on B-spaces, we see that a topological space is a B-space iff its topology is an $\mathcal{A}$-lattice (but not necessarily an A-topology). By the $\vee$*-spectrum* of a lattice, we mean the set of its $\vee$*-prime* (i.e. $\mathcal{F}$-compact) elements. Frequently, one is working with the dual notion of $\wedge$*-prime* elements and $\wedge$*-spectra*. Now to the *Third Representation Theorem* (cf. [E16], [Hf4], [La1]):

2.20 Theorem *An ordered set is algebraic iff it is the* $\vee$*-spectrum of a superalgebraic lattice (which is isomorphic to the lattice of all Scott-closed sets). Conversely, a lattice is superalgebraic iff it is isomorphic to the lattice of all Scott-closed sets of an algebraic ordered set.*

Indeed, by 2.18, any algebraic ordered set Q is a sober B-space with respect to its Scott topology, and consequently, Q is isomorphic to the $\vee$-spectrum of the superalgebraic lattice of all Scott-closed sets. Conversely, it is not difficult to check that the $\vee$-spectrum of any superalgebraic lattice is always an algebraic ordered set, being closed under directed joins.

A similar reasoning, involving the Representation Theorem 2.16, leads to the following analogous result for compactly generated ordered sets:

2.21 Corollary *An ordered set Q is compactly generated iff it is a* $\bigvee$*-dense subset of some superalgebraic lattice L such that directed joins in Q agree with those in L.*

3 Categories of compactly generated and algebraic ordered sets

As in similar situations where certain mathematical structures are expressed by means of others, it appears now desirable to extend the representation theory for compactly generated and algebraic ordered sets, respectively, from the object level to appropriate morphism classes, in order to obtain so-called *categorical equivalence theorems.* The importance of such equivalences is well-known from other contexts: they enable us to translate each result on one category into the language of the other; and, while the arguments may be quite simple in one of the categories under consideration, they may lead to rather profound conclusions in the other.

We assume the reader to be familiar with basic categorical notions such as *object, morphism, functor, natural transformation, opposite category* etc. However, for readers having only minor experience with category theory, we recall two central notions occurring very frequently in the sequel: an *isomorphism* between categories $\mathbf{A}$ and $\mathbf{B}$ is a pair of functors $\mathcal{G} : \mathbf{A} \to \mathbf{B}$ and $\mathcal{H} : \mathbf{B} \to \mathbf{A}$ such that the composite functor $\mathcal{H} \circ \mathcal{G}$ is the identity functor $1_{\mathbf{A}}$ on $\mathbf{A}$ and $\mathcal{G} \circ \mathcal{H}$ is the identity functor $1_{\mathbf{B}}$ on $\mathbf{B}$; sometimes in this situation either of the functors $\mathcal{G}$ and $\mathcal{H}$ is also referred to as a *(functorial) isomorphism.* More generally, two functors $\mathcal{G}$ and $\mathcal{H}$ establish an *equivalence* between the categories $\mathbf{A}$ and $\mathbf{B}$ if there are natural isomorphisms $\eta : 1_{\mathbf{A}} \to \mathcal{H} \circ \mathcal{G}$ and $\varepsilon : \mathcal{G} \circ \mathcal{H} \to 1_{\mathbf{B}}$ (this means that for each $\mathbf{A}$-object A, there is an $\mathbf{A}$-isomorphism $\eta_A : A \to \mathcal{H}\mathcal{G}A$ with $\mathcal{H}\mathcal{G}\varphi \circ \eta_A = \eta_{A'} \circ \varphi$ for each $\mathbf{A}$-morphism $\varphi : A \to A'$, and similarly for ε.) The existence of such an equivalence has the following important consequences (cf. [AHS]):

(1) $\mathcal{G}$ and $\mathcal{H}$ are *(isomorphism-)dense* functors; that is, each $\mathbf{B}$-object is isomorphic to the image of some $\mathbf{A}$-object under $\mathcal{G}$, and each $\mathbf{A}$-object is isomorphic to the image of some $\mathbf{B}$-object under $\mathcal{H}$.

(2) $\mathcal{G}$ and $\mathcal{H}$ are *full* and *faithful*; that is, for any two $\mathbf{A}$-objects A and A', $\mathcal{G}$ induces a bijection between the set of all $\mathbf{A}$-morphisms from A to A' and the set of all $\mathbf{B}$-morphisms from $\mathcal{G}A$ to $\mathcal{G}A'$, and analogously for $\mathcal{H}$. Conversely, any dense, full and faithful functor yields an equivalence.

On account of these facts, a categorical equivalence theorem may also be regarded as a pair of representation theorems: objects and morphisms of one category are expressed, "up to isomorphism", in terms of objects and morphisms of the other category and the mediating functors. In what follows, we shall be frequently confronted with such situations or their arrow-reversing counterparts, so-called *dualities*: a *dual isomorphism* (resp. *dual equivalence*) between two categories $\mathbf{A}$ and $\mathbf{B}$ is an isomorphism (resp. equivalence) between $\mathbf{A}$ and the opposite category $\mathbf{B}^{op}$ (where morphism arrows are reversed). The existence of such a duality entails, among other things, that products in one category correspond to coproducts in the other category.

One typical equivalence theorem, relating topological with order-theoretical structures, has been established in [E10]; roughly speaking, it states that the category $\mathbf{CS}_0$ of T_0 closure spaces and continuous maps is equivalent to the category $\mathbf{CG}_{\bigvee}$ of complete lattices with distinguished $\bigvee$-generators as objects and maps preserving joins and the selected $\bigvee$-generators as morphisms. For our present purposes, it will be convenient to generalize this equivalence to the setting of so-called *ordinary spaces* on the one hand and ordered sets with selected $\bigvee$-generators on the other hand. By an *ordinary space* or, if no confusion is likely, by a *space*, we mean a pair $(S, \mathcal{X})$ consisting of a set S and a point closure system $\mathcal{X}$ on S. The classical topological notion of continuity is extended by calling a map φ between spaces $(S, \mathcal{X})$ and $(S', \mathcal{X}')$ *continuous* if

(C1) for each $Y' \in \mathcal{X}'$, the inverse image $\varphi^{-1}[Y']$ belongs to $\mathcal{X}$,

and φ is said to be *weakly closed* if

(C2) for each $Y \in \mathcal{X}$, the hull $\bar{\varphi}Y = \bigcap\{Y' \in \mathcal{X}' \mid \varphi[Y] \subseteq Y'\}$ belongs to $\mathcal{X}'$.

Of course, the second condition is automatically fulfilled if $\mathcal{X}'$ happens to be a closure system, but also in many other concrete situations encountered at the borderline between order and topology (see, for example, [E8] and Section 6). For a fruitful theory of ordinary spaces, it turns out that a reasonable class of morphisms should satisfy not only the continuity condition (C1) but also the weak closedness property (C2); such maps will be called *strongly continuous.* Apparently our choice of morphisms satisfies the necessary categorical axioms: identity maps are strongly continuous, and the class of (strongly) continuous maps is closed under composition.

We know that every space $(S, \mathcal{X})$ carries a natural quasiorder, the specialization $\leq_{\mathcal{X}}$ defined by $x \leq_{\mathcal{X}} y$ iff x belongs to the closure of y. It is easy to see that every continuous map φ between spaces $(S, \mathcal{X})$ and $(S', \mathcal{X}')$ preserves specialization; in other words, it is an isotone map between the quasiordered sets $(S, \leq_{\mathcal{X}})$ and $(S', \leq_{\mathcal{X}'})$. Since the closure of a point x is just the principal ideal generated by x, every space $(S, \mathcal{X})$ admits a natural "*principal ideal embedding*"

$$\eta_S = \eta_S^{\mathcal{X}} : S \longrightarrow \mathcal{X}, \quad x \longmapsto {\downarrow}x.$$

Notice that η_S is continuous as a map from $(S, \mathcal{X})$ into the space $(\mathcal{X}, \mathcal{M}\mathcal{X})$, since

$$\eta_S{}^{-1}[\mathcal{P}Y \cap \mathcal{X}] = \{x \in S \mid {\downarrow}x \subseteq Y\} = Y \text{ for each } Y \in \mathcal{X}.$$

At this point it appears opportune to recall a few basic facts from residuation theory (for more background, see e.g. [AHS], [BJ] or [Com]). A map between quasiordered sets Q and Q' is *residuated* iff inverse images of principal ideals are principal ideals; in other words, iff it is continuous as a map between the spaces $(Q, \mathcal{M}Q)$ and $(Q', \mathcal{M}Q')$. Dualizing the concept of residuated maps, one calls a map between quasiordered sets *residual* if inverse images of principal filters (= dual principal ideals) are principal filters. Residuated and residual maps between ordered sets always occur in pairs: an *adjoint pair* consists of two maps $\varphi : Q \to Q'$ and $\psi : Q' \to Q$ between ordered sets such that for all $x \in Q$ and $x' \in Q'$,

$$\varphi x \leq x' \iff x \leq \psi x'.$$

In this case, φ is called the *left* or *lower adjoint* of ψ, and ψ is called the *right* or *upper adjoint* of φ. These two maps determine each other uniquely, by the equations

$$\varphi x = \min\{x' \in Q' \mid x \leq \psi x'\},$$

$$\psi x' = \max\{x \in Q \mid \varphi x \leq x'\}.$$

It is an easy exercise to show that a map is residuated iff it has an upper adjoint, and residual iff it has a lower adjoint. Furthermore, any lower (upper) adjoint preserves joins (meets), and conversely, any join- (meet-) preserving map between complete lattices is a lower (upper) adjoint.

On account of these remarks, the category $\mathbf{C}_\nabla$ of complete lattices and join-preserving maps is a full subcategory of $\mathbf{O}_\nabla$, the category of ordered sets and residuated maps. Dually, the category $\mathbf{C}^\Delta$ of complete lattices and meet-preserving maps is a full subcategory of the category $\mathbf{O}^\Delta$ of ordered sets and residual maps. Moreover, the categories $\mathbf{O}_\nabla$ and $\mathbf{O}^\Delta$ (respectively, $\mathbf{C}_\nabla$ and $\mathbf{C}^\Delta$) are dually isomorphic via the "adjoining" functors $\mathcal{U}$ and $\mathcal{L}$ which act identically on the objects and provide the passage from lower to upper adjoints and vice versa; thus $\mathcal{U}\varphi$ denotes the upper adjoint of a residuated map φ, and $\mathcal{L}\psi$ denotes the lower adjoint of a residual map ψ. Since $\mathcal{L}\circ\mathcal{U}$ and $\mathcal{U}\circ\mathcal{L}$ are the identity functors on $\mathbf{O}_\nabla$ and $\mathbf{O}^\Delta$, respectively, we see that $\mathcal{U} : \mathbf{O}_\nabla \to \mathbf{O}^\Delta$ and $\mathcal{L} : \mathbf{O}^\Delta \to \mathbf{O}_\nabla$ are in fact mutually inverse functorial isomorphisms.

The following fundamental connection between topological and certain order-theoretical morphisms has been proved in [E19] (see also [E10]):

3.1 Proposition *A map φ between spaces $(S,\mathcal{X})$ and $(S',\mathcal{X}')$ is strongly continuous iff there exists a unique residuated map $\bar{\varphi} : \mathcal{X} \to \mathcal{X}'$ with $\bar{\varphi}\circ\eta_S = \eta_{S'}\circ\varphi$, viz.*

$$\bar{\varphi}X = \bigcap\{X' \in \mathcal{X}' \mid \varphi[X] \subseteq X'\} \qquad (X \in \mathcal{X}).$$

Moreover, if ψ is a strongly continuous map from $(S',\mathcal{X}')$ to $(S,\mathcal{X})$ and φ is lower adjoint to ψ then $\bar{\varphi}$ is lower adjoint to $\bar{\psi}$.

For the intended order-theoretical representation of *ordinary T_0 spaces*, i.e. ordinary spaces in which distinct points have distinct closures, we introduce the following category $\mathbf{OG}_\nabla$ of "*ordered sets with $\bigvee$-generators*": objects are pairs (Q,G) where Q is an ordered set and G is a $\bigvee$-generator (i.e. $\bigvee$-dense subset) of Q. Morphisms between two objects (Q,G) and (Q',G') are residuated maps $\varphi : Q \to Q'$ preserving the selected generators, i.e. satisfying $\varphi[G] \subseteq G'$.

For any space $(S,\mathcal{X})$, the set

$$G_{S,\mathcal{X}} = \{\downarrow x \mid x \in S\}$$

of all point closures is a $\bigvee$-generator for the point closure system $\mathcal{X}$, regarded as an ordered set with respect to inclusion (indeed, each member of $\mathcal{X}$ is a union of point closures). Hence, by Proposition 3.1, we have a "generic functor" $\mathcal{G}$ from the category $\mathbf{OS}$ of (ordinary) spaces with strongly continuous maps to the category $\mathbf{OG}_\nabla$, assigning to each space $(S,\mathcal{X})$ the pair$(\mathcal{X}, G_{S,\mathcal{X}})$ and to each strongly continuous map φ the residuated map $\mathcal{G}\varphi = \bar{\varphi} : \mathcal{X} \to \mathcal{X}'$. This is actually an $\mathbf{OG}_\nabla$-morphism between $\mathcal{G}(S,\mathcal{X})$ and $\mathcal{G}(S',\mathcal{X}')$ since each point closure $\downarrow x$ is mapped onto the point closure $\downarrow\varphi x$.

On the other hand, there is a functor $\mathcal{H}$ from the category $\mathbf{OG}_\nabla$ to the category $\mathbf{OS}$, associating with any $\mathbf{OG}_\nabla$-object (Q,G) the space $(G,\mathcal{X}_{Q,G})$ where

$$\mathcal{X}_{Q,G} = \{G \cap \downarrow y \mid y \in Q\}.$$

On the morphism level, $\mathcal{H}$ simply acts by restriction to the $\bigvee$-generators. We have to verify that for any $\mathbf{OG}_\nabla$-morphism $\Phi : (Q,G) \to (Q',G')$, the restriction $\varphi = \mathcal{H}\Phi : G \to G'$ is actually strongly continuous; to this aim, we use the upper adjoint Ψ of Φ and observe that for $y' \in Q'$, the inverse image $\varphi^{-1}[G' \cap \downarrow y']$ coincides with the set $G \cap \downarrow \Psi y'$. Thus φ satisfies condition (C1).

Concerning the weak closedness condition (C2), we show that for $X = G \cap \downarrow y \in \mathcal{X}_{Q,G}$, the set $X' = G' \cap \downarrow \Phi y$ is the smallest member of $\mathcal{X}_{Q',G'}$ containing the image $\varphi[X]$. Indeed, $x \in X$ means $x \in G$ and $x \leq y$, whence $\varphi x = \Phi x \in G'$ and $\Phi x \leq \Phi y$, i.e. $\varphi x \in X'$. On the other hand, if $G' \cap \downarrow y'$ is any member of $\mathcal{X}_{Q',G'}$ containing $\varphi[X]$ then y' is an upper bound of $\varphi[X] = \Phi[G \cap \downarrow y]$, and as Φ preserves arbitrary joins, it follows that $\Phi y = \Phi(\bigvee(G \cap \downarrow y)) = \bigvee \Phi[G \cap \downarrow y] \leq y'$, so that $X' = G' \cap \downarrow \Phi y \subseteq G' \cap \downarrow y'$. □

If Q is a complete lattice and G consists of $\vee$-prime elements only, then $\mathcal{X}_{Q,G}$ is the collection of all closed sets in the so-called *hull-kernel topology* of G, motivating the use of the letter $\mathcal{H}$ for the corresponding functor (see e.g. [HM1]). Since

$$\mathcal{H}(Q, G) = (G, \mathcal{X}_{Q,G})$$

is always a T_0 space, we may regard $\mathcal{H}$ as a functor from $\mathbf{OG}_\nabla$ to the category $\mathbf{OS}_0$ of ordinary T_0 spaces and strongly continuous maps.

On account of the above remarks on residuated maps, we have a modified dual isomorphism $\mathcal{U}$ between the category $\mathbf{OG}_\nabla$ and the category $\mathbf{OG}^\Delta$ of ordered sets with $\bigvee$-generators and residual maps whose lower adjoints preserve these generators; the inverse functor $\mathcal{L}$ assigns to each $\mathbf{OG}^\Delta$-morphism its lower adjoint. The composite functor $\mathcal{U} \circ \mathcal{G}$ associates with a strongly continuous map φ between ordinary T_0 spaces $(S, \mathcal{X})$ and $(S', \mathcal{X}')$ the inverse image map $\varphi^{-1} : \mathcal{X}' \to \mathcal{X}$, which is the upper adjoint of $\bar{\varphi} : \mathcal{X} \to \mathcal{X}'$.

Of course, the functor $\mathcal{G}$ restricts to a functor between the category $\mathbf{CS}$ (resp. $\mathbf{CS}_0$) of closure spaces (resp. T_0 closure spaces) with (strongly) continuous maps and the category $\mathbf{CG}_\nabla$ of complete lattices with $\bigvee$-generators and maps preserving joins and the selected generators. In the opposite direction, the hull-kernel functor $\mathcal{H}$ restricts to a functor from $\mathbf{CG}_\nabla$ to $\mathbf{CS}_0 \subseteq \mathbf{CS}$. Furthermore, $\mathcal{U}$ and $\mathcal{L}$ establish a duality between the category $\mathbf{CG}_\nabla$ and the category $\mathbf{CG}^\Delta$ of complete lattices with $\bigvee$-generators and meet-preserving maps whose lower adjoints preserve the selected generators.

For any T_0 space $(S, \mathcal{X})$, the corestricted principal ideal embedding

$$\eta_{(S,\mathcal{X})} : S \longrightarrow G_{S,\mathcal{X}}, \quad x \longmapsto \downarrow x$$

is an isomorphism, and an easy verification shows that this defines a natural isomorphism η between the identity functor on $\mathbf{OS}_0$ and the composite functor $\mathcal{H} \circ \mathcal{G}$.

On the other hand, for any $\mathbf{OG}_\nabla$-object (Q, G), the map

$$\varepsilon_{(Q,G)} : \mathcal{X}_{Q,G} \longrightarrow Q, \quad Y \longmapsto \bigvee Y$$

is an isomorphism because G is $\bigvee$-dense in Q. Again, it is not hard to see that in this way a natural isomorphism ε between the composite functor $\mathcal{G} \circ \mathcal{H}$ and the identity functor on $\mathbf{OG}_\nabla$ is obtained.

In all, we have shown the following equivalence theorem (cf. [E10], [E19]):

3.2 Theorem *The generic functor $\mathcal{G}$ and the hull-kernel functor $\mathcal{H}$ establish an equivalence between the category* $\mathbf{OS}_0$ *of ordinary T_0 spaces and the category* $\mathbf{OG}_\nabla$ *of ordered sets with $\bigvee$-generators; moreover, these functors induce an equivalence between the category* $\mathbf{CS}_0$ *of T_0 closure spaces and the category* $\mathbf{CG}_\nabla$ *of complete lattices with $\bigvee$-generators.*

After these general considerations on categories of spaces and ordered sets, let us return to the study of compactly generated and algebraic ordered sets. Recall that by Theorem 2.16, the compactly generated orderd sets are, up to isomorphism, the inductive point closure systems. More precisely, for any compactly generated ordered set Q and any $\bigvee$-dense subset G consisting of compact elements, the T_0 point closure system $\mathcal{X}_{Q,G}$ is inductive and isomorphic to Q via the map $\varepsilon_{(Q,G)}$. Conversely, starting with a T_0 space $(S, \mathcal{X})$ whose point closure system $\mathcal{X}$ is inductive, we obtain a compactly generated ordered set $\mathcal{X}$, and the point closures form a $\bigvee$-generator of $\mathcal{X}$ consisting of compact elements.

These considerations motivate the following definitions. Let $\mathcal{D}\mathbf{OG}_\nabla$ (resp. $\mathcal{D}\mathbf{CG}_\nabla$) denote that full subcategory of $\mathbf{OG}_\nabla$ whose objects (Q, G) have the property that Q is up-complete (resp. a complete lattice) and the $\bigvee$-generator G consists of ($\mathcal{D}$-)compact elements only, so that Q is compactly generated. On the other hand, let $\mathcal{D}\mathbf{OS}_0$ (resp. $\mathcal{D}\mathbf{CS}_0$) denote the category of T_0 spaces (resp. closure spaces) with inductive, i.e. $\mathcal{D}$-union complete systems. Then, combining the above remarks, we arrive at the following categorical improvement of 2.16:

3.3 Theorem *The generic functor $\mathcal{G} : \mathbf{OS}_0 \to \mathbf{OG}_\nabla$ and the hull-kernel functor $\mathcal{H} : \mathbf{OG}_\nabla \to \mathbf{OS}_0$ induce equivalences between the categories $\mathcal{D}\mathbf{OG}_\nabla$ and $\mathcal{D}\mathbf{OS}_0$, respectively, between the categories $\mathcal{D}\mathbf{CG}_\nabla$ and $\mathcal{D}\mathbf{CS}_0$.*

In case of *algebraic* ordered sets, the situation is a bit simpler, because an algebraic ordered set is uniquely determined, up to isomorphism, by the ordered subset of its compact elements. Therefore, we may "forget" the generator when passing from an arbitrary ordered set P to its up-completion $\mathcal{D}^\wedge P$, the algebraic ordered set of all directed lower sets of P. We make this assignment functorial as follows. Any isotone, i.e. order-preserving map $\varphi : P \to P'$ extends to a map

$$\mathcal{D}^\wedge\varphi : \mathcal{D}^\wedge P \longrightarrow \mathcal{D}^\wedge P', \quad Y \mapsto {\downarrow}\varphi[Y]$$

which preserves directed unions, i.e. joins, and sends principal ideals to principal ideals. As the principal ideals of P are precisely the compact members of $\mathcal{D}^\wedge P$, we can say that *$\mathcal{D}^\wedge P$ preserves compactness.* Thus $\mathcal{D}^\wedge$ may be regarded as a functor from the category $\mathbf{O}$ of ordered sets with isotone maps to the category $\mathbf{AO}$ of algebraic ordered sets with maps preserving directed joins and compactness.

In the other direction, we have a functor $\mathcal{K}$ from the category $\mathbf{AO}$ to the category $\mathbf{O}$, restricting objects and morphisms to the ordered subsets of all compact elements. For any algebraic ordered set Q, the isomorphism

$$\varepsilon_Q : \mathcal{D}^\wedge \mathcal{K} Q \longrightarrow Q, \quad Y \longmapsto \bigvee Y$$

is natural in the categorical sense. Similarly, we obtain a natural isomorphism

$$\eta_P : P \longrightarrow \mathcal{M} P, \quad x \longmapsto \downarrow x$$

between the identity functor on $\mathbf{O}$ and the composite functor $\mathcal{K}\mathcal{D}^\wedge$.

The following equivalence theorem is now immediate:

3.4 Theorem *The functor $\mathcal{D}^\wedge$ induces an equivalence between the category* $\mathbf{O}$ *of ordered sets and the category* $\mathbf{AO}$ *of algebraic ordered sets with maps preserving directed joins and compactness.*

Sometimes, one is interested in stronger types of morphisms. For example, in case of algebraic complete lattices one might wish that *arbitrary* joins are preserved. In order to include this situation in our theory, we have also to strengthen the morphisms of the category $\mathbf{O}$. The appropriate definition has been prepared by our preceding considerations on ordinary spaces. We only have to add the remark that *a lower adjoint map between algebraic ordered sets preserves compactness iff its upper adjoint preserves directed joins.* A general $\mathcal{Z}$-version of this fact will be proved in 4.11. Thus we have (cf. [GG]):

3.5 Proposition *The functors $\mathcal{U}$ and $\mathcal{L}$ induce mutually inverse dual isomorphisms between the category $\mathbf{AO}_\nabla$ of algebraic ordered sets with residuated maps preserving compactness and the category $\mathbf{AO}^\Delta$ of algebraic ordered sets with residual maps preserving directed joins. Moreover, $\mathcal{U}$ and $\mathcal{L}$ restrict to dual isomorphisms between the full subcategories $\mathbf{AL}_\nabla$ and $\mathbf{AL}^\Delta$ whose objects are algebraic lattices.*

By definition, an $\mathbf{AL}_\nabla$-morphism preserves arbitrary joins and compactness, while an $\mathbf{AL}^\Delta$-morphism preserves arbitrary meets and directed joins (cf. [Com]).

In accordance with our general notion of continuity, we call a map φ between ordered sets $P = (S, \leq)$ and $P' = (S', \leq')$ *$\mathcal{D}$-continuous* if inverse images of directed lower sets are again directed lower sets; in other words, if it is continuous as a map between the spaces $(S, \mathcal{D}^\wedge P)$ and $(S', \mathcal{D}^\wedge P')$. Since any such map is certainly isotone, it maps directed sets to directed sets and is therefore automatically weakly closed, hence strongly continuous. From 3.1 we infer that for any $\mathcal{D}$-continuous map φ there is a residuated map $\bar{\varphi} : \mathcal{D}^\wedge P \to \mathcal{D}^\wedge P'$ with $\bar{\varphi} \circ \eta_P = \eta_{P'} \circ \varphi$. This map $\bar{\varphi}$ coincides with the above defined lifted map $\mathcal{D}^\wedge \varphi$. Thus $\mathcal{D}^\wedge$ restricts to a functor from the category $\mathbf{O}\mathcal{D}$ of ordered sets with $\mathcal{D}$-continuous maps as morphisms to the category of algebraic ordered sets with residuated maps preserving compactness. A map between $\perp$-$\vee$-semilattices is $\mathcal{D}$-continuous iff it preserves finite joins (see 6.5). Hence we have the following variant of Theorem 3.4:

3.6 Theorem *The up-completion $\mathcal{D}^\wedge$ induces functorial equivalences between*

(a) *the category* **O**$\mathcal{D}$ *of ordered sets with* $\mathcal{D}$*-continuous maps and the category* $\mathbf{AO}_\nabla$ *of algebraic ordered sets with residuated maps preserving compactness,*
(b) *the category* $\mathbf{JS}_\perp$ *of* $\perp$-$\vee$*-semilattices* ($= \mathcal{F}$*-complete ordered sets*) *with maps preserving finite joins and the category* $\mathbf{AL}_\nabla$ *of algebraic lattices with maps preserving arbitrary joins and compactness.*

Combining this result with Proposition 3.5, we arrive at the following duality which is well known at least for the case of join-semilattices (see [Com]):

3.7 Corollary *The category* **O**$\mathcal{D}$ *is dually equivalent to the category* $\mathbf{AO}^\Delta$ *of algebraic ordered sets with residual maps preserving directed joins, and the full subcategory* $\mathbf{JS}_\perp$ *is dually equivalent to the category* $\mathbf{AL}^\Delta$ *of algebraic lattices with maps preserving directed joins and arbitrary meets.*

On the object level, Theorem 3.6 states that an ordered set Q is algebraic iff it is isomorphic to the up-completion $\mathcal{D}^\wedge P$ of some ordered set P which is uniquely determined, up to isomorphism, by Q: in fact, P must be isomorphic to $\mathcal{K}Q$. On the morphism level, Theorems 3.4 and 3.6 tell us that a map φ between ordered sets is isotone (respectively, $\mathcal{D}$-continuous) iff there exists a unique $\mathcal{D}$-join preserving (respectively, residuated) map $\bar{\varphi}$ sending compact elements to compact elements such that the following diagram commutes:

$$\begin{tikzcd} P \arrow[r, "\varphi"] \arrow[d, "\eta_P"'] & P' \arrow[d, "\eta_{P'}"] \\ \mathcal{D}^\wedge P \arrow[r, "\bar{\varphi}"'] & \mathcal{D}^\wedge P' \end{tikzcd}$$

Next, we make the one-to-one correspondence between algebraic ordered sets and sober B-spaces functorial (see Theorem 2.18). For this purpose, we have to restrict suitably the morphism class on the topological side (it turns out that continuity is not enough). Thus we call a map φ between spaces *core-continuous* if it is continuous and inverse images of cores are cores; the latter condition means that φ is a residual map with respect to the associated specialization orders, while continuity with respect to the Scott topologies means preservation of directed joins. In case of sober B-spaces, this type of morphisms admit a particularly convenient description:

3.8 Lemma *For a map* ψ *between sober B-spaces, the following conditions are equivalent:*
(a) ψ *is core-continuous.*

(b) *Inverse images of open cores under ψ are open cores.*
(c) *ψ is a $\mathcal{D}$-join preserving residual map between the associated algebraic ordered sets.*

Proof The equivalence (a)$\Longleftrightarrow$(c) has been explained before, and (a)$\Longrightarrow$(b) is clear. (b)$\Longrightarrow$(c): Since the open cores form a basis for any B-space, we infer from (b) that ψ is continuous (with respect to the Scott topologies); hence ψ preserves $\mathcal{D}$-joins as a map between the associated algebraic ordered sets Q' and Q. It remains to show that ψ has a lower adjoint $\varphi : Q \to Q'$. For this, observe first that for each compact element $x \in Q$, the principal dual ideal $\uparrow x$ is an open core with respect to the Scott topology of Q, and consequently $\psi^{-1}[\uparrow x]$ is an open core with respect to the Scott topology of Q'; in other words, there exists a unique compact element $x' \in Q'$ such that $\psi^{-1}[\uparrow x] = \uparrow x'$. Apparently, the assignment $x \mapsto x'$ yields an isotone map from $\mathcal{K}Q$ to $\mathcal{K}Q'$. For any $y \in Q$, the set $\mathcal{K}Q \cap \downarrow y$ is directed and has join y. Accordingly, the image $\{x' \mid x \in \mathcal{K}Q \cap \downarrow y\}$ is again directed and has therefore a join in Q'; denoting this join by φy, we obtain a map $\varphi : Q \to Q'$. In order to see that this map is in fact lower adjoint to ψ, consider any element $z \in Q'$. If $\varphi y \leq z$ then $x' \leq z$ for all $x \in \mathcal{K}Q \cap \downarrow y$. As $x' \leq z$ means $z \in \uparrow x' = \psi^{-1}[\uparrow x]$, i.e. $x \leq \psi z$, it follows that $y = \bigvee(\mathcal{K}Q \cap \downarrow y) \leq \psi z$. Conversely, $y \leq \psi z$ implies $x \leq \psi z$, i.e. $x' \leq z$ for all $x \in \mathcal{K}Q \cap \downarrow y$, and then $\varphi y \leq z$. □

In order to translate the order-theoretical preservation of compactness into the language of topology, we use again the fact that an element of an up-complete ordered set is compact iff it generates a Scott-open principal filter. Hence the following definition appears adequate: a map φ between closure spaces $(S, \mathcal{X})$ and $(S', \mathcal{X}')$ is called *quasiopen* iff for each open set U in the first space, the *saturation* $\uparrow\varphi[U]$, i.e., the intersection of all open sets containing the image $\varphi[U]$, is open in the second space. The following order-theoretical characterization of such maps is easily checked:

3.9 Lemma *A map between sober B-spaces is quasiopen iff it is isotone and preserves compactness as a map between the corresponding algebraic ordered sets.*

This together with 3.8 and the Second Representation Theorem 2.18 yields

3.10 Theorem (1) *The category of sober B-spaces and continuous quasiopen maps is isomorphic to the category* **AO** *of algebraic ordered sets and maps preserving directed joins and compactness.*
(2) *The category of sober B-spaces and core-continuous maps is isomorphic to the category* $\mathbf{AO}^{\Delta}$ *of algebraic ordered sets and residual maps preserving directed joins, hence dually isomorphic to the category* $\mathbf{AO}_{\nabla}$ *of algebraic ordered sets and residuated maps preserving compactness.*

At the end of this section, let us formulate a categorical version of the Third Representation Theorem 2.20, providing an equivalence between the category **AO** of algebraic ordered sets and the category $\mathbf{A}_{\nabla}$ of $\mathcal{A}$-lattices (= superalgebraic lattices) and maps preserving joins and supercompactness (where the latter means that $\bigvee$-prime elements are mapped onto $\bigvee$-prime elements). A similar argument as for 3.5 shows that the adjunction functors $\mathcal{U}$ and $\mathcal{L}$ induce a dual isomorphism between $\mathbf{A}_\nabla$ and the category $\mathbf{A}^\Delta$ of $\mathcal{A}$-lattices and complete homomorphisms (see again 4.11). On account of 3.1, we have for any $\mathcal{D}$-join preserving, i.e. Scott-continuous function φ between algebraic ordered sets Q and Q' a residuated, i.e. join-preserving map

$$\mathcal{D}^\vee\varphi = \bar{\varphi} : \mathcal{D}^\vee Q \longrightarrow \mathcal{D}^\vee Q', \quad Y \longmapsto \varphi[Y]^-.$$

Using soberness of the Scott topology on the algebraic ordered set Q, one concludes that the supercompact elements of the $\mathcal{A}$-lattice $\mathcal{D}^\vee Q$ are precisely the principal ideals generated by compact elements of Q. Hence, φ preserves compactness iff $\bar{\varphi}$ preserves supercompactness (recall that point closures are mapped onto point closures). Thus we arrive at our final equivalence theorem for algebraic ordered sets (cf. [La1], [HM2]):

3.11 Theorem *The Scott functor $\mathcal{D}^\vee$ gives rise to an equivalence between the category* **AO** *of algebraic ordered sets with maps preserving directed joins and compactness and the category $\mathbf{A}_\nabla$ of $\mathcal{A}$-lattices with maps preserving arbitrary joins and supercompactness. Hence* **AO** *is dual to the category $\mathbf{A}^\Delta$ of $\mathcal{A}$-lattices and complete homomorphisms.*

4 $\mathcal{Z}$-compactly generated ordered sets and $\mathcal{Z}$-sober spaces

It is now time to develop a general "$\mathcal{Z}$-theory", replacing the subset selection $\mathcal{D}$ by an (almost) arbitrary selection $\mathcal{Z}$. It turns out that a great part of the previous considerations extend, without any restriction, to the general $\mathcal{Z}$-setting. However, from time to time one has to inspect carefully the arguments working for the specific selection $\mathcal{D}$ of directed subsets: for example, at certain points one needs the fact that directed unions or isotone images of directed sets are again directed. To ensure that analogous conclusions work for our general subset selections $\mathcal{Z}$, we shall introduce properties like *union completeness, $\mathcal{Z}$-quasiclosedness* etc. Fortunately, most of the required properties are shared by all subset selections we are interested in. Thus, for example, many of the results of Section 2 and 3 translate from $\mathcal{D}$ to selections like $\mathcal{A}$ (arbitrary lower sets), $\mathcal{E}$ (singletons), or $\mathcal{F}$ (finite subsets).

Recall that an element x of a $\mathcal{Z}$-complete ordered set Q is said to be *$\mathcal{Z}$-compact* or *$\mathcal{Z}$-prime* if for each $Z \in \mathcal{Z}Q$ with $x \leq \bigvee Z$, there is some $y \in Z$ such that $x \leq y$. We denote by $\mathcal{K}_\mathcal{Z} Q$ the set of all $\mathcal{Z}$-compact elements of Q and call it the *$\mathcal{Z}$-spectrum* of Q; of course, this name is motivated by the special case $\mathcal{Z} = \mathcal{F}$ ($\vee$-prime elements!), and not by the case $\mathcal{Z} = \mathcal{D}$ (compact elements!); if no confusion is likely to arise, we simply write P for $\mathcal{K}_\mathcal{Z} Q$, considered as an ordered set with the order induced from Q. If the $\mathcal{Z}$-spectrum is $\bigvee$-dense in Q then we call Q a *$\mathcal{Z}$-compactly generated ordered set.* By a *$\mathcal{Z}$-lattice*, we mean a $\mathcal{Z}$-compactly generated complete lattice. Thus, in accordance with earlier definitions, the $\mathcal{A}$-lattices are the superalgebraic lattices, while the $\mathcal{C}$- resp. $\mathcal{D}$-lattices are the algebraic (= compactly generated complete) lattices. $\mathcal{B}$- resp. $\mathcal{F}$-lattices are sometimes referred to as *spatial coframes*, because they are, up to isomorphism, just the lattices of closed sets of topological spaces (see e.g. [Pa] or [Bü]).

The first result of this section is very easy but basic for the representation theory of $\mathcal{Z}$-compactly generated ordered sets (cf. [E4], [E19]).

4.1 Lemma *The following conditions on a point closure system $\mathcal{X}$ are equivalent:*

(a) *$\mathcal{X}$ is a $\mathcal{Z}$-compactly generated ordered set, and $\mathcal{Z}$-joins in $\mathcal{X}$ are $\mathcal{Z}$-unions.*

(b) *$\mathcal{X}$ is $\mathcal{Z}$-$\bigcup$-complete, i.e. $\mathcal{Y} \in \mathcal{Z}\mathcal{X}$ implies $\bigcup \mathcal{Y} \in \mathcal{X}$.*

(c) *$\mathcal{X}$ is $\mathcal{Z}$-complete, and each point closure is a $\mathcal{Z}$-compact member of $\mathcal{X}$.*

(d) *$\mathcal{X}$ is $\mathcal{Z}$-complete, and each element of $\mathcal{X}$ is a union of $\mathcal{Z}$-compact members of $\mathcal{X}$.*

Proof The implication chain (a) $\Longrightarrow$ (b) $\Longrightarrow$ (c) $\Longrightarrow$ (d) is obvious, and for (d) $\Longrightarrow$ (a), one only has to observe that for $\mathcal{Y} \in \mathcal{Z}\mathcal{X}$, the join $Z = \bigvee \mathcal{Y}$ in $\mathcal{X}$ must be (contained in) the union $\bigcup \mathcal{Y}$, because each $x \in \bigvee \mathcal{Y}$ belongs to some $\mathcal{Z}$-compact $X \subseteq \bigvee \mathcal{Y}$, whence $x \in X \subseteq Y$ for some $Y \in \mathcal{Y}$. □

4.2 Corollary *A closure system $\mathcal{X}$ is $\mathcal{Z}$-$\bigcup$-complete iff each point closure is a $\mathcal{Z}$-compact member of $\mathcal{X}$ iff $\mathcal{X}$ is a $\mathcal{Z}$-lattice in which $\mathcal{Z}$-joins agree with $\mathcal{Z}$-unions.*

We have seen that for the topological representation theory of algebraic ordered sets, the notion of soberness is of particular importance, because sober spaces are determined, up to homeomorphism, by their lattices of open resp. closed sets. When dealing with other types of closure spaces which are not topological, one has to consider a suitable generalization, viz. the notion of *$\mathcal{Z}$-soberness*. We have reported about this fruitful concept on the occasion of several conferences and colloquium talks, e.g. at Bremen 1984, Durban 1985, and L'Aquila 1986. (Independently, a similar general type of sober spaces was studied by B. Banaschewski and G. Bruns in their paper on "*The fundamental duality of partially ordered sets*" [BB]). Motivated by the above remarks, we call an ordinary space $(S, \mathcal{X})$ or its point closure system *$\mathcal{Z}$-sober* if $\mathcal{X}$ is $\mathcal{Z}$-$\bigcup$-complete and each $\mathcal{Z}$-compact member of $\mathcal{X}$ is the closure of a unique point. By 4.1, this is equivalent to saying that $\mathcal{X}$ is T_0 (distinct points have distinct closures), $\mathcal{Z}$-complete, and the $\mathcal{Z}$-compact members of $\mathcal{X}$ are precisely the point closures. In particular, a closure space (resp. its closure system) is $\mathcal{Z}$-sober iff it is T_0 and the $\mathcal{Z}$-compact closed sets are precisely the point closures. Thus the $\mathcal{F}$-sober closure spaces are nothing but the topological sober spaces (closed version). Similarly, it is not hard to see that $\mathcal{A}$-sober ($= \mathcal{P}$-sober) spaces are just the T_0-A-spaces. For $\mathcal{D}$-sober closure spaces, the subsequent lemma yields a somewhat surprising *algebraic* characterization, although the notion of $\mathcal{Z}$-soberness was initially motivated by *topological* questions (for a related remark on $\mathcal{D}$-soberness, see [BB]).

4.3 Lemma *A closure system $\mathcal{X}$ is $\mathcal{D}$-sober iff it is the system of all ideals of a (unique)* *join-semilattice with least element. Similarly, a closure system is $\mathcal{E}$-sober iff it is the system of all principal ideals of a (unique) complete lattice.*

Proof Let S be any join-semilattice with least element. The ideal system $\mathcal{I}S$ is an inductive closure system whose compact members are precisely the principal ideals; indeed, any compact ideal is finitely generated, and in a join-semilattice with least element, finitely generated ideals are principal; in other words, they are the point closures with respect to the system $\mathcal{I}S$.

Conversely, assume $(S, \mathcal{X})$ is an arbitrary $\mathcal{D}$-sober closure space. We consider S as an ordered set with respect to the specialization order $\leq_{\mathcal{X}}$. Then the principal ideals are just the point closures, i.e. the compact members of $\mathcal{X}$, and it follows that $\mathcal{X}$ is $\mathcal{D}$-$\bigcup$-complete. Therefore any directed union of principal ideals, that is, any directed lower set, belongs to $\mathcal{X}$. In order to prove the reverse inclusion $\mathcal{X} \subseteq \mathcal{D}^{\wedge}S$, we only have to observe that $\mathcal{X}$ is an algebraic lattice, so that each member of $\mathcal{X}$ is a directed union of compact elements of $\mathcal{X}$, i.e., of principal ideals. Finally, we know that each finitely generated member of an inductive closure system is compact, and in case of a $\mathcal{D}$-sober system, it must be a principal ideal. But this condition forces S to be a join-semilattice with least element: if the closure of a finite set F is the principal ideal $\downarrow x$ then x must be the join of F. $\mathcal{E}$-soberness is treated analogously. □

The previous lemma is easily generalized to so-called *m-ideals* and *m-compact elements*, where m denotes any cardinal number (cf. [E3], [Gr1], [Ro]): let $\mathcal{P}_m Q$ denote the collection of all subsets of Q with less than m elements; if each $Z \in \mathcal{P}_m Q$ has a join then Q is called *m-complete* (rather than *$\mathcal{P}_m$-complete*). A subset D of Q is *m-directed* if each $Z \in \mathcal{P}_m Q$ has an upper bound in D, and the collection of all m-directed subsets is denoted by $\mathcal{D}_m Q$. An element x of an ordered set Q is called *m-compact* if it belongs to every *m-ideal*, i.e. to every m-directed lower set Y with $x \in \Delta Y$. Notice that a lower subset of an m-complete ordered set is an m-ideal iff it is closed under m-joins, i.e. joins of subsets with less than m elements. A $\mathcal{D}_m$-complete ordered set in which every element is the join of an m-directed set of m-compact elements is called *m-algebraic*.

Now similar arguments as for 4.3 yield:

4.4 Lemma *A closure system is $\mathcal{D}_m$-sober iff it is the system of all m-ideals of a (unique) m-complete ordered set.*

In contrast to this observation, ($\mathcal{F}$-)sober spaces are not determined by their specialization order; for example, every Hausdorff space is sober, and its specialization order is always the identity relation.

We are now in a position to prove a general "$\mathcal{Z}$-version" of the Representation Theorem 2.16 for compactly generated ordered sets. By an *invariant subset selection*, we mean a subset selection $\mathcal{Z}$ such that for any isomorphism φ between ordered sets Q and Q', $Z \in \mathcal{Z}Q$ implies $\varphi[Z] \in \mathcal{Z}Q'$. This property is shared by almost all subset selections occurring in mathematical practice; on the other hand, it is evident that reasonable structural results can be expected only for invariant subset selections.

4.5 Theorem *Let $\mathcal{Z}$ and $\mathcal{Z}'$ be invariant subset selections. Then an ordered set Q is $\mathcal{Z}$-compactly generated and dually $\mathcal{Z}'$-complete iff it is isomorphic to a $\mathcal{Z}$-$\bigcup$-complete (respectively, $\mathcal{Z}$-sober) and $\mathcal{Z}'$-$\bigcap$-complete point closure system.*

Proof By 4.1, any $\mathcal{Z}$-$\bigcup$-complete and $\mathcal{Z}'$-$\bigcap$-complete point closure system is a $\mathcal{Z}$-compactly generated and dually $\mathcal{Z}'$-complete ordered set, and so is each isomorphic copy of it, by the invariance assumption on $\mathcal{Z}$ and $\mathcal{Z}'$.

Conversely, assume there is given an arbitrary $\mathcal{Z}$-compactly generated and dually $\mathcal{Z}'$-complete ordered set Q. Since the set $P = \mathcal{K}_{\mathcal{Z}}Q$ of all $\mathcal{Z}$-compact elements is $\bigvee$-dense in Q, we have an isomorphism

$$\kappa_Q^{\mathcal{Z}} : Q \longrightarrow \mathcal{X}_{\mathcal{Z}}Q = \{P \cap \downarrow y \mid y \in Q\}, \quad y \longmapsto P \cap \downarrow y$$

between Q and the system $\mathcal{X}_{\mathcal{Z}}Q$. The latter is a $\mathcal{Z}$-sober point closure system because $\kappa_Q^{\mathcal{Z}}$ induces a one-to-one correspondence between the $\mathcal{Z}$-compact elements of Q and the principal ideals of P, i.e. the point closures of the ordinary space $(P, \mathcal{X}_{\mathcal{Z}}Q)$. In particular, the system $\mathcal{X}_{\mathcal{Z}}Q$ is $\mathcal{Z}$-$\bigcup$-complete, and the equation

$$\bigcap \kappa_Q^{\mathcal{Z}}[Z] = \kappa_Q^{\mathcal{Z}}(\bigwedge Z) \qquad (Z \in \mathcal{Z}'Q)$$

shows that it is also $\mathcal{Z}'$-$\bigcap$-complete (again, we have to use the invariance of $\mathcal{Z}$ and $\mathcal{Z}'$). □

Taking for $\mathcal{Z}'$ the subset selection $\mathcal{P}$ of all subsets, or, on the other extreme, the selection $\mathcal{E}$ of all singletons, we see that 4.5 includes the following representation theorem (cf. [Bü], [E4]):

4.6 Corollary *The $\mathcal{Z}$-$\bigcup$-complete (respectively, $\mathcal{Z}$-sober) closure systems are, up to isomorphism, just the $\mathcal{Z}$-lattices. Similarly, $\mathcal{Z}$-$\bigcup$-complete ($\mathcal{Z}$-sober) point closure systems represent $\mathcal{Z}$-compactly generated ordered sets.*

The special choice $\mathcal{Z} = \mathcal{D}$ amounts to Theorem 2.16 and its variants for semilattices etc. For example, combining 4.3 with 4.6, we see that the algebraic lattices (i.e., the $\mathcal{D}$-lattices) are, up to isomorphism, the ideal systems of $\bot$-$\vee$-semilattices. Let us have a look at some other relevant specializations.

$\mathcal{Z} = \mathcal{A}$ (or $\mathcal{Z} = \mathcal{P}$): Recall that an $\mathcal{A}$-compact element is usually called *supercompact* or *completely join-prime*. By 4.6, *the $\mathcal{A}$-lattices (= superalgebraic lattices) are, up to isomorphism, precisely the T_0-A-topologies.*

$\mathcal{Z} = \mathcal{B}$ (or $\mathcal{Z} = \mathcal{F}$): By 4.6, *the $\mathcal{F}$-lattices, i.e. the spatial coframes are, up to isomorphism, just the (sober) topological closure systems.* While this result belongs to the folklore of lattice-theoretical topology, its non-complete versions are not so common: for example, taking $\mathcal{Z} = \mathcal{Z}' = \mathcal{B}$, we infer from 4.5 that *the $\vee$-primely generated lattices are,*

up to isomorphism, those point closure systems which are set lattices , i.e. closed under binary unions and intersections. Such lattices are always distributive and play a considerable role in the spectral theory of various algebraic structures. However, not every distributive lattice is $\vee$-primely generated (counterexamples: atomless Boolean algebras), although by the Prime Ideal Theorem (PIT), every distributive lattice is isomorphic to some set lattice. Notice that in contrast to the "constructive" Representation Theorem 4.5, PIT requires some kind of choice principle weaker than AC (see e.g. [Ha]), and it is also needed for the proof of (indeed, is equivalent to) the fact that every algebraic distributive lattice is isomorphic to the topology of a *Stone space*, i.e., a sober space with a basis of compact open sets (cf. [BD], [Gr2], [Com]): for this, one must know that each element of a distributive algebraic lattice is a meet of $\wedge$-primes, and then the dual of 4.5 applies with $\mathcal{Z} = \mathcal{F}$. The aforementioned meet-decomposition property is an immediate consequence of 2.15 and the remark that $\bigwedge$-irreducible elements in distributive lattices are $\wedge$-prime; however, 2.15 requires the full strength of AC or some equivalent principle like Zorn's Lemma, while the existence of meet-decompositions into $\wedge$-prime elements follows from the logically weaker PIT and the representation of algebraic lattices as ideal lattices. We shall return to these basic ingredients of the famous *Stone duality* at a later point.

Next, we are pursuing the trace to a $\mathcal{Z}$-theorem generalizing the equivalence between ordered sets with $\bigvee$-generators of compact elements and ordinary spaces with inductive point closure systems. Our approach via the notion of $\mathcal{Z}$-soberness also enables us to derive an equivalence theorem for $\mathcal{Z}$-compactly generated ordered sets, without mentioning specific $\bigvee$-generators.

Denote by $\mathcal{Z}\mathbf{OS}_0$ that full subcategory of $\mathbf{OS}_0$, the category of ordinary T_0 spaces and strongly continuous maps, whose objects have $\mathcal{Z}$-union complete point closure systems. Similarly, let $\mathcal{Z}\mathbf{OG}_\nabla$ denote that full subcategory of $\mathbf{OG}_\nabla$, the category of ordered sets with $\bigvee$-generators, whose objects are pairs (Q, G) such that the $\bigvee$-generator G consists of $\mathcal{Z}$-compact elements only (cf. Section 3). The full subcategories $\mathcal{Z}\mathbf{CS}_0$ (T_0 closure spaces with $\mathcal{Z}$-$\bigcup$-complete closure systems) and $\mathcal{Z}\mathbf{CG}_\nabla$ (complete lattices with $\bigvee$-generators of $\mathcal{Z}$-compact elements) are defined analogously. Then the same arguments as for the special case $\mathcal{Z} = \mathcal{D}$ (see 3.3 and [E10]) lead to the following:

4.7 Theorem *The generic functor $\mathcal{G}$ and the hull-kernel functor $\mathcal{H}$ induce equivalences between the categories $\mathcal{Z}\mathbf{OS}_0$ and $\mathcal{Z}\mathbf{OG}_\nabla$ (resp. $\mathcal{Z}\mathbf{CS}_0$ and $\mathcal{Z}\mathbf{CG}_\nabla$).*

These equivalences become a bit more succinct when restricted to $\mathcal{Z}$-sober spaces. Let $\mathcal{Z}\mathbf{SOS}$ (resp. $\mathcal{Z}\mathbf{SCS}$) denote the category of $\mathcal{Z}$-sober ordinary spaces (resp. closure spaces) and strongly continuous maps. On the other hand, denote by $\mathcal{Z}\mathbf{CGO}_\nabla$ (resp. $\mathcal{Z}\mathbf{CGC}_\nabla$) the category of $\mathcal{Z}$-compactly generated ordered sets (resp. $\mathcal{Z}$-lattices) and residuated maps preserving $\mathcal{Z}$-compactness, i.e., sending $\mathcal{Z}$-compact elements to $\mathcal{Z}$-compact elements. The categories $\mathcal{Z}\mathbf{CGO}_\nabla$ and $\mathcal{Z}\mathbf{CGC}_\nabla$ may be regarded as full subcategories of $\mathcal{Z}\mathbf{OG}_\nabla$, by identifying a $\mathcal{Z}$-compactly generated ordered set Q with the $\mathcal{Z}\mathbf{OG}_\nabla$-object $(Q, \mathcal{K}_\mathcal{Z}Q)$. Bearing this identification in mind, we have a *generic functor*

$$\mathcal{G}_\mathcal{Z} : \mathcal{Z}\mathbf{SOS} \longrightarrow \mathcal{Z}\mathbf{CGO}_\nabla$$

assigning to each $\mathcal{Z}$-sober space $(S, \mathcal{X})$ the point closure system $\mathcal{X}$, considered as an ordered set with respect to inclusion, and to each morphism $\varphi : (S, \mathcal{X}) \to (S', \mathcal{X}')$ the residuated map $\bar{\varphi} : \mathcal{X} \to \mathcal{X}'$. In the converse direction, the modified *hull-kernel functor*

$$\mathcal{H}_{\mathcal{Z}} : \mathcal{Z}\mathbf{CGO}_{\nabla} \longrightarrow \mathcal{Z}\mathbf{SOS}$$

assigns to each $\mathcal{Z}$-compactly generated ordered set Q the ordinary space

$$\mathcal{H}_{\mathcal{Z}}Q = (\mathcal{K}_{\mathcal{Z}}Q, \mathcal{X}_{\mathcal{Z}}Q).$$

(cf. Section 3 and the proof of 4.5). Indeed, the point closure system

$$\mathcal{X}_{\mathcal{Z}}Q = \{\mathcal{K}_{\mathcal{Z}}Q \cap {\downarrow} y \mid y \in Q\}$$

is $\mathcal{Z}$-sober since the isomorphism

$$\kappa_Q^{\mathcal{Z}} : Q \longrightarrow \mathcal{X}_{\mathcal{Z}}Q, \quad y \longmapsto \mathcal{K}_{\mathcal{Z}}Q \cap {\downarrow} y$$

induces a one-to-one correspondence between the $\mathcal{Z}$-compact elements of Q and the principal ideals of $\mathcal{K}_{\mathcal{Z}}Q$, i.e., the point closures of the space $\mathcal{H}_{\mathcal{Z}}Q$.

Thus our general Equivalence Theorem 3.2 for ordinary spaces and ordered sets with $\bigvee$-generators amounts to

4.8 Theorem *The categories* $\mathcal{Z}\mathbf{SOS}$ *and* $\mathcal{Z}\mathbf{CGO}_{\nabla}$ *(resp.* $\mathcal{Z}\mathbf{SCS}$ *and* $\mathcal{Z}\mathbf{CGC}_{\nabla}$*) are equivalent via the functors* $\mathcal{G}_{\mathcal{Z}}$ *and* $\mathcal{H}_{\mathcal{Z}}$.

In order to turn this equivalence into a duality, we simply have to apply the adjunction functors $\mathcal{U}$ and $\mathcal{L}$ (see Section 3): the category $\mathcal{Z}\mathbf{CGO}_{\nabla}$ is dually isomorphic to the category $\mathcal{Z}\mathbf{CGO}^{\Delta}$ with the same objects ($\mathcal{Z}$-compactly generated ordered sets) but residual maps whose lower adjoints preserve $\mathcal{Z}$-compactness as morphisms. Again, the objects of the full subcategory $\mathcal{Z}\mathbf{CGC}^{\Delta}$ are the $\mathcal{Z}$-lattices.

4.9 Corollary *The categories* $\mathcal{Z}\mathbf{SOS}$ *and* $\mathcal{Z}\mathbf{CGO}^{\Delta}$ *(resp.* $\mathcal{Z}\mathbf{SCS}$ *and* $\mathcal{Z}\mathbf{CGC}^{\Delta}$*) are dually equivalent via the functors* $\mathcal{U} \circ \mathcal{G}_{\mathcal{Z}}$ *and* $\mathcal{H}_{\mathcal{Z}} \circ \mathcal{L}$.

For many subset selections, the morphisms of the categories $\mathcal{Z}\mathbf{CGO}^{\Delta}$ and $\mathcal{Z}\mathbf{CGC}^{\Delta}$ admit a more handy direct description. For this and many other purposes, it is convenient to associate with each subset selection $\mathcal{Z}$ the *$\mathcal{Z}$-ideal extension* $\mathcal{Z}^{\wedge}$ (cf. [E15], [Me]), where

$$\mathcal{Z}^{\wedge}P = \{{\downarrow} Z \mid Z \in \mathcal{E}P \cup \mathcal{Z}P\}$$

is the collection of all *$\mathcal{Z}$-ideals* of P; these are either principal ideals or lower sets generated by members of $\mathcal{Z}P$. For example, we have $\mathcal{A} = \mathcal{P}^{\wedge}$, $\mathcal{M} = \mathcal{E}^{\wedge}$, and $\mathcal{Z}^{\wedge\wedge} = \mathcal{Z}^{\wedge}$ for arbitrary subset selections $\mathcal{Z}$. Now we borrow some well-worn definitions from topology and call a map φ between ordered sets P and P'

$$\textit{$\mathcal{Z}$-continuous} \text{ if } \varphi^{-1}[Z'] \in \mathcal{Z}^{\wedge}P \text{ for all } Z' \in \mathcal{Z}^{\wedge}P',$$

$$\mathcal{Z}\text{-}\textit{quasiclosed} \text{ if } \downarrow\varphi[Z'] \in \mathcal{Z}^\wedge P' \text{ for all } Z \in \mathcal{Z}^\wedge P,$$

$$\mathcal{Z}\text{-}\textit{quasiopen} \text{ if } P' \setminus \uparrow\varphi[P \setminus Z] \in \mathcal{Z}^\wedge P' \text{ for all } Z \in \mathcal{Z}^\wedge P$$

(cf. Section 3). The following remark is very helpful in the present context (cf. [E19]):

4.10 Lemma *Let $\varphi : P \to P'$ be the lower adjoint of $\psi : P' \to P$. Then φ is $\mathcal{Z}$-continuous iff ψ is $\mathcal{Z}$-quasiclosed, and ψ is $\mathcal{Z}$-continuous iff φ is $\mathcal{Z}$-quasiopen.*

For the proof, one has to observe the identities $\varphi^{-1}[\downarrow Z'] = \downarrow\psi[Z']$ for $Z' \subseteq P'$ and $\psi^{-1}[\uparrow Y] = \uparrow\varphi[Y]$ for $Y \subseteq P$. In case of maps between $\mathcal{Z}$-compactly generated ordered sets, this result is supplemented by the following:

4.11 Proposition *A residual map between $\mathcal{Z}$-compactly generated ordered sets is $\mathcal{Z}$-quasiclosed and preserves $\mathcal{Z}$-joins iff its lower adjoint is $\mathcal{Z}$-continuous and preserves $\mathcal{Z}$-compactness.*

Proof Suppose $\varphi : Q \to Q'$ is the lower adjoint of a map $\psi : Q' \to Q$, and φ preserves $\mathcal{Z}$-compactness. Take $Z' \in \mathcal{Z}Q'$ and consider any $\mathcal{Z}$-compact element $x \in Q$ with $x \leq \psi(\bigvee Z')$. As φx is $\mathcal{Z}$-compact and satisfies $\varphi x \leq \bigvee Z'$, it follows that $\varphi x \leq y'$ for some $y' \in Z'$ and then $x \leq \psi y' \leq \bigvee \psi[Z']$. Since $\psi(\bigvee Z')$ is a join of $\mathcal{Z}$-compact elements, we conclude that $\psi(\bigvee Z') \leq \bigvee \psi[Z']$, and the other inequality is clear because ψ is isotone. Thus ψ preserves $\mathcal{Z}$-joins.

Conversely, assume that ψ is a $\mathcal{Z}$-quasiclosed map preserving $\mathcal{Z}$-joins, and let x be a $\mathcal{Z}$-compact element of Q. For $Z' \in \mathcal{Z}Q'$, we find a $Z \in \mathcal{Z}Q$ with $\downarrow Z = \downarrow\psi[Z']$. Now $\varphi x \leq \bigvee Z'$ implies $x \leq \psi(\bigvee Z') = \bigvee \psi[Z'] = \bigvee Z$, and by $\mathcal{Z}$-compactness of x, it follows that $x \in \downarrow Z = \downarrow\psi[Z']$; as φ is lower adjoint to ψ, this means $\varphi x \in \downarrow Z'$. Hence φx is $\mathcal{Z}$-compact, too. □

Special instances of this useful result have been mentioned in [GG] and [Com]. The general $\mathcal{Z}$-situation was treated in [BE1]. Now let us call a subset selection $\mathcal{Z}$ $\mathbf{O}^\Delta$-*invariant* if every residual map (i.e., every $\mathbf{O}^\Delta$-morphism) is $\mathcal{Z}$-quasiclosed; by 4.10, this is equivalent to saying that every residuated map is $\mathcal{Z}$-continuous. We shall come back to this and related invariance properties in Section 6. For the moment, it suffices to mention that many subset selections like $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}, \mathcal{F}, \mathcal{P}$ and $\mathcal{W}$, but also the MacNeille completion $\mathcal{N}$, the Frink ideal completion $\mathcal{I}$, and the Scott completion $\mathcal{D}^\vee$, are all $\mathbf{O}^\Delta$-invariant. As an immediate consequence of 4.11, we can now give a simplified characterization of $\mathcal{Z}\mathbf{CGO}^\Delta$-morphisms in case of "good" subset selections.

4.12 Corollary *Let $\mathcal{Z}$ be an $\mathbf{O}^\Delta$-invariant subset selection. Then the morphisms of the category $\mathcal{Z}\mathbf{CGO}^\Delta$ are the $\mathcal{Z}$-join preserving residual maps between $\mathcal{Z}$-compactly generated ordered sets; and the morphisms of the category $\mathcal{Z}\mathbf{CGC}^\Delta$ are those maps between $\mathcal{Z}$-lattices which preserve $\mathcal{Z}$-joins and arbitrary meets.*

Let us pause for a moment with the general $\mathcal{Z}$-theory and have a look at some special subset selections. For $\mathcal{Z} = \mathcal{A}$ resp. $\mathcal{Z} = \mathcal{P}$, 4.8 (combined with 4.12) states that the category of T_0-A-spaces and continuous maps is equivalent to the category of superalgebraic lattices ($= \mathcal{A}$-lattices) and maps preserving arbitrary joins and supercompactness, hence dual to the category of superalgebraic lattices and complete homomorphisms (preserving arbitrary joins and meets).

For $\mathcal{Z} = \mathcal{B}_\perp$ and for $\mathcal{Z} = \mathcal{F}$, we get an equivalence between the category of sober spaces and the category of spatial coframes ($= \mathcal{F}$-lattices) with maps preserving arbitrary joins and $\vee$-primes, which in turn is dual to the category of spatial coframes and coframe homomorphisms, i.e. maps preserving finite joins and arbitrary meets. Passing to the "open version", this result is converted into the better known equivalence between sober spaces and *spatial locales*, respectively, the duality between sober spaces and *spatial frames.*

For $\mathcal{Z} = \mathcal{C}$ and for $\mathcal{Z} = \mathcal{D}$, 4.8 in connection with 4.3 yields the "classical" equivalence between the category $\mathbf{JS}_\perp$ of $\perp$-$\vee$-semilattices and the category $\mathbf{AL}_\nabla$ of algebraic lattices, which is dual to the category $\mathbf{AL}^\Delta$ (see 3.6, 3.7 and 4.12).

More generally, let m be an arbitrary cardinal number. From 4.4 we know that the $\mathcal{D}_m$-sober closure spaces are just the m-complete ordered sets, equipped with the system $\mathcal{D}_m{}^\wedge P$ of m-ideals. Therefore, the *specialization functor* $\mathcal{Q}$ which assigns to each ordinary space $(S, \mathcal{X})$ the quasiordered set $(S, \leq_\mathcal{X})$ and acts identically on strongly continuous maps between ordinary spaces, induces a functorial isomorphism

$$\mathcal{Q} : \mathcal{D}_m\mathbf{SCS} \longrightarrow \mathcal{P}_m\mathbf{CO}\mathcal{D}_m$$

between the category of $\mathcal{D}_m$-sober closure spaces (with continuous maps) and the category of m-complete ordered sets and $\mathcal{D}_m$-continuous maps, i.e. isotone maps with the property that inverse images of m-ideals are m-ideals; but this simply means that m-joins, i.e. joins of sets with less than m elements, are preserved (for a more general result, see 6.5). Thus Theorem 4.8 also provides the categorical framework for a known representation theorem on *m-algebraic lattices*, i.e. m-compactly generated complete lattices (see [E3], [Gr1], [Ro]):

4.13 Corollary *The category $\mathcal{D}_m\mathbf{SCS}$ of $\mathcal{D}_m$-sober closure spaces is isomorphic to the category $\mathcal{P}_m\mathbf{CO}\mathcal{D}_m$ of m-complete ordered sets with maps preserving m-joins. The m-ideal functor $\mathcal{D}_m{}^\wedge$ induces an equivalence between $\mathcal{P}_m\mathbf{CO}\mathcal{D}_m$ and the category $\mathcal{D}_m\mathbf{CGC}_\nabla$ of m-algebraic lattices with maps preserving arbitrary joins and m-compactness, which in turn is dually isomorphic to the category $\mathcal{D}_m\mathbf{CGC}^\Delta$ of m-algebraic lattices with maps preserving arbitrary meets and $\mathcal{D}_m$-joins.*

Notice that the equivalence between $\mathcal{P}_m\mathbf{CO}\mathcal{D}_m$ and $\mathcal{D}_m\mathbf{CGC}_\nabla$ extends to an equivalence between the larger category $\mathcal{P}_m\mathbf{CO}$ of m-complete ordered sets with isotone maps and the larger category $\mathcal{D}_m\mathbf{CGC}$ of m-algebraic lattices with maps preserving $\mathcal{D}_m$-joins and m-compactness. Indeed, for any isotone map φ between m-complete ordered sets P and P', the lifted map

$$\mathcal{D}_m{}^\wedge\varphi = \bar{\varphi} : \mathcal{D}_m{}^\wedge P \longrightarrow \mathcal{D}_m{}^\wedge P', \quad Y \longmapsto {\downarrow}\varphi[Y]$$

preserves $\mathcal{D}_m$-joins (= $\mathcal{D}_m$-unions) and maps m-compact elements (= principal ideals) to m-compact elements. However, the category $\mathcal{D}_m\mathbf{CGC}$ does not possess a "natural dual" as it exists for the subcategory $\mathcal{D}_m\mathbf{CGC}_\nabla$.

The specialization functor $\mathcal{Q}$ induces an isomorphism between the category $\mathcal{D}_m\mathbf{SCs}$ of $\mathcal{D}_m$-sober closure spaces with specialization-preserving maps and the category $\mathcal{P}_m\mathbf{CO}$. In all, we obtain the following commutative diagram of equivalences ($\simeq$) and dualities ($\simeq^*$):

$$\begin{tikzcd}
\boxed{\mathcal{D}_m\mathbf{CGC}} \arrow[rrr, equal, "="] \arrow[ddd, shift right, "\mathcal{H}_{\mathcal{D}_m}"'] \arrow[ddd, phantom, "\simeq"] & & & \boxed{\mathcal{D}_m\mathbf{CGC}} \arrow[ddd, shift right] \arrow[ddd, phantom, "\simeq"] \\
& \boxed{\mathcal{D}_m\mathbf{CGC}_\nabla} \arrow[ul, hook] \arrow[r, shift left, "\mathcal{U}"] \arrow[r, phantom, "\simeq^*"] \arrow[d, shift right] \arrow[d, phantom, "\simeq"] \arrow[dr, shift right, "\mathcal{K}_{\mathcal{D}_m}"'] \arrow[dr, phantom, "\simeq"] & \boxed{\mathcal{D}_m\mathbf{CGC}^\Delta} \arrow[l, shift left, "\mathcal{L}"] \arrow[d, shift left] \arrow[d, phantom, "\simeq^*"] & \\
& \boxed{\mathcal{D}_m\mathbf{SCS}} \arrow[dl, hook] \arrow[u, shift right] \arrow[r, shift left] \arrow[r, phantom, "\simeq"] & \boxed{\mathcal{P}_m\mathbf{CO}\,\mathcal{D}_m} \arrow[dr, hook] \arrow[l, shift left] \arrow[u, shift left] \arrow[ul, shift right, "\mathcal{D}_m^\wedge"'] & \\
\boxed{\mathcal{D}_m\mathbf{SCs}} \arrow[uuu, shift right, "\mathcal{G}_{\mathcal{D}_m}"'] \arrow[rrr, shift left, "Q"] \arrow[rrr, phantom, "\simeq"] & & & \boxed{\mathcal{P}_m\mathbf{CO}} \arrow[lll, shift left, "Q^{-1}"] \arrow[uuu, shift right, "\mathcal{D}_m^\wedge"']
\end{tikzcd}$$

5 A symmetric generalization of the Stone duality

We have now all tools in hand for a nice "symmetric" duality between certain categories of $\mathcal{Z}$-sober spaces, generalizing, among many other dualities, the classical Stone duality between Boolean lattices and Boolean spaces, respectively, between bounded distributive lattices and Stone spaces whose basis of compact open sets is closed under finite intersections.

Let $\mathcal{Z}$ and $\mathcal{Z}'$ be $\mathbf{O}^\Delta$-invariant subset selections and denote by $\mathcal{Z}\mathbf{SCS}\mathcal{Z}'$ the following category: objects are the $\mathcal{Z}$-sober closure spaces with a basis of $\mathcal{Z}'$-compact *open* sets (thus the open sets, i.e. the complements of closed sets, form a $\mathcal{Z}'$-lattice), and morphisms are so-called *$\mathcal{Z}'$-proper* maps, i.e. maps between these spaces such that inverse images of $\mathcal{Z}'$-compact open sets are $\mathcal{Z}'$-compact and open. Of course, such morphisms are always continuous. In the classical Stone duality, $\mathcal{D}$-proper maps between Stone spaces are simply called *proper*. As the $\mathcal{A}$-compact (= $\mathcal{P}$-compact) open sets are precisely the open cores, $\mathcal{A}$-proper maps are also called *core-continuous* (see 3.8). $\mathcal{E}$-properness is simply continuity.

On the lattice-theoretical side, consider the category $\mathcal{Z}\mathbf{CGC}\mathcal{Z}'$ of $\mathcal{Z}$-compactly and dually $\mathcal{Z}'$-compactly generated complete lattices, together with maps preserving arbitrary joins, $\mathcal{Z}'$-meets, and $\mathcal{Z}$-compactness, as morphisms. By 4.12 and its dual, this class of morphisms is closed under composition, and we have a duality functor $\mathcal{U}^* : \mathcal{Z}\mathbf{CGC}\mathcal{Z}' \to \mathcal{Z}'\mathbf{CGC}\mathcal{Z}$, assigning to each $\mathcal{Z}\mathbf{CGC}\mathcal{Z}'$-object L the dual lattice L^* and to each $\mathcal{Z}\mathbf{CGC}\mathcal{Z}'$-morphism $\varphi : L \to L'$ its upper adjoint $\mathcal{U}\varphi$, regarded as a morphism between L'^* and L^*. Furthermore, from 4.8, we know that the functors $\mathcal{G}_\mathcal{Z}$ and $\mathcal{H}_\mathcal{Z}$ induce an equivalence between the categories $\mathcal{Z}\mathbf{SCS}\mathcal{Z}'$ and $\mathcal{Z}\mathbf{CGC}\mathcal{Z}'$. Putting all pieces together, we arrive at the following

Generalized Stone Duality:

5.1 Theorem *Let $\mathcal{Z}$ and $\mathcal{Z}'$ be $\mathbf{O}^\triangle$-invariant subset selections. Then one has the following commutative diagram of equivalences ($\simeq$) and dualities ($\simeq^*$), respectively:*

$$
\begin{tikzcd}[column sep=huge, row sep=huge]
\mathcal{Z}\mathbf{CGC}\mathcal{Z}' \arrow[r, shift left, "\mathcal{U}^*"] \arrow[r, phantom, "\simeq^*"] \arrow[d, shift right, "\mathcal{H}_\mathcal{Z}"'] \arrow[d, phantom, "\simeq"] & \mathcal{Z}'\mathbf{CGC}\mathcal{Z} \arrow[l, shift left, "\mathcal{U}^*"] \arrow[d, shift left, "\mathcal{H}_{\mathcal{Z}'}"] \arrow[d, phantom, "\simeq"] \\
\mathcal{Z}\mathbf{SCS}\mathcal{Z}' \arrow[u, shift right, "\mathcal{G}_\mathcal{Z}"'] \arrow[r, shift left, "\mathcal{D}_{\mathcal{Z},\mathcal{Z}'}"] \arrow[r, phantom, "\simeq^*"] & \mathcal{Z}'\mathbf{SCS}\mathcal{Z} \arrow[u, shift left, "\mathcal{G}_{\mathcal{Z}'}"] \arrow[l, shift left, "\mathcal{D}_{\mathcal{Z}',\mathcal{Z}}"]
\end{tikzcd}
$$

The duality functor

$$\mathcal{D}_{\mathcal{Z},\mathcal{Z}'} = \mathcal{H}_{\mathcal{Z}'} \circ \mathcal{U}^* \circ \mathcal{G}_\mathcal{Z}$$

assigns to each $\mathcal{Z}\mathbf{SCS}\mathcal{Z}'$-object $(S, \mathcal{X})$ the $\mathcal{Z}'$-spectrum of $\mathcal{X}^*$, that is, the set $\mathcal{K}_{\mathcal{Z}'}\mathcal{X}^*$ of all dually $\mathcal{Z}'$-compact members of $\mathcal{X}$, equipped with the closure system

$$\mathcal{X}_{\mathcal{Z}'}\mathcal{X}^* = \{\{Z \in \mathcal{K}_{\mathcal{Z}'}\mathcal{X}^* \mid X \subseteq Z\} \mid X \in \mathcal{X}\},$$

and to any $\mathcal{Z}\mathbf{SCS}\mathcal{Z}'$-morphism $\varphi : (S, \mathcal{X}) \to (S', \mathcal{X}')$ the $\mathcal{Z}'\mathbf{SCS}\mathcal{Z}$-morphism

$$\varphi^{-1} : \mathcal{K}_{\mathcal{Z}'}\mathcal{X}'^* \to \mathcal{K}_{\mathcal{Z}'}\mathcal{X}^*.$$

We should now take our time to pick a basket of fruits from the ramified tree growing out of the General Stone Duality (GSD), by considering special choices of the subset selections $\mathcal{Z}$ and $\mathcal{Z}'$.

Let us start with the minimal choice $\mathcal{Z}' = \mathcal{E}$. In this situation, $\mathcal{Z}\mathbf{SCS}\mathcal{E}$ is simply the category of $\mathcal{Z}$-sober closure spaces (and continuous maps), while $\mathcal{E}\mathbf{SCS}\mathcal{Z}$ is the category of $\mathcal{E}$-sober spaces with bases of $\mathcal{Z}$-compact open sets and continuous maps such that these bases are preserved under inverse images. But we know from 4.3 that the $\mathcal{E}$-sober spaces are merely the complete lattices (endowed with the system of principal ideals), and that the continuous maps between them are the residuated ones, i.e. those which preserve arbitrary joins. An open set, that is, the complement of a principal ideal $\downarrow x$, is $\mathcal{Z}$-compact iff x is dually $\mathcal{Z}$-compact in the underlying complete lattice L. Hence the $\mathcal{Z}$-compact open sets form a basis for the open sets iff each element of L is a meet of dually $\mathcal{Z}$-compact elements; in other words, iff L is a dual $\mathcal{Z}$-lattice. Furthermore, a residuated map φ between dual $\mathcal{Z}$-lattices has the property that inverse images of $\mathcal{Z}$-compact open sets are $\mathcal{Z}$-compact iff its upper adjoint preserves dual $\mathcal{Z}$-compactness. Hence, the duality functor $\mathcal{U}^*$ yields a dual isomorphism between $\mathcal{E}\mathbf{SCS}\mathcal{Z}$ and the category $\mathcal{Z}\mathbf{CGC}_\nabla$ of $\mathcal{Z}$-lattices and residuated maps preserving $\mathcal{Z}$-compactness, whence by 4.12, $\mathcal{E}\mathbf{SCS}\mathcal{Z}$ is isomorphic to the category $\mathcal{Z}\mathbf{CGC}^\Delta$ of $\mathcal{Z}$-lattices and residual maps preserving $\mathcal{Z}$-joins (our general hypothesis is that $\mathcal{Z}$ is $\mathbf{O}^\Delta$-invariant).

The role of the categories $\mathcal{Z}\mathbf{CGC}\mathcal{E}$ and $\mathcal{E}\mathbf{CGC}\mathcal{Z}$ is obvious in this context: $\mathcal{Z}\mathbf{CGC}\mathcal{E}$ is just the aforementioned category $\mathcal{Z}\mathbf{CGC}_\nabla$, while $\mathcal{E}\mathbf{CGC}\mathcal{Z}$ is the category of dual $\mathcal{Z}$-lattices and residuated maps preserving $\mathcal{Z}$-meets; hence this category is isomorphic to the category $\mathcal{Z}\mathbf{CGC}^\Delta$: in fact, it is obtained from the latter by dualizing the order relations on the objects but keeping fixed the underlying maps of the morphisms. In all, we see that our GSD includes the Equivalence Theorem 4.8 and the associated duality given by 4.12. In particular, the equivalence between sober spaces and spatial locales, respectively, the duality between sober spaces and spatial frames, is a special instance of GSD.

Next, let us consider the other extreme, where $\mathcal{Z}'$ is the Alexandroff completion $\mathcal{A}$ (or the power set selection $\mathcal{P}$):

$\mathcal{Z}\mathbf{SCS}\mathcal{A}$ is the category of $\mathcal{Z}$-sober closure spaces with a basis of open cores (recall that these are just the supercompact open sets!) In [E19], such spaces have been baptized $\mathcal{Z}$-sober *basic* spaces; if $\mathcal{Z}$ includes the subset selection $\mathcal{F}$ then these are precisely the $\mathcal{Z}$-sober B-spaces (see Section 2). Morphisms in the category $\mathcal{Z}\mathbf{SCS}\mathcal{A}$ are the core-continuous maps (inverse images of open cores are open cores). On the other hand, $\mathcal{A}\mathbf{SCS}\mathcal{Z}$ is the category of $\mathrm{T_0}$-A-spaces with $\mathcal{Z}$-proper maps. Notice that any A-space has a basis of $\mathcal{Z}$-compact open sets, namely of open cores.

The lattice-theoretical counterparts of these categories are described as follows: $\mathcal{Z}\mathbf{CGC}\mathcal{A}$ is the category of superalgebraic lattices (= $\mathcal{A}$-lattices) and complete homomorphisms preserving $\mathcal{Z}$-compactness, while $\mathcal{A}\mathbf{CGC}\mathcal{Z}$ is the category of superalgebraic lattices and maps preserving arbitrary joins, $\mathcal{Z}$-meets and supercompactness. By GSD, *the category $\mathcal{Z}\mathbf{SCS}\mathcal{A}$ is equivalent to $\mathcal{Z}\mathbf{CGC}\mathcal{A}$ and dual to the categories $\mathcal{A}\mathbf{SCS}\mathcal{Z}$ and $\mathcal{A}\mathbf{CGC}\mathcal{Z}$.* Moreover, the category $\mathcal{A}\mathbf{SCS}\mathcal{Z}$ admits a purely order-theoretical interpretation, since the $\mathrm{T_0}$-A-spaces are nothing but the ordered sets, endowed with the closure system of lower sets; thus the open sets are just the upper sets. Modulo this identification, the morphisms of the category $\mathcal{A}\mathbf{SCS}\mathcal{Z}$ are those isotone maps between ordered sets which have the property that inverse images of $\mathcal{Z}$-compact upper sets are again $\mathcal{Z}$-compact. For the specific choice $\mathcal{Z} = \mathcal{P}_m$ (m any cardinal number), an upper set is $\mathcal{Z}$-compact iff it is m-filtered (i.e. m-directed with respect to the dual order; see [E18]). Thus $\mathcal{A}\mathbf{SCS}\mathcal{P}_m$ may be regarded as the category of ordered sets and isotone maps such that inverse images of m-filtered upper sets are m-filtered. Passing to the dual objects, we see that the category $\mathbf{O}\mathcal{D}_m$ of ordered sets and $\mathcal{D}_m$-continuous maps is isomorphic to the category $\mathcal{A}\mathbf{SCS}\mathcal{P}_m$ of T_0-A-spaces and $\mathcal{P}_m$-proper maps.

On the other hand, a similar reasoning as for 3.6 shows that the m-ideal functor $\mathcal{D}_m{}^\wedge$ induces an equivalence between the category $\mathbf{O}\mathcal{D}_m$ and the category $\mathcal{D}_m\mathbf{AO}_\nabla$ of m-algebraic ordered sets and residuated maps preserving m-compactness, which is dually isomorphic to the category $\mathcal{D}_m\mathbf{AO}^\Delta$ of m-algebraic ordered sets and residual maps preserving $\mathcal{D}_m$-joins (compare the diagram at the end of Section 4!)

In all, we have collected together the following equivalences and dualities:

5.2 Corollary *For any cardinal number m, the category $\mathbf{O}\mathcal{D}_m$ of ordered sets and $\mathcal{D}_m$-continuous maps is equivalent to each of the following categories:*

$\mathcal{D}_m\mathbf{AO}_\nabla$:	*m-algebraic ordered sets and residuated maps preserving m-compactness,*
$\mathcal{A}\mathbf{SCS}\mathcal{P}_m$:	*T_0-A-spaces and $\mathcal{P}_m$-proper maps,*
$\mathcal{A}\mathbf{CGC}\mathcal{P}_m$:	*$\mathcal{A}$-lattices and maps preserving joins, $\mathcal{P}_m$-meets and supercompactness,*

and dual to each of the following categories:

$\mathcal{D}_m\mathbf{AO}^\Delta$:	*m-algebraic ordered sets and residual maps preserving $\mathcal{D}_m$-joins,*
$\mathcal{P}_m\mathbf{SCS}\mathcal{A}$:	*$\mathcal{P}_m$-sober basic closure spaces and core-continuous maps,*
$\mathcal{P}_m\mathbf{CGC}\mathcal{A}$:	*$\mathcal{A}$-lattices and complete homomorphisms preserving $\mathcal{P}_m$-primes.*

$$\begin{tikzcd}
\boxed{\mathcal{A}\mathbf{CGC}\mathcal{P}_m} \arrow[rrrr, dash, "\cong^*"] \arrow[dd, dash, "\cong"'] \arrow[dr, dash, "\cong"] \arrow[drr, dash, "\cong"] & & & & \boxed{\mathcal{P}_m\mathbf{CGC}\mathcal{A}} \arrow[dd, dash, "\cong"] \arrow[dll, dash, "\cong^*"'] \arrow[dl, dash, "\cong"] \\
& \boxed{\mathcal{D}_m\mathbf{AO}_\nabla} \arrow[r, dash, "\cong"] & \boxed{\mathbf{O}\mathcal{D}_m} \arrow[r, dash, "\cong^*"] & \boxed{\mathcal{D}_m\mathbf{AO}^\Delta} & \\
\boxed{\mathcal{A}\mathbf{SCS}\mathcal{P}_m} \arrow[rrrr, dash, "\cong^*"] \arrow[ur, dash, "\cong"] \arrow[urr, dash, "\cong"'] & & & & \boxed{\mathcal{P}_m\mathbf{SCS}\mathcal{A}} \arrow[ull, dash, "\cong^*"] \arrow[ul, dash, "\cong"']
\end{tikzcd}$$

Let us have a closer look at the smallest infinite cardinal $m = \omega$ (where $\mathcal{P}_m = \mathcal{F}$):

$\mathcal{A}$**SCS**$\mathcal{F}$ is the category of T_0-A-spaces with $\mathcal{F}$-proper maps. On the other hand, the dual category $\mathcal{F}$**SCS**$\mathcal{A}$ is the category of sober B-spaces together with core-continuous maps. By Theorem 3.10, this category is isomorphic to the category $\mathbf{AO}^{\triangle}$ of algebraic ordered sets and residual maps preserving directed joins, and the latter is dually isomorphic to the category $\mathbf{AO}_{\nabla}$ of algebraic ordered sets and residuated maps preserving compactness.

$\mathcal{A}$**CGC**$\mathcal{F}$ is the category of superalgebraic lattices and frame homomorphisms preserving supercompactness, while $\mathcal{F}$**CGC**$\mathcal{A}$ is the category of superalgebraic lattices and complete homomorphisms preserving $\vee$-primes. Now, Corollary 5.2 yields for $m = \omega$ the following improvement of 3.10:

5.3 Corollary *The category* **O**$\mathcal{D}$ *of ordered sets and* $\mathcal{D}$*-continuous maps is equivalent to each of the categories*

$$\mathbf{AO}_{\nabla}, \mathcal{A}\mathbf{SCS}\mathcal{F}, \mathcal{A}\mathbf{CGC}\mathcal{F},$$

and dual to each of the categories

$$\mathbf{AO}^{\triangle}, \mathcal{F}\mathbf{SCS}\mathcal{A}, \mathcal{F}\mathbf{CGC}\mathcal{A}.$$

A similar reasoning leads to:

5.4 Corollary *The following categories are self-dual and mutually equivalent:*

$\mathbf{O}_{\nabla}$:	*ordered sets with residuated maps,*
$\mathbf{O}^{\triangle}$:	*ordered sets with residual maps,*
$\mathcal{A}\mathbf{SCS}\mathcal{A}$:	*T_0-A-spaces and core-continuous maps,*
$\mathcal{A}\mathbf{CGC}\mathcal{A}$:	*A-lattices and complete homomorphisms preserving supercompactness.*

Another interesting specialization is the case $\mathcal{Z} = \mathcal{D}$ (or $\mathcal{C}$):

The objects of the category $\mathcal{D}$**SCS**$\mathcal{A}$ are the $\mathcal{D}$-sober basic spaces; by 4.3, they may be regarded as $\perp$-$\vee$-semilattices with the property that their ideal lattice is superalgebraic. It turns out that this is the case iff the semilattice in question has the property that each element is the join of a *finite* number of $\vee$-primes (see the end of Section 7). Such semilattices are sometimes called *freely generated*, because every isotone map from the set of $\vee$-primes into an arbitrary $\perp$-$\vee$-semilattice extends uniquely to a $\perp$-$\vee$-homomorphism on the whole semilattice. The morphisms in $\mathcal{D}$**SCS**$\mathcal{A}$ are the core-continuous maps; in terms of the underlying semilattices, core-continuity means that for any $\bigwedge$-prime ideal, the inverse image is again a $\bigwedge$-prime ideal. But since the ideal lattice is superalgebraic, hence a coframe, the $\bigwedge$-prime ideals are precisely the *completely irreducible* ones, i.e. those ideals which cannot be represented as an intersection of other ideals. Hence a $\mathcal{D}$**SCS**$\mathcal{A}$-morphism is characterized by the property that completely irreducible ideals are preserved under the formation of inverse images. A few more computations show that this holds iff the map in question is residual and preserves finite joins. In other words, $\mathcal{D}$SCS$\mathcal{A}$ is isomorphic to the category of freely generated join-semilattices with residual $\perp$-$\vee$-homomorphisms, hence dual to the category of freely generated join-semilattices with residuated maps preserving $\vee$-primes. The dual category $\mathcal{A}$SCS$\mathcal{D}$ is the category of T_0-A-spaces and proper maps. $\mathcal{A}$CGC$\mathcal{D}$ is the category of superalgebraic lattices and maps preserving supercompactness, arbitrary joins, and filtered meets, while the dual category $\mathcal{D}$CGC$\mathcal{A}$ has the same objects, but the morphisms preserve compactness, arbitrary joins and arbitrary meets. By GSD, we have the following

5.5 Corollary *The category of freely generated join-semilattices with residuated maps preserving $\vee$-primes is equivalent to the categories $\mathcal{A}$SCS$\mathcal{D}$ and $\mathcal{A}$CGC$\mathcal{D}$, and dual to the categories $\mathcal{D}$SCS$\mathcal{A}$ and $\mathcal{D}$CGC$\mathcal{A}$.*

The case $\mathcal{Z} = \mathcal{E}$ is similar. The objects of the category $\mathcal{E}$SCS$\mathcal{A}$ are the $\mathcal{E}$-sober spaces with a basis of supercompact open sets, and as we have seen, these spaces are simply the (dually) superalgebraic lattices together with the closure systems of principal ideals. In complete analogy to 5.2, we have:

5.6 Corollary *The category $\mathcal{E}$SCS$\mathcal{A}$ is isomorphic to the category $\mathcal{E}$CGC$\mathcal{A}$ of superalgebraic lattices and complete homomorphisms, hence dually isomorphic to the category $\mathcal{A}$CGC$\mathcal{E}$ of superalgebraic lattices and maps preserving supercompactness and arbitrary joins; the latter is equivalent to the category $\mathcal{A}$SCS$\mathcal{E}$ of T_0-A-spaces and continuous maps.*

Let m denote any *regular* cardinal number. Then the previous observations are easily generalized to m-complete ordered sets such that every element is a join of less than m elements which are *m-$\vee$-prime*, i.e. *$\mathcal{P}_m$-prime* (the alternative name *$\mathcal{P}_m$-compact* is a bit misleading in the present context). By reasons to be explained later on, such ordered sets might be called *$\mathcal{P}_m$-algebraic.* With the help of Lemma 4.4, one can show that the category $\mathcal{D}_m$SCS$\mathcal{A}$ of $\mathcal{D}_m$-sober basic closure spaces is isomorphic (via specialization) to the category of $\mathcal{P}_m$-algebraic ordered sets and maps with the property that completely irreducible m-ideals are preserved under the formation of inverse images. A bit more involved is the proof of the fact that these morphisms are precisely the residual maps preserving m-joins.

The dual category $\mathcal{A}$SCS$\mathcal{D}_m$ of T_0-A-spaces and $\mathcal{D}_m$-proper maps turns out to be isomorphic to the category $\mathbf{O}\mathcal{P}_m$ of ordered sets and *$\mathcal{P}_m$-continuous* maps, i.e. isotone maps $\varphi : P \to P'$ such that $\varphi^{-1}[Z'] \in \mathcal{P}_m{}^\wedge P$ for all $Z' \in \mathcal{P}_m{}^\wedge P'$. The isomorphism is established by assigning to each T_0-A-space the ordered set P whose order relation is *dual* to the specialization order; the members of $\mathcal{P}_m{}^\wedge P$ are precisely the m-compact open sets of the original space, so $\mathcal{D}_m$-properness is in fact tantamount to $\mathcal{P}_m$-continuity. The regularity assumption on m ensures that $\mathcal{P}_m{}^\wedge P$ is always a $\mathcal{P}_m$-algebraic ordered set (being $\mathcal{P}_m$-sober; see 7.8). Moreover, $\mathcal{P}_m{}^\wedge$ yields an equivalence between $\mathbf{O}\mathcal{P}_m$ and the category $\mathcal{P}_m\mathbf{AO}_\nabla$ of $\mathcal{P}_m$-algebraic ordered sets with residuated maps preserving m-$\vee$-primes (see 7.13).

These observations together with GSD yield the following bunch of equivalences and dualities (observe the nice "complementarity" between $\mathcal{P}_m$ and $\mathcal{D}_m$ in 5.2 and 5.7):

5.7 Corollary *Let m be a regular cardinal number. Then the category* $\mathbf{O}\mathcal{P}_m$ *of ordered sets and* $\mathcal{P}_m$*-continuous maps is equivalent to the categories*

$\mathcal{P}_m\mathbf{AO}_\nabla$:	$\mathcal{P}_m$*-algebraic ordered sets and residuated maps preserving m-*$\vee$*-primes,*
$\mathcal{A}\mathbf{SCS}\mathcal{D}_m$:	T_0*-A-spaces and* $\mathcal{D}_m$*-proper maps,*
$\mathcal{A}\mathbf{CGC}\mathcal{D}_m$:	$\mathcal{A}$*-lattices and maps preserving arbitrary joins, m-filtered meets and supercompactness,*

and dual to the categories:

$\mathcal{P}_m\mathbf{AO}^\Delta$:	$\mathcal{P}_m$*-algebraic ordered sets and residual maps preserving m-joins,*
$\mathcal{D}_m\mathbf{SCS}\mathcal{A}$:	$\mathcal{D}_m$*-sober basic closure spaces and core-continuous maps,*
$\mathcal{D}_m\mathbf{CGC}\mathcal{A}$:	$\mathcal{A}$*-lattices and complete homomorphisms preserving m-compactness.*

$$
\begin{tikzcd}
\boxed{\mathcal{A}\mathbf{CGC}\mathcal{D}_m} \arrow[rrrr, dash, "\simeq^*"'] \arrow[dd, dash, "\simeq"'] \arrow[dr, dash, "\simeq"'] \arrow[drr, dash, "\simeq"] & & & & \boxed{\mathcal{D}_m\mathbf{CGC}\mathcal{A}} \arrow[dd, dash, "\simeq"] \arrow[dl, dash, "\simeq"] \arrow[dll, dash, "\simeq^*"'] \\
& \boxed{\mathcal{P}_m\mathbf{AO}_\nabla} \arrow[r, dash, "\simeq"'] \arrow[dl, dash, "\simeq"'] & \boxed{\mathbf{O}\mathcal{P}_m} \arrow[r, dash, "\simeq^*"'] \arrow[dll, dash, "\simeq"] \arrow[drr, dash, "\simeq^*"'] & \boxed{\mathcal{P}_m\mathbf{AO}^\Delta} \arrow[dr, dash, "\simeq"] & \\
\boxed{\mathcal{A}\mathbf{SCS}\mathcal{D}_m} \arrow[rrrr, dash, "\simeq^*"] & & & & \boxed{\mathcal{D}_m\mathbf{SCS}\mathcal{A}}
\end{tikzcd}
$$

The next combination, $\mathcal{Z} = \mathcal{F}$ and $\mathcal{Z}' = \mathcal{D}$, leads to the classical Stone duality (cf. [St1-2], [BD], [Gr2], [Jo1]):

The objects of the category $\mathcal{F}\mathbf{SCS}\mathcal{D}$ are the *Stone spaces*, i.e. sober spaces with a basis of compact open sets, and the morphisms are the proper maps (inverse images of compact open sets are compact and open). On the other hand, the objects of the category $\mathcal{D}\mathbf{SCS}\mathcal{F}$ are the $\mathcal{D}$-sober closure spaces with a basis of $\vee$-prime open sets; from 4.3 we know that the $\mathcal{D}$-sober closure spaces are just the $\perp$-$\vee$-semilattices, endowed with the closure system of all ideals. Thus the existence of a basis of $\vee$-prime open sets is tantamount to the fact that each closed set, i.e. each ideal, is an intersection of ($\wedge$-)prime ideals. By Banaschewski's Prime Element Theorem [Ba4], which is logically equivalent to the Prime Ideal Theorem (PIT), the representation of ideals of a $\perp$-$\vee$-semilattice as intersections of prime ideals is possible if and only if the ideal lattice of the given semilattice S is distributive, and this is well-known to be equivalent to distributivity of S itself: for all $x, y, z \in S$ with $x \leq y \vee z$, there exist

$y' \leq y$ and $z' \leq z$ such that $x = y' \vee z'$ (cf. [E7], [Gr2], [Ka]). In this sense, the morphisms of the category $\mathcal{D}$SCS$\mathcal{F}$ are the *prime ideal continuous* maps between distributive join-semilattices, i.e. maps such that inverse images of prime ideals are again prime ideals. It was shown in [DE] that a map φ between join-semilattices S and S' is prime-ideal continuous iff it preserves finite joins and satisfies the following condition (where $X_\downarrow$ denotes the set of all lower bounds of X): for all finite $F \subseteq S$ and all $x \in \varphi[F]_\downarrow$, there is a $y \in F_\downarrow$ with $x \leq \varphi(y)$. In case of a bounded *lattice*, the latter condition simply means that φ preserves finite meets. Hence, for bounded distributive lattices, the prime-ideal continuous maps are simply the 0-1-lattice homomorphisms. The latter fact is known to be equivalent to PIT, while the extension of the Stone duality to join-semilattices seems to be less common on the morphism level.

5.8 Corollary *The category of distributive join-semilattices with least elements and prime-ideal continuous maps is equivalent to the category $\mathcal{F}$CGC$\mathcal{D}$ of algebraic distributive lattices with frame homomorphisms preserving compactness; the latter is dually isomorphic to the category $\mathcal{D}$CGC$\mathcal{F}$ of dually algebraic distributive lattices and maps preserving $\vee$-spectra, arbitrary joins and filtered meets, which in turn is equivalent to $\mathcal{D}$SCS$\mathcal{F}$, the category of Stone spaces and proper maps. Hence $\mathcal{F}$SCS$\mathcal{D}$ and $\mathcal{D}$SCS$\mathcal{F}$ are duals of each other.*

The next case to be considered is $\mathcal{Z} = \mathcal{Z}' = \mathcal{F}$:

A topological space is called *strongly connected* [Hf5] or *ultraconnected* [SS] if it is nonempty and not representable as a union of two nonempty open subsets; thus an open subset of a space is strongly connected (as a subspace) iff it is a $\vee$-prime member of the lattice of open sets. A topological space with a basis of strongly connected open subsets is said to be *strongly locally connected* (cf. [Hf5]). Thus $\mathcal{F}$SCS$\mathcal{F}$ is the category of strongly locally connected sober spaces and $\mathcal{F}$-proper maps (i.e. continuous maps such that inverse images of strongly connected open sets are strongly connected). Its lattice-theoretical counterpart is the category $\mathcal{F}$CGC$\mathcal{F}$ whose objects are both spatial frames and spatial coframes, i.e. dual $\mathcal{F}$-lattices and $\mathcal{F}$-lattices; to have a short name, we call them *bispatial frames.* Morphisms in the category $\mathcal{F}$CGC$\mathcal{F}$ are frame homomorphisms preserving $\vee$-spectra (i.e. $\mathcal{F}$-compactness). From GSD, we infer:

5.9 Corollary *The category $\mathcal{F}$SCS$\mathcal{F}$ of strongly locally connected sober spaces with $\mathcal{F}$-proper maps and the category $\mathcal{F}$CGC$\mathcal{F}$ of bispatial frames with frame homomorphisms preserving $\vee$-spectra are self-dual categories and equivalent to each other.*

An interesting full subcategory of $\mathcal{F}$SCS$\mathcal{F}$ is the category SCS$\mathcal{F}$ of sober C-spaces (see Section 2) with $\mathcal{F}$-proper maps. As was demonstrated in [Hf4] and [La1], the sober C-spaces, endowed with their specialization order, are precisely the *continuous ordered sets*, equipped with their Scott topology (cf. the corresponding Theorem 2.18 on sober B-spaces); here, continuous ordered sets enter as a natural generalization of algebraic ordered sets: they are up-complete ordered sets such that for each element x there is a smallest directed lower set, called the *way-below ideal of* x, whose join dominates x. The present theory of ($\mathcal{Z}$-)algebraic ordered sets may be extended nicely to the "non-discrete" setting of ($\mathcal{Z}$-)continuous ordered sets, but by reasons of limited space, we have to reserve this topic to a forthcoming paper. The reader interested in the theory of continuous ordered sets and lattices may consult, for example, the *Compendium of Continuous Lattices* [Com] and its successors [BH], [HH] and [La2]; for some $\mathcal{Z}$-generalizations, see [BE1] and [Ve].

By the previous remarks, the category **SCS**$\mathcal{F}$ is isomorphic to the category **CO**$\mathcal{F}$ of continuous ordered sets and maps such that inverse images of Scott-open filters are again Scott open filters. J. Lawson [La1] was the first to observe that **CO**$\mathcal{F}$ *is a self-dual category*, and consequently, the same holds for **SCS**$\mathcal{F}$; again this result is obtained by specializing the above ($\mathcal{F}$,$\mathcal{F}$)-duality: the functor $\mathcal{G}_{\mathcal{F}}$ induces an equivalence between the category **SCS**$\mathcal{F}$ and the category **CD**$\mathcal{F}$ of completely distributive lattices with frame homomorphisms preserving $\bigvee$-spectra (cf. [HM2]). As complete distributivity is a self-dual property, we see that **CD**$\mathcal{F}$ is a self-dual full subcategory of $\mathcal{F}$**CGC**$\mathcal{F}$, and consequently, the categories **CO**$\mathcal{F}$ and **SCS**$\mathcal{F}$ are self-dual as well.

Passing to the full subcategory **SBS**$\mathcal{F}$ of sober B-spaces and $\mathcal{F}$-proper maps, we obtain the following self-duality of algebraic ordered sets, originally also due to Lawson [La1]:

5.10 Corollary *The category* **AO**$\mathcal{F}$ *of algebraic ordered sets and maps preserving Scott-open filters under inverse images is self-dual, isomorphic to the category* **SBS**$\mathcal{F}$ *of sober B-spaces and* $\mathcal{F}$*-proper maps, and equivalent to the self-dual category* **A**$\mathcal{F}$ *of* $\mathcal{A}$*-lattices and frame homomorphisms preserving* $\bigvee$*-spectra.*

It is quite interesting to observe the effect of changing the morphism classes: compare the last result with the ($\mathcal{A}$,$\mathcal{F}$)-duality discussed before and with the Duality Theorem 3.10!

The final case we are discussing is $\mathcal{Z} = \mathcal{Z}' = \mathcal{D}$:

The category $\mathcal{D}$**SCS**$\mathcal{D}$ has as objects (up to the aforementioned identification) the $\perp$-$\bigvee$-semilattices with the property that each ideal is an intersection of dually compact ideals; we call them *semilattices with duality.* Morphisms in this category are characterized by the condition that dually compact ideals are preserved under inverse images. Any such morphism preserves finite joins and is, therefore, a semilattice homomorphism.

The objects of the category $\mathcal{D}$**CGC**$\mathcal{D}$ are the *bicompactly generated lattices* (cf. [At]), and the morphisms preserve compactness, arbitrary joins and filtered meets.

Thus, in this specific setting, GSD amounts to the following self-duality:

5.11 Corollary *The category* $\mathcal{D}$**SCS**$\mathcal{D}$ *of semilattices with duality and the category* $\mathcal{D}$**CGC**$\mathcal{D}$ *of bicompactly generated lattices are self-dual equivalent categories.*

We hope that this broad spectrum of applications will convince the reader of the power of the General Stone Duality. A similar approach was proposed by Banaschewski and Bruns in their paper "*The Fundamental Duality of Partially Ordered Sets*" [BB]. However, the objects involved in that duality are more complicated, and some of the applications are less direct.

6 $\mathcal{Z}$-ideal extensions and completions

We are now well prepared to extend the theory of algebraic ordered sets to the general setting of an arbitrary subset selection $\mathcal{Z}$ instead of the selection $\mathcal{D}$ of directed sets. As a convenient tool for the manifold applications of this general approach, we first present two types of "ideal extensions", associating with a given subset selection $\mathcal{Z}$ two others in a natural and constructive way. These and other "global" constructions have been discussed at some length in [E8] and in [E14-18]; it will suffice here to recall the main definitions and facts. The simplest ideal extension has already been used in the previous two sections, viz. the *$\mathcal{Z}$-ideal extension* $\mathcal{Z}^\wedge$ defined by

$$\mathcal{Z}^\wedge Q = \{\downarrow Z \mid Z \in \mathcal{E}Q \cup \mathcal{Z}Q\}.$$

Subset selections of the form $\mathcal{Z}^\wedge$ are also referred to as *global standard extensions*, because they assign to each (quasi-)ordered set a certain extension in a global manner; similarly, a *global standard completion* assigns to each (quasi-)ordered set Q a standard completion of Q, that is, a closure system of lower sets including all principal ideals (cf. [E8], [EW]). As mentioned in Section 1, the *Alexandroff completion* $\mathcal{A} = \mathcal{P}^\wedge$ is the largest global standard extension, and $\mathcal{M} = \mathcal{E}^\wedge$ is the smallest global standard extension, while the *MacNeille completion* $\mathcal{N}$ is the smallest global standard completion.

The passage from $\mathcal{Z}$ to $\mathcal{Z}^\wedge$ does not change the main notions and definitions of the $\mathcal{Z}$-theory; for example,

$\mathcal{Z}^\wedge$-($\bigcup$-)*complete* means $\mathcal{Z}$-($\bigcup$-)*complete,*
$\mathcal{Z}^\wedge$-*compact* means $\mathcal{Z}$-*compact,* and
$\mathcal{Z}^\wedge$-*join preserving* means $\mathcal{Z}$-*join preserving and isotone.*

Recall that an isotone map φ between ordered sets P and P' is said to be *$\mathcal{Z}$-quasiclosed* if $Z \in \mathcal{Z}^\wedge P$ implies $\downarrow\varphi[Z] \in \mathcal{Z}^\wedge P'$. Equivalently, one could postulate this conclusion for all $Z \in \mathcal{Z}P$. Given any category **C** of ordered sets or, more generally, any class **C** of isotone maps, we call the subset selection $\mathcal{Z}$ **C**-*invariant* if every **C**-morphism is $\mathcal{Z}$-quasiclosed.

In this context, the most important classes, respectively, categories are:

O	(ordered sets and) order-preserving, i.e. isotone maps
E	(ordered sets and) order-embeddings
I	(ordered sets and) isomorphisms
$\mathbf{O}^\triangle$	(ordered sets and) residual maps
$\mathbf{O}_\nabla$	(ordered sets and) residuated maps.

We remark that each of the subset selections $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}, \mathcal{F}, \mathcal{P}, \mathcal{W}$, and the associated ideal extensions $\mathcal{Z}^\wedge$ are $\mathbf{O}$-invariant (*a fortiori* $\mathbf{E}$-, $\mathbf{I}$-, $\mathbf{O}^\Delta$- and $\mathbf{O}_\nabla$-invariant). Notice that a subset selection $\mathcal{Z}$ is $\mathbf{I}$-invariant iff the associated ideal extension $\mathcal{Z}^\wedge$ is invariant in the sense of Section 4. Every "reasonable" subset selection has this property, while $\mathbf{O}$-invariance does not occur so frequently. Therefore we adopt the following convention:

Throughout this section, all subset selections are $\mathbf{I}$*-invariant.*

The earlier "$\mathcal{Z}$-literature" (see, e.g., [Me], [Ne], [WWT]) is mostly restricted to so-called *subset systems* (a commonly used but not very instructive name); these are subset selections $\mathcal{Z}$ such that for any isotone map $\varphi : Q \to Q'$ and each $Z \in \mathcal{Z}Q$, the image $\varphi[Z]$ is a member of $\mathcal{Z}Q'$ (some authors postulate, in addition, that all singletons should belong to $\mathcal{Z}Q$). It is clear that any subset system $\mathcal{Z}$ and the associated global standard extension $\mathcal{Z}^\wedge$ are $\mathbf{O}$-invariant. Conversely, every $\mathbf{O}$-invariant global standard extension (*"ideal subset system"*) arises in this way from a suitable subset system, as was observed by Meseguer [Me]; his article continues the basic work by Wright, Wagner and Thatcher [WWT] and contains some of the most important ideas concerning subset systems, regarded from a categorical point of view.

Prominent examples of subset systems are $\mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}, \mathcal{F}, \mathcal{P}$ and $\mathcal{W}$. However, it should be emphasized that *the only* $\mathbf{O}$*- resp.* $\mathbf{E}$*-invariant global standard completion is the Alexandroff completion*, and that this is not a subset system. Since global standard completions are of more and more increasing interest for order-theoretical methods in topology, algebra and computer sciences, one certainly would not like to exclude that important class of subset selections from the considerations *a priori*. This is one of the reasons why we have introduced various degrees of "invariance" for the morphisms in question; one has to test from case to case which type will be adequate for the given situation. For example, we shall see that many global standard completions happen to be $\mathbf{O}^\Delta$-invariant, so that the General Stone Duality applies to these situations.

Any subset selection $\mathcal{Z}$ gives rise to several associated global standard completions (see [E8], [E15] and [E18]); one of them is the *$\mathcal{Z}$-join ideal completion* (*$\mathcal{Z}^\vee$-ideal completion*) $\mathcal{Z}^\vee$, where

$$\mathcal{Z}^\vee Q = \{Y \in \mathcal{A}Q \mid \text{ for all } Z \in \mathcal{Z}Q \text{ with } Z \subseteq Y, \quad x = \bigvee Z \text{ implies } x \in Y\}$$

consists of all *$\mathcal{Z}$-$\bigvee$-closed* lower sets of Q (cf. [Sch3]). Note the equations

$$\mathcal{Z}^{\vee\wedge} = \mathcal{Z}^{\wedge\vee} = \mathcal{Z}^\vee.$$

Clearly, $\mathcal{E}^\vee$ is the Alexandroff completion $\mathcal{A}$. In the completion theory for ordered sets and join-preserving maps, the *$\bigvee$-ideal completion* $\mathcal{P}^\vee$ plays a central role. Of even greater importance for us is the *Scott completion* $\mathcal{D}^\vee$, as we have seen in Sections 2 and 3. The *$\vee$-ideal completion* $\mathcal{F}^\vee$ is well-known at least for join-semilattices S, where the members of $\mathcal{F}^\vee S$ are the ideals in the usual sense. Notice also that, more generally, for m-complete ordered sets P, the m-ideals are precisely the members of $\mathcal{P}_m{}^\vee$.

Proceeding to the definition of suitable morphisms within the $\mathcal{Z}$-theory, we recall that a map φ between ordered sets is said to be *$\mathcal{Z}$-continuous* if inverse images of $\mathcal{Z}$-ideals under φ are $\mathcal{Z}$-ideals; it is called *weakly $\mathcal{Z}$-continuous* if inverse images of principal ideals are $\mathcal{Z}$-ideals. In particular, every residuated map is weakly $\mathcal{Z}$-continuous. In case of an arbitrary (I-invariant) subset selection $\mathcal{Z}$, it is important to distinguish between $\mathcal{Z}$-continuity and weak $\mathcal{Z}$-continuity, because of the following simple but effective remark (cf. [E8]):

6.1 Lemma *The principal ideal embeddings*

$$\eta_P^{\mathcal{Z}} : P \to \mathcal{Z}^\wedge P, \quad x \longmapsto {\downarrow} x$$

are always weakly $\mathcal{Z}$-continuous, while they are $\mathcal{Z}$-continuous if and only if $\mathcal{Z}^\wedge P$ is $\mathcal{Z}$-$\bigcup$-complete.

If the latter condition is fulfilled for all ordered sets P then the subset selection $\mathcal{Z}$ is said to be *union complete.*

Weakening the notion of $\mathcal{Z}$-quasiclosedness, we call a map φ between ordered sets P and P' *weakly $\mathcal{Z}$-closed* if for each $Z \in \mathcal{Z}P$, the set

$$\bar{\varphi} Z = (\varphi[Z])^- = \bigcap \{Y' \in \mathcal{Z}^\wedge P' \mid \varphi[Z] \subseteq Y'\}$$

is a member of $\mathcal{Z}^\wedge P'$. Clearly, this weak closedness condition is fulfilled for all isotone maps whenever we are dealing with an O-invariant subset selection, and for all maps between ordered sets in case of a global standard completion. A $\mathcal{Z}$-continuous and weakly $\mathcal{Z}$-closed map will be called *strongly $\mathcal{Z}$-continuous.* By definition, a map φ between ordered sets P and P' is weakly $\mathcal{Z}$-continuous iff it is continuous as a map between the ordinary spaces $(P, \mathcal{Z}^\wedge P)$ and $(P', \mathcal{M}P')$, while φ is (strongly) $\mathcal{Z}$-continuous iff it is (strongly) continuous as a map between $(P, \mathcal{Z}^\wedge P)$ and $(P', \mathcal{Z}^\wedge P')$ (see Section 3). From this it is clear that the class of strongly $\mathcal{Z}$-continuous maps is always closed under composition. Hence these maps are suited to constitute the morphism class for a category, denoted by $\mathbf{O}\mathcal{Z}$, whose objects are arbitrary ordered sets. By 3.2, every $\mathbf{O}\mathcal{Z}$-morphism (in [E8] and [E15]: *$\mathcal{Z}$-morphism*) $\varphi : P \to P'$ extends to a residuated map

$$\mathcal{Z}^\wedge \varphi = \bar{\varphi} : \mathcal{Z}^\wedge P \longrightarrow \mathcal{Z}^\wedge P'.$$

In all, we have shown:

6.2 Proposition *For any subset selection $\mathcal{Z}$, the $\mathcal{Z}$-ideal extension $\mathcal{Z}^\wedge$ gives rise to a functor from the category $\mathbf{O}\mathcal{Z}$ of ordered sets and strongly $\mathcal{Z}$-continuous maps to the category $\mathbf{O}_\nabla$ of ordered sets and residuated maps. Moreover, if $\mathcal{Z}$ is union complete then $\mathcal{Z}^\wedge$ may be regarded as a functor from $\mathbf{O}\mathcal{Z}$ to the category of $\mathcal{Z}$-compactly generated ordered sets and residuated maps.*

Unfortunately, it is not clear whether the lifted maps $\mathcal{Z}^\wedge\varphi$ do preserve $\mathcal{Z}$-compactness in general. However, this is certainly the case if $\mathcal{Z}$ is $\mathbf{O}^\Delta$-invariant, because the upper adjoint φ^{-1} preserves $\mathcal{Z}$-joins (in fact, $\mathcal{Z}$-unions; see 4.12.) Hence we may note the following:

6.3 Corollary *Every union-complete and* $\mathbf{O}^{\Delta}$*-invariant subset selection* $\mathcal{Z}$ *gives rise to a functor* $\mathcal{Z}^{\wedge}$ *from the category* $\mathbf{O}\mathcal{Z}$ *to the category* $\mathcal{Z}\mathbf{CGO}_{\nabla}$ *of* $\mathcal{Z}$*-compactly generated ordered sets and residuated maps preserving* $\mathcal{Z}$*-compactness.*

For many important subset selections like $\mathcal{A}, \mathcal{D}, \mathcal{E}, \mathcal{F}, \mathcal{I}, \mathcal{N}, \mathcal{P}, \mathcal{F}^{\vee}$ and $\mathcal{D}^{\vee}$, all three notions of $\mathcal{Z}$-continuity are equivalent. For example,

(*weakly, strongly*) $\mathcal{A}$*-continuous = isotone,*
(*weakly, strongly*) $\mathcal{E}$*-continuous = residuated.*

A subset selection $\mathcal{Z}$ such that every weakly $\mathcal{Z}$-continuous map is already $\mathcal{Z}$-continuous is said to be *compositive*, because these selections are characterized by the property that the class of weakly $\mathcal{Z}$-continuous maps is closed under composition (this follows easily from the observation that a map $\varphi : P \to P'$ is $\mathcal{Z}$-continuous iff the composite map $\eta^{\mathcal{Z}}_{P'} \circ \varphi$ is weakly $\mathcal{Z}$-continuous; see [E8] and [E15]).

If $\mathcal{Z}$ is a standard extension with the property that every weakly $\mathcal{Z}$-continuous map is even strongly $\mathcal{Z}$-continuous then $\mathcal{Z}$ is called a *standard construction.* Hence every standard construction is compositive, and every compositive standard completion is a standard construction. However, there exist compositive standard extensions which fail to be standard constructions; for an example, see [E15].

For readers familiar with the categorical notion of *monads* (see e.g. [AHS], [ML]; in [Go]: *standard constructions*), we mention that *every order-theoretical standard construction* $\mathcal{Z}$ *gives rise to a monad* $(\mathcal{Z}, \eta^{\mathcal{Z}}, \mu^{\mathcal{Z}})$, where the "multiplication maps"

$$\mu^{\mathcal{Z}}_{P} : \mathcal{Z}\mathcal{Z}P \longrightarrow \mathcal{Z}P, \quad \mathcal{Y} \longmapsto \bigcup \mathcal{Y}$$

are well-defined because $\mathcal{Z}$ is union complete, by 6.1. Slightly modified, this result may be restated as a *reflection theorem* (see [E8] and [E15]):

6.4 Theorem *Every standard construction* $\mathcal{Z}$ *gives rise to a reflector from the category* $\mathbf{O}\mathcal{Z}$ *of ordered sets and* (*strongly*) $\mathcal{Z}$*-continuous maps to the category* $\mathcal{Z}\mathbf{CO}_{\nabla}$ *of* $\mathcal{Z}$*-complete ordered sets and residuated maps, with reflection morphisms* $\eta^{\mathcal{Z}}_{P} : P \to \mathcal{Z}P$*. Thus* $\eta^{\mathcal{Z}}_{P}$ *is universal in the following sense: for every* $\mathcal{Z}$*-continuous map from* P *into a* $\mathcal{Z}$*-complete ordered set* Q*, there exists a unique residuated map* $\varphi^{\vee} : \mathcal{Z}P \to Q$ *with* $\varphi = \varphi^{\vee} \circ \eta^{\mathcal{Z}}_{P}$*; this extension* $\varphi^{\vee}$ *is given by* $\varphi^{\vee}Z = \bigvee \varphi[Z]$*.*

The next result, proved in [E8] (see also [Sch3]), provides a large class of standard constructions.

6.5 Lemma *For any* $\mathbf{O}$*-invariant subset selection* $\mathcal{Z}$ *and any isotone map* φ*, the following properties are equivalent:*

(a) *φ preserves $\mathcal{Z}$-joins.*
(b) *φ is weakly $\mathcal{Z}^\vee$-continuous.*
(c) *φ is $\mathcal{Z}^\vee$-continuous.*
(d) *φ is strongly $\mathcal{Z}^\vee$-continuous.*

The first two statements are equivalent for arbitrary subset selections $\mathcal{Z}$.

Notational remark: In the literature on subset systems, $\mathcal{Z}^\vee$-*continuous* maps are usually called $\mathcal{Z}$-*continuous* (cf. [Me], [Ne]); our notation is more flexible and underscores the topological aspects of the theory.

The next result clarifies the position of compositive subset selections.

6.6 Lemma *Every compositive subset selection is union complete and $\mathbf{O}^\triangle$-invariant. Conversely, every union-complete and $\mathbf{O}$-invariant subset selection is compositive.*

Proof By Lemma 6.1, a compositive subset selection $\mathcal{Z}$ is union complete. Since every residuated map is weakly $\mathcal{Z}$-continuous, it is $\mathcal{Z}$-continuous for compositive $\mathcal{Z}$, so 4.10 applies to show that $\mathcal{Z}$ is $\mathbf{O}^\triangle$-invariant.

Conversely, if $\mathcal{Z}$ is union complete and $\mathbf{O}$-invariant then for any weakly $\mathcal{Z}$-continuous map $\varphi : P \to P'$, we have a well-defined isotone map

$$\Phi : P' \longrightarrow \mathcal{Z}^\wedge P, \quad x \longmapsto \varphi^{-1}[\downarrow x].$$

Hence for $Z' \in \mathcal{Z}^\wedge P'$, the prolonged image

$$\downarrow\Phi[Z'] = \{Z \in \mathcal{Z}^\wedge P \mid Z \subseteq \varphi^{-1}[\downarrow x] \text{ for some } x \in Z'\}$$

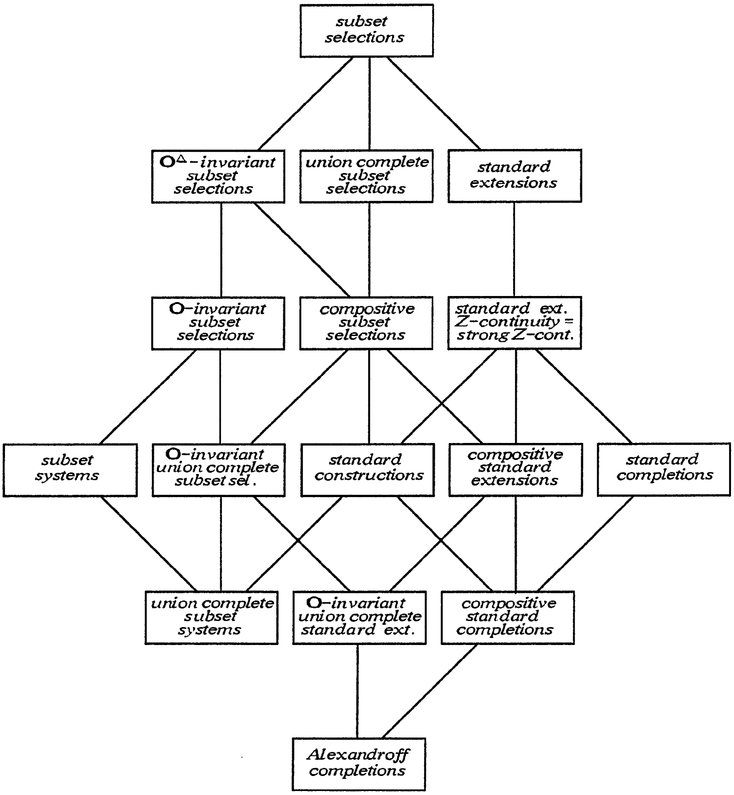
belongs to $\mathcal{Z}^\wedge\mathcal{Z}^\wedge P$, and by union completeness, its union $\varphi^{-1}[Z']$ is a member of $\mathcal{Z}^\wedge P$. This proves $\mathcal{Z}$-continuity of φ. □

6.7 Corollary *For any union-complete and $\mathbf{O}$-invariant subset selection $\mathcal{Z}$, the associated $\mathcal{Z}$-ideal extension $\mathcal{Z}^\wedge$ is a standard construction.*

On the other hand, a direct application of 6.5 (combined with 6.4) yields:

6.8 Corollary *For any $\mathbf{O}$-invariant subset selection $\mathcal{Z}$, the ideal completion $\mathcal{Z}^\vee$ is a standard construction, hence compositive, union complete and $\mathbf{O}^\triangle$-invariant. Thus $\mathcal{Z}^\vee$ is a reflector from the category $\mathbf{O}\mathcal{Z}^\vee$ of ordered sets with maps preserving $\mathcal{Z}$-joins to the category of complete lattices and join-preserving maps.*

Let us represent the hierarchy of the main types of subset selections in an implication diagram, with properties of increasing strength from top to bottom (cf. [E15]):



6.9 Examples Compositive O-invariant subset selections are $\mathcal{A}, \mathcal{D}, \mathcal{E}, \mathcal{F}$, and $\mathcal{P}$, but neither $\mathcal{B}$ nor $\mathcal{C}$ nor $\mathcal{W}$! For $\mathcal{B}$, union completeness is obviously violated, whereas it might be tempting to conjecture that $\mathcal{C}$ be union complete, because the union of a nonempty system of chains which is totally ordered by inclusion is again a chain. However, if $\mathcal{Y}$ is a system of lower sets generated by chains and $\mathcal{Y}$ is itself a chain with respect to inclusion, then the union $\bigcup\mathcal{Y}$ need not be generated by a chain. For example, if P is the lattice $\mathcal{F}\Omega$ of all finite subsets of the first uncountable ordinal Ω, then the collection $\{\mathcal{F}\alpha \mid \alpha < \Omega\}$ is a chain of countable ideals in P, and each countable ideal is generated by a chain. But the union $\mathcal{F}\Omega = \bigcup\{\mathcal{F}\alpha \mid \alpha < \Omega\}$ cannot be generated by a chain because any chain $\mathcal{Y}$ in $\mathcal{F}\Omega$ must be countable (the cardinality function being an injection into ω), and therefore its union cannot give the uncountable set Ω. In all, we arrive at the somewhat surprising conclusion that $\mathcal{D}$ but not $\mathcal{C}$ is union complete, although $\mathcal{C}$-$\bigcup$-completeness and $\mathcal{D}$-$\bigcup$-completeness are equivalent properties, as we have seen in Section 1. These remarks show that "good" subset selections should be compositive and, therefore, union complete; but, helas, there are also familiar subset selections like $\mathcal{C}$ and $\mathcal{W}$ which fail to be union complete.

For later applications, it will be convenient to call a standard extension $\mathcal{X}$ of an ordered set P *$\mathcal{Z}$-quasiclosed* if so is the principal ideal embedding

$$\eta = \eta_P^{\mathcal{X}} : P \longrightarrow \mathcal{X}, \quad x \longmapsto \downarrow x,$$

in other words, if for each $Z \in \mathcal{Z}P$, the lower set $\downarrow_{\mathcal{X}}\eta[Z]$ belongs to $\mathcal{Z}^{\wedge}\mathcal{X}$.

Recall that every $\mathcal{Z}$-sober standard extension is $\mathcal{Z}$-$\bigcup$-complete. Conversely, we show:

6.10 Lemma *Every $\mathcal{Z}$-$\bigcup$-complete and $\mathcal{Z}$-quasiclosed standard extension $\mathcal{X}$ of an ordered set P with $\mathcal{X} \subseteq \mathcal{Z}^{\wedge}P$ is $\mathcal{Z}$-sober and coincides with $\mathcal{Z}^{\wedge}P$.*

Proof We know that for any $\mathcal{Z}$-compact member X of $\mathcal{X}$, the lower set $\downarrow_{\mathcal{X}}\eta[X]$ is a $\mathcal{Z}$-ideal of $\mathcal{X}$ whose union is X. By $\mathcal{Z}$-compactness, it follows that X belongs to this $\mathcal{Z}$-ideal and is therefore a member of $\eta[X]$; hence X must be a principal ideal, i.e. a point closure. For the second claim, use $\mathcal{Z}$-$\bigcup$-completeness. □

Applied to the $\mathcal{Z}$-join ideal completion $\mathcal{Z}^{\vee}$ instead of $\mathcal{Z}$, the last lemma provides an interesting characterization of $\mathcal{Z}^{\vee}$-sober standard completions, at least in case of $\mathbf{O}$-invariant subset selections:

6.11 Proposition *Let $\mathcal{Z}$ be an $\mathbf{O}$-invariant subset selection and $\mathcal{X}$ a $\mathcal{Z}^{\vee}$-quasiclosed standard completion of an ordered set P. Then the following statements are equivalent:*

(a) $\mathcal{X} = \mathcal{Z}^{\vee}P$.
(b) *$\mathcal{X}$ is $\mathcal{Z}^{\vee}$-$\bigcup$-complete and contained in $\mathcal{Z}^{\vee}P$.*
(c) *$\mathcal{X}$ is $\mathcal{Z}^{\vee}$-sober.*

Each of these conditions implies that P is $\mathcal{Z}$-complete. Conversely, if P is a complete lattice then every standard completion of P is $\mathcal{Z}^{\vee}$-quasiclosed, and consequently, $\mathcal{Z}^{\vee}P$ is the unique $\mathcal{Z}^{\vee}$-sober standard completion of P.

Proof (a) $\Longrightarrow$ (b): By 6.8, $\mathcal{Z}^{\vee}$ is union complete.
(b) $\Longrightarrow$ (c): See 6.10.
(c) $\Longrightarrow$ (a): First, we show that the closure $\Gamma Z = \bigcap\{X \in \mathcal{X} \mid Z \subseteq X\}$ of $Z \in \mathcal{Z}P$ is a $\mathcal{Z}^{\vee}$-compact member of $\mathcal{X}$. For the $\mathcal{Z}$-quasiclosed embedding $\eta = \eta_P^{\mathcal{X}}$, we obtain $\downarrow_{\mathcal{X}}\eta[Z] \in \mathcal{Z}^{\wedge}\mathcal{X}$, and if $\mathcal{Y}$ is a member of $\mathcal{Z}^{\vee}\mathcal{X}$ with $\Gamma Z \subseteq \bigvee\mathcal{Y} = \bigcup\mathcal{Y}$ then $\downarrow_{\mathcal{X}}\eta[Z]$ is entirely contained in the lower set $\mathcal{Y}$. Hence $\Gamma Z = \bigvee \downarrow_{\mathcal{X}}\eta[Z] \in \mathcal{Y}$.

Now, by $\mathcal{Z}^{\vee}$-soberness, ΓZ must be a point closure $\downarrow x$, and in particular, x is the join of Z in P (because $\mathcal{X}$ contains all principal ideals). Thus P is $\mathcal{Z}$-complete. Moreover, for $Y \in \mathcal{X}$ and $Z \in \mathcal{Z}P$ with $Z \subseteq Y$, $x = \bigvee Z$ implies $x \in \Gamma Z \subseteq Y$, so that Y is actually a $\mathcal{Z}^{\vee}$-ideal. This proves the inclusion $\mathcal{X} \subseteq \mathcal{Z}^{\vee}P$. Concerning the converse inclusion, we remark that for $Y \in \mathcal{Z}^{\vee}P$ the lower set $\downarrow_{\mathcal{X}}\eta[Y]$ belongs to $\mathcal{Z}^{\vee}\mathcal{X}$, and by $\mathcal{Z}^{\vee}$-$\bigcup$-completeness of $\mathcal{X}$, it follows that $Y = \bigcup \downarrow_{\mathcal{X}} \eta[Y] \in \mathcal{X}$.

Finally, if P is a complete lattice then $\eta : P \to \mathcal{X}$ is a residual map with lower adjoint

$$\varepsilon : \mathcal{X} \longrightarrow P, \quad X \longmapsto \bigvee X.$$

Hence, by 6.8, η is $\mathcal{Z}^{\vee}$-quasiclosed. □

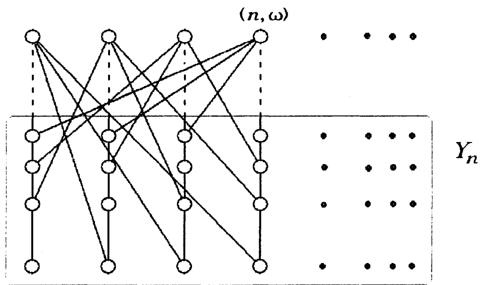
A first application of this result is Lemma 4.4, stating that any m-complete ordered set P has exactly one $\mathcal{D}_m$-sober standard completion, namely the m-ideal completion $\mathcal{D}_m{}^{\wedge}P = \mathcal{P}_m{}^{\vee}P$. Indeed, this follows at once from 6.11 because $\mathcal{P}_m$ as well as $\mathcal{D}_m$ are **O**-invariant subset selections. Moreover, we infer from 6.11 that m-completeness of P is not only sufficient but also necessary for the existence of a $\mathcal{D}_m$-sober standard completion.

Another interesting consequence of 6.11 is obtained for the selection $\mathcal{D}$ of all directed sets:

6.12 Corollary *A complete lattice L has exactly one $\mathcal{D}^{\vee}$-sober standard completion, namely the Scott completion $\mathcal{D}^{\vee}L$.*

Observing that every finitely generated lower set is Scott closed, we see that $\mathcal{F}^{\wedge}P$ is always contained in $\mathcal{D}^{\vee}P$, and consequently, every $\mathcal{D}^{\vee}$-compact element is $\mathcal{F}$-prime, i.e. $\vee$-prime. Therefore, *if the topological closure system $\mathcal{D}^{\vee}P$ is sober then it is $\mathcal{D}^{\vee}$-sober* (being $\mathcal{D}^{\vee}$-$\bigcup$-complete by 6.8). However, the converse implication fails, as we shall see below. In his short note "*Scott is not always sober*" [Jo2], Johnstone has given the following example of an up-complete ordered set with non-sober Scott topology:

6.13 Example Order the set $\omega \times (\omega + 1)$ by $(n, m) \leq (n', m')$ iff ($n = n'$ and $m \leq m'$) or ($m \leq n'$ and $m' = \omega$), as indicated in the diagram below.



This ordered set P cannot be sober in its Scott topology because the whole set P is a $\vee$-prime closed set but not a point closure. Moreover, P is even $\mathcal{D}^{\vee}$-compact in $\mathcal{D}^{\vee}P$: Consider any $\mathcal{Y} \in \mathcal{D}^{\vee}\mathcal{D}^{\vee}P$ with $P = \bigcup \mathcal{Y}$. For $n \in \omega$, the principal ideal $\downarrow(n, \omega)$ is a member of $\mathcal{Y}$ (since $(n,\omega) \in Y$ for some $Y \in \mathcal{Y}$ and $\mathcal{Y}$ is a lower set in $\mathcal{D}^\vee P$). Hence the sets

$$Y_n = \{(m,n') \mid m \in \omega, n' \leq n\} \subseteq \downarrow(n,\omega)$$

form an ascending chain in $\mathcal{Y}$ whose join in $\mathcal{D}^\vee P$ is P, and this join must belong to $\mathcal{Y}$. Therefore, *$\mathcal{D}^\vee P$ is not $\mathcal{D}^\vee$-sober.*

In contrast to this example, Corollary 6.12 shows that a complete lattice is always "Scott-sober". This is all the more surprising since Isbell [Is2] succeeded in refining Johnstone's example to obtain a *complete lattice whose Scott topology is not sober.* By 6.12, this cannot happen with "*$\mathcal{D}^\vee$-sober*" instead of "sober".

In connection with various kinds of $\mathcal{Z}$-compactness, it is certainly reasonable to have a look at ordered sets consisting of $\mathcal{Z}$-compact elements only.

6.14 Proposition *The following three statements on a $\mathcal{Z}$-complete ordered set Q are equivalent:*

(a) *$Q = \mathcal{K}_{\mathcal{Z}}Q$, i.e. every element of Q is $\mathcal{Z}$-compact.*
(b) *$\mathcal{M}Q = \mathcal{Z}^\wedge Q$, i.e. every $\mathcal{Z}$-ideal of Q is principal.*
(c) *$\mathcal{A}Q = \mathcal{Z}^\vee Q$, i.e. every lower set of Q is a $\mathcal{Z}^\vee$-ideal.*

Proof (a) $\Longrightarrow$ (b): For $Y \in \mathcal{Z}^\wedge Q$, choose some $Z \in \mathcal{Z}Q$ with $Y = \downarrow Z$. Then, by (a), the join $x = \bigvee Y = \bigvee Z$ exists and is $\mathcal{Z}$-compact, whence $Y = \downarrow Z = \downarrow x \in \mathcal{M}Q$.
(b) $\Longrightarrow$ (c): $\mathcal{M}Q = \mathcal{Z}^\wedge Q$ implies $\mathcal{A}Q = \mathcal{M}^\vee Q = \mathcal{Z}^{\wedge\vee}Q = \mathcal{Z}^\vee Q$.
(c) $\Longrightarrow$ (a): If $x \notin \downarrow Z$ for some $Z \in \mathcal{Z}Q$ then $Z \subseteq Q \backslash \uparrow x \in \mathcal{A}Q = \mathcal{Z}^\vee Q$, whence $x \not\leq \bigvee Z$. □

Notice that each of the equations

$$\mathcal{M}Q = \mathcal{W}^\wedge Q, \quad \mathcal{M}Q = \mathcal{C}^\wedge Q, \quad \mathcal{M}Q = \mathcal{D}^\wedge Q$$

means that Q satisfies the Ascending Chain Condition (ACC). Grätzer asked in [Gr2] whether a lattice isomorphic to its ideal lattice would necessarily satisfy the ACC. A positive answer was given by Higgs [Hi], but his proof involved transfinite tools, including the Maximal Principle and ordinal numbers. In [E14], we gave a "choice-free" proof for a more general result, based on Bourbaki's Fixpoint Lemma. In our present terminology, this result reads as follows:

6.15 Theorem *Let $\mathcal{Z}$ be an* **I**-*invariant union complete subset selection. Then the following three statements on an ordered set Q are equivalent:*

(a) *$\mathcal{W}^\wedge Q \subseteq \mathcal{Z}^\wedge Q \simeq Q$.*
(b) *$\mathcal{M}Q = \mathcal{W}^\wedge Q = \mathcal{Z}^\wedge Q$.*
(c) *Q satisfies the ACC and is isomorphic to $\mathcal{Z}^\wedge Q$.*

Proof (a) $\Longrightarrow$ (b): Let φ be any isomorphism between Q and $\mathcal{Z}^\wedge Q$. By union completeness of $\mathcal{Z}$, the ordered set $\mathcal{Z}^\wedge Q$ is $\mathcal{Z}$-complete, and so is its isomorphic copy Q. Since φ^{-1}

is an isomorphism and

$$\eta : Q \longrightarrow \mathcal{Z}^\wedge Q, \quad x \longmapsto \downarrow x$$

is an embedding, so is

$$\psi = \varphi^{-1} \circ \eta : Q \longrightarrow Q.$$

For any non-$\mathcal{Z}$-compact element $x_0 \in Q$, the element

$$x_1 = \bigvee \varphi x_0$$

cannot be $\mathcal{Z}$-compact either, because φx_0 is a $\mathcal{Z}$-ideal but not principal (otherwise, it would be $\mathcal{Z}$-compact in $\mathcal{Z}^\wedge Q$). Furthermore, we have $x_0 < \psi x_1$ since φx_0 is a proper subset of $\varphi \psi x_1 = \eta x_1$. Repeating this argument, we obtain a properly ascending sequence

$$x_0 < \psi x_1 < \psi^2 x_2 < \ldots < \psi^n x_n < \ldots$$

whose join is not $\mathcal{Z}$-compact either, as $\mathcal{W}^\wedge Q \subseteq \mathcal{Z}^\wedge Q$. Hence we obtain a fixpoint free extensive selfmap on the $\mathcal{W}$-complete set N of all non-$\mathcal{Z}$-compact elements of Q, by assigning to each $x_0 \in N$ the element

$$\bigvee \{\psi^n x_n \mid n \in \omega\}.$$

But by Bourbaki's Fixpoint Lemma, such an ordered set must be empty. It follows that $\mathcal{M}Q = \mathcal{Z}^\wedge Q$, and then the inclusion $\mathcal{M}Q \subseteq \mathcal{W}^\wedge Q \subseteq \mathcal{Z}^\wedge Q$ forces $\mathcal{M}Q = \mathcal{W}^\wedge Q$.

The first part of the proof also yields the implication (c) $\Longrightarrow$ (b). That (b) implies (a) and (c) is obvious. □

In accordance with our introductory remarks on the exchange between (well-ordered) chains and directed sets, we notice that a union-complete subset selection $\mathcal{Z}$ with

$$\mathcal{W}^\wedge Q \subseteq \mathcal{Z}^\wedge Q$$

for *all* ordered sets Q must satisfy the seemingly stronger inclusion

$$\mathcal{D}^\wedge Q \subseteq \mathcal{Z}^\wedge Q.$$

To see this, observe that for $\mathcal{Y} \in \mathcal{W}\mathcal{Z}^\wedge Q$, we have

$$\downarrow \mathcal{Y} = \{Z \in \mathcal{Z}^\wedge Q \mid Z \subseteq Y \text{ for some } Y \in \mathcal{Y}\} \in \mathcal{W}^\wedge \mathcal{Z}^\wedge Q \subseteq \mathcal{Z}^\wedge \mathcal{Z}^\wedge Q$$

and then $\bigcup \mathcal{Y} = \bigcup \downarrow \mathcal{Y} \in \mathcal{Z}^\wedge Q$. Thus $\mathcal{Z}^\wedge Q$ is $\mathcal{W}$- and, therefore, $\mathcal{D}$-$\bigcup$-complete (see 1.13). But for $D \in \mathcal{D}Q$, the set $\{\downarrow x \mid x \in D\}$ belongs to $\mathcal{D}\mathcal{Z}^\wedge Q$, whence $\downarrow D = \bigcup \{\downarrow x \mid x \in D\}$ is a member of $\mathcal{Z}^\wedge Q$, proving the inclusion $\mathcal{D}^\wedge Q \subseteq \mathcal{Z}^\wedge Q$.

Let us list a few applications of Theorem 6.15. For the last one (see 6.16(d)), we need the notion of consistent sets (cf. [DP]), where in the present context, we mean by a *consistent* subset of an ordered set Q a subset S such that every finite subset of S has an upper bound in Q (thus every directed subset is consistent, but not conversely). The corresponding subset selection $\mathcal{C}o$ is $\mathbf{O}$-invariant, union complete, and satisfies the inclusion $\mathcal{D}Q \subseteq \mathcal{C}oQ$, so that 6.15 actually applies to this selection. The other three subset selections we take into consideration are: $\mathcal{A}$ (lower sets; cf. Dilworth and Gleason [DG]), $\mathcal{A}_0$ (nonempty lower sets), and $\mathcal{D}^\wedge$ (directed lower sets).

6.16 Corollary *Let Q be any ordered set.*

(1) *Q is never isomorphic to $\mathcal{A}Q$.*
(2) *$Q \simeq \mathcal{A}_0 Q$ iff Q is dually well-ordered.*
(3) *$Q \simeq \mathcal{D}^\wedge Q$ iff Q satisfies the ACC.*
(4) *$Q \simeq \mathcal{C}o^\wedge Q$ iff Q is a forest satisfying the ACC.*

As the subset selection $\mathcal{C}$ of all chains is *not* union complete, we do not know whether an ordered set Q isomorphic to $\mathcal{C}^\wedge Q$ must satisfy the ACC.

Finally, a few comments are in order about the hypothesis

$$\mathcal{W}^\wedge Q \subseteq \mathcal{Z}^\wedge Q$$

in 6.15. There do exist I-invariant and union complete subset selections $\mathcal{Z}$ such that $\mathcal{W}^\wedge Q \not\subseteq \mathcal{Z}^\wedge Q$, and nevertheless $Q \simeq \mathcal{Z}^\wedge Q$ implies $\mathcal{M}Q = \mathcal{Z}^\wedge Q$, for example the Dedekind-MacNeille completion $\mathcal{N}$. On the other hand, it turns out that, without the above hypothesis, it may happen very well that an ordered set Q is isomorphic to $\mathcal{Z}^\wedge Q$ although not every $\mathcal{Z}$-ideal is principal. For example, in case of the Scott completion $\mathcal{D}^\vee$, it is clear that an ordered set Q isomorphic to $\mathcal{D}^\vee Q$ must be a $\mathcal{D}^\vee$-lattice, in particular a distributive complete lattice, since $\mathcal{D}^\vee Q$ is a dual topology. Moreover, from [E14], we cite the following fact, demonstrating the abundance of lattices isomorphic to their own Scott completion:

6.17 Theorem *Every finite lattice is the image of a completely distributive lattice $L \simeq \mathcal{D}^\vee L$ under a map preserving directed joins and arbitrary meets. Moreover, every completely distributive lattice is the image of a completely distributive lattice $L \simeq \mathcal{D}^\vee L$ under a complete homomorphism (preserving arbitrary joins and meets).*

From 6.15 it is clear that a lattice L with $L \simeq \mathcal{D}^\vee L$ cannot satisfy the ACC; moreover, that any non-$\mathcal{D}^\vee$-compact element (in particular, the least element) of such a lattice must have a properly ascending chain of upper bounds. The simplest lattice with this property is the chain $\omega + 1$, and this chain is in fact isomorphic to its own Scott completion. A non-linearly ordered example of a lattice L with $L \simeq \mathcal{D}^\vee L$ is the ordinal sum $\omega + \mathcal{P}2 + \omega + 1$, where $\mathcal{P}2$ is regarded as a four-element Boolean lattice (cf. [E14]).

The following question remains open:

Is any lattice L with $L \simeq \mathcal{D}^\vee L$ completely distributive?

The answer is affirmative in case of algebraic lattices and, more generally, of continuous lattices; indeed, an up-complete ordered set is continuous if and only if its Scott completion is completely distributive (see [Com] and [E6]).

Originally, Scott had "invented" continuous lattices in search of a model for the λ-calculus (see e.g. [Sc1]). For this purpose, he constructed, via certain projective limits, a continuous lattice isomorphic to its own function space, the lattice of all Scott-continuous, i.e. $\mathcal{D}$-join preserving selfmaps of the underlying lattice. For a thorough investigation of similar projective limit constructions leading to lattices isomorphic to their own $\mathcal{Z}$-ideal extensions (for suitable standard completions $\mathcal{Z}$), see [E13].

7 $\mathcal{Z}$-Inductive and $\mathcal{Z}$-algebraic ordered sets

We are now turning towards a $\mathcal{Z}$-generalization of algebraic ordered sets and their representation as ideal systems. For the case of subset systems, such a uniform approach was initiated by Wright, Wagner and Thatcher [WWT], with the intention of certain applications to computer sciences. In the much more general situation of an arbitrary subset selection $\mathcal{Z}$, it is not entirely evident what might be the "best" definition of $\mathcal{Z}$-algebraic ordered sets. We offer three slightly different definitions and leave it to the reader to make his own choice of his favorite notion. Of course, this choice may depend, from case to case, on the problem to be solved. Let Q be a $\mathcal{Z}$-compactly generated ordered set and $P = \mathcal{K}_{\mathcal{Z}}Q$ its $\mathcal{Z}$-spectrum, the set of all $\mathcal{Z}$-compact elements. Recall that the map

$$\kappa_Q^{\mathcal{Z}} : Q \longrightarrow \mathcal{X}_{\mathcal{Z}}Q = \{P\cap{\downarrow}y \mid y \in Q\}, \quad y \longmapsto P\cap{\downarrow}y$$

is an isomorphism (see 4.5). Now we call the $\mathcal{Z}$-compactly generated ordered set Q

$\mathcal{Z}$-inductive	if $\mathcal{X}_{\mathcal{Z}}Q = \mathcal{Z}^{\wedge}P$,
$\mathcal{Z}$-prealgebraic	if $\mathcal{X}_{\mathcal{Z}}{}^{\wedge}Q \subseteq \mathcal{Z}^{\wedge}Q$ (i.e. ${\downarrow}(P\cap{\downarrow}y) \in \mathcal{Z}^{\wedge}Q$ for all $y \in Q$),
$\mathcal{Z}$-algebraic	if $\mathcal{X}_{\mathcal{Z}}Q \subseteq \mathcal{Z}^{\wedge}P$ and the inclusion map from P into Q is $\mathcal{Z}$-quasiclosed.

These three properties are related as follows:

7.1 Lemma *An ordered set is $\mathcal{Z}$-algebraic iff it is $\mathcal{Z}$-inductive and $\mathcal{Z}$-prealgebraic.*

Proof It is clear that a $\mathcal{Z}$-algebraic ordered set Q is $\mathcal{Z}$-prealgebraic. In order to show that Q is also $\mathcal{Z}$-inductive, it remains to prove the equation $Z = P\cap{\downarrow}y$ for $Z \in \mathcal{Z}^{\wedge}P$ and $y = \bigvee Z$ (the join exists because Q is $\mathcal{Z}$-compactly generated, in particular $\mathcal{Z}$-complete). The inclusion $Z \subseteq P\cap{\downarrow}y$ is obvious. For the converse inclusion, consider any $x \in P\cap{\downarrow}y$; as x is $\mathcal{Z}$-compact in Q and $x \leq \bigvee Z = \bigvee {\downarrow}Z$, we infer that $x \in {\downarrow}Z$, because $Z \in \mathcal{Z}^{\wedge}P$ implies ${\downarrow}Z \in \mathcal{Z}^{\wedge}Q$. But Z is a lower set in P, so it follows that $x \in P\cap{\downarrow}Z = Z$.

Now assume Q is $\mathcal{Z}$-inductive and $\mathcal{Z}$-prealgebraic. Then $Z \in \mathcal{Z}^{\wedge}P$ means $Z = P\cap{\downarrow}y$ for some $y \in Q$, and we conclude that ${\downarrow}Z = {\downarrow}(P\cap{\downarrow}y)$ belongs to $\mathcal{Z}^{\wedge}Q$; in other words, the inclusion map from P into Q is $\mathcal{Z}$-quasiclosed. □

7.2 Corollary *Let Q be an ordered set such that the inclusion map from $P = \mathcal{K}_{\mathcal{Z}}Q$*

into Q is $\mathcal{Z}$-quasiclosed (this is automatically fulfilled if $\mathcal{Z}$ is **E**-*invariant). The following statements are equivalent:*

(a) *Q is $\mathcal{Z}$-algebraic.*
(b) *Q is $\mathcal{Z}$-inductive.*
(c) *Q is $\mathcal{Z}$-compactly generated and $\mathcal{X}_{\mathcal{Z}}Q \subseteq \mathcal{Z}^\wedge P$.*
(d) *Q is $\mathcal{Z}$-complete, and for each $y \in Q$, there is a $Z \in \mathcal{Z}^\wedge P$ with $y = \bigvee Z$.*

The last of these four conditions is usually the most convenient one to work with. For many important subset selections like $\mathcal{Z} = \mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}, \mathcal{F}, \mathcal{P}$, or $\mathcal{W}$, all three notions coincide:

$\mathcal{Z}$-inductive = $\mathcal{Z}$-prealgebraic = $\mathcal{Z}$-algebraic.

For this coincidence, it is sufficient that $\mathcal{Z}$ be "*relativizable*" in the following sense: for all ordered sets P, Q and all $Z \subseteq P \cap Q$, $\downarrow_P Z \in \mathcal{Z}^\wedge P$ iff $\downarrow_Q Z \in \mathcal{Z}^\wedge Q$.

However, arbitrary subset selections $\mathcal{Z}$ do not behave so nicely: neither is a $\mathcal{Z}$-inductive ordered set always $\mathcal{Z}$-prealgebraic, nor is a $\mathcal{Z}$-prealgebraic ordered set always $\mathcal{Z}$-inductive.

7.3 Examples (1) The ordered set $P = (\mathbb{Z}, \sqsubseteq)$ from Example 2.1 satisfies the ACC. We show that $L = \mathcal{A}P = \mathcal{D}^\vee P$ is a $\mathcal{D}^\vee$-inductive complete lattice whose $\mathcal{D}^\vee$-compact elements are the principal ideals of P. Since P is an algebraic ordered set, we know from 2.18 that $\mathcal{D}^\vee P$ is sober. In particular, every $\mathcal{D}^\vee$-compact member of $\mathcal{D}^\vee P$ is a point closure (being $\mathcal{F}$-compact = $\vee$-prime); conversely, it is clear that principal ideals are $\bigvee$-prime, hence $\mathcal{D}^\vee$-compact in $\mathcal{A}P$. Thus $\mathcal{D}^\vee P$ is a $\mathcal{D}^\vee$-sober closure system, and as we shall see in 7.4, this implies that $L = \mathcal{D}^\vee P$ is $\mathcal{D}^\vee$-inductive. But for $x_0 = 0$ and $Y = \mathbb{Z} \setminus \{x_0\} \in \mathcal{A}P = L$, the lower set $\mathcal{Y} = \downarrow_L(\mathcal{M}P \cap \downarrow_L Y)$ contains the ascending chain

$$\{Y_n = \{x \in \mathbb{Z} \mid -n \leq x < 0\} \mid n \in \omega\}$$

but not its union. Hence $\mathcal{Y}$ is not a member of $\mathcal{D}^\vee L$, and L cannot be $\mathcal{D}^\vee$-prealgebraic, because the inclusion map from $\mathcal{M}P$ to L fails to be $\mathcal{D}^\vee$-quasiclosed.

(2) For arbitrary ordered sets Q, let $\bar{\mathcal{B}}Q$ denote the set of all upper bounded subsets of Q. Then $\bar{\mathcal{B}}$ is an **O**-invariant but not relativizable subset selection. Any power set lattice $\mathcal{P}S$ is $\bar{\mathcal{B}}$-prealgebraic, and $P = \mathcal{E}S$ is the set of all $\bar{\mathcal{B}}$-compact elements of $\mathcal{P}S$. While each subset of $\mathcal{P}S$ is bounded, the set $P = P \cap \downarrow S$ is not upper bounded in P if S has more than one element. Hence $\mathcal{P}S$ is not $\bar{\mathcal{B}}$-inductive.

(3) For $\mathcal{Z} = \mathcal{A}$ (or $\mathcal{Z} = \mathcal{P}$), we have

$\mathcal{Z}$-compactly generated = $\mathcal{A}$-inductive = $\mathcal{A}$-(pre)algebraic = superalgebraic = $\mathcal{A}$-lattice.

(4) Similarly, for any cardinal number m, we have

$\mathcal{P}_m$*-inductive* $=$ $\mathcal{P}_m$*-(pre)algebraic* (cf. 5.7)

while a $\mathcal{P}_m$-primely generated ordered set need not be $\mathcal{P}_m$-inductive. In particular, for the selection $\mathcal{F} = \mathcal{P}_\omega$, the $\mathcal{F}$-inductive, i.e. $\mathcal{F}$-algebraic ordered sets are the freely generated join-semilattices, while an $\mathcal{F}$-lattice need not be $\mathcal{F}$-algebraic; for example, in the $\mathcal{F}$-lattice of all Scott-closed subsets of the real plane, the half-plane $\{(x,y) \mid x+y \leq 0\}$ cannot be represented as a join of *finitely* many $\vee$-prime Scott-closed sets, i.e. principal ideals.

(5) For any cardinal number m, we have

$\mathcal{D}_m$*-inductive* $=$ $\mathcal{D}_m$*-(pre)algebraic* $=$ m*-algebraic* (cf. 5.2)

and the case $m = \omega$ yields

$\mathcal{D}$*-inductive* $=$ $\mathcal{D}$*-(pre)algebraic* $=$ *algebraic.*

But we have seen in 2.3(1) that a ($\mathcal{D}$-)compactly generated ordered set need not be ($\mathcal{D}$-)algebraic. Nevertheless, a $\mathcal{D}_m$-compactly generated *complete lattice* is already m-algebraic, because for any element y of such a lattice, the set of all m-compact elements dominated by y is m-directed.

One of the main results in the pioneer paper by Wright, Wagner and Thatcher [WWT] states that for any union complete subset system $\mathcal{Z}$, the $\mathcal{Z}$-inductive ordered sets are, up to isomorphism, precisely the $\mathcal{Z}$-ideal extensions $\mathcal{Z}^\wedge P$. In order to obtain a similar representation of $\mathcal{Z}$-inductive ordered sets for arbitrary I-invariant subset selections $\mathcal{Z}$, we must restrict suitably the class of ordered sets under consideration. It turns out that $\mathcal{Z}$-soberness is again the right ingredient. Thus we call an ordered set P *$\mathcal{Z}$-induced* if the point closure system $\mathcal{Z}^\wedge P$ is $\mathcal{Z}$-sober; in other words, if $\mathcal{Z}^\wedge P$ is $\mathcal{Z}$-$\bigcup$-complete and each $\mathcal{Z}$-compact member of $\mathcal{Z}^\wedge P$ is a principal ideal.

7.4 Proposition *The following statements on two ordered sets P and Q are equivalent:*

(a) *P is $\mathcal{Z}$-induced, and Q is isomorphic to $\mathcal{Z}^\wedge P$.*
(b) *Q is $\mathcal{Z}$-inductive, and P is isomorphic to $\mathcal{K}_\mathcal{Z} Q$.*

Proof (a) $\Longrightarrow$ (b): We may assume $Q = \mathcal{Z}^\wedge P$. Then Q is $\mathcal{Z}$-compactly generated, as $\mathcal{Z}^\wedge P$ is $\mathcal{Z}$-($\bigcup$-)complete and $\mathcal{M}P = \mathcal{K}_\mathcal{Z} Q$ is $\bigvee$-dense in Q. For $Y \in Q$, we have

$$\mathcal{M}P \cap {\downarrow_Q} Y = \{ {\downarrow_P} y \mid y \in Y\} = \eta_P^{\mathcal{M}}[Y],$$

where

$$\eta_P^{\mathcal{M}} : P \longrightarrow \mathcal{M}P, \quad y \longmapsto {\downarrow_P} y$$

is an isomorphism. Thus we obtain

$$\mathcal{X}_\mathcal{Z} Q = \{\mathcal{K}_\mathcal{Z} Q \cap {\downarrow_Q} Y \mid Y \in Q\} = \{\eta_P^{\mathcal{M}}[Y] \mid Y \in \mathcal{Z}^\wedge P\} = \mathcal{Z}^\wedge \mathcal{M}P = \mathcal{Z}^\wedge \mathcal{K}_\mathcal{Z} Q.$$

Hence Q is $\mathcal{Z}$-inductive.

(b) $\Longrightarrow$ (a): We may assume $P = \mathcal{K}_{\mathcal{Z}}Q$. Then the map

$$\kappa_Q^{\mathcal{Z}} : Q \longrightarrow \mathcal{Z}^{\wedge}P, \quad y \longmapsto P \cap {\downarrow_Q} y$$

is an isomorphism. As Q is $\mathcal{Z}$-complete, so is $\mathcal{Z}^{\wedge}P$, and

$$\mathcal{M}P = \kappa_Q^{\mathcal{Z}}[P] = \mathcal{K}_{\mathcal{Z}}\mathcal{Z}^{\wedge}P.$$

This shows that the point closure system $\mathcal{Z}^{\wedge}P$ is $\mathcal{Z}$-sober. □

In the present context, it is an obvious question to ask which ordered sets are isomorphic to their own $\mathcal{Z}$-spectra. In case of suitable subset selections $\mathcal{Z}$ and $\mathcal{Z}$-inductive ordered sets, this can happen only trivially, namely if *each* element is $\mathcal{Z}$-compact, as the following modification of Theorem 6.15 shows:

7.5 Corollary *Suppose $\mathcal{Z}$ is an* **I**-*invariant union-complete subset selection with $\mathcal{W}^{\wedge}Q \subseteq \mathcal{Z}^{\wedge}Q$ for all ordered sets Q. Then a $\mathcal{Z}$-inductive ordered set Q can be isomorphic to $\mathcal{K}_{\mathcal{Z}}Q$ only if $Q = \mathcal{K}_{\mathcal{Z}}Q$.*

Proof By 7.4, we have $Q \simeq \mathcal{Z}^{\wedge}P$ for $P = \mathcal{K}_{\mathcal{Z}}Q \simeq Q$, whence $P \simeq \mathcal{Z}^{\wedge}P$, and by 6.15, $\mathcal{M}P = \mathcal{Z}^{\wedge}P$. In other words, each member of $\mathcal{Z}^{\wedge}P$ is $\mathcal{Z}$-compact, and the same holds for the isomorphic copy Q. □

The hypothesis of $\mathcal{Z}$-inductivity cannot be weakened to $\mathcal{Z}$-completeness, as was shown in [E14] for the cases $\mathcal{Z} = \mathcal{A}_0$ and $\mathcal{Z} = \mathcal{D}$, by an example of a complete chain C with $C \simeq \mathcal{K}C$ but $C \neq \mathcal{K}C$.

If the inclusion $\mathcal{W}^{\wedge}Q \subseteq \mathcal{Z}^{\wedge}Q$ is dropped, it can happen very well that a $\mathcal{Z}$-inductive ordered set is isomorphic but not equal to its $\mathcal{Z}$-spectrum: for example, the chain $\omega + 1$ is $\mathcal{D}^{\vee}$-inductive and isomorphic to the proper subset $\omega + 1 \setminus \{0\}$ of all $\mathcal{D}^{\vee}$-compact elements.

7.6 Corollary *The $\mathcal{Z}$-inductive ordered sets are, up to isomorphism, precisely the $\mathcal{Z}$-ideal extensions of $\mathcal{Z}$-induced ordered sets; and on the other hand, the $\mathcal{Z}$-induced ordered sets are precisely the $\mathcal{Z}$-spectra of $\mathcal{Z}$-inductive ordered sets (together with the induced order).*

The latter conclusion justifies the attribute "*$\mathcal{Z}$-induced*" for such ordered sets. For an analogous representation of $\mathcal{Z}$-algebraic ordered sets, we have to strengthen a bit the property of being $\mathcal{Z}$-induced. Thus we call an ordered set P *$\mathcal{Z}$-adequate* if the standard extension $\mathcal{Z}^{\wedge}P$ is $\mathcal{Z}$-$\bigcup$-complete and $\mathcal{Z}$-quasiclosed. By 6.10, any such P is $\mathcal{Z}$-induced. Of course, *every* ordered set is $\mathcal{Z}$-adequate if $\mathcal{Z}$ happens to be union complete and **E**-invariant (e.g. for each of the subset selections $\mathcal{Z} = \mathcal{A}, \mathcal{B}, \mathcal{D}, \mathcal{E}, \mathcal{F}$, or $\mathcal{P}$). Although global standard *completions* distinct from $\mathcal{A}$ are *not* **E**-invariant, many of them are compositive, hence union complete and $\mathbf{O}^{\Delta}$-invariant, as we have seen in 6.6. But for $\mathbf{O}^{\Delta}$-invariant $\mathcal{Z}$, every standard completion $\mathcal{X}$ of a complete lattice L is $\mathcal{Z}$-quasiclosed (because the embedding $\eta_P^{\mathcal{X}} : L \to \mathcal{X}$ is residual with lower adjoint $\varepsilon_P^{\mathcal{X}} : \mathcal{X} \to L$, $X \mapsto \bigvee X$). Thus we conclude:

7.7 Corollary *If $\mathcal{Z}$ is a compositive subset selection then every complete lattice is $\mathcal{Z}$-adequate.*

Nevertheless, there exist rather "good" compositive standard completions $\mathcal{Z}$ for which not all $\mathcal{Z}$-complete ordered sets are $\mathcal{Z}$-adequate. For example, the algebraic ordered set P in 2.1 is certainly not $\mathcal{D}^{\vee}$-adequate because the principal embedding $\eta_P^{\mathcal{D}^{\vee}} : P \to \mathcal{D}^{\vee}P$ fails to be $\mathcal{D}^{\vee}$-quasiclosed (see 7.3(1)). In analogy to 7.4, we have now the following one-to-one correspondence between (isomorphism classes of) $\mathcal{Z}$-adequate ordered sets and $\mathcal{Z}$-algebraic ordered sets:

7.8 Proposition *The following statements on two ordered sets P and Q are equivalent:*

(a) *P is $\mathcal{Z}$-adequate, and Q is isomorphic to $\mathcal{Z}^{\wedge}P$.*
(b) *Q is $\mathcal{Z}$-algebraic, and P is isomorphic to $\mathcal{K}_{\mathcal{Z}}Q$.*

Proof By 7.4, it suffices to verify that an ordered set P is $\mathcal{Z}$-adequate iff $Q = \mathcal{Z}^{\wedge}P$ is $\mathcal{Z}$-algebraic. But this is an immediate consequence of the following three observations:
(1) P is $\mathcal{Z}$-adequate iff P is $\mathcal{Z}$-induced and $\eta_P^{\mathcal{Z}}$ is $\mathcal{Z}$-quasiclosed.
(2) Q is $\mathcal{Z}$-algebraic iff Q is $\mathcal{Z}$-inductive and the inclusion map $\iota : \mathcal{M}P \to Q$ is $\mathcal{Z}$-quasiclosed.
(3) The maps $\eta_P^{\mathcal{Z}}$ and ι are linked by the isomorphism $\eta_P^{\mathcal{M}} : P \to \mathcal{M}P$ via the equation $\eta_P^{\mathcal{Z}} = \iota \circ \eta_P^{\mathcal{M}}$. □

7.9 Corollary *The $\mathcal{Z}$-algebraic ordered sets are, up to isomorphism, the $\mathcal{Z}$-ideal extensions of $\mathcal{Z}$-adequate ordered sets. On the other hand, the $\mathcal{Z}$-adequate ordered sets are precisely the $\mathcal{Z}$-spectra of $\mathcal{Z}$-algebraic ordered sets.*

7.10 Corollary *For any union-complete and* E*-invariant subset selection $\mathcal{Z}$, the $\mathcal{Z}$-algebraic ordered sets are, up to isomorphism, the $\mathcal{Z}$-ideal extensions of arbitrary ordered sets.*

In particular, this includes the following known facts:
(A) *The $\mathcal{A}$-lattices are, up to isomorphism, the Alexandroff completions of ordered sets.*
(B) *The freely generated join-semilattices are, up to isomorphism, the $\mathcal{F}$-ideal extensions of ordered sets.*
(C) *The algebraic ordered sets are, up to isomorphism, the $\mathcal{D}$-ideal extensions (= up-completions) of ordered sets.*

Another application of 7.8 is obtained for compositive subset selections (use 7.7!):

7.11 Corollary *For compositive subset selections $\mathcal{Z}$, the $\mathcal{Z}$-ideal extension $\mathcal{Z}^{\wedge}L$ of any complete lattice L is $\mathcal{Z}$-algebraic.*

Thus, for example, the Scott completion of any complete lattice (but not of all up-complete ordered sets) is a $\mathcal{D}^\vee$-algebraic lattice, and conversely, every $\mathcal{D}^\vee$-algebraic lattice is isomorphic to the Scott completion of some ordered set (namely that of its $\mathcal{D}^\vee$-compact elements).

For a categorical reformulation of the above results, denote by $\mathcal{Z}\mathbf{iO}\mathcal{Z}$ the category of $\mathcal{Z}$-induced ordered sets with strongly $\mathcal{Z}$-continuous maps as morphisms, and by $\mathcal{Z}\mathbf{aO}\mathcal{Z}$ the full subcategory of $\mathcal{Z}$-adequate ordered sets. On the other hand, let $\mathcal{Z}\mathbf{IO}_\nabla$ and $\mathcal{Z}\mathbf{AO}_\nabla$ denote the category of all $\mathcal{Z}$-inductive, respectively, $\mathcal{Z}$-algebraic ordered sets and residuated maps preserving $\mathcal{Z}$-compactness. Now combining the previous representation theorems with the Equivalence Theorem 4.8 for $\mathcal{Z}$-compactly generated ordered sets, we arrive at the following modification:

7.12 Theorem *The $\mathcal{Z}$-ideal extension $\mathcal{Z}^\wedge$ gives rise to an equivalence between the categories $\mathcal{Z}\mathbf{iO}\mathcal{Z}$ and $\mathcal{Z}\mathbf{IO}_\nabla$, respectively, between the categories $\mathcal{Z}\mathbf{aO}\mathcal{Z}$ and $\mathcal{Z}\mathbf{AO}_\nabla$. The inverse equivalence is obtained by restriction to the $\mathcal{Z}$-spectra.*

If $\mathcal{Z}$ is union complete and $\mathbf{E}$-invariant then $\mathcal{Z}\mathbf{iO}\mathcal{Z} = \mathcal{Z}\mathbf{aO}\mathcal{Z}$ is simply the category $\mathbf{O}\mathcal{Z}$ of ordered sets and strongly $\mathcal{Z}$-continuous maps. In this case, $\mathbf{O}\mathcal{Z}$ is dual to the category $\mathcal{Z}\mathbf{IO}^\Delta = \mathcal{Z}\mathbf{AO}^\Delta$ of $\mathcal{Z}$-algebraic ordered sets and residual maps preserving $\mathcal{Z}$-joins.

In particular, we obtain for $\mathcal{Z} = \mathcal{D}$ the known equivalence between the category $\mathbf{O}\mathcal{D}$ and the category $\mathbf{AO}_\nabla$ (see 3.6), respectively, the duality between $\mathbf{O}\mathcal{D}$ and $\mathbf{AO}^\Delta$ (see 3.7).

For $\mathcal{Z} = \mathcal{F}$, Theorem 7.12 states that the category $\mathbf{O}\mathcal{F}$ of ordered sets and $\mathcal{F}$-continuous maps (finitely generated lower sets are preserved under inverse images) is equivalent to the category $\mathcal{F}\mathbf{IO}_\nabla = \mathcal{F}\mathbf{AO}_\nabla$ of freely generated join-semilattices and residuated maps preserving $\vee$-spectra.

Of course, for $\mathcal{Z} = \mathcal{A}$ (or $\mathcal{P}$), $\mathcal{Z}\mathbf{iO}\mathcal{Z} = \mathcal{Z}\mathbf{aO}\mathcal{Z}$ is simply the category $\mathbf{O}$ of ordered sets and isotone maps, while $\mathcal{Z}\mathbf{IO}_\nabla = \mathcal{Z}\mathbf{AO}_\nabla$ is the category $\mathbf{A}_\nabla$ of superalgebraic lattices and maps preserving joins and supercompactness.

Enlarging the morphism classes of the above categories, we denote by $\mathcal{Z}\mathbf{AO}$ the category of $\mathcal{Z}$-algebraic ordered sets and isotone maps preserving $\mathcal{Z}$-joins and $\mathcal{Z}$-compactness. In order to ensure that this is actually a category, we must guarantee that the composition of $\mathcal{Z}$-join preserving isotone maps again preserves $\mathcal{Z}$-joins, and this is certainly the case if $\mathcal{Z}$ is $\mathbf{O}$-invariant. For the minimal choice $\mathcal{Z} = \mathcal{E}$, we see that $\mathcal{E}\mathbf{AO}$ is just the category $\mathbf{O}$ of ordered sets and isotone maps, while for the maximal choice $\mathcal{Z} = \mathcal{A}$ (or $\mathcal{Z} = \mathcal{P}$), we obtain the category $\mathbf{A}_\nabla = \mathcal{A}\mathbf{AO}$ of superalgebraic lattices. Now to our final equivalence theorem:

7.13 Theorem *Let $\mathcal{Z}$ be a union-complete and $\mathbf{O}$-invariant subset selection.*
(1) *The $\mathcal{Z}$-ideal extension $\mathcal{Z}^\wedge$ establishes an equivalence between the categories $\mathbf{O}$ and $\mathcal{Z}\mathbf{AO}$.*
(2) *The $\mathcal{Z}$-join ideal completion $\mathcal{Z}^\vee$ induces an equivalence between $\mathcal{Z}\mathbf{AO}$ and $\mathbf{A}_\nabla$.*
(3) *The composite functor $\mathcal{Z}^\vee \circ \mathcal{Z}^\wedge$ is naturally isomorphic to the Alexandroff completion.*

Proof (1) is straightforward and has been proved in [WWT] and [E15].

(3) The union map

$$\nu_P : \mathcal{Z}^\vee \mathcal{Z}^\wedge P \longrightarrow \mathcal{A}P, \quad \mathcal{Y} \longmapsto \bigcup \mathcal{Y}$$

turns out to be a natural isomorphism (see [E8]).

(2) By (1), $\mathcal{Z}^\vee$ is naturally isomorphic to the composite functor $\mathcal{Z}^\vee \circ \mathcal{Z}^\wedge \circ \mathcal{K}_{\mathcal{Z}}$, which in turn is naturally isomorphic to the functor $\mathcal{A} \circ \mathcal{K}_{\mathcal{Z}}$, by (3). Again by (1), applied to $\mathcal{A}$ instead of $\mathcal{Z}$, $\mathcal{A} \circ \mathcal{K}_{\mathcal{Z}}$ is a functorial equivalence between the categories $\mathcal{Z}\mathbf{AO}$ and $\mathbf{A}_\nabla$, and consequently, the same is true for $\mathcal{Z}^\vee$. □

It appears rather plausible that the "inverse" equivalence between $\mathbf{A}_\nabla$ and $\mathcal{Z}\mathbf{AO}$ is given by the $\mathcal{Z}^\vee$-spectrum functor $\mathcal{K}_{\mathcal{Z}^\vee}$. However, a closer investigation reveals some obstacles and shows that for a smooth theory, $\mathcal{Z}^\vee$-compact elements must be replaced with so-called *$\mathcal{Z}$-hypercompact* elements, i.e. elements x such that for arbitrary Y with $x \le \bigvee Y$, there is some $Z \in \mathcal{Z}L$ with $Z \subseteq \downarrow Y$ and $x \le \bigvee Z$. More about that modified notion of compactness in a forthcoming note. For the moment, it will suffice to observe that at least the most important subset selections, namely $\mathcal{D}_m$ for arbitrary cardinals m and $\mathcal{P}_m$ for regular cardinals m, have the pleasant property that $\mathcal{Z}$-hypercompactness is equivalent to $\mathcal{Z}^\vee$-compactness. Hence, in these cases, the $\mathcal{Z}^\vee$-spectrum functor actually provides an equivalence between the categories $\mathbf{A}_\nabla$ and $\mathcal{Z}\mathbf{AO}$. In particular, we have the following commuting diagram of equivalence functors between the four most important categories occurring in connection with questions of compact generation:

$$\begin{tikzcd}
\text{\textit{Algebraic ordered sets}} & \boxed{\mathbf{AO} = \mathcal{D}\mathbf{AO}} \arrow[r, shift right, "\mathcal{D}^\vee"'] \arrow[d, shift right, "\mathcal{K}_{\mathcal{D}}"'] & \boxed{\mathbf{A}_\nabla = \mathcal{A}\mathbf{AO}} \arrow[l, shift right, "\mathcal{K}_{\mathcal{F}} = \mathcal{K}_{\mathcal{D}^\vee}"'] \arrow[d, shift right, "\mathcal{K}_{\mathcal{D}}"'] \arrow[dl, shift left, "\mathcal{K}_{\mathcal{A}}"] & \text{\textit{$\mathcal{A}$-lattices}} \\
\text{\textit{ordered sets}} & \boxed{\mathbf{O} = \mathcal{E}\mathbf{AO}} \arrow[u, shift right, "\mathcal{D}^\wedge"'] \arrow[r, shift right, "\mathcal{F}"'] \arrow[ur, shift left, "\mathcal{A}"] & \boxed{\mathcal{F}\mathbf{AO}} \arrow[l, shift right, "\mathcal{K}_{\mathcal{F}}"'] \arrow[u, shift right, "\mathcal{D}^\wedge"'] & \text{\textit{Freely generated join-semilattices}}
\end{tikzcd}$$

On account of our final Theorem 7.13, we arrive at the somewhat surprising conclusion that all "nice" categories of $\mathcal{Z}$-algebraic ordered sets are mutually equivalent!

8 Prospect: towards relative $\mathcal{Z}$-compactness

Though we have presented in the preceding sections a few highlights of the general "$\mathcal{Z}$-theory", this should not be the end of the story. However, for reasons of limited space (and perhaps limited acceptance by the patient reader), we had to exclude entirely some interesting but more involved chapters of this theory.

For example, we left it open how to generalize the topological representation of algebraic ordered sets by sober spaces with minimal basis to the general setting of $\mathcal{Z}$-(pre-)algebraic ordered sets. The key for this is Proposition 6.11, ensuring that under rather reasonable assumptions, an ordered set P admits precisely one $\mathcal{Z}^\vee$-sober standard completion, namely the $\mathcal{Z}$-join ideal completion $\mathcal{Z}^\vee P$. As indicated shortly in Section 7, $\mathcal{Z}^\vee$-compactness has to be replaced with $\mathcal{Z}$-hypercompactness in order to make the program successful. For example, it can be shown that under rather mild assumptions on $\mathcal{Z}$, the $\mathcal{Z}$-(pre-)algebraic ordered sets, endowed with the system of $\mathcal{Z}$-join ideals, are precisely the $\mathcal{Z}$-hypersober closure spaces together with their specialization order, where a T_0 closure space is said to be *$\mathcal{Z}$-hypersober* iff the point closures are precisely the $\mathcal{Z}$-hypercompact closed sets.

Of still much greater interest is the extension of the theory from $\mathcal{Z}$-compactly generated ordered sets to so-called *$\mathcal{Z}$-distributive* ordered sets (see [BE2], [E4], [E17]), respectively, from $\mathcal{Z}$-algebraic ordered sets to *$\mathcal{Z}$-continuous* ordered sets (see [BE1], [No], [Ve]). Both types of ordered sets are described most easily in terms of the *$\mathcal{Z}$-below relation* $\ll_Z$, defined by $x \ll_Z y$ iff x is "*relatively $\mathcal{Z}$-compact in y*", i.e. x belongs to every $\mathcal{Z}$-ideal whose cut closure contains the point y. Thus $x \ll_Z y$ means that x is a member of the *$\mathcal{Z}$-below ideal*

$$\Downarrow_Z y = \bigcap\{Z \in \mathcal{Z}^\wedge Q \mid y \in \Delta Z\}.$$

A $\mathcal{Z}$-complete ordered set Q is *$\mathcal{Z}$-distributive* iff each element of Q is the join of its $\mathcal{Z}$-below ideal, and Q is *$\mathcal{Z}$-continuous* iff the $\mathcal{Z}$-below ideal of each $y \in Q$ is a $\mathcal{Z}$-ideal with join y. While Q is $\mathcal{Z}$-complete iff the principal ideal embedding $\eta_Q^{\mathcal{Z}}$ has a lower adjoint, viz. the join map

$$\varepsilon_Q^{\mathcal{Z}} : \mathcal{Z}^\wedge Q \longrightarrow Q, \quad Z \longmapsto \bigvee Z,$$

the existence of a lower adjoint for the latter is equivalent to $\mathcal{Z}$-continuity of Q (cf. [No]).

In the theory of ($\mathcal{D}$-)continuous ordered sets (cf. [BH], [La1-2] etc.), the so-called *interpolation property*, i.e., the idempotency of the $\mathcal{D}$-below relation for any continuous ordered set, is of fundamental importance. Unfortunately, for arbitrary subset selections $\mathcal{Z}$, the $\mathcal{Z}$-below relation is not necessarily idempotent (see [BE2]); however, union complete and **O**-invariant subset selections behave nicely also in this respect: for such subset selections $\mathcal{Z}$, every $\mathcal{Z}$-continuous ordered set has the interpolation property (cf. [BE1]). Therefore, let us assume for simplicity that in the subsequent concluding remarks, $\mathcal{Z}$ always denotes a union complete and **O**-invariant subset selection.

Let us indicate briefly how $\mathcal{Z}$-distributive and $\mathcal{Z}$-continuous ordered sets enter into the picture as natural generalizations of $\mathcal{Z}$-compactly generated resp. $\mathcal{Z}$-algebraic ordered sets.

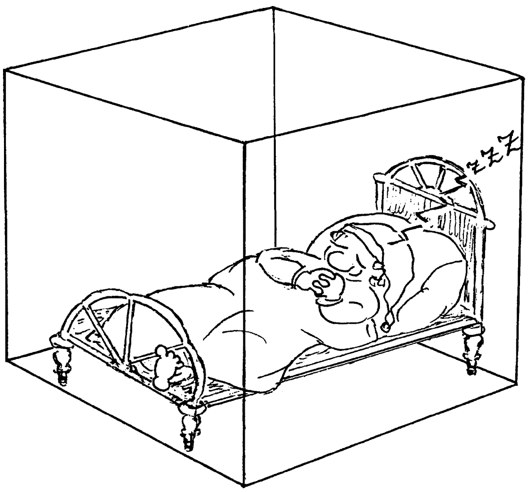
(1) By definition, an element x is $\mathcal{Z}$-compact iff it is relatively $\mathcal{Z}$-compact in itself,

and a $\mathcal{Z}$-complete ordered set Q is $\mathcal{Z}$-compactly generated (resp. $\mathcal{Z}$-algebraic) iff for each $y \in Q$, the set $\kappa_{\mathcal{Z}} y$ of all $\mathcal{Z}$-compact elements below y has join y (and $\downarrow\kappa_{\mathcal{Z}} y$ is a $\mathcal{Z}$-ideal). Analogously, Q is $\mathcal{Z}$-distributive iff for each $y \in Q$, the set $\twoheaddownarrow_{\mathcal{Z}} y$ of all elements which are relatively $\mathcal{Z}$-compact in y, has join y (and is a $\mathcal{Z}$-ideal).

(2) A complete lattice is $\mathcal{Z}$-compactly generated iff it is embeddable in a "discrete cube" $\{0,1\}^P$ under preservation of $\mathcal{Z}$-joins and arbitrary meets (see 4.6).
QUESTION: Which complete lattices admit a $\mathcal{Z}$-join and meet-preserving embedding in a "continuous cube" $[0,1]^P$?
ANSWER: The $\mathcal{Z}$-distributive complete lattices!

(3) The $\mathcal{Z}$-hypersober B-spaces, endowed with their specialization order, are precisely the $\mathcal{Z}$-algebraic ordered sets, equipped with the system of $\mathcal{Z}$-join ideals.
QUESTION: How to characterize $\mathcal{Z}$-hypersober C-spaces (see Section 2) ?
ANSWER: They are precisely the $\mathcal{Z}$-continuous ordered sets together with the system of $\mathcal{Z}^{\vee}$-ideals!

(4) On account of Theorem 7.13, the $\mathcal{Z}$-join ideal completion functor $\mathcal{Z}^{\vee}$ yields a categorical equivalence between $\mathcal{Z}$-algebraic ordered sets and superalgebraic lattices. Furthermore, there is a similar equivalence (via Scott completion) between ($\mathcal{D}$-)continuous ordered sets and completely distributive lattices.
QUESTION: Is there an analogous equivalence between categories of $\mathcal{Z}$-continuous ordered sets and suitable categories of completely distributive lattices?
ANSWER: Yes, but things become more complicated. The patient reader is kindly asked to wait for the announced paper on $\mathcal{Z}$-continuous ordered sets.



Zet's ME, embedded in a cube

References

[AB] Abian, S., and Brown, A.B., A theorem on partially ordered sets, with applications to fixed point theorems, *Canad. J. Math.* **13** (1961), 78-82.

[AHS] Adámek, J., Herrlich, H., and Strecker, G., *Abstract and Concrete Categories*, Wiley-Interscience, New York 1990.

[Al] Alexandroff, P., Diskrete Räume. *Mat. Sb.* (*N.S.*) **2** (1937), 501-518.

[At] Atherton, C.R., Concerning intrinsic topologies on Boolean algebras and certain bicompactly generated lattices, *Glasgow Math. J.* **11** (1970), 156-161.

[BD] Balbes, R., and Dwinger, Ph., *Distributive Lattices,* University of Missouri Press, Columbus, MO, 1974.

[Ba1] Banaschewski, B., Über den Satz von Zorn, *Math. Nachr.* **10** (1953), 181-186.

[Ba2] Banaschewski, B., Hüllensysteme und Erweiterungen von Quasi-Ordnungen, *Z. Math. Logik Grundlag. Math.* **2** (1956), 369-377.

[Ba3] Banaschewski, B., Essential extensions of T_0-spaces, *General Topology and Appl.* **7** (1977), 1-22.

[Ba4] Banaschewski, B., Prime elements from prime ideals, *Order* **2** (1985), 211-213.

[BB] Banaschewski, B., and Bruns, G., The fundamental duality of partially ordered sets, *Order* **5** (1988), 61-74.

[BH] Banaschewski, B., and Hoffmann, R.-E., (eds.), *Continuous Lattices, Proc. Bremen* 1979; Lecture Notes in Math. **871**, Springer-Verlag, Berlin – Heidelberg – New York 1981.

[BE1] Bandelt, H.-J., and Erné, M., The category of $\mathcal{Z}$-continuous posets, *J. Pure Appl. Algebra* **30** (1983), 219-226.

[BE2] Bandelt, H.-J., and Erné, M., Representations and embeddings of $\mathcal{M}$-distributive lattices, *Houston J. Math.* **10** (1984), 315-324.

[Bt] Batbedat, A., Des schemas en demi-groupes commutatifs, *Semigroup Forum* **16** (1978), 473-481.

[Bi] Birkhoff, G., *Lattice Theory,* 3rd ed., Amer. Math. Soc. Colloq. Publ. **25**, Providence, RI, 1973.

[BF] Birkhoff, G., and Frink, O., Representation of lattices by sets, *Trans. Amer. Math. Soc.* **64** (1948), 299-316.

[BJ] Blyth, T.S., and Janowitz, M.F., *Residuation Theory,* Pergamon Press, Oxford 1972.

[Bo] Bourbaki, N., Sur le théorème de Zorn, *Arch Math.* (*Basel*) **2** (1949/50), 434-437.

[Bü] Büchi, J.R., Representations of complete lattices by sets, *Portugal. Math.* **11** (1952), 151-167.

[CD] Crawley, P., and Dilworth, R.P., *Algebraic Theory of Lattices,* Prentice-Hall, Englewood Cliffs, NJ, 1973.

[Co] Cohn, P.M., *Universal Algebra,* Harper and Row, New York 1965.

[Com] Gierz, G., Hofmann, K.H., Keimel, K., Lawson, J.D., Mislove, M., and Scott, D.S., *A Compendium of Continuous Lattices,* Springer-Verlag, Berlin – Heidelberg – New York 1980.

[DP] Davey, B., and Priestley, H., *Introduction to Lattices and Order,* Cambridge University Press, Cambridge 1990.

[DE] David, E., and Erné, M., Ideal completion and Stone representation of ideal-distributive ordered sets, in: *Topology and its Applications, Proc. Oxford Symposium* (P. J. Collins, ed.), to appear.

[DG] Dilworth, R.P., and Gleason, A.M., A generalized Cantor theorem, *Proc. Amer. Math. Soc.* **13** (1962), 704-705.

[E1] Erné, M., Order-topological lattices, *Glasgow J. Math.* **21** (1980), 57-68.

[E2] Erné, M., Verallgemeinerungen der Verbandstheorie, I: Halbgeordnete Mengen und das Prinzip der Vervollständigungsinvarianz, Preprint No. **109**, Inst. für Mathematik, Univ. Hannover 1980.

[E3] Erné, M., Verallgemeinerungen der Verbandstheorie, II: m-Ideale in halbgeordneten Mengen und Hüllenräumen, Habilitationsschrift, Inst. für Mathematik, Univ. Hannover 1980.

[E4] Erné, M., Homomorphisms of $\mathcal{M}$-generated and $\mathcal{M}$-distributive posets, Preprint No. **125**, Inst. für Mathematik, Univ. Hannover 1981.

[E5] Erné, M., A completion-invariant extension of the concept of continuous lattices, in [BH], 43-60.

[E6] Erné, M., Scott convergence and Scott topology in partially ordered sets, in [BH], 61-96.

[E7] Erné, M., Distributivgesetze und die Dedekind'sche Schnittvervollständigung, *Abh. Braunschweig. Wiss. Ges.* **33** (1982), 117-145.

[E8] Erné, M., Adjunctions and standard constructions for partially ordered sets, in: *Contributions to General Algebra 2. Proc. Klagenfurt Conf. 1982* (G. Eigenthaler et al., eds.), Hölder-Pichler-Tempsky, Wien 1983, 77-106.

[E9] Erné, M., On the existence of decompositions in lattices, *Algebra Universalis* **16** (1983), 338-343.

[E10] Erné, M., Lattice representations for categories of closure spaces, in: *Categorical Topology, Proc. Toledo, Ohio 1983* (H.L. Bentley et al., eds.), Heldermann, Berlin 1984, 197-222.

[E11] Erné, M., Chains, directed sets and continuity, Preprint No. **175**, Inst. für Mathematik, Univ. Hannover 1984.

[E12] Erné, M., Compactly generated and continuous closure systems, *Memo Sem. Cont. Semilattices* 5-7-84 (1984).

[E13] Erné, M., Fixed point constructions for standard completions, *Memo Sem. Cont. Semilattices* 3-10-85 (1985).

[E14] Erné, M., Posets isomorphic to their extensions, *Order* **2** (1985), 199-210.

[E15] Erné, M., Order extensions as adjoint functors, *Quaestiones Math.* **9** (1986), 146-206.

[E16] Erné, M., Compact generation in partially ordered sets, *J. Austral. Math. Soc.* **42** (1987), 69-83.

[E17] Erné, M., The Dedekind-MacNeille completion as a reflector, *Order* **8** (1991), 159-173.

[E18] Erné, M., Bigeneration in complete lattices and principal separation in partially ordered sets, *Order* **8** (1991), 197-221.

[E19] Erné, M., The *ABC* of order and topology, in: *Category Theory at Work* (H. Herrlich and H.-E. Porst, eds.), Heldermann, Berlin 1981, 57-83.

[EW] Erné, M., and Wilke, G., Standard completions for quasiordered sets, *Semigroup Forum* **27** (1983), 351-376.

[Fl] Fleischer, I., Even every join-extension solves a universal problem, *J. Austral. Math. Soc.* **21** (1976), 220-223.

[Fr] Frink, O., Ideals in partially ordered sets, *Amer. Math. Monthly* **61** (1954), 223-234.

[GG] Geissinger, L., and Graves, W., The category of complete algebraic lattices, *J. Combin. Theory Ser.A* **13** (1972), 332-338.

[Go] Godement, R., *Théorie des faisceaux,* Hermann, Paris 1958.

[Gr1] Grätzer, G., On the family of certain subalgebras of a universal algebra, *Indag. Math.* **27** (1965), 790-802.

[Gr2] Grätzer, G., *General Lattice Theory,* Birkhäuser, Basel 1978.

[Gu1] Gunter, C., *Profinite Solutions for Recursive Domain Equations,* Ph. D. Thesis, University of Wisconsin, Madison 1985.

[Gu2] Gunter, C., Comparing categories of domains, in: *Mathematical foundations of programming semantics* (A. Melton, ed.), Lecture Notes in Comput. Sci. **239**, Springer-Verlag, Berlin – Heidelberg – New York 1985, 101-121.

[Ha] Halpern, J.D., The independence of the axiom of choice from the Boolean prime ideal theorem, *Fund. Math.* **55** (1964), 57-66.

[Hi] Higgs, D., Lattices isomorphic to their ideal lattices, *Algebra Universalis* **1** (1971), 71-72.

[Hf1] Hoffmann, R.-E., Sobrification of partially ordered sets, *Semigroup Forum* **17** (1979), 123-138.

[Hf2] Hoffmann, R.-E., Continuous posets and adjoint sequences, *Semigroup Forum* **18** (1979), 173-188.

[Hf3] Hoffmann, R.-E., Projective sober spaces, in [BH], 125-158.

[Hf4] Hoffmann, R.-E., Continuous posets, prime spectra of completely distributive complete lattices, and Hausdorff compactifications, in [BH], 159-208.

[Hf5] Hoffmann, R.-E., Topological spaces admitting a dual, in: *Categorical Topology* (H. Herrlich and G. Preuß, eds.), Lecture Notes in Math. **719**, Springer-Verlag, Berlin – Heidelberg – New York 1979, 157-166.

[HH] Hoffmann, R.-E., and Hofmann, K.H. (eds.), *Continuous Lattices and Their Applications, Proc. Bremen 1982,* Lecture Notes in Pure and Appl. Math. **101**, Marcel Decker, New York 1985.

[HM1] Hofmann, K.H., and Mislove, M., Local compactness and continuous lattices, in [BH], 209-248.

[HM2] Hofmann, K.H., and Mislove, M., Free objects in the category of completely distributive lattices, in [HH], 129-151.

[Is1] Isbell, J.R., Directed unions and chains, *Proc. Amer. Math. Soc.* **17** (1966), 1467-1468.

[Is2] Isbell, J.R., Completion of a construction of Johnstone, *Proc. Amer. Math. Soc.* **85** (1982), 333-334.

[Iw] Iwamura, T., A lemma on directed sets, *Zenkoku Shijo Sugaku Danwakai* **262** (1944), 107-111 [Japanese].

[Jo1] Johnstone, P.T., *Stone Spaces,* Cambridge University Press, Cambridge 1982.

[Jo2] Johnstone, P.T., Scott is not always sober, in [BH], 282-283.

[Ju] Jung, A., *Cartesian Closed Categories of Domains,* Ph. D. Thesis, Technische Hochschule Darmstadt 1988.

[Ka] Katriňák, T., Pseudokomplementäre Halbverbände, *Mat. Časopis* **18** (1968), 121-143.

[Kr] Krasner, M., Un type d'ensembles semi-ordonnés et ses rapports avec une hypothèse de M. A. Weil, *Bull. Soc. Math. France* **67** (1939), 162-176.

[La1] Lawson, J.D., The duality of continuous posets, *Houston J. Math.* **5** (1979), 357-386.

[La2] Lawson, J.D., The versatile continuous order, in: Lecture Notes in Comput. Sci. **298**, Springer-Verlag, Berlin – Heidelberg – New York 1987, 134-160.

[Le] Levy, A., Axioms of multiple choice, *Fund. Math.* **50** (1962), 475-483.

[Ma] Maeda, F., *Kontinuierliche Geometrien,* Springer-Verlag, Berlin – Göttingen – Heidelberg 1958.

[Ml] MacLane, S., *Categories for the Working Mathematician,* Springer-Verlag, Berlin – Heidelberg – New York 1971.

[My] Markowsky, G., Chain-complete posets and directed sets with applications, *Algebra Universalis* **6** (1976), 53-68.

[MS] Mayer-Kalkschmidt, J., and Steiner, E., Some theorems in set theory and applications in the ideal theory of partially ordered sets, *Duke Math. J.* **31** (1964), 287-290.

[MMT] McKenzie, R.N., McNulty, G.F., and Taylor, W.F., *Algebras, Lattices and Varieties I,* Wadsworth and Brooks, Monterey, CA. 1987.

[Me] Meseguer, J., Order completion monads, *Algebra Universalis* **16** (1983), 63-82.

[Mo] Moore, G.H., *Zermelo's Axiom of Choice,* Springer-Verlag, Berlin – Heidelberg – New York 1982.

[Ms] Mostowski, A., Über die Unabhängigkeit des Wohlordnungssatzes vom Ordnungsprinzip, *Fund. Math.* **32** (1939), 201-259.

[Ne] Nelson, E., $\mathcal{Z}$-continuous algebras, in [BH], 315-334.

[No] Novak, D., Generalization of continuous posets, *Trans. Amer. Math. Soc.* **272** (1982), 645-667.

[Pa] Papert, S., Which distributive lattices are lattices of closed sets? *Proc. Cambridge Philos. Soc.* **55** (1959), 172-176.

[Pl] Plotkin, G., The category of complete partial orders: a tool for making meanings, in: *Proc. Summer School on Foundations of Artificial Intelligence and Computer Science*, Istituto di Scienze dell' Informazione, Univ. di Pisa 1978.

[Ro] Rosický, J., On a characterization of the lattice of m-ideals of an ordered set, *Arch. Math. (Brno)* **8** (1972), 137-142.

[Sch1] Schmidt, J., Über die Rolle der transfiniten Schlußweisen in einer allgemeinen Idealtheorie, *Math. Nachr.* **7** (1952), 165-182.

[Sch2] Schmidt, J., Einige grundlegende Begriffe und Sätze aus der Theorie der Hüllenoperatoren, *Ber. Math. Tagung Berlin 1953*, 21-48.

[Sch3] Schmidt, J., Each join-completion of a partially ordered set is the solution of a universal problem, *J. Austral. Math. Soc.* **17** (1974), 406-413.

[Sc1] Scott, D.S., Continuous lattices, in: *Toposes, Algebraic Geometry and Logic* (F.W. Lawvere, ed.), Lecture Notes in Math. **274**, Springer-Verlag, Berlin – Heidelberg – New York 1972.

[Sc2] Scott, D.S., Domains for denotational semantics, in: *ICALP 82* (M. Nielsen and E. M. Schmidt, eds.), Lecture Notes in Comput. Sci. **140**, Springer-Verlag, Berlin – Heidelberg – New York 1982.

[Sm] Smyth, M.B., The largest cartesian closed category of domains, *Theoret. Comput. Sci.* **27** (1983), 109-119.

[SP] Smyth, M.B., and Plotkin, G., The category theoretic solution of recursive domain equations, *SIAM J. Comput.* **11** (1982), 761-783.

[SS] Steen, L.A., and Seebach, J.A., *Counterexamples in Topology,* Holt, Rinehart & Winston, New York 1970.

[St1] Stone, M.H., The theory of representations for Boolean algebras, *Trans. Amer. Math. Soc.* **40** (1936), 37-111.

[St2] Stone, M.H., Topological representation of distributive lattices and Brouwerian logics, *Časopis Pešt. Mat.* **67** (1937), 1-25.

[Sty] Stoy, J.E., *Denotational Semantics: The Scott - Strachey Approach to Programming Language Theory,* MIT Press, Cambridge, MA. 1977.

[Sz] Szpilrajn, E., Sur l'extension de l'ordre partiel, *Fund. Math.* **16** (1930), 386-389.

[Ve] Venugopalan, P., $\mathcal{Z}$-continuous posets, *Houston J. Math.* **12** (1986), 275-294.

[We] Weihrauch, K., *Computability,* EATCS Monogr. Theoret. Comp. Sci. **9**, Springer-Verlag, Berlin – Heidelberg – New York 1987.

[Wi] Witt, E., Beweisstudien zum Satz von M. Zorn, *Math. Nachr.* **4** (1951), 434-438.

[WWT] Wright, J.B., Wagner, E.G., and Thatcher, J.W., A uniform approach to inductive posets and inductive closure, *Theoret. Comput. Sci.* **7** (1978), 57-77.

[Ze1] Zermelo, E., Beweis, daß jede Menge wohlgeordnet werden kann, *Math. Ann.* **59** (1904), 514-516.

[Ze2] Zermelo, E., Neuer Beweis für die Möglichkeit einer Wohlordnung, *Math. Ann.* **65** (1908), 107-128.

A Boolean Formalization of Predicate Calculus

Isidore FLEISCHER
Department of Mathematics and Statistics
University of Windsor
Windsor, Ontario, N9B 3P4
Canada

Abstract

It is proposed that a more convenient formalization of predicate calculus is as a free Boolean algebra with extrema for the subsets of variable renaming, these extrema functioning as the quantifiers. In support of this proposal, an ab initio development of the calculus is sketched, a comparison with the standard treatment (which in effect construes the quantifiers as certain closure operators) is made and a proof of the Gödel completeness theorem based on this formalization is presented.

A predicate calculus is intended to deal symbolically with a species of relational structure.

For example, if one wishes to symbolize the notion of a partially ordered set, say qua set equipped with an irreflexive transitive relation, one could write down the axioms

$$\neg(x < x)$$

$$(x < y) \mathbin{\&} (y < z) \rightarrow x < z.$$

To be more explicit, one should precede these formulae with universal quantifiers — thus the first should read $\forall x \neg(x < x)$.

Examining these formulae, one sees that one has need for an "alphabet" of symbols for the order: $<$, variables: $x, y, z, \ldots$, propositional connectives: $\neg, \&, \rightarrow$, and quantifiers: $\forall x, \forall y, \ldots$.

In general a relational structure is a set equipped with a family of relations or predicates, each of a finite number of arguments. The "type" of the structure is determined by this family (with the number of arguments in each relation specified). To formulate statements such as axioms one requires symbols, one for each of the predicates in the family, as well as variables and logical operators as above. From these materials one constructs formulae — "correctly": i.e. so as to be interpretable meaningfully in the relational structures of this type; although this correctness can be determined without reference to the structures, i.e. via rules of formula construction, thus "syntactically". Each predicate symbol

I. G. Rosenberg and G. Sabidussi (eds.), Algebras and Orders, 193–198.

comes equipped with its finite number of argument places; filling these with (not necessarily distinct) variables results in an "atomic formula"; closing the atomic formulae under (free, or formal) composition with the logical operators (the quantifiers and $\neg$ operating as unary, & and $\rightarrow$ as binary, operators) yields the system of all formulae. Construing this system as the absolutely free algebra AF on the atomic formulae as generators is a slight deformation of usual practice, in which formulae are identified with expressions in some specific symbolism which represent the elements of this algebra, often defined recursively to avoid reference to a completed totality.

One interprets these formulae ("semantics") in relational structures of the appropriate type: i.e. in sets equipped with relations corresponding to the formal predicates and which have the right number of arguments. To obtain a "truth-value" for a formula $x < y$ in a set N equipped with a binary relation $<$, one must still substitute specific elements of N for the variables x and y: e.g. if 3 is substituted for x and 2 for y then $x < y$ obtains the value "false" in the natural numbers. Composite formulae obtain truth-values in accord with the "truth-tables" of their generating connectives: e.g. $(x < y)\&(y < z)$ is true iff both $x < y$ and $y < z$ are, $\neg(x < y)$ iff $x < y$ is not, and $\forall y(x < y)$ iff $x < y$ is for every substitution for y (that for x being held fixed).

It is convenient to substitute simultaneously for all variables: Call *substitution* any map of the variables V to a set M equipped with relational interpretations for the formal predicates. Every such assigns truth-values to all the formulae in AF. Formulae with the same truth-value for every substitution (in every relationally equipped M) are *semantically equivalent*; such formulae have the same semantic content and one may as well define substitution and interpretation for their equivalence class rather than for them. More generally, one formula *semantically implies another* if for every substitution, the truth of the former entails the truth of the latter; this defines a partial order on the set of all formulae modulo the semantic equivalence K. This set AF/K has in addition a quite elaborate algebraic structure to whose description we now turn.

That propositional connectives are evaluated by truth-tables comes to each substitution acting as a homomorphism (for these connectives) to the two-element Boolean algebra: hence K validates all Boolean identities and AF/K is a Boolean algebra (a subalgebra of the powerset of substitutions); semantic implication is just its order relation. Since quantification respects semantic implication, the quantifiers pass to this "semantic Lindenbaum" algebra B as isotone selfmaps. One has moreover $\forall x\varphi \leq \varphi$, with equality if φ is "independent of" x: i.e. has the same truth-value on substitutions differing only at x. The inequality will also hold on replacing x at all its occurrences in (some formula mapping on) φ by another variable y not used in its generating atomics; since $\forall x\varphi$ is still independent of y, one even has $\forall x\varphi = \forall y\forall x\varphi \leq \forall y\varphi^*$ for the result φ^* of the replacement, hence equality by symmetry. $\forall x\varphi$ is even the infimum of the φ^* for any infinite set of the y not used in generating (some formula in the class) φ: For if ψ is any common lower bound, then since it must be independent of one of these $y, \psi = \forall y\psi \leq \forall y\varphi^* = \forall x\varphi$ — dually, $\exists x\varphi$ is their supremum.

The "dummy variable" property, $\forall x\varphi = \forall y\varphi^*$, permits extending variable replacement to all of B (whence $\forall x\varphi \leq$ all replacements for x in φ, hence = their inf). Let (x/y) denote the operation of replacing variable x in a formula (insofar as it occurs, else identity) by variable y. For atomic formulae, this is unproblematic: change every x filling an argument place in the predicate symbol to y, leaving the other variables unchanged. This operation is extended to commute with propositional connectives and quantifiers in variables $z \neq x, y$; on the formulae generated by these (i.e. in which neither x nor y is quantified) it preserves semantic equivalence, and since every class includes one of these, it passes to an everywhere defined selfmap on the algebra of equivalence classes. (That (x/y) cannot be defined directly on formulae quantified in y can be seen e.g. by considering $\exists y(x < y)$.) Since it commutes with the propositional connectives, this induced selfmap is a Boolean endomorphism which commutes with quantifiers in variables other than x, y; and the action preserves equalities between finite compositions of replacements. One now has the operator equation $\forall x = \bigwedge_u (x/u)$ for any infinite set of u; and dually for $\exists x$.

In summary: B is a Boolean algebra, equipped with an endomorphic monoid action by the transformation monoid generated by the (x/y), with common extrema for $\{(x/y)\varphi\}$ for any infinite set of y (for each x and φ), having each φ fixed except for replacement of finitely many x, and having (x/y) commute with extrema of $(z/\cdot)$ for $z \neq x, y$. Actually, since the immediately preceding conditions entail the dummy variable operator identity $\forall z'(z/z') = \forall z(z'/z)$, it suffices to have it commute for some $z \neq x, y$ in order to have it commute for all such. Moreover, this commuting is equivalent to (x/y) preserving these extrema: e.g. $\bigwedge_u (x/y)(z/u) = \bigwedge_{u \neq x}(z/u)(x/y) = (x/y)\forall z$; whence also to preserve the extrema for quantifiers in x and y: applied to φ independent of $z \neq x, y$, $(x/y)\bigwedge_u (x/u)(z/x) = (x/y)\bigwedge_u (z/u)(x/z) = \bigwedge_u (x/u)$ and $(x/y)\bigwedge_u (y/u)(z/y) = (x/y)\bigwedge_u (z/u)(y/z) = \bigwedge_u (x/y)(y/u)$. Thus the (x/y), as a consequence of commuting with extrema of $(z/\cdot)$ for infinitely many z, are *complete* endomorphisms in that they preserve the extrema which yield the quantifiers, as well as the Boolean operations. This completeness yields the commutativity of like quantifiers: e.g. $\forall x \forall y = \forall y \forall x$.

The goal of an "axiomatization" of the predicate calculus is a *syntactic* description of the semantic equivalence: i.e. of the kernel of the map from the system of formulae to the algebra of semantic equivalence classes, which does not appeal to their interpretation. This is usually presented by means of "axioms" and "rules of inference" which are used to generate a quotient partial order from the axioms. The effect is to make the quotient a Boolean algebra with the additional unary selfmaps induced by the quantifiers — although the passage to the quotient is usually not carried through, the system presented remaining that of the formulae with variable dependence described by a recursively defined "occurrence" and the attained preorder a non-antisymmetric relation of "implication". The axiomatization is called "sound" if this syntactic implication is semantically valid — i.e. if the kernel of interdeducibility is contained in the semantic kernel — and "complete" if it even coincides with it.

The axiomatization proposed here is the kernel of the map to the *free partial infinitary Boolean algebra* on the atomic formulae as generators. "Partial infinitary" refers to the presence of infima and suprema for the subsets generated by the replacements (which will be seen to extend to the algebra from the generators by freeness) of a single variable in an element by every other; "free" refers to the algebra's admitting unique extension, of every set map of the atomics into any complete algebra, to a Boolean morphism preserving these extrema. Since the formulae are the absolutely free algebra on the atomic formulae as generators, the identity on them extends uniquely to a morphism (which converts the propositional connectives to Boolean operations and the quantifiers to extrema) from the formulae to this free algebra. This is the map whose kernel is the proposed axiomatization.

Soundness has in essence been shown: since B is a Boolean image of the formulae, which has extrema for the images of the subsets of replacement instances, the universal property of the free algebra entails that the map to it from the formulae factors that to B, whence the former's kernel is contained in the latter's. (To invoke the universal property, embed B in its MacNeille completion, which is a complete Boolean extension preserving all existent extrema.) To prove completeness of the axiomatization, it must be shown that the quotient map is injective which (since it is a Boolean morphism) comes to showing that no non-zero element is sent on zero: i.e. holds for some substitution. This is the principal content of the Gödel completeness theorem, the major result in the subject (which will be proved below).

Here is the construction of this free algebra: Start with the set of atomic formulae — i.e. with the predicate symbols filled in all possible ways with variables — and form the free (finitary) Boolean algebra they generate. This algebra F is characterized as admitting unique extension to a homomorphism of every set map from the generators into any Boolean algebra. Consequently the variable replacements, which are selfmaps on the generators, extend uniquely to endomorphisms of F, thus equipping it with an action by variable replacement. This F should now be completed so as to have extrema for the subsets of variable replacement of every individual variable in each of its elements — and freely: i.e. this further Boolean extension F' is characterized as admitting unique extension, to a morphism which preserves these extrema, of every morphism mapping F into any complete algebra.

One must still verify that the supremum created for a variable replacement subset is also one for any infinitely many of its terms: i.e. that any upper bound in F', for infinitely many replacements of a variable in a $\varphi \in F$, already bounds all of them. If suffices to see this for upper bounds which are finite joins of the extrema used to generate F' over F (since every element is a finite meet of such); moreover the infs in this finite join can be replaced by an arbitrary one of their terms, by infinite distributivity. Now if the sup $\bigvee$ of all replacement instances in a $\psi \in F$ were $\geq \varphi$ then so would be some sup of finitely many — else the finite sups would generate an ideal in F not containing φ which would yield a $\mathbf{2}$-valued morphism sending φ on 1 and extend to F' to send this sup on 0. Replace the variable in φ by one of the infinitely many not initially in φ or ψ, and replacement by which is dominated by $\bigvee$. The dominating finite sup of instances of ψ may of course involve this variable but all replacement instances of the latter are bounded by the infinite sup $\bigvee$, hence so are all the replacement instances of this variable in φ.

To extend (x/y) to F', observe that it is absorbed by the (x/u) and commutes with the (z/u) for $z \neq x, y$ except for $u = x$ which it converts to y — thus (x/y) extends to F' (by the universal property) so as to be absorbed by $\bigvee_u(x/u)$ and to commute with $\bigvee_u(z/u)$ for $z \neq x, y$. Uniqueness ensures that composition of the extensions continues to respect transformational composition; $\bigvee(x/u)\varphi$ absorbs (x/y), as it should, and all the replacements that φ absorbed.

The process of forming F' from F may now be applied to F' to yield F'' — of course retaining the extrema created in F' (definable by equations by distributivity); only morphisms preserving these extrema are supposed extendible to F'' — and so on; the union F^* of the ascending chain $F \subset F' \subset F'' \subset \ldots$ is the desired algebra — it is a Boolean algebra, closed both for the action of the variable replacements and for the quantifiers construed as extrema.

The inductive build-up of this free algebra from the atomics corresponds exactly to the free inductive generation of all formulae from the atomics by alternate formal application of propositional connectives and quantifiers. In fact, the free algebra might just as well serve in place of the formulae, which become dispensable, their sole function having been to fix the assertions admitted by the predicate calculus, a service accomplished more effectively by the free algebra. This is the "Boolean formalization" referred to in the title.

We pause to recall the usual axiomatization and to compare it with this one. In the former, one has to keep track of when a variable "occurs freely" in a formula. This relation is determined recursively via the formal generation of the formulae from the atomics by the logical operators. The variables which occur freely in an atomic formula are just those which fill the argument places of its predicate symbol; free occurrence is preserved by the action of propositional connectives and quantifiers in other variables, destroyed by quantification over that variable.

Replacement of one variable by another is authorized when the free occurrences of the replaced variable (if any) would not cause these occurrences of the replacing variable to become bound — thus just when they are not in the scope of a quantifier of the replacing variable. It is defined so as to commute with propositional connectives and quantification over other variables.

Besides converting the action of the propositional connectives to Boolean modulo equivalence, the standard axiomatization imposes: $(x/y)\varphi$ (where authorized) $\leq \exists x\varphi$ and $\varphi \leq \psi$, for a ψ in which x does not occur free, entails $\exists x\varphi \leq \psi$. With $y = x$ one obtains $\varphi \leq \exists x\varphi$ and with $\varphi = \psi$, $\exists x\psi \leq \psi$; i.e. $\exists x$ is absorbed by the ψ free of x, hence $\exists x$ is a closure operator with image the Boolean subalgebra of the ψ free of x. Since $\exists y\psi$ is still free of x, the composite $\exists y\exists x$ is idempotent, hence a closure operator equal to $\exists x\exists y$, which maps on the elements free of x and y. Observe that it is also possible to change a quantified variable x to a z not appearing in $\varphi : \exists z(x/z)\varphi \leq \exists z\exists x\varphi = \exists x\varphi = \exists x(z/x)(x/z)\varphi \leq \exists z(x/z)\varphi$: in particular, every variable replacement is authorized for some equivalent formula.

For the comparison: That $\bigvee_u(x/u)$ functions as a closure operator to the Boolean subalgebra of elements ψ which absorb x-replacements follows since $\psi \geq \varphi$ entails $\psi = (x/u)\psi \geq (x/u)\varphi$ for each u hence $\geq \exists x\varphi$. Conversely, with each $\exists x$ a closure operator $\geq (x/u)$ to the x-fixed elements, if $\psi \geq (x/z)\varphi$ for a z whose replacements are absorbed by both ψ and φ then $\psi \geq \exists z(x/z)\varphi = \exists x\varphi$.

It remains to see that the axiomatization is "complete": that is, that the canonical Boolean morphism, from the free partial infinitary algebra F^* to the semantic algebra B, is an isomorphism. This is the principal content of the Gödel completeness theorem; it amounts to showing that every non-zero element of F^* holds at some substitution to a structure. The proof below is in essence that given in [L, pp. 51–55] with the simplifications that the Boolean formulation brings.

The sought-for structure will be built on the variables but it may be necessary to enlarge these. Adjoining new variables enlarges the collection of free generators of the algebra; since the extrema which represent the quantifiers can be calculated using any infinite set of variable replacements, they retain their value under variable adjunction and the original algebra is embedded with preservation of all its structure. A filter Φ in the original algebra, which includes an existential quantification $\exists x\varphi(x)$, meets every replacement instance $(x/x')\varphi$ by any of the new variables x' – for if some $\psi \in \Phi$ were disjoint from $(x/x')\varphi$ then $0 = \bigvee_u (x'/u)0 = \bigvee_u (x'/u)[\psi \wedge (x/x')\varphi] = \bigvee_u \psi \wedge (x/u)\varphi = \psi \wedge \exists x\varphi(x)$. Thus Φ and $(x/x')\varphi$ generate a proper filter in the enlarged algebra.

It will suffice to show that every $\varphi \neq 0$ in F is included in an ultrafilter, in a possibly variable enlarged algebra, which includes with every existential quantification $\exists x\psi$ some replacement instance $(x/x')\psi$ – for such an ultrafilter determines a structure on the variables for which it consists of the elements holding at the identity substitution.

Start with the principal filter generated by φ and expand it to an ultrafilter Φ. If $\exists x\psi \in \Phi$ has no replacement instance in Φ, adjoin a new variable and expand Φ to a filter having such an instance; continue doing so (taking unions of filters to synthesize infinitely many adjunctions) till all $\exists x\psi \in \Phi$ have replacement instances in Φ. Expand the Φ attained to an ultrafilter in the variable-expanded algebra and repeat. There results an increasing sequence of free algebras each with an ultrafilter, which contains the preceding one, and includes a replacement instance for each of the latter's existential quantifications. Their union will be an ultrafilter, in the union of the algebras, which includes replacement instances for each of its existential quantifications.

References

[H] Halmos, P.R., *Lectures on Boolean Algebras*, Van Nostrand, Princeton, 1963.

[HA] Hilbert, D., Ackermann, W., *Principles of Mathematical Logic*, New York, 1950.

[L] Lyndon, R.C., *Notes on Logic*, Van Nostrand, Princeton, 1966.

[R] Rieger, L., *Algebraic Methods of Mathematical Logic*, Academia, Prague, 1967.

Lectures on Free Lattices

Ralph FREESE*
Department of Mathematics
University of Hawaii, Honolulu
Hawaii 96822
U.S.A

Abstract

These lectures notes present the theory of free lattices, assuming only a basic understanding of lattice theory. They begin with Whitman's solution to the word problem and his canonical form. The well known consequences of these are given as well as several lesser known consequences, such as the continuity of free lattices, the existence of a fixed point free unary polynomial on a free lattice, and the fact that finite sublattices of a free lattice satisfy a nontrivial lattice equation. The theory of covers in free lattices is developed and some of the consequences explored. Tschantz's Theorem and a new characterization of semisingular elements are discussed and some important consequences of these results are given such as the existence of dense maximal chains in intervals of a free lattice.

These are expanded lectures notes to a series of five lectures given at the séminaire de mathématiques supérieures: Algèbres et ordres, at the Université de Montréal during the summer of 1991. A much more thorough treatment of free lattices is given in the forthcoming monograph *Free Lattices*, written by J. Ježek, J. B. Nation, and the author, [19].

There are several reasons for studying free lattices. Free lattices are in some sense the most general lattices. Namely, every lattice is a homomorphic image of a free lattice. For this reason, free lattices are important in the study of lattice structure theory. Free lattice techniques are an important tool in lattice theory. For example, Dilworth's famous result [11] that every lattice can be embedded into a uniquely complemented lattice uses free lattice techniques. Naturally free lattices are closely associated with equations in lattices and the study of lattice varieties is closely allied with free lattices. R. McKenzie's important study of lattice varieties, [25], is a good example of this connection.

Free lattices are also important in the study of algebra. Splitting equations, a concept invented by McKenzie in his study of lattice varieties, are important in the study of congruence lattices of algebras, particularly their equational properties, see [2], [5], [6], [9], and

*The author would like to thank Jennifer Hyndman upon whose careful notes these lecture notes are based. She made numerous suggestions which greatly improved the readability of these notes. She is also responsible for many of the TEXnical details of this manuscript.

The author would like to thank J. B. Nation and J. Ježek for many helpful suggestions. He would also like to thank Ivo Rosenberg, the Université de Montréal, and NATO for inviting him to this conference.

This research was supported by NSF Grant no. DMS89–01756.

I. G. Rosenberg and G. Sabidussi (eds.), Algebras and Orders, 199–237.

[17]. The concept of a bounded homomorphism of a free lattice finds its way into algebra, see, for example, [16].

But perhaps the most important reason for studying free lattices is that they are intrinsically interesting. At first glance free lattices would appear to be insurmountably difficult, but as we study them and they reveal more of their structure, they become more and more fascinating. Standing on the shoulders of others who have worked in this area, notably Whitman and Jónsson, it is possible to uncover some of the mysteries of free lattices. Hopefully these notes will convey this fascination to the reader.

1 Basics

In this section we present some basic notations about free lattices and Alan Day's doubling construction. The notation of these notes follows the conventions in [26].

If $a < b$ are elements in a lattice $\mathbf{L}$ and there is no $c \in L$ with $a < c < b$, then we say that *a is covered by b*, and we write $a \prec b$. In this situation we also say that *b covers a* and write $b \succ a$. In addition we say that b is an *upper cover* of a and that a is a *lower cover* of b. We also define a nameless equivalence relation on L by saying c is equivalent to d if there is a finite sequence $c = c_0, c_1, \ldots, c_n = d$ such that c_i either covers or is covered by c_{i+1}. The blocks of the equivalence relation are called the *connected components* of the covering relation of $\mathbf{L}$. The *connected component* of $a \in L$ is the block containing a.

A useful construction for free lattice theory is Alan Day's doubling construction. We will use this construction in this section to derive one of the basic properties of free lattices, known as Whitman's condition, following Day's approach using doubling [4]. The doubling construction also plays an crucial role in the proof of Day's important result [7] that free lattices are weakly atomic.

Let $\mathbf{L}$ be a lattice. A subset C of L is *convex* if whenever a and b are in C and $a \le c \le b$, then $c \in C$. Of course an interval of a lattice is a convex set as are lower and upper pseudo-intervals. A subset C of L is a *lower pseudo interval* if it is a finite union of intervals, all with the same least element. An *upper pseudo interval* is of course the dual concept.

Let C be a convex subset of a lattice $\mathbf{L}$ and let $L[C]$ be the disjoint union $(L-C)\cup(C\times 2)$. Order $L[C]$ by $x \le y$ if one of the following holds.

1. $x, y \in L - C$ and $x \le y$ holds in $\mathbf{L}$,
2. $x, y \in C \times 2$ and $x \le y$ holds in $C \times 2$,
3. $x \in L - C$, $y = \langle u, i\rangle \in C \times 2$, and $x \le u$ holds in $\mathbf{L}$, or
4. $x = \langle v, i\rangle \in C \times 2$, $y \in L - C$, and $v \le y$ holds in $\mathbf{L}$.

Let $\lambda : L[C] \to L$ be defined by

$$\lambda(x) = \begin{cases} x & \text{if } x \in L - C \\ v & \text{if } x = \langle v, i\rangle \in C \times 2 \end{cases}$$

The next theorem shows that, under this order, $L[C]$ is a lattice, denoted $\mathbf{L}[C]$.

Theorem 1.1 *Let C be a convex subset of a lattice* $\mathbf{L}$. *Then* $\mathbf{L}[C]$ *is a lattice and* $\lambda : \mathbf{L}[C] \to \mathbf{L}$ *is a lattice epimorphism.*

Proof Routine calculations show that $\mathbf{L}[C]$ is a partially ordered set. Let $x_i \in L - C$ for $i = 1, \ldots, n$ and let $\langle u_j, k_j\rangle \in C \times 2$ for $j = 1, \ldots, m$. Let $v = \bigvee x_i \vee \bigvee u_j$ in $\mathbf{L}$ and let $k = \bigvee k_j$ in $\mathbf{2}$. Of course, if $m = 0$, then $k = 0$. Then in $\mathbf{L}[C]$,

$$(1) \qquad x_1 \vee \cdots \vee x_n \vee \langle u_1, k_1\rangle \vee \cdots \vee \langle u_m, k_m\rangle = \begin{cases} v & \text{if } v \in L - C \\ \langle v, k\rangle & \text{if } v \in C \end{cases}$$

To see this let y be the right side of the above equation, i.e., let $y = v$ if $v \in L - C$ and $y = \langle v, k\rangle$ if $v \in C$. It is easy to check that y is an upper bound for each x_i and each $\langle u_j, k_j\rangle$. Let z be another upper bound. First suppose $z = \langle a, r\rangle$ where $a \in C$. Since z is an upper bound, it follows from the definition of the ordering that $a \geq v$ and $r \geq k$ and this implies $z \geq y$. Thus y is the least upper bound in this case. The case when $z \notin C$ is even easier. The formula for meets is of course dual. Thus $\mathbf{L}[C]$ is a lattice and it follows from equation (1) and its dual that λ is a homomorphism which is clearly onto $\mathbf{L}$. □

Free lattices

Since the lattice operations are both associative, we define *lattice terms* over a set X, and their associated lengths, in a manner analogous to the way they are defined for rings. Each element of X is a term of length 1. Terms of length 1 are called *variables.* If $t_1, \ldots, t_n$ are terms of lengths $k_1, \ldots, k_n$, then $(t_1 \vee \cdots \vee t_n)$ and $(t_1 \wedge \cdots \wedge t_n)$ are terms with length $1 + k_1 + \cdots + k_n$. When we write a term we usually omit the outermost parentheses. Notice that if x, y, and $z \in X$ then

$$x \vee y \vee z \qquad x \vee (y \vee z) \qquad (x \vee y) \vee z$$

are all terms (which always represent the same element when interpreted into any lattice) but the length of $x \vee y \vee z$ is 4, while the other two terms are both of length 5. Thus our length function gives preference to the first expression, i.e., it gives preference to expressions where unnecessary parentheses are removed. Also note that the length of a term is the number of variables, counting repetitions, plus the number of pairs of parentheses (i.e., the number of left parentheses). The length of a term is also called its *rank.* The set of *subterms* of a term t is defined in the usual way: if t is a variable then t is the only subterm of t and, if $t = t_1 \vee \cdots \vee t_n$ or $t = t_1 \wedge \cdots \wedge t_n$, then the subterms of t consist of t together with the subterms of $t_1, \ldots, t_n$.

By the phrase '$t(x_1, \ldots, x_n)$ is a term' we mean that t is a term and $x_1, \ldots, x_n$ are (pairwise) distinct variables including all variables occurring in t. If $t(x_1, \ldots, x_n)$ is a term and $\mathbf{L}$ is a lattice, then $t^{\mathbf{L}}$ denotes the interpretation of t in $\mathbf{L}$, i.e., the induced n–ary operation on $\mathbf{L}$. If $a_1, \ldots, a_n \in L$, we will usually abbreviate $t^{\mathbf{L}}(a_1, \ldots, a_n)$ by $t(a_1, \ldots, a_n)$. Very often in the study of free lattices X will be a subset of L. In this case we will use $t^{\mathbf{L}}$ to denote $t^{\mathbf{L}}(x_1, \ldots, x_n)$.

If $s(x_1, \ldots, x_n)$ and $t(x_1, \ldots, x_n)$ are terms and $\mathbf{L}$ is a lattice in which $s^{\mathbf{L}} = t^{\mathbf{L}}$ as functions, then we say the *equation* $s \approx t$ *holds* in $\mathbf{L}$.

Let $\mathbf{F}$ be a lattice and $X \subseteq F$. We say that $\mathbf{F}$ is *freely generated* by X if X generates $\mathbf{F}$ and every map from X into any lattice $\mathbf{L}$ extends to a lattice homomorphism of $\mathbf{F}$ into $\mathbf{L}$. Since X generates $\mathbf{F}$, such an extension is unique. It follows easily that if $\mathbf{F}_1$ is freely generated by X_1 and $\mathbf{F}_2$ is freely generated by X_2 and $|X_1| = |X_2|$, then $\mathbf{F}_1$ and $\mathbf{F}_2$ are isomorphic. Thus if X is a set, a lattice freely generated by X is unique up to isomorphism. We will see that such a lattice always exists. It is referred to as the *free lattice over* X and is denoted $\mathbf{FL}(X)$. If n is a cardinal number, $\mathbf{FL}(n)$ denotes a free lattice whose free generating set has size n.

To construct $\mathbf{FL}(X)$, let $\mathrm{T}(X)$ be the set of all terms over X. $\mathrm{T}(X)$ can be viewed as an algebra with two binary operations. Define an equivalence relation $\sim$ on $\mathrm{T}(X)$ by $s \sim t$ if and only if the equation $s \approx t$ holds in all lattices. It is not difficult to verify that $\sim$ restricted to X is the equality relation, that $\sim$ is a congruence relation on $\mathrm{T}(X)$, and that $\mathrm{T}(X)/\sim$ is a lattice freely generated by X, provided we identify each element of $x \in X$ with its singleton set, $\{x\}$. This is the standard construction of free algebras, see, for example, [1] or [26].

This construction is much more useful if we have an effective procedure which determines, for arbitrary lattice terms s and t, if $s \sim t$. The problem of finding such a procedure is informally known as the *word problem* for free lattices. In [31], Whitman gave an efficient solution to this word problem. Virtually all work on free lattices is based on his solution.

If $w \in \mathbf{FL}(X)$, then w is an equivalence class of terms. Each term of this class is said to *represent* w and is called a *representative* of w. More generally, if $\mathbf{L}$ is a lattice generated by a set X, we say that a term $t \in \mathrm{T}(X)$ represents $a \in L$ if $t^{\mathbf{L}} = a$.

Lemma 1.2 *Let* $\mathbf{FL}(X)$ *be the free lattice generated by* X *and let* Y *be a finite, nonempty subset of* X. *Then*

1. $\bigwedge Y$ *is join prime in* $\mathbf{FL}(X)$, *and*

2. *every element of* $\mathbf{FL}(X)$ *is either above* $\bigwedge Y$ *or below* $\bigvee Z$, *for some finite* $Z \subseteq X - Y$.

If X *is finite, every element of* $\mathbf{FL}(X)$ *is either in the filter* $1/\bigvee Y$ *or the ideal* $\bigvee(X - Y)/0$.

Proof Let $f : X \to \mathbf{2}$ be given by $f(y) = 1$ if $y \in Y$ and $f(z) = 0$ if $z \in X - Y$. Since X is a free generating set, f can be extended to a homomorphism which we also denote by f.

The idea of the proof is to show that $f^{-1}(1)$ is the filter $1/\bigwedge Y$ and $f^{-1}(0)$ is the ideal generated by $X - Y$. So suppose that $f(w) = 0$ and let t be a term of minimal rank representing w. We show by induction on the rank of t that $w \leq \bigvee Z$ for some finite subset $Z \subseteq X - Y$. If t is a variable then this is immediated from the definition of f. If $t = t_1 \vee \cdots \vee t_n$ and $w_i = t_i^{\mathbf{FL}(X)}$ then clearly $f(w_i) = 0$ for all i, and so by induction, $w_i \leq Z_i$ for some finite subset Z_i of $X - Y$. In this case we let $Z = \cup Z_i$ and clearly $w \leq \bigvee Z$. If $t = t_1 \wedge \cdots \wedge t_n$ then $f(w_i) = 0$ for some i. In this case let $Z = Z_i$. Then $w \leq w_i \leq \bigvee Z_i = \bigvee Z$.

The fact that if $f(w) = 1$, then $w \geq \bigwedge Y$ follows from a similar argument. Part (2) of the lemma follows from this and part (1) follows from part (2). The final statement also follows from part (2). □

The next lemma will be used below to characterize when a subset of a lattice generates a free lattice.

Lemma 1.3 *Let* $\mathbf{L}$ *be a lattice generated by a set* X*. Let* $x \in X$ *and suppose that for every finite subset* S *of* X*,*

$$x \le \bigvee S \quad \textit{implies} \quad x \le s \quad \textit{for some } s \in S.$$

Then (1.3) *holds for all finite subsets of* L*.*

Proof Let $\mathbf{FL}(X)$ be the free lattice generated by X and let $x \in X$. By definition of a free lattice, the identity map on X can be extended to a homomorphism $h : \mathbf{FL}(X) \to \mathbf{L}$. Let $f : X \to \mathbf{2}$ be the epimorphism with $f(x) = 1$ and $f(z) = 0$ if $z \in X - \{x\}$. By the previous lemma with $Y = \{x\}$, the kernel ψ of f consists of two blocks: the filter $1/x$ and the ideal generated by the elements $\bigvee Z$ for $Z \subseteq X - \{x\}$, Z finite. We wish to show that $\ker h \subseteq \psi$. If this were not the case then there would be elements u and $v \in \mathbf{FL}(X)$ such that $h(u) = h(v)$, $u \ge x$, and $v \le \bigvee Z$ for some finite $Z \subseteq X - \{x\}$. But then

$$h(x) = h(x \wedge u) = h(x \wedge v) \le h(\bigvee Z),$$

and, since h is the identity on X, this implies that $x \le \bigvee Z$ holds in $\mathbf{L}$, contrary to hypothesis. Thus $\ker h \le \psi$ and, by one of the standard isomorphisms theorems of algebra, there is a homomorphism $g : \mathbf{L} \to \mathbf{2}$ such that $gh = f$.

Now suppose $x \le \bigvee S$ for some $S \subseteq L$. Applying g to this inequality, we see that $g(s) = 1$ for some $s \in S$. Since $\mathbf{L}$ is generated by X, there is a term t over X such that $s = t^{\mathbf{L}}$. Then

$$1 = g(s) = g(t^{\mathbf{L}}) = gh(t^{\mathbf{FL}(X)}) = f(t^{\mathbf{FL}(X)}).$$

As in the proof of the last lemma, this implies $t^{\mathbf{FL}(X)} \ge x$. Applying h we obtain $s \ge x$, as desired. □

The next theorem shows that the free generators of a free lattice are join prime. Actually this result, Lemma 1.2, and Corollary 1.5 hold for the generators of any relatively free lattice, as was shown by Jónsson in [23].

Theorem 1.4 *Let* $\mathbf{FL}(X)$ *be the free lattice generated by* X*. If* x *and* $y \in X$ *then* $x \le y$ *if and only if* $x = y$*. Moreover, each* $x \in X$ *is join and meet prime.*

Proof Suppose $x \le y$ and let $f : \mathbf{FL}(X) \to \mathbf{2}$ be the homomorphism determined by the map $f(x) = 1$ and $f(z) = 0$ if $z \ne x$. If $x \ne y$ then, since $x \le y$, we have $1 = f(x) \le f(y) = 0$, a contradiction.

The second statement of the theorem follows from Lemma 1.2. □

Unlike the situation for groups and Boolean algebras, the free generating set of a free lattice is uniquely determined, as the next corollary shows.

Corollary 1.5 1. *If* Y *generates* $\mathbf{FL}(X)$ *then* $X \subseteq Y$*.*

2. *The automorphism group of* $\mathbf{FL}(X)$ *is isomorphic to the full symmetric group on* X*.*

Proof Each $x \in X$ is both join and meet irreducible. Hence if x is in the sublattice generated by Y, it must be in Y. This proves the first statement. The second statement follows easily from the first. □

The next corollary gives some of the basic coverings in free lattices, including the atoms and coatoms. It follows immediately from Lemma 1.2 and the fact that $0_{\mathbf{FL}(X)} = \bigwedge X$, when X is finite.

Corollary 1.6 *Let X be finite and let Y be a nonempty subset of X. Then in $\mathbf{FL}(X)$ we have the following covers:*

$$\bigvee Y \prec \bigvee Y \vee \bigwedge(X - Y) \qquad \textit{and} \qquad \bigwedge Y \succ \bigwedge Y \wedge \bigvee(X - Y).$$

The atoms of $\mathbf{FL}(X)$ are the elements $\bigwedge(X - \{x\})$, for $x \in X$. The coatoms are the elements $\bigvee(X - \{x\})$, for $x \in X$. □

Now we turn to Whitman's condition, which is the crux of the solution of the word problem for free lattices.

Theorem 1.7 *The free lattice $\mathbf{FL}(X)$ satisfies the following condition:*

(W) If $v = v_1 \wedge \cdots \wedge v_r \le u_1 \vee \cdots \vee u_s = u$ then either, $v_i \le u$ for some i, or $v \le u_j$ for some j.

Proof Suppose $v = v_1 \wedge \cdots \wedge v_r \le u_1 \vee \cdots \vee u_s = u$ but that $v_i \le u$ and $v \le u_j$ hold for no i and no j. If $v \le x \le u$ for some $x \in X$, then since x is meet prime, $v_i \le x \le u$ for some i, contrary to our assumption. Let I be the interval u/v and let $\mathbf{FL}(X)[I]$ be the lattice obtained by doubling I, see Figures 1 and 2. By the above remarks, none of the generators is doubled. This implies that X is a subset of $\mathbf{FL}(X)[I]$ and so the identity map on X extends to a homomorphism $f : \mathbf{FL}(X) \to \mathbf{FL}(X)[I]$. Since $x \notin I$, $\lambda(x) = x$, where λ is the epimorphism defined by (1). Hence $\lambda(f(w)) = w$ for all $w \in \mathbf{FL}(X)$ and this implies $f(w) = w$ if $w \notin I$. Thus it follows from (1) and its dual that

$$\begin{aligned} f(v) = f(v_1) \wedge \cdots \wedge f(v_r) &= v_1 \wedge \cdots \wedge v_r = \langle v, 1 \rangle \\ &\not\le \langle u, 0 \rangle = u_1 \vee \cdots \vee u_s \\ &= f(u_1) \vee \cdots \vee f(u_s) = f(u), \end{aligned}$$

contradicting the fact that $v \le u$. □

The condition (W) is known as *Whitman's condition.* Notice that it does not refer to the generating set and so it is inherited by sublattices. Also note that Day's doubling is a procedure for correcting (W)–failures. As such it has many uses, see [4].

Corollary 1.8 *Every sublattice of a free lattice satisfies* (W). *Every element of a lattice which satisfies* (W) *is either join or meet irreducible.* □

The next theorem gives a slight variant from [14] of Whitman's condition which is more efficient for computation and also useful in theoretical arguments.

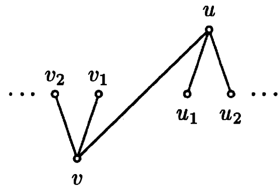


Figure 1.

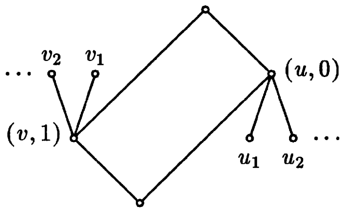


Figure 2.

Theorem 1.9 *The free lattice* $\mathbf{FL}(X)$ *satisfies the following condition:*

(W+) If $v = v_1 \wedge \cdots \wedge v_r \wedge x_1 \wedge \cdots \wedge x_n \leq u_1 \vee \cdots \vee u_s \vee y_1 \vee \cdots \vee y_m = u$, where x_i and $y_j \in X$, then either, $x_i = y_j$ for some i and j, or $v_i \leq u$ for some i, or $v \leq u_j$ for some j.

Proof Suppose we apply (W) to the inequality $v \leq u$ and obtain $x_i \leq u = u_1 \vee \cdots \vee u_s \vee y_1 \vee \cdots \vee y_m$. Then either $x_i \leq y_j$ for some j, or $x_i \leq u_k$ for some k. The former implies $x_i = y_j$ and the latter implies $v \leq u_k$. Thus (W+) holds in either case. The other possibilities are handled by similar arguments. □

Notice that (W+) replaces the test $x_i \leq u$ with the test $x_i \in \{y_1, \ldots, y_m\}$, which is of course much easier. Also notice that in applying (W+) it is permitted that some of the v_i's and u_j's are in X. Thus, for example, if $v = v_1 \wedge \cdots \wedge v_r \wedge x_1 \wedge \cdots \wedge x_n$ and $u = u_1 \vee \cdots \vee u_s$, then $v \leq u$ if and only if $v_i \leq u$ for some i or $v \leq u_j$ for some j.

Theorems 1.4 and 1.7 combine to give a recursive procedure for deciding, for terms s and t, if $s^{\mathbf{FL}(X)} \leq t^{\mathbf{FL}(X)}$ known as Whitman's solution to the word problem. In Chapter XII we will give a presentation of this algorithm more suitable for a computer (rather than for a human) and study its time and space complexity.

Theorem 1.10 *If* $s = s(x_1, \ldots, x_n)$ *and* $t = t(x_1, \ldots, x_n)$ *are terms and* $x_1, \ldots, x_n \in X$, *then the truth of*

$$s^{\mathbf{FL}(X)} \leq t^{\mathbf{FL}(X)} \tag{2}$$

can be determined by applying the following rules.

1. *If* $s = x_i$ *and* $t = x_j$, *then* (2) *holds if and only* $x_i = x_j$.

2. *If $s = s_1 \vee \cdots \vee s_k$ is a formal join then* (2) *holds if and only if $s_i^{\mathbf{FL}(X)} \leq t^{\mathbf{FL}(X)}$ holds for all i.*

3. *If $t = t_1 \wedge \cdots \wedge t_k$ is a formal meet then* (2) *holds if and only if $s^{\mathbf{FL}(X)} \leq t_i^{\mathbf{FL}(X)}$ holds for all i.*

4. *If $s = x_i$ and $t = t_1 \vee \cdots \vee t_k$ is a formal join, then* (2) *holds if and only if $x_i \leq t_j^{\mathbf{FL}(X)}$ for some j.*

5. *If $s = s_1 \wedge \cdots \wedge s_k$ is a formal meet and $t = x_i$, then* (2) *holds if and only if $s_j^{\mathbf{FL}(X)} \leq x_i$ for some j.*

6. *If $s = s_1 \wedge \cdots \wedge s_k$ is a formal meet and $t = t_1 \vee \cdots \vee t_m$ is a formal join, then* (2) *holds if and only if $s_i^{\mathbf{FL}(X)} \leq t^{\mathbf{FL}(X)}$ holds for some i, or $s^{\mathbf{FL}(X)} \leq t_j^{\mathbf{FL}(X)}$ holds for some j.*

Proof It is easy to see that all possibilities are covered by (1)–(6) and that each of these leads to a genuine reduction (except for (1), which gives the answer directly). □

We are now in a position to give an easy criterion to determine if a subset of a lattice generates a sublattice isomorphic to a free lattice.

Corollary 1.11 *Let* $\mathbf{L}$ *be a lattice which satisfies* (W) *and let* X *generate* $\mathbf{L}$*. Then* $\mathbf{L}$ *is isomorphic to* $\mathbf{FL}(X)$ *if and only if the following condition and its dual hold for all* $x \in X$ *and all finite subsets* $Y \subseteq X$*.*

$$x \leq \bigvee Y \quad \textit{implies} \quad x \in Y.$$

Proof If x and $y \in X$ satisfy $x \leq y$ then the condition with $Y = \{y\}$ implies that $x = y$ and by Lemma 1.3 each $x \in X$ is join and meet prime. Of course the identity map on X extends to a homomorphism of $\mathbf{FL}(X)$ onto $\mathbf{L}$. Since X in $\mathbf{L}$ satisfies (1)–(6) of Theorem 1.10, if s and t are terms then $s^{\mathbf{L}} \leq t^{\mathbf{L}}$ if and only if $s^{\mathbf{FL}(X)} \leq t^{\mathbf{FL}(X)}$ and thus this map must be an isomorphism. □

Corollary 1.12 *A subset* S *of a free lattice* $\mathbf{FL}(X)$ *generates a sublattice isomorphic to a free lattice if and only if for all* $s \in S$ *and all finite subsets* $Y \subseteq S$*,*

$$s \leq \bigvee Y \quad \textit{implies} \quad s \in Y,$$

and the dual condition holds. □

2 Canonical forms

The canonical form of an element plays an important role in free lattice theory. In this section we show that each element w of a free lattice has a term of least rank representing it, unique up to commutativity. This term is called the *canonical form* of w. The phrase 'unique up to commutativity' can be made precise by defining *equivalent under commutativity* to be the equivalence relation, $s \equiv t$, given by recursively applying the following rules.

1. $s, t \in X$ and $s = t$.
2. $s = s_1 \vee \cdots \vee s_n$ and $t = t_1 \vee \cdots \vee t_n$ and there is a permutation σ of $\{1, \ldots, n\}$ such that $s_i \equiv t_{\sigma(i)}$.
3. The dual of (2) holds.

The next theorem shows that if two terms both represent the same element of $\mathbf{FL}(X)$ and both have minimal rank among all such representatives, then they are equivalent under commutativity.

Later in this section we will derive the semidistributive laws from the existence of the canonical form. We also show that there is a strong connection between the canonical forms and the arithmetic of free lattices. This will allow us to define canonical form in terms of lattice theoretic properties.

The following concept is very important in lattice theory, particularly free lattice theory. Let $\mathbf{L}$ be a lattice and let A and B be finite subsets of L. We say that *a join refines B* and we write $A \ll B$ if for each $a \in A$ there is a $b \in B$ with $a \leq b$. The dual notion is called *meet refinement* and is denoted $A \gg B$. Note, however, that $A \ll B$ does not imply $B \gg A$. Also note that if A and B are both antichains, $A \ll B$, and $B \ll A$, then $A = B$. We use the term 'join refinement' because if $u = \bigvee A = \bigvee B$ and $A \ll B$ then $u = \bigvee A$ is a better join representation of u than $u = \bigvee B$ in that its elements are further down in the lattice. We will see that in free lattices there is a unique best join representation of each element, i.e., a representation that join refines all other join representations.

Theorem 2.1 *For each $w \in \mathbf{FL}(X)$ there is a term of minimal rank representing w, unique up to commutativity. This term is called the canonical form of w.*

Proof Suppose that s and t are both terms of minimal rank that represent the same element w in $\mathbf{FL}(X)$. If either s or t is in X, then clearly $s = t$.

Suppose that $t = t_1 \vee \cdots \vee t_n$ and $s = s_1 \vee \cdots \vee s_m$. If some t_i is formally a join, we could lower the rank of t by removing the parentheses around t_i. Thus each t_i is not formally a join. Now $t_i \leq s_1 \vee \cdots \vee s_m$. Applying (W) if t_i is formally a meet and using join primality if $t_i \in X$, we conclude that either $t_i \leq s_j$ for some j, or $t_i = \bigwedge t_{ij}$ and $t_{ij} \leq s$ for some j. In the second case, since s and t represent the same element, we have $t_i \leq t_{ij} \leq t$ and thus we could replace t_i with t_{ij} in t, producing a shorter term still representing w, violating the minimality of the term t. Hence in all cases there is a j such that $t_i \leq s_j$. Thus $\{t_1, \ldots, t_n\} \ll \{s_1, \ldots, s_m\}$. By symmetry, $\{s_1, \ldots, s_n\} \ll \{t_1, \ldots, t_m\}$. Since both are antichains (by the minimality) they represent the same set of elements of $\mathbf{FL}(X)$. Thus $m = n$ and after renumbering $s_i \sim t_i$. Now by induction s_i and t_i are the same up to commutativity.

If $t = t_1 \vee \cdots \vee t_n$ and $s = s_1 \wedge \cdots \wedge s_m$, then (W) implies that either $t_i \sim t$ for some i or $s_j \sim s$ for some j, violating the minimality.

The remaining cases can be handled by duality. □

Naturally we say a term is in *canonical form* if it is the canonical form of the element it represents. The following theorem gives an effective procedure for transforming a term into canonical form, i.e., if one of the conditions below fails, then t can be transformed into a term of smaller rank representing the same element. So repeated use of these conditions will transform a term into canonical form.

Theorem 2.2 *A term $t = t_1 \vee \cdots \vee t_n$, with $n > 1$, is in canonical form if and only if*

1. *each t_i is either in X or formally a meet,*
2. *each t_i is in canonical form,*
3. *$t_i \not\leq t_j$ for all $i \neq j$ (the t_i's form an antichain),*
4. *if $t_i = \bigwedge t_{ij}$ then $t_{ij} \not\leq t$ for all j.*

A term $t = t_1 \wedge \cdots \wedge t_n$, with $n > 1$, is in canonical form if and only if the duals of the above conditions hold. A term $x \in X$ is always in canonical form.

Proof All of these conditions are clearly necessary. For the converse we need to show that if t satisfies (1)–(4) then it has minimal rank among the terms which represent the same element of $\mathbf{FL}(X)$ as t. Suppose that $s = s_1 \vee \cdots \vee s_n$ is a term of minimal rank representing the same element of $\mathbf{FL}(X)$ as t. Now using (1)–(4) and the arguments of the last theorem we can show that

$$\begin{aligned} \{t_1, \ldots, t_n\} &\ll \{s_1, \ldots, s_m\} \\ \{s_1, \ldots, s_n\} &\ll \{t_1, \ldots, t_m\} \end{aligned}$$

Since both are antichains, we have that $n = m$ and after renumbering $s_i \sim t_i$, $i = 1, \ldots, n$. The proof can now easily be completed with the aid of induction. □

Let $w \in \mathbf{FL}(X)$ be join reducible and suppose $t = t_1 \vee \cdots \vee t_n$ (with $n > 1$) is the canonical form of w. Let $w_i = t_i^{\mathbf{FL}(X)}$. Then $\{w_1, \ldots, w_n\}$ are called the *canonical joinands* of w. We also say $w = w_1 \vee \cdots \vee w_n$ canonically and that $w_1 \vee \cdots \vee w_n$ is the *canonical join representation* (or *canonical join expression*) of w. If w is join irreducible, we define the canonical joinands of w to be the set $\{w\}$. Of course the *canonical meetands* of an element in a free lattice are defined dually. More generally, u is called a *subelement* of w if it is the element of $\mathbf{FL}(X)$ corresponding to some subterm of the canonical representation of w. Although the other terms defined in this paragraph are standard, the term *subelement* is new. When speaking loosely, one could use 'subterm' in place of 'subelement,' but this is obviously not correct. Notice that according to this definition (and the definition of subterm), $x \vee y$ is not a subelement of $x \vee y \vee z$. A join representation $a = a_1 \vee \cdots \vee a_n$ in an arbitrary lattice is said to be a *nonrefinable join representation* if $a = b_1 \vee \cdots \vee b_m$ and $\{b_1, \ldots, b_m\} \ll \{a_1, \ldots, a_n\}$ imply $\{a_1, \ldots, a_n\} \subseteq \{b_1, \ldots, b_m\}$.

The next theorem shows a strong connection between the syntactical canonical form and the arithmetic of the free lattice. It shows that the canonical join representation of an element of a free lattice is the best way to write it as a join in that any other join representation is an easy consequence of it, see Figure 3.

Theorem 2.3 *Let $w = w_1 \vee \cdots \vee w_n$ canonically. If $w = u_1 \vee \cdots \vee u_m$ then*

$$\{w_1, \ldots, w_n\} \ll \{u_1, \ldots, u_m\}.$$

Thus $w = w_1 \vee \cdots \vee w_n$ is the unique, nonrefinable join representation of w.

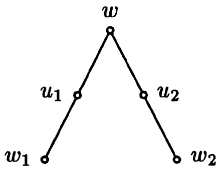


Figure 3.

Proof The arguments used to prove Theorem 2.1 easily give this result. □

A lattice is called *join semidistributive* if it satisfies the following condition.

(SD$_\wedge$) $\quad a \vee b = a \vee c \quad$ implies $\quad a \vee b = a \vee (b \wedge c).$

Meet semidistributivity is defined dually and denoted (SD$_\wedge$). A lattice is *semidistributive* is it satisfies both of these conditions.

Theorem 2.4 *Free lattices are semidistributive.*

Proof Suppose $a \vee b = a \vee c = w$ and let $w = w_1 \vee \cdots \vee w_n$ canonically. By renumbering, we may assume that $w_i \leq a$ for $i \leq k$ and $w_i \not\leq a$ for $i > k$. By Theorem 2.3, $\{w_1, \ldots, w_n\} \ll \{a, b\}$. Thus we must have $w_i \leq b$ for $i > k$. By the same reasoning, $w_i \leq c$ for $i > k$ and thus $w_i \leq b \wedge c$ for $i > k$. This implies $w = a \vee (b \wedge c)$, as desired. □

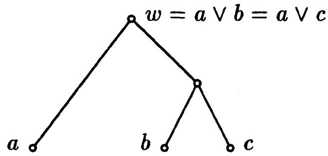


Figure 4.

3 Continuity

A lattice is said to be *lower continuous* if whenever

(3) $$a_0 \geq a_1 \geq a_2 \geq \cdots$$

is a descending chain having a greatest lower bound $a = \bigwedge a_i$, then, for any b,

$$\bigwedge (a_i \vee b) = a \vee b.$$

Upper continuous is defined dually. A lattice is *continuous* if it is both lower and upper continuous. Often it is assumed that continuous lattices are complete, but we do not make that assumption here. In this section we prove Whitman's result that free lattices are continuous.

Theorem 3.1 *Free lattices are continuous.*[1]

Proof Suppose that $a = \bigwedge a_i$ for a descending chain as in (3) and let b be arbitrary. Clearly $a \vee b$ is a lower bound for $\{a_i \vee b : i \in \omega\}$. Suppose that it is not the greatest lower bound and let c be an element of minimal rank such that

$$(4) \qquad c \not\leq a \vee b, \qquad c \leq a_i \vee b \quad \text{for all } i.$$

If $c \leq a_i$ for infinitely many i's, then $c \leq a$ which contradicts (4). Thus, by removing finitely many a_i's, we may assume that $c \not\leq a_i$ for all i. If $c = \bigvee c_j$ is a proper join, then for some j, $c_j \not\leq a \vee b$ and thus c_j violates the minimality of c. If $c = \bigwedge c_j$ then we apply (W) to $c = \bigwedge c_j \leq a_i \vee b$. We obtain a violation of (4) if $c \leq b$ and we have already ruled out the possibility that $c \leq a_i$. Hence for each i, there is a j such that $c_j \leq a_i \vee b$. Since there are only finitely many j's, there must be a fixed j such that $c_j \leq a_i \vee b$ for infinitely many i's. This implies $c_j \leq a_i \vee b$ for all i (since the $a_i \wedge b$ form a descending chain) and thus (4) holds with c_j in place of c, contradicting the minimality of the rank of c. If c is a generator then $c \leq a_i \vee b$ implies $c \leq a_i$ or $c \leq b$ both of which cannot occur. □

4 Fixed point free polynomials

A lattice **L** is called *order polynomial complete* if every order preserving map on **L** can be represented as a unary polynomial, see [28]. Wille was able to characterize finite polynomial complete lattices in [32], but the problem of characterizing infinite polynomial complete lattices remains open, and a solution does not appear to be near. Anne Davis Morel [3] has shown that if **L** is not a complete lattice, then it has an order preserving map f without a fixed point, i.e., $f(u) \neq u$ for all $u \in L$. (Tarski had proved the converse [29].) Thus if it were true that every unary polynomial on a lattice had a fixed point then this would imply that every order polynomial complete lattice is complete. It is not easy to construct a unary polynomial on a lattice without a fixed point. An example of a modular lattice with such a polynomial is given in [12]. Here we exhibit a polynomial on **FL**(3) without a fixed point. Using this polynomial we give an example, due to Whitman, of an ascending chain in **FL**(3) without a least upper bound, showing that free lattices are not complete.

The problem of characterizing those lattices in which every unary polynomial has a fixed point remains open. It is discussed in [12] where it is pointed out that every locally complete lattice has this property. A lattice is *locally complete* if every finitely generated sublattice is complete.

Let

$$p(u, x, y, z) = (((((u \wedge y) \vee z) \wedge x) \vee y) \wedge z) \vee x$$

[1]Actually every finitely presented lattice is continuous, see [15].

It is an interesting historical fact that Whitman worried that this theorem might be vacuous, i.e., he did not know if a free lattice could contain a descending chain with a greatest lower bound. That such a chain exists was first established by R. A. Dean, who showed, in unpublished work, that $x \vee (y \wedge z)$ has no upper cover and hence is the meet of a descending chain. Whitman did construct a descending chain without a meet, see below.

and let $q(u,x,y,z)$ be defined dually. Define unary polynomials

$$\begin{aligned} f(u) = f_1(u) &= p(u,x,y,z) \\ f_2(u) &= q(u,y,z,x) \\ f_3(u) &= p(u,z,x,y) \\ f_4(u) &= q(u,x,y,z) \\ f_5(u) &= p(u,y,z,x) \\ f_6(u) &= q(u,z,x,y) \end{aligned}$$

Theorem 4.1 *f is fixed point free on* **FL**(3)*, i.e., $f(w) = w$ for no $w \in$* **FL**(3)*.*

Proof Suppose that s is an element of minimal rank with $f(s) = s$. If $f_i(t) = t$ then clearly t can be transformed by some permutation of x, y, and z and possibly duality into a fixed point of f. Thus, since such a transformation leaves the rank invariant, the rank of t must be at least that of s.

Let $r = ((((s \wedge y) \vee z) \wedge x) \vee y) \wedge z$, so that $s = f(s) = r \vee x$. Also

$$\begin{aligned} (5) \qquad f_6(r) &= (((((((((((s \wedge y) \vee z) \wedge x) \vee y) \wedge z) \vee x) \wedge y) \vee z) \wedge x) \vee y) \wedge z \\ &= ((((f(s) \wedge y) \vee z) \wedge x) \vee y) \wedge z \\ &= ((((s \wedge y) \vee z) \wedge x) \vee y) \wedge z \\ &= r \end{aligned}$$

We will show that r is a canonical joinand of s and thus the above equation violates the minimality of s.

Since $s \leq 1 = x \vee y \vee z$, $s = f(s) \leq f(1) = ((((y \vee z) \wedge x) \vee y) \wedge z) \vee x$ and thus

$$(6) \qquad s \ngeq y, \qquad s \ngeq z.$$

All the elements in the range of f lie above x, so $x \leq s$. Now $r \nleq x$, since otherwise $s = r \vee x = x$, but one easily checks that $f(x) \neq x$. Thus r and x are incomparable since the other inequality would imply $z \geq x$. So $s = r \vee x$ is a proper join. Let $s = v_1 \vee \cdots \vee v_n$ be the canonical form of s. By Theorem 2.3, $\{v_1, \ldots, v_n\} \ll \{r, x\}$, i.e., for each i, either $v_i \leq x$ or $v_i \leq r$. Also, since $x \leq s = \bigvee v_i$ and generators are join prime, $x \leq v_i$ for some i. We take $i = 1$. If $v_1 \leq r$ then $x \leq r$, a contradiction. Thus $v_1 \leq x$ and hence $x = v_1$. This in turn implies that $v_i \leq r$ for $i > 1$.

Now apply (W) to

$$[(((s \wedge y) \vee z) \wedge x) \vee y] \wedge z = r \leq s = \bigvee v_i.$$

If $z \leq s$, we contradict (6). If $(((s \wedge y) \vee z) \wedge x) \vee y \leq s$, then $y \leq s$, again violating (6). Hence $r \leq v_i$ for some i. Since $r \nleq x = v_1$, we must have $i > 1$. But then $v_i \leq r$, and hence $r = v_i$, showing that r is a canonical joinand of s. As pointed out earlier this, together with (5), violates the minimality of s. □

Open Question 1 *Which unary polynomials on free lattices are fixed point free?*

Let f be the unary polynomial on $\mathbf{FL}(3)$ defined above. Clearly $x \leq f(x)$ and thus

$$x \leq f(x) \leq f^2(x) \leq \cdots$$

is an ascending chain which we denote C. We claim that f does not have a least upper bound. It follows from continuity and an easy inductive argument that *if g is any unary polynomial on a free lattice and $a_0 \leq a_1 \leq a_2 \leq \cdots$ is an ascending chain with a least upper bound $\bigvee a_i$, then $g(\bigvee a_i) = \bigvee g(a_i)$.* Applying this to C we have

$$f(\bigvee f^i(x)) = \bigvee f^{i+1}(x) = \bigvee f^i(x),$$

which implies that f has a fixed point. This contradiction shows that C does not have a least upper bound. We leave it as an exercise for the reader to show that if $\{x, y, z\} \subseteq X$ then C does not have a least upper bound in $\mathbf{FL}(X)$. Hence $\mathbf{FL}(X)$ *is not a complete lattice unless* $|X| \leq 2$.

5 Sublattices of free lattices

In this section we prove Whitman's theorem that $\mathbf{FL}(\omega)$ is a sublattice of $\mathbf{FL}(3)$ and Jónsson and Kiefer's theorem that there is a nontrivial equation satisfied by all finite sublattices of a free lattice. We begin with the following result of Galvin and Jónsson [21].

Theorem 5.1 *Every free lattice* $\mathbf{FL}(X)$*, and hence every sublattice of a free lattice, is a countable union of antichains. Thus free lattices, and sublattices of free lattices, contain no uncountable chains.*

Proof The result is obvious if X is finite, so assume X is infinite and that X_0 is a countable subset of X. Let $\mathbf{G}$ be the group of automorphisms of $\mathbf{FL}(X)$ which are induced from the permutations of X fixing all but finitely many $x \in X$. For u and $v \in \mathbf{FL}(X)$, let $u \sim_{\mathbf{G}} v$ denote the fact that u and v lie in the same orbit, i.e., there is a $\sigma \in G$ such that $\sigma(u) = v$. Notice that $u \sim_{\mathbf{G}} v$ means that u and v can be represented by the same term except that the variables are changed. For every element $u \in \mathbf{FL}(X)$, there is a v in the sublattice generated by X_0 with $u \sim_{\mathbf{G}} v$. Thus $\mathbf{FL}(X)$ has only countably many orbits under $\mathbf{G}$. (An orbit is just an equivalence class of $\sim_{\mathbf{G}}$.)

Let $\sigma \in G$ and suppose $u < \sigma(u)$. Then by applying σ to this inequality we obtain

$$u < \sigma(u) < \sigma^2(u) < \cdots. \tag{7}$$

Each element of $\mathbf{G}$ has finite order and thus $\sigma^n(u) = u$ for some positive n. But then (7) implies $u < u$, a contradiction. Thus each orbit is an antichain and there are only countably orbits, which proves the theorem. □

Theorem 5.2 $\mathbf{FL}(3)$ *contains a sublattice isomorphic to* $\mathbf{FL}(\omega)$.

Proof Let $f = f_1, \ldots, f_6$ be the unary polynomials defined above. Notice that f_4 is the dual of f. Let $c_n = f^n(x)$ be the n^{th} element of the ascending chain C of Whitman's example above and let $d_n = f_4^n(x)$ be the element dual to c_n. So the d_n's form a descending chain $d_0 = x > d_1 > d_2 > \cdots$. For $n \geq 1$, define

$$w_n = z \vee (c_n \wedge (d_n \vee y)).$$

We will show, via Corollary 1.12, that $w_1, w_2, \ldots$ generate a free lattice. Our development follows [1]. First an easy lemma.

Lemma 5.3 *The following are true for all $n \geq 1$.*

1. $y \not\leq c_n$
2. $y \not\leq w_n$
3. $c_n \not\leq z$
4. $x \not\leq d_n$
5. $x \not\leq w_n$

Proof For (1), $c_n = f^n(x)$ and it was shown in the proof of Theorem 3.1 that nothing in the range of f lies above y. For (2), it is easy to see that $y \leq w_n$ implies $y \leq c_n$, which we have already eliminated. Since $x \leq c_n$, (3) holds and (4) follows from $d_n < x$. (5) follows from (W+), (4) and two applications of the fact that x is join prime. □

Suppose that $w_m \leq w_n$. Then $c_m \wedge (d_m \vee y) \leq z \vee (c_n \wedge (d_n \vee y)) = w_n$ and we apply (W+). Using the lemma all cases can be eliminated easily except $c_m \wedge (d_m \vee y) \leq c_n \wedge (d_n \vee y)$. If this holds then both of the following hold.

$$\begin{aligned} c_m \wedge (d_m \vee y) &\leq c_n = [((((c_{n-1} \wedge y) \vee z) \wedge x) \vee y) \wedge z] \vee x \\ c_m \wedge (d_m \vee y) &\leq d_n \vee y \end{aligned}$$

Applying (W+) to the first inequality and using Theorem 1.4 and Lemma 5.3 we conclude $c_m \leq c_n$ and hence $m \leq n$. Similarly, the second inequality leads to $d_m \vee y \leq d_n \vee y$.

One easily checks from the definition that $f_4^k(x) \vee y = f_5^k(x \vee y)$. Since f_5 has no fixed points and $x \vee y \geq f_5(x \vee y)$, we see that $f_5^k(x \vee y)$ forms a descending chain, i.e., $f_5^m(x \vee y) \leq f_5^n(x \vee y)$ if and only if $m \geq n$. Since $f_5^m(x \vee y) = d_m \vee y \leq d_n \vee y = f_5^n(x \vee y)$, $m \geq n$. Hence $m = n$ and thus the w_n's form an antichain.

Now suppose that $w_m \leq w_{n_1} \vee \cdots \vee w_{n_k}$ for some distinct, positive $m, n_1, \ldots, n_k$. Then $c_m \wedge (d_m \vee y) \leq w_{n_1} \vee \cdots \vee w_{n_k}$ and we apply (W). By Lemma 5.3, neither x nor y lie below $w_{n_1} \vee \cdots \vee w_{n_k}$, and it follows that the only possibility is $c_m \wedge (d_m \vee y) \leq w_{n_i}$, for some i. But, by joining both sides with z, this gives $w_m \leq w_{n_i}$, a contradiction.

Now suppose $w_{n_1} \wedge \cdots \wedge w_{n_k} \leq w_m$. No $w_n \leq z$ and hence $w_{n_1} \wedge \cdots \wedge w_{n_k} \not\leq z$. This fact, and the incomparability of the w_n's, imply that

$$z \leq w_{n_1} \wedge \cdots \wedge w_{n_k} \leq c_m \wedge (d_m \vee y) \leq d_m \vee y,$$

which cannot occur. Thus, by Theorem 1.12, the sublattice generated by w_n, $n = 1, 2, \ldots$, is isomorphic to $\mathbf{FL}(\omega)$. □

Next we present Jónsson and Kiefer's theorem.

Theorem 5.4 *Let* $\mathbf{L}$ *be a lattice satisfying* (W). *Suppose elements* a_1, a_2, a_3, *and* $v \in L$ *satisfy*

1. $a_i \not\leq a_j \vee a_k \vee v$ *whenever* $\{i, j, k\} = \{1, 2, 3\}$,
2. $v \not\leq a_i$ *for* $i = 1, 2, 3$,
3. v *is meet irreducible.*

Then $\mathbf{L}$ *contains a sublattice isomorphic to* $\mathbf{FL}(3)$.

Proof For $\{i, j, k\} = \{1, 2, 3\}$, let $b_i = a_i \vee [(a_j \vee v) \wedge (a_k \vee v)]$. If $b_i \leq b_j \vee b_k$ then $a_i \leq b_i \leq b_j \vee b_k \leq a_j \vee a_k \vee v$, contradicting (1). Thus the b_i's are join irredundant. In particular they are pairwise incomparable. Suppose $b_1 \wedge b_2 \leq b_3 = a_3 \vee [(a_1 \vee v) \wedge (a_2 \vee v)]$, and apply (W). Neither b_1 nor b_2 is below b_3 and if $b_1 \wedge b_2 \leq a_3$, then $v \leq a_3$, contradicting (2). Hence $b_1 \wedge b_2 \leq [(a_1 \vee v) \wedge (a_2 \vee v)] \leq a_1 \vee v$, and we apply (W) again. Since $v \leq b_1 \wedge b_2$, the inequality $b_1 \wedge b_2 \leq a_1$ would imply $v \leq a_1$ and so cannot occur. If $b_1 \wedge b_2 \leq v$, then $v = b_1 \wedge b_2$ and so would be a proper meet, which contradicts (3). If $b_2 \leq a_1 \vee v$, then $a_2 \leq a_1 \vee v$, which violates (1). Thus we must have $b_1 \leq a_1 \vee v$ which implies

$$(a_2 \vee v) \wedge (a_3 \vee v) \leq a_1 \vee v.$$

By (1) neither meetand is contained in $a_1 \vee v$ and by (2) $(a_2 \vee v) \wedge (a_3 \vee v) \not\leq a_1$. The last possibility gives $v = (a_2 \vee v) \wedge (a_3 \vee v)$, contradicting (3). □

A lattice $\mathbf{L}$ is said to have *breadth at most n* if whenever $a \in L$ and S is a finite subset of L such that $a = \bigvee S$, there is a subset T of S, with $a = \bigvee T$ and $|T| \leq n$. The breadth of a lattice is the least n such that it has breadth at most n. The reader can verify that this concept is self dual.

Corollary 5.5 *If* $\mathbf{L}$ *is a finite lattice satisfying* (W), *then the breadth of* $\mathbf{L}$ *is at most* 4. *The variety generated by finite lattices which satisfy* (W) *is not the variety of all lattices. In particular, finite sublattices of a free lattice have breadth at most four and satisfy a nontrivial lattice equation.*

Proof Suppose $\mathbf{L}$ is a finite lattice satisfying (W) and $a = a_1 \vee \cdots \vee a_n$ holds for some $a \in L$ and some $n > 4$ and that this join is irredundant. Then, since every element of a lattice which satisfies (W) must be either join or meet irreducible, $v = a_4 \vee \cdots \vee a_n$ is meet irreducible. By Theorem 5.4, $\mathbf{L}$ has a sublattice isomorphic to $\mathbf{FL}(3)$ and hence is infinite, a contradiction. Thus $\mathbf{L}$ has breadth at most 4.

If B_4 is the class of all lattices of breadth at most 4, then, by Jónsson's Theorem [22], every subdirectly irreducible lattice in $\mathbf{V}(\mathrm{B}_4)$ lies in $\mathbf{HSP}_u(\mathrm{B}_4)$. But it is easy to see that B_4 is closed under these operators. Thus if $\mathbf{V}\mathrm{B}_4$ were all lattices, every subdirectly irreducible lattice would have breadth at most 4, which is not the case. □

It is not hard to see (either directly or using the theory of covers of chapter III) that $a_i = \bigwedge_{j \neq i} x_j$ are the atoms of $\mathbf{FL}(x_1, \ldots, x_n)$.

Corollary 5.6 *If* $n \geq 5$, *the sublattice of* $\mathbf{FL}(n)$ *generated by the atoms is infinite.* □

The sublattice generated by the atoms of $\mathbf{FL}(3)$ is the 8 element Boolean algebra. The sublattice generated by the atoms of $\mathbf{FL}(4)$ has 22 elements and is diagrammed in Figure 5. To verify this, one needs to label all the elements and show that all the joins and meets are correct. This is left as an exercise for the reader.

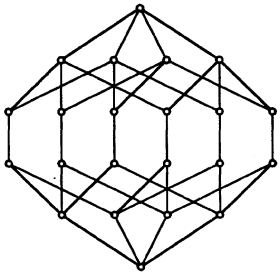


Figure 5.

6 Covers

In this section we study covers in a finitely generated free lattice $\mathbf{FL}(X)$ and give an algorithm which finds all the lower covers of a given element. Of course, upper covers can be treated dually. We begin with elementary material.

Most of the results in this section are from Freese and Nation's paper *Covers in free lattices* [18]. Throughout this section let X be a finite set with at least three elements.

6.1 Basic results on covers

A join irreducible element w of $\mathbf{FL}(X)$ is completely join irreducible if and only if it has a lower cover; this lower cover is then unique and will be denoted by w_*. Dually, the upper cover of a completely meet irreducible element w will be denoted by w^*.

Theorem 6.1 *Let w be a completely join irreducible element of $\mathbf{FL}(X)$. Then there exists a largest element v such that $v \geq w_*$ but $v \not\geq w$. Moreover, v is the unique canonical meetand of w_* which is not above w.*

Proof If w_* is join reducible, then it is meet irreducible by (W), and $v = w_*$ clearly satisfies the conclusions of the theorem.

Now suppose w_* is join irreducible. There must be a canonical meetand v of w_* such that $v \not\geq w$. Of course, $w \wedge v = w_*$ as $w \succ w_*$. By the refinement property of Theorem 2.3, every canonical meetand of w_* is either above w or above v. This means that v is the only canonical meetand not above w. If u is an element such that $u \geq w_*$ but $u \not\geq w$ then $w \wedge u = w_*$ and so, again by the refinement property of canonical representations, $v \geq u$. □

If w is a completely join irreducible element of $\mathbf{FL}(X)$, then the unique element v from Theorem 6.1 will be denoted by $\kappa(w)$. For w not completely join irreducible, $\kappa(w)$ is undefined. It is possible for $w_* = \kappa(w)$, but this occurs only near the top and bottom of free lattices, see [20]. A particularly useful formulation of the previous theorem is the following.

Corollary 6.2 *Suppose w is completely join irreducible in $\mathbf{FL}(X)$. If $u \geq w_*$ then either $u \geq w$ or $u \leq \kappa(w)$. Thus the interval $1/w_*$ is the disjoint union of $1/w$ and $\kappa(w)/w_*$.* □

Dually, if u is completely meet irreducible, then $\kappa'(u)$ denotes the unique canonical joinand of u^* not below u.

It is easy to see that if w is a completely join irreducible element of $\mathbf{FL}(X)$, then $\kappa(w)$ is completely meet irreducible and

$$w_* = w \wedge \kappa(w), \qquad \kappa(w)^* = w \vee \kappa(w).$$

The mapping κ is a bijection of the set of completely join irreducible elements onto the set of completely meet irreducible elements of $\mathbf{FL}(X)$, and κ' is its inverse.

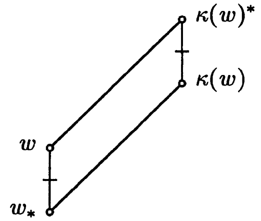


Figure 6.

Theorem 6.3 *Let w be a completely join irreducible element of $\mathbf{FL}(X)$ and let v be a canonical meetand of w. If v is not a generator, then there is exactly one canonical joinand of v not below w.*

Proof Since $w \leq v$, we have $v \nleq \kappa(w)$ and so there is at least one canonical joinand u of v not below $\kappa(w)$. Then $w \leq w_* \vee u$ by Corollary 6.2. Applying Whitman's condition (W) to this inequality and taking into account the canonical representation of w, we obtain, by Theorem 2.3 on canonical forms, that $v' \leq w_* \vee u$ for some canonical meetand v' of w. But then $v' \leq w_* \vee u \leq v$ and hence $v' = v$, since two distinct canonical meetands cannot be comparable. We have proved $w_* \vee u = v$, which implies that each canonical joinand of v is either below (and hence equal to) u or below w_*. □

Theorem 6.4 *Let w be a join reducible element of $\mathbf{FL}(X)$. There is a bijection f of the set of lower covers of w onto the set of the canonical joinands of w that are completely join irreducible. The bijection f can be defined as follows:*

1. *If v is a lower cover of w, then $f(v)$ is the unique canonical joinand w_i of w such that $w_i \not\leq v$; moreover, the interval $w/0$ is a disjoint union of w/w_i and $v/0$.*

2. *If w_i is a completely join irreducible canonical joinand of w, then $f^{-1}(w_i) = \kappa(w_i) \wedge w$ and $f^{-1}(w_i)$ is the only lower cover of w not above w_i.*

Proof Let $v \prec w$. Then at least one canonical joinand w_i of w is not below v. By join semidistributivity, this w_i is unique. We claim that $w_i \wedge v \prec w_i$. If $w_i \wedge v < u < w_i$ for an element u, then since $u \not\leq v$, $u \vee v = w$ and consequently every canonical joinand of w must be either below u or below v. But w_i is neither.

Conversely, let w_i be a canonical joinand of w with a lower cover w_{i*}. Put $v = \kappa(w_i) \wedge w$. Denote by t the join of all the canonical joinands of w other than w_i. We have $w_{i*} \vee t < w$, so that $w_{i*} \vee t \not\geq w_i$ and consequently $w_{i*} \vee t \leq \kappa(w_i)$. This shows that every canonical joinand of w other than w_i is below v. If $v < u < w$, then $u \geq w_{i*}$ but $u \not\leq \kappa(w_i)$, so that $u \geq w_i$ and hence $u \geq w_i \vee t = w$, a contradiction. This proves $v \prec w$. If v' is any other lower cover of w not above w_i, then by (1), $w_{i*} = w_i \wedge v = w_i \wedge v'$; by semidistributivity, $w_{i*} = w_i \wedge (v \vee v') = w_i \wedge w = w_i$, a contradiction. □

An element w is called *lower atomic* if for every element u such that $u < w$ there exists an element v with $u < v \prec w$. An *upper atomic* element is defined dually.

Corollary 6.5 *Let $w = w_1 \vee \cdots \vee w_n$ be the canonical form of a join reducible element w. Then w has at most n lower covers; the number of the lower covers of w coincides with the number of those canonical joinands of w that are completely join irreducible.* □

Corollary 6.6 *A join reducible element has a lower cover in* $\mathbf{FL}(X)$ *if and only if at least one of its canonical joinands is completely join irreducible.* □

Corollary 6.7 *Let $w = w_1 \vee \cdots \vee w_n$ be the canonical form of a join reducible element w. The element w is lower atomic if and only if it has precisely n lower covers.*

Proof Let the canonical joinands be all completely join irreducible. If $u < w$, then $u \not\geq w_i$ for some i and then $u \leq \kappa(w_i) \wedge w \prec w$. □

We see now that the question, which elements of $\mathbf{FL}(X)$ have a lower cover, is decidable if only the same question for join irreducible elements of $\mathbf{FL}(X)$ is decidable. Also, in order to find all the lower covers of a join reducible element, by Theorem 6.4 it is sufficient to know how to decide which join irreducible elements have a lower cover and how to construct $\kappa(w)$ for the completely join irreducible elements w. This is what we shall accomplish later in this chapter.

Example 6.8 $\kappa(x) = \bigvee(X - \{x\})$, $x_* = x \wedge \kappa(x)$, $\kappa'(x) = \bigwedge(X - \{x\})$ and $x^* = x \vee \kappa'(x)$. If w is a meet of generators, $w = \bigwedge Y$ for $\emptyset \subset Y \subset X$, then $\kappa(w) = \bigvee(X - Y)$ and $w_* = w \wedge \kappa(w)$; the element w is upper atomic and its upper covers are the elements $w \vee \bigwedge(X - \{y\})$ for each $y \in Y$, see Lemma 1.2.

6.2 J-closed sets and the standard epimorphism

For any element $w \in \mathbf{FL}(X)$ we define a set $\mathrm{J}(w)$ of join irreducible subelements of w in this way:

1. if u is a proper meet, then $u \in \mathrm{J}(w)$ if and only if u is a subelement of w;
2. if u is a generator, then $u \in \mathrm{J}(w)$ if and only if either $u = w$ or u is a canonical joinand of a subelement of w which is a proper join.

We see that the set $\mathrm{J}(w)$ is contained in the set of join irreducible subelements of w and can differ from this set in the generators only. It can be also defined recursively in this way:

$$\mathrm{J}(w) = \begin{cases} \{w\} & \text{if } w \text{ is a meet of generators,} \\ \{w\} \cup \bigcup_{i,j} \mathrm{J}(w_{ij}) & \text{if } w = \bigwedge_i \bigvee_j w_{ij} \wedge \bigwedge_k x_k \text{ canonically.} \\ \bigcup_i \mathrm{J}(w_i) & \text{if } w = \bigvee_i w_i \text{ canonically.} \end{cases}$$

A subset A of $\mathbf{FL}(X)$ is said to be *J-closed* if it is a set of join irreducible elements and $w \in A$ implies $\mathrm{J}(w) \subseteq A$.

Clearly, $\mathrm{J}(w)$ is a finite J-closed set; it is the least J-closed set containing the element w.

For a meet irreducible element w we define $\mathrm{M}(w)$ dually; by an *M-closed set* we mean a set of meet irreducible elements of $\mathbf{FL}(X)$ such that $w \in A$ implies $\mathrm{M}(w) \subseteq A$.

A *join cover* of an element u in a lattice is a subset A such that $u \leq \bigvee A$. A join cover A of u is said t be *minimal* if whenever B is a join cover of a and $B \ll A$, then $A \subseteq B$. We say that a subset A of a lattice $\mathbf{F}$ has the *join cover refinement property* if A is a set of join irreducible elements and for each $a \in A$, every join cover of a can be refined to a join cover of a contained in A. We shall show that in free lattices, the join cover refinement property and the property of being J-closed are the same. First we prove a lemma which will be used below.

Lemma 6.9 *Let $w = w_1 \wedge \cdots \wedge w_m$ be the canonical form of an element in* $\mathbf{FL}(X)$*, and let $\{w_{i1}, \ldots, w_{in}\}$ be the canonical joinands of w_i for some i. Then $\{w_{i1}, \ldots, w_{in}\}$ is a minimal join cover of w.*

Proof For i fixed, $\{w_{i1}, \ldots, w_{in}\}$ is clearly a join cover of w. Let C be a refinement of it such that $w \leq \bigvee C$. By (W) there is an i' with $w_{i'} \leq \bigvee C$. But $\bigvee C \leq w_i$; we get $i' = i$ and $\bigvee C = \bigvee_j w_{ij}$. The latter expression is canonical, which means that the join cover consisting of the elements w_{ij} refines C from which it follows that $\{w_{i1}, \ldots, w_{in}\} \subseteq C$ by Theorem 2.3. □

Theorem 6.10 *A set A of join irreducible elements of* $\mathbf{FL}(X)$ *is J-closed if and only if it has the join cover refinement property.*

Proof Let A be J-closed, $w = \bigwedge_i \bigvee_j w_{ij} \wedge \bigwedge_k x_k$ (canonically) be an element of A and C be a join cover of w. We shall show by induction on the length of w that C can be refined to a join cover contained in A. If C is a trivial join cover, i.e., if $w \leq u$ for some $u \in C$, then $\{w\}$ is a refinement of C contained in A. Now let C be nontrivial. So by (W), either

$w_i \le \bigvee C$ for some i or $x_k \le \bigvee C$ for some k. If $x_k \le \bigvee C$, then $w \le x_k \le c$ for some $c \in C$ by Theorem 1.4, and C is a trivial join cover in this case. So we may assume that there is an i such that $w_{ij} \le \bigvee C$ for all j. Take one such i fixed. By induction, for each j there exists a join cover C_j of w_{ij} refining C and contained in A. The union of all these join covers C_j is a join cover of $\bigvee_j w_{ij}$, so that it is a join cover of w; it refines C and is contained in A.

Conversely, let A have the join cover refinement property. Let $w \in A$, $w = \bigwedge_i \bigvee_j w_{ij} \wedge \bigwedge_k x_k$ canonically. By Lemma 6.9, for each i, $\{w_{i1}, \dots, w_{in}\}$ is a minimal join cover of w, and hence must be contained in A, and from this it follows that A is J–closed. □

For any subset A of a lattice $\mathbf{F}$ with zero 0 we denote by $A^\vee$ the subset of $\mathbf{F}$ consisting of all joins of finite subsets of A, including $\bigvee \emptyset = 0$. If A is finite, then $A^\vee$ is a finite lattice, with the join operation coinciding with that of the lattice $\mathbf{F}$ and the meet operation $\wedge'$ defined by $a \wedge' b = \bigvee\{c \in A : c \le a \wedge b \text{ in } \mathbf{F}\}$. If A is a finite set of join irreducible elements of $\mathbf{F}$, then A is the set of join irreducible elements of the lattice $A^\vee$.

For any finite subset A of a lattice $\mathbf{F}$ with zero we define a mapping $f : F \to A^\vee$, called the *standard mapping* of $\mathbf{F}$ onto $A^\vee$, by

$$f(u) = \bigvee\{a \in A : a \le u\}$$

for any $u \in F$. Clearly, f is an order-preserving mapping, the restriction of f to $A^\vee$ is the identity and $f(u) \le u$ for all $u \in F$. This means that each element of $A^\vee$ is the least preimage of itself under f. Thus if f is a homomorphism, then f is lower bounded. Also note that if $a \in A$, then $a \le f(u)$ if and only if $a \le u$.

The standard mapping f is always a meet homomorphism:

$$\begin{aligned} f(u) \wedge' f(v) &= \bigvee\{a \in A : a \le f(u) \wedge f(v)\} \\ &= \bigvee\{a \in A : a \le u \wedge v\} = f(u \wedge v). \end{aligned}$$

On the other hand, the standard mapping need not to be a join homomorphism. If it is, we call f the *standard epimorphism.*

Theorem 6.11 *Let $\mathbf{F}$ be a lattice with zero and A be a finite set of join irreducible elements of $\mathbf{F}$ with the join cover refinement property. Then the standard mapping f of $\mathbf{F}$ onto $A^\vee$ is a homomorphism. Thus f is the standard epimorphism and is lower bounded.*

Proof All we need to do is to show that f is a join homomorphism. First note that if $a \in A$ and $a \le u \vee v$ for u and $v \in F$, then $\{u, v\}$ is a join cover of a. Thus there is a $C \subseteq A$ with $C \ll \{u, v\}$ and $a \le \bigvee C$. Since $f(c) = c$ for all $c \in C$, and every $c \in C$ is either below u or v, we have

$$a \le \bigvee C = \bigvee_{c \in C} f(c) \le f(u) \vee f(v).$$

Hence,

$$f(u \vee v) = \bigvee\{a \in A : a \le u \vee v\} \le f(u) \vee f(v).$$

Since f is order preserving, $f(u \vee v) = f(u) \vee f(v)$. □

Theorem 6.12 *Let f be a lower bounded homomorphism of a lattice* $\mathbf{F}$ *onto a finite lattice* $\mathbf{L}$. *Denote by A the set of the elements $\beta_f(c)$, where c runs over the join irreducible elements of* $\mathbf{L}$. *Then A has the join cover refinement property and* $\mathbf{L} \cong A^{\vee}$.

Proof Let $u = \beta_f(c) \in A$, where c is join irreducible in $\mathbf{L}$. If $u = p \vee q$ in $\mathbf{F}$, then $c = f(p) \vee f(q)$ in $\mathbf{L}$, so that either $c = f(p)$ or $c = f(q)$; but then either $u = \beta_f(c) = \beta_f f(p) \leq p$ or, similarly, $u \leq q$. We see that A is a finite set of join irreducible elements of $\mathbf{F}$. Since β_f preserves joins, it is easy to see that β_f is an isomorphism of $\mathbf{L}$ onto $A^{\vee}$.

To see that A has the join cover refinement property, let $u \in A$ and let C be a join cover of u in $\mathbf{F}$. Then $u \leq \bigvee C$ implies $f(u) \leq \bigvee f(C)$, and since $\mathbf{L}$ is finite, the join cover $f(C)$ of $f(u)$ refines to a join cover D consisting of join irreducible elements of $\mathbf{L}$. Put $C' = \beta_f(D) = \{\beta_f(d) : d \in D\}$. Then C' is a refinement of C, C' is contained in A and $u = \beta_f f(u) \leq \beta_f(\bigvee D) = \bigvee \beta_f(D) = \bigvee C'$, so that C' is a join cover of u. □

6.3 Finite lower bounded lattices

Theorem 6.13 *A finite lattice* $\mathbf{L}$ *is lower bounded if and only if $L \cong A^{\vee}$ for a finite, J-closed subset A of a finitely generated free lattice.*

Proof This is a consequence of Theorems 6.10, 6.11 and 6.12. □

Theorem 6.14 *Let A be a finite, J-closed subset of* $\mathbf{FL}(X)$. *If a subset B of A is also J-closed in* $\mathbf{FL}(X)$, *then the lattice $B^{\vee}$ is a homomorphic image of $A^{\vee}$. More specifically, the restriction to $A^{\vee}$ of the standard epimorphism of* $\mathbf{FL}(X)$ *onto $B^{\vee}$ is a homomorphism of $A^{\vee}$ onto $B^{\vee}$. Moreover, every homomorphic image of $A^{\vee}$ is isomorphic to $B^{\vee}$ for a J-closed subset B of A.*

Proof Let $B \subseteq A$ be J-closed. By Theorem 6.11, there are the standard epimorphisms $f : \mathbf{FL}(X) \to A^{\vee}$ and $g : \mathbf{FL}(X) \to B^{\vee}$. We need to show that $\ker f \subseteq \ker g$.

Let $u, v \in \mathbf{FL}(X)$ be two elements such that $f(u) = f(v)$. By the definition of standard epimorphism, for $w \in A$ we have $w \leq u$ if and only if $w \leq f(u)$ and, similarly, $w \leq v$ if and only if $w \leq f(v)$. Since $f(u) = f(v)$, we get $w \leq u$ if and only if $w \leq v$. As $B \subseteq A$, this is true also for every $w \in B$ and hence

$$g(u) = \bigvee\{w \in B : w \leq u\} = \bigvee\{w \in B : w \leq v\} = g(v).$$

Thus $\ker f \subseteq \ker g$, which means that there is a (unique) homomorphism h of $A^{\vee}$ onto $B^{\vee}$ with $g = hf$. For $a \in A^{\vee}$ we have $h(a) = h(f(a)) = g(a)$ and so h is a restriction of g.

Now let g be a homomorphism of $A^{\vee}$ onto a lattice $\mathbf{L}$. Because f is lower bounded and every homomorphism between finite lattices is bounded, the epimorphism $gf : \mathbf{FL}(X) \to \mathbf{L}$ is lower bounded. Denote by B the set of the elements $\beta_{gf}(a)$, where a runs over the join irreducible elements of $\mathbf{L}$. Each $\beta_{gf}(a)$ belongs to A, as $\beta_{gf}(a) = \beta_f(\beta_g(a))$ and the least preimage of a join irreducible element must be join irreducible. We get $B \subseteq A$. By Theorem 6.12, $\mathbf{L} \cong B^{\vee}$ and B has the join cover refinement property; by Theorem 6.10, B is J-closed. □

Theorem 6.15 *The set of the congruences ϕ of $\mathbf{FL}(X)$ for which $\mathbf{FL}(X)/\phi$ is a finite, lower bounded lattice is a filter in the congruence lattice of $\mathbf{FL}(X)$. This filter is dually isomorphic to the distributive lattice of finite J-closed subsets of $\mathbf{FL}(X)$ (which is a lattice with respect to the operations of union and intersection).*

Proof The first assertion follows from the facts that a homomorphic image of a finite, lower bounded lattice is again finite and lower bounded and that a subdirect product of finitely many lower bounded lattices is lower bounded. On the other hand, the set of finite J-closed subsets of $\mathbf{FL}(X)$ is clearly closed under finite unions and intersections, so that it is a distributive lattice. The dual isomorphism between the two lattices can be described in the following way. For a finite J-closed set A, the corresponding congruence is the kernel of the standard epimorphism of $\mathbf{FL}(X)$ onto $A^{\vee}$. Conversely, for a congruence ϕ such that $\mathbf{FL}(X)/\phi$ is a finite, lower bounded lattice (so that the canonical epimorphism of $\mathbf{FL}(X)$ onto $\mathbf{FL}(X)/\phi$ is lower bounded), the corresponding finite J-closed subset A is the set of the least elements in the join irreducible congruence classes of ϕ (this set is J-closed by Theorem 6.12). □

It also follows that the subdirect decompositions of a given finite, lower bounded lattice $A^{\vee}$ correspond to the decompositions of the set A into unions of J-closed subsets in the following sense:

Corollary 6.16 *If A_i $(1 \leq i \leq n)$ are finite J-closed subsets of $\mathbf{FL}(X)$, then the lattice $B^{\vee}$, where $B = \bigcup_{i=1}^{n} A_i$, is a subdirect product of the lattices $A_i^{\vee}$. Conversely, if A is a J-closed subset of $\mathbf{FL}(X)$ and the lattice $A^{\vee}$ is isomorphic to a subdirect product of lattices $\mathbf{L}_i$ $(1 \leq i \leq n)$, then there exist finite J-closed subsets A_i such that $A = \bigcup_{i=1}^{n} A_i$ and $\mathbf{L}_i \cong A_i^{\vee}$ for $1 \leq i \leq n$.* □

Theorem 6.17 *Every lower bounded lattice is join semidistributive.*

Proof Let f be a lower bounded homomorphism of $\mathbf{FL}(n)$ onto a lattice $\mathbf{L}$. Then $u = a \vee b = a \vee c$ in $\mathbf{L}$ implies

$$\beta_f(u) = \beta_f(a) \vee \beta_f(b) = \beta_f(a) \vee \beta_f(c) = \beta_f(a) \vee (\beta_f(b) \wedge \beta_f(c)),$$

since $\mathbf{FL}(n)$ is join semidistributive. Thus

$$u = f\beta_f(u) = f(\beta_f(a) \vee (\beta_f(b) \wedge \beta_f(c))) = a \vee (b \wedge c).$$

□

On the other hand, a finite and lower bounded lattice is not necessarily meet semidistributive. Our next goal is to prove that a finite, lower bounded lattice is meet semidistributive if and only if it is bounded. This result was proved in [8],[10],[24] and [27]. Here we shall follow the proof given in [27]. First we need to introduce some auxiliary notation and to prove six lemmas.

Let $\mathbf{L}$ be a finite lattice and a be a join irreducible element of $\mathbf{L}$. The unique lower cover of a in $\mathbf{L}$ will be denoted by a_*. Of course a_* depends on $\mathbf{L}$ (so that it might be different in a sublattice). If there exists an element $b \in L$ such that the interval $1/a_*$ is the disjoint union of $1/a$ and b/a_*, then this uniquely determined element b will be denoted by $\kappa_L(a)$.For a meet irreducible element a define a^* and $\kappa'_L(a)$ dually.

Lemma 6.18 *A finite lattice* **L** *is meet semidistributive if and only if* $\kappa_L(a)$ *exists for any join irreducible element* a *of* **L**.

Proof Let **L** be meet semidistributive and a be a join irreducible element of **L**. If $u \geq a_*$ and $u \not\geq a$, then $u \wedge a = a_*$. By meet semidistributivity, $a \wedge \bigvee\{u \in L : u \geq a_* \text{ and } u \not\geq a\} = a_*$, and clearly this join is the largest such element. Thus

$$\kappa_L(a) = \bigvee\{u \in L : u \geq a_* \text{ and } u \not\geq a\}.$$

Conversely, suppose that **L** fails meet semidistributivity with $d = a \wedge b = a \wedge c < a \wedge (b \vee c)$. Let e be an element of **L** which is minimal with respect to $e \leq a \wedge (b \vee c)$ but $e \not\leq d$. Clearly, e is join irreducible. By the minimal property of e, $e_* \leq d$. Using $e \leq a$, we calculate $e \wedge b = e \wedge c = e_*$ whereas $e \wedge (b \vee c) = e$, which means that $\kappa_L(e)$ does not exist. □

Now let **L** be a finite semidistributive lattice. One can easily see that κ_L is a bijection of the set of join irreducible elements of **L** onto the set of meet irreducible elements and that κ'_L is its inverse. For a meet irreducible element a we have $a_* = a \wedge \kappa_L(a)$ and $\kappa_L(a)^* = a \vee \kappa_L(a)$.

The following important result is due to A. Day, [8]. Nation has a very elegant proof in [27].

Theorem 6.19 (A. Day) *A finite, lower bounded lattice is bounded if and only if it is meet semidistributive.*

Theorem 6.20 *Let A be a finite, J-closed set of join irreducible elements of* $\mathbf{FL}(X)$. *Then the following are equivalent:*

1. *the lattice $A^\vee$ is bounded;*
2. *the lattice $A^\vee$ is meet semidistributive;*
3. *every element of A has a lower cover in* $\mathbf{FL}(X)$.

Proof By Theorem 6.13, the lattice $A^\vee$ is lower bounded; thus (1) is equivalent to (2) by Theorem 6.19.

Next, let us show that (1) implies (3). If $w \in A$, then w is also join irreducible in the finite lattice $A^\vee$, so w has a unique lower cover u in $A^\vee$. The standard epimorphism $f : \mathbf{FL}(X) \to A^\vee$ is bounded and we have the associated mappings $\alpha_f, \beta_f : A^\vee \to \mathbf{FL}(X)$. Since $w = \beta_f f(w)$, if $v < w$ in $\mathbf{FL}(X)$ then $f(v) \leq u$, whence $v \leq \alpha_f(u)$. Thus $w \succ w \wedge \alpha_f(u)$.

It remains to show that (3) implies (2). Assume that every element $w \in A$ has a lower cover w_* in $\mathbf{FL}(X)$. Let $w \in A$, denote again by u the unique lower cover of w in $A^\vee$ and let $K = \{s \in A^\vee : s \geq u \text{ and } s \not\geq w\}$.

We shall show first that $a, b \in K$ implies $a \vee b \in K$. Let $f : \mathbf{FL}(X) \to A^\vee$ be the standard epimorphism, so that f is lower bounded. We have $\beta_f(w) = w$, $f(a) = a$, $f(b) = b$ and $f(w_*) = u$. Therefore $f(w_* \vee a) = a$ and $f(w_* \vee b) = b$, so that $w \not\leq w_* \vee a$ and $w \not\leq w_* \vee b$ in $\mathbf{FL}(X)$. Thus $w_* = w \wedge (w_* \vee a) = w \wedge (w_* \vee b)$, whence by meet semidistributivity $w_* = w \wedge (w_* \vee a \vee b)$. Thus $w \not\leq w_* \vee a \vee b$. In particular, $w \not\leq a \vee b$, so $a \vee b \in K$, as desired.

This means that for every join irreducible element w of the lattice $A^\vee$ the element $\kappa_{A^\vee}(w)$ exists; the lattice is then meet semidistributive by Lemma 6.18. □

6.4 The lattice $\mathbf{L}^{\vee}(w)$

For a join irreducible element w of $\mathbf{FL}(X)$, we denote by $\mathbf{L}^{\vee}(w)$ the lattice $\mathrm{J}(w)^{\vee}$. The lower cover of w in this lattice will be denoted by $w_{\dagger}$. Thus

$$w_{\dagger} = \bigvee\{u \in \mathrm{J}(w) : u < w\}.$$

For a meet irreducible element w, the lattice $\mathbf{L}^{\wedge}(w) = \mathrm{M}(w)^{\wedge}$ and the element $w^{\dagger}$ are defined dually.

Theorem 6.21 *Let w be a join irreducible element of $\mathbf{FL}(X)$. Then $\mathbf{L}^{\vee}(w)$ is a finite, lower bounded and subdirectly irreducible lattice with $w/w_{\dagger}$ as a critical prime quotient. The kernel of the standard epimorphism $f : \mathbf{FL}(X) \to \mathbf{L}^{\vee}(w)$ is the unique largest congruence ϕ of $\mathbf{FL}(X)$ with the property that $(u, w) \notin \phi$ whenever $u < w$.*

Proof The lattice is lower bounded by Theorem 6.13. If ϕ is a nontrivial congruence of $\mathbf{L}^{\vee}(w)$, then by Theorem 6.15 ϕ is the kernel of a homomorphism g of $\mathbf{L}^{\vee}(w)$ onto $A^{\vee}$ for a J-closed set A properly contained in $\mathrm{J}(w)$ and $g(u) = \bigvee\{v \in A : v \leq u\}$ for all u. Since A is a J-closed proper subset of $\mathrm{J}(w)$, $w \notin A$ and so

$$g(w) = \bigvee\{v \in A : v \leq w\} = \bigvee\{v \in A : v < w\} = g(w_{\dagger}).$$

Thus $\mathbf{L}^{\vee}(w)$ is subdirectly irreducible with $w/w_{\dagger}$ as a critical prime quotient.

Since w is the least preimage of itself under f, we have $(u, w) \notin \ker f$ whenever $u < v$. Now, $\ker f$ is a maximal congruence with this property, since $\mathbf{FL}(X)/\ker f$ is subdirectly irreducible with the critical prime quotient given above. It remains to show that the join of an arbitrary family of congruences ϕ with $(u, v) \notin \phi$ for all $u < w$ has again this property. However, this is an easy consequence of the well-known fact that if $u < w$ and $(u, w) \in \bigvee \phi_i$, then there is a finite chain $u = u_0 < \ldots < u_n = w$ such that each pair (u_{k-1}, u_k) belongs to some ϕ_i. □

Let us recall that for any prime quotient u/v of $\mathbf{FL}(X)$ (or of any lattice) there exists a largest congruence separating the elements u, v; this congruence is denoted by $\psi(u, v)$.

The following result was proved in [25].

Theorem 6.22 *Let w be a completely join irreducible element of $\mathbf{FL}(X)$. The congruence $\psi(w, w_*) = \psi(\kappa(w)^*, \kappa(w))$ is the kernel of the standard epimorphism $f : \mathbf{FL}(X) \to \mathbf{L}^{\vee}(w)$ and at the same time it is the kernel of the dual standard epimorphism $g : \mathbf{FL}(X) \to \mathbf{L}^{\wedge}(\kappa(w))$. The lattice $\mathbf{FL}(X)/\psi(w, w_*)$ is a splitting lattice and it is isomorphic to $\mathbf{L}^{\vee}(w)$.*

Proof Since $w \wedge \kappa(w) = w_*$ and $w \vee \kappa(w) = \kappa(w)^*$, the quotients w/w_* and $\kappa(w)^*/\kappa(w)$ are projective and a congruence separates w from w_* if and only if it separates $\kappa(w)^*$ from $\kappa(w)$. Hence $\psi(w, w_*) = \psi(\kappa(w)^*, \kappa(w))$. It follows from Theorem 6.21 and its dual that this congruence is the kernel of f and at the same time the kernel of g and that the factor is isomorphic to both $\mathbf{L}^{\vee}(w)$ and $\mathbf{L}^{\wedge}(\kappa(w))$, which lattice is then both lower bounded and upper bounded and, of course, subdirectly irreducible and finite. □

Theorem 6.23 *The following are equivalent for a join irreducible element $w \in \mathbf{FL}(X)$:*

1. *w is completely join irreducible (i.e., w has a lower cover in* $\mathbf{FL}(X)$*);*

2. *every element of* $\mathrm{J}(w)$ *is completely join irreducible;*

3. *every subelement of w is lower atomic;*

4. *the lattice* $\mathbf{L}^{\vee}(w)$ *is meet semidistributive.*

Proof Clearly, (3) implies (2) and (2) implies (1). By Theorem 6.20, (2) is equivalent to (4). By Theorem 6.22 (1) implies (4), since a splitting lattice is necessarily meet semidistributive. As every subelement of w not belonging to $\mathrm{J}(w)$ is either a generator or a join reducible element whose all canonical joinands belong to $\mathrm{J}(w)$, (2) implies (3) by Corollary 6.7. □

Combining this result with Corollary 6.7 one easily obtains the following characterization of lower atomic elements.

Corollary 6.24 *The following are equivalent for an element $w \in$* $\mathbf{FL}(X)$*:*

1. *w is lower atomic in* $\mathbf{FL}(X)$*);*

2. *every element of* $\mathrm{J}(w)$ *is completely join irreducible;*

3. *every subelement of w is lower atomic;*

4. *the number of lower covers of w equals the number of canonical joinands of w.* □

6.5 Syntactic algorithms

Let w be a join irreducible element of $\mathbf{FL}(X)$ and let

$$\mathrm{K}(w) = \{v \in \mathrm{J}(w) : w_{\dagger} \vee v \not\geq w\}$$

where $w_{\dagger}$ is defined by (6.4). The join $\bigvee \mathrm{K}(w)$ is an element of $\mathbf{L}^{\vee}(w)$ with the property that for every $v \in \mathbf{L}^{\vee}(w)$ with $v \geq w_{\dagger}$, either $v \geq w$ or $v \leq \bigvee \mathrm{K}(w)$. This means that if $\bigvee \mathrm{K}(w) \not\geq w$, then $\bigvee \mathrm{K}(w) = \kappa_{\mathbf{L}^{\vee}(w)}(w)$.

If w is completely join irreducible, then for an arbitrary element v of $\mathbf{FL}(X)$ we have $w_{\dagger} \vee v \geq w$ if and only if $w_{*} \vee v \geq w$. Indeed, when $f : \mathbf{FL}(X) \to \mathbf{L}^{\vee}(w)$ is the standard epimorphism, $w_{*} \vee v \geq w$ implies $w_{\dagger} \vee v \geq w_{\dagger} \vee f(v) = f(w_{*}) \vee f(v) = f(w_{*} \vee v) \geq f(w) = w$. This is true in particular for all $v \in \mathrm{J}(w)$ and so we obtain $\bigvee \mathrm{K}(w) = f(\kappa(w))$.

Theorem 6.25 *Let w be a join irreducible element of* $\mathbf{FL}(X)$*. Then w has a lower cover in* $\mathbf{FL}(X)$ *if and only if the following two conditions are satisfied:*

1. *every $u \in \mathrm{J}(w) - \{w\}$ has a lower cover in* $\mathbf{FL}(X)$*;*

2. $w \not\leq \bigvee \mathrm{K}(w)$.

Proof The direct implication is a consequence of Theorem 6.23. Conversely, assume that the two conditions are satisfied. By Theorem 6.23 and Lemma 6.18, we need only to show that $\kappa_{\mathbf{L}^\vee(w)}(u)$ exists for each $u \in \mathrm{J}(w)$. For $u = w$ it follows from (2) that $\bigvee \mathrm{K}(w) = \kappa_{\mathbf{L}^\vee(w)}(w)$. Let $u \neq w$. Since $\mathbf{L}^\vee(u)$ is meet semidistributive, $\kappa_{\mathbf{L}^\vee(u)}(u)$ exists; denote this element by q. Denote by h the homomorphism of $\mathbf{L}^\vee(w)$ onto $\mathbf{L}^\vee(u)$ that is a restriction of the standard epimorphism (see Theorem 6.14). We shall show that $\alpha_h(q)$, the largest preimage of q under h, is $\kappa_{\mathbf{L}^\vee(w)}(u)$. If $v \in \mathbf{L}^\vee(w)$ is above the unique lower cover of u in $\mathbf{L}^\vee(w)$, then $h(v)$ is above the unique lower cover of u in $\mathbf{L}^\vee(u)$ and so either $h(v) \geq u$ or $h(v) \leq q$. If $h(v) \geq u$, then $v \geq u$, since u is the least preimage of itself under h. If $h(v) \leq q$, then clearly $v \leq \alpha_h(q)$. □

Theorem 6.26 *Let w be a completely join irreducible element of* $\mathbf{FL}(X)$. *Then* $\mathrm{M}(\kappa(w)) = \{\kappa(u) : u \in \mathrm{J}(w)\}$.

Proof There are the standard epimorphism $f : \mathbf{FL}(X) \to \mathbf{L}^\vee(w)$ and the dual standard epimorphism $g : \mathbf{FL}(X) \to \mathbf{L}^\wedge(\kappa(w))$. By Theorem 6.22, $\ker f = \psi(w, w_*) = \psi(\kappa(w)^*, \kappa(w)) = \ker g$.

If $u \notin X$ is a canonical meetand of w and U is the set of canonical joinands of u, then U is a minimal nontrivial join cover of w in $\mathbf{FL}(X)$ by Lemma 6.9. Clearly, U is then also a minimal nontrivial join cover of w in $\mathbf{L}^\vee(w)$, and we have $\beta_f(u) = u$ for all $u \in U$. Hence each canonical meetand of w is either a generator or an element of the form $\bigvee_{u \in U} \beta_f(u)$ for a minimal nontrivial join cover U of w in $\mathbf{L}^\vee(w)$.

Dually, each canonical joinand of $\kappa(w)$ is either a generator or an element of the form $\bigwedge_{u \in U} \alpha_g(u)$ for a minimal nontrivial meet cover U of $\kappa(w)$ in $\mathbf{L}^\wedge(\kappa(w))$. Now, the lattices $\mathbf{L}^\vee(w)$ and $\mathbf{L}^\wedge(\kappa(w))$ are isomorphic, as both are isomorphic to $\mathbf{FL}(X)/\psi(w, w_*) = \mathbf{FL}(X)/\psi(\kappa(w)^*, \kappa(w))$, and we get that each canonical joinand of $\kappa(w)$ is either a generator or an element of the form $\bigwedge_{u \in U} \alpha_f(u)$ for a minimal nontrivial meet cover U of the element $\kappa_{\mathbf{L}^\vee(w)}(w)$ in $\mathbf{L}^\vee(w)$.

On the other hand, if v is a meet irreducible element of $\mathbf{L}^\vee(w)$, then $\alpha_f(v) = \kappa(u)$ for some $u \in \mathrm{J}(w)$. Indeed, we can take $u = \kappa'_{\mathbf{L}^\vee(w)}(v)$; to see that then $\alpha_f(v) = \kappa(u)$, one can observe that $\alpha_f(v)$ is completely meet irreducible in $\mathbf{FL}(X)$ and $f(u_* \vee \alpha_f(v)) = f(u_*) \vee v = v$, so that $u_* \leq \alpha_f(v)$.

We have proved that each canonical joinand of $\kappa(w)$ is either a generator or an element of the form $\bigwedge_{u \in U} \kappa(u)$ with $U \subseteq \mathrm{J}(w)$. Using this, one can easily see that $\mathrm{M}(\kappa(w))$ is contained in $\{\kappa(u) : u \in \mathrm{J}(w)\}$. The reverse inclusion is a matter of duality. □

Theorem 6.27 *Let w be a completely join irreducible element of* $\mathbf{FL}(X)$. *Then*

$$\kappa(w) = \bigvee\{x \in X : w_\dagger \vee x \not\geq w\} \vee \bigvee\{k^\dagger \wedge \kappa(v) : v \in \mathrm{J}(w) - \{w\},\ w \not\leq \kappa(v)\}$$

where

$$k^\dagger = \bigwedge\{\kappa(v) : v \in \mathrm{J}(w) - \{w\},\ \kappa(v) \geq \bigvee \mathrm{K}(w)\}.$$

Proof Put $A = \{\kappa(v) : v \in \mathrm{J}(w) - \{w\}\}$, so that by Theorem 6.26 the set A is M-closed and $A = \mathrm{M}(\kappa(w)) - \{\kappa(w)\}$. We have the standard epimorphism $f : \mathbf{FL}(X) \to \mathbf{L}^\vee(w)$, the dual standard epimorphism $g : \mathbf{FL}(X) \to \mathbf{L}^\wedge(\kappa(w))$ and the dual standard epimorphism $h : \mathbf{FL}(X) \to A^\wedge$. Since A is a proper subset of $\mathrm{M}(\kappa(w))$, $\ker h$ properly contains the congruence $\ker g = \ker f$. For this reason, if $\kappa(v) \in A$ is an element such that $\kappa(v) \geq \bigvee \mathrm{K}(w) = \kappa_{\mathbf{L}^\vee(w)}(w)$, then

$$\kappa(v) = h(\kappa(v)) \geq h(\kappa_{\mathbf{L}^\vee(w)}(w)) = h(\kappa(w)) = h(\kappa(w)^*) \geq \kappa(w)^*.$$

So, for an element $\kappa(v) \in A$ we have $\kappa(v) \geq \bigvee \mathrm{K}(w)$ if and only if $\kappa(v) \geq \kappa(w)^*$ (the converse implication is obvious). This shows that $k^\dagger = \alpha_g g(\kappa(w)^*) = \alpha_f f(\kappa(w)^*)$.

To prove that $\kappa(w)$ is equal to the big join, observe first that all the joinands are below $\kappa(w)$. We need only to prove that the big join is above any canonical joinand of $\kappa(w)$. Let u be a canonical joinand of $\kappa(w)$. If $u \in X$, then $u \in \{x \in X : w_\dagger \vee x \ngeq w\}$. So, let $u \notin X$. By the dual of Theorem 6.3 there is a unique canonical meetand of u not above $\kappa(w)$. Of course, this canonical meetand belongs to $\mathrm{M}(\kappa(w)) - \{\kappa(w)\}$ and thus equals $\kappa(v)$ for some $v \in \mathrm{J}(w) - \{w\}$. Since $\kappa(v) \ngeq \kappa(w)^*$, we have $\kappa(v) \ngeq w$. So, $u \leq k^\dagger \wedge \kappa(v)$ where $v \in \mathrm{J}(w) - \{w\}$ and $w \nleq \kappa(v)$. □

Lemma 6.28 *If w is completely join irreducible, then w_* is not a canonical joinand of any canonical meetand of w.*

Proof Suppose the lemma fails and let w be a counterexample of minimal rank. Then w is a completely join irreducible element and w_* is a canonical joinand of some canonical meetand w_1 of w. It follows from Theorem 6.3, that $w_1 = w_{11} \vee w_*$ canonically for some w_{11}. This implies $w_* \in \mathrm{J}(w)$ and so is completely join irreducible by Theorem 6.23 and w_{**} and $\kappa(w_*)$ exist. Moreover, since w_* is a subelement of w, it has lower rank.

First suppose that $\kappa(w_*) \leq \kappa(w)$. Now, by Corollary 6.2 applied to w_*, either $w_{**} \vee w_{11} \leq \kappa(w_*)$ or $w_{**} \vee w_{11} \geq w_*$. In the former case, $w_{11} \leq \kappa(w_*)$ and so

$$w \leq w_1 = w_{11} \vee w_* \leq \kappa(w_*) \vee w_* \leq \kappa(w),$$

a contradiction. In the latter case,

$$w_{11} \vee w_{**} = w_{11} \vee w_* \vee w_{**} = w_1.$$

Since $w_1 = w_{11} \vee w_*$ canonically, this contradicts the fact that canonical expression cannot be refined, see Theorem 2.3.

So we may assume that $\kappa(w_*) \nleq \kappa(w)$. Now $w \succ w_* \succ w_{**}$ and since w is join irreducible, the interval w/w_{**} contains only these three elements. Thus $w \wedge \kappa(w_*) = w_{**}$. Let $\kappa(w) = v_1 \vee \cdots \vee v_m$ canonically. By Theorem 6.3, we may assume that $v_i \leq w_{**}$, for $i \geq 2$. Applying (W) to

$$w_{**} = w \wedge \kappa(w_*) \leq \kappa(w) = \bigvee v_i$$

yields that $w_{**} \leq v_i$ for some i. If $i = 1$, then v_1 is above all other v_i's. This implies that $\kappa(w) = v_1$. Of course v_1 is join irreducible and $\kappa(w)$ is meet irreducible, and so $\kappa(w)$ must be a generator x. By Corollary 1.6, $w = \kappa'(\kappa(w)) = \kappa'(x) = \bigwedge X - \{x\} \succ 0$, and so w_{**} cannot exit.

Thus we must have that $w_{**} \leq v_i$ for some $i \geq 2$. But then $v_i = w_{**}$, i.e., w_{**} is a canonical joinand of $\kappa(w)$, which is a canonical meetand of w_*. Thus w_* is also a counterexample to the lemma. Since w_* has lower rank than w, this is a contradiction. □

The following theorem will prove to be very important in our study of chains of covers and finite intervals in free lattices.

Theorem 6.29 *Let w be a completely join irreducible element of $\mathbf{FL}(X)$, $w = w_1 \wedge \cdots \wedge w_m$ canonically and let w_* be join irreducible. Then $\{\kappa(w)\} \cup \{w_i : w_i \not\geq \kappa(w)\}$ is the set of canonical meetands of w_*.*

Proof By Theorem 6.1, $\kappa(w)$ is the unique canonical meetand of w_* not above w. Denote by $u_1, \ldots, u_k$ all the remaining canonical meetands of w_* and let $w_1, \ldots, w_n$ be all the canonical meetands of w not above $\kappa(w)$. Clearly, the element $w_1 \wedge \cdots \wedge w_n$ is in canonical form. Since $w_* = w \wedge \kappa(w) = w_1 \wedge \cdots \wedge w_n \wedge \kappa(w)$, we have $\{u_1, \ldots, u_k, \kappa(w)\} \gg \{w_1, \ldots, w_n, \kappa(w)\}$ and so $\{u_1, \ldots, u_k\} \gg \{w_1, \ldots, w_n\}$. In order to show that the two sets are equal, it remains to prove $u_1 \wedge \cdots \wedge u_k = w_1 \wedge \cdots w_n$, as this will then imply $\{w_1, \ldots, w_n\} \gg \{u_1, \ldots, u_k\}$. Suppose, on the contrary, that $u_1 \wedge \cdots \wedge u_k \not\leq w_i$ for some $i \leq n$. Then the condition (W) applied to the inequality $u_1 \wedge \cdots \wedge u_k \wedge \kappa(w) = w_* \leq w_i$, where w_i is to be expressed as the join of its canonical joinands, gives us that there exists a canonical joinand of w_i above w_*. However, by Theorem 6.3 all but one canonical joinands are below w_*; consequently, w_* is a canonical joinand of w_i. This of course contradicts Lemma 6.28. □

Theorem 6.30 *Let Y be a subset of X and w be a join irreducible element of the lattice $\mathbf{FL}(Y)$, which we can consider to be the the sublattice of $\mathbf{FL}(X)$ generated by Y. Then w has a lower cover in $\mathbf{FL}(Y)$ if and only if it has a lower cover in $\mathbf{FL}(X)$.*

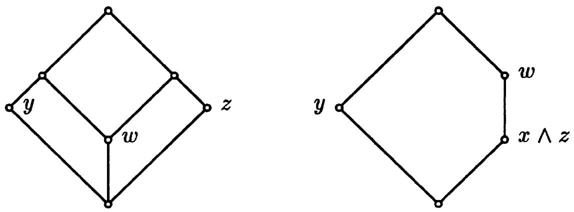
Proof This is a consequence of Theorem 6.23, as the lattice $\mathbf{L}^{\vee}(w)$ depends only on the elements in the sublattice of $\mathbf{FL}(X)$ generated by the elements of X that occur in the canonical expression of w. □

It is useful to realize in this connection that an element w of $\mathbf{FL}(Y)$, with Y a proper subset of X, is join irreducible in $\mathbf{FL}(X)$ if and almost always only if it is join irreducible in $\mathbf{FL}(Y)$. The only exception to this rule is the element $\bigwedge Y$, which is join irreducible in $\mathbf{FL}(X)$ but not join irreducible in $\mathbf{FL}(Y)$. This exceptional case could have been avoided by working in the variety of $(0,1)$-lattices, but this is not conventional for the study of free lattices.

Let us remark that while the existence of a lower cover of w in $\mathbf{FL}(X)$ depends only on the set $\mathrm{var}(w)$ of generators occurring in w, the element w_* actually covered by w does depend on the set X. Using Theorem 6.27, it is not hard to see that if $\kappa(w) = p(x_1, \ldots, x_n)$ in $\mathbf{FL}(\mathrm{var}(w))$, then in $\mathbf{FL}(X)$ the new $\kappa(w)$ is given (not necessarily in canonical form) by $p(x_1 \vee s, \ldots, x_n \vee s)$ where $s = \bigvee(X - \mathrm{var}(w))$.

7 Examples

Example 7.1 In $\mathbf{FL}(X)$ with $X = \{x, y, z\}$ take $w = x \wedge (y \vee z)$. Then $\mathrm{J}(w) = \{x \wedge (y \vee z), y, z\}$ and $\mathbf{L}^{\vee}(w)$ is the lattice drawn in Figure 7(1). This lattice fails meet semidistributivity, so as we concluded in Theorem 6.23, w does not have a lower cover. (This first known example of an element in $\mathbf{FL}(X)$ without lower cover is an unpublished result of R. A. Dean.) By the dual of Theorem 6.4, the upper covers of w are of the form $w \vee \kappa'(w_i)$ for each w_i in the canonical meet representation of w which has an upper cover. In this case, $w_1 = x$ and $\kappa'(x) = y \wedge z$, yielding the upper cover $(x \wedge (y \vee z)) \vee (y \wedge z)$; and $w_2 = y \vee z$ with $\kappa'(y \vee z) = x$, yielding the upper cover $(x \wedge (y \vee z)) \vee x = x$.



(1) $\mathbf{L}^{\vee}(x \wedge (y \vee z))$ (2) $\mathbf{L}^{\vee}(x \wedge (y \vee (x \wedge z)))$

Figure 7.

Example 7.2 Again in $\mathbf{FL}(X)$ with $X = \{x, y, z\}$ take $w = x \wedge (y \vee (x \wedge z))$. Then $\mathrm{J}(w) = \{w, y, x \wedge z\}$. The lattice $\mathbf{L}^{\vee}(w)$, which is drawn in Figure 7(2), is a five-element nonmodular lattice. This lattice is meet semidistributive, so we conclude that w has a lower cover. To find $\kappa(w)$, apply Theorem 6.27: $\kappa(w) = z \vee ((x \vee z) \wedge y)$. Then $w_* = w \wedge \kappa(w) = x \wedge (y \vee (x \wedge z)) \wedge (z \vee ((x \vee z) \wedge y))$. (This element is in canonical form.)

As in the preceding example, we obtain an upper cover of w in $w \vee \kappa'(x) = (x \wedge (y \vee (x \wedge z))) \vee (y \wedge z)$. However, this is the only upper cover, since the second canonical meetand of w, the element $y \vee (x \wedge z)$, has no upper cover by the argument dual to that given in Example 7.1. This means that the element w has no upper cover below x.

Example 7.3 Let $w = x \wedge (y \vee (x \wedge z)) \wedge (z \vee (x \wedge y)) \in \mathbf{FL}(x, y, z)$. Then $\mathbf{L}^{\vee}(w)$ is given in Figure 8 and $\kappa(w) = (y \wedge (x \vee z)) \vee (z \wedge (x \vee y))$.

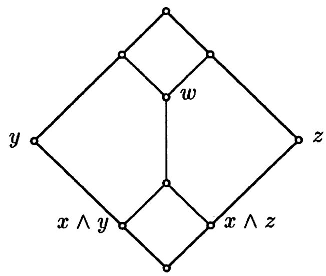


Figure 8.

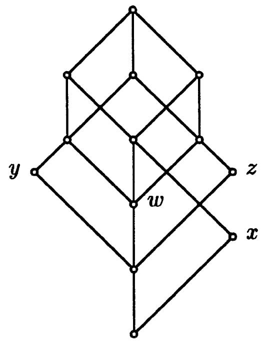


Figure 9.

Example 7.4 Let $w = (x \vee (y \wedge z)) \wedge (y \vee z)$. The $\mathbf{L}^{\vee}(w)$ is given in Figure 9. Since this lattice is not semidistributive, w has no lower cover.

Example 7.5 An element $w \in \mathbf{FL}(X)$ is called *coverless* if it has no lower and no upper covers. Here is an example of a coverless element in $\mathbf{FL}(X)$ with $X = \{x, y, z\}$:

$$w = ((x \wedge (y \vee z)) \vee (y \wedge z)) \wedge ((y \wedge (x \vee z)) \vee (x \wedge z)).$$

To see that w has no lower cover, take the subelement $x \wedge (y \vee z)$, which is without lower cover by Example 7.1, and apply Theorem 6.23. On the other hand, by the dual of (3) in Example 7.4, $(x \wedge (y \vee z)) \vee (y \wedge z)$ has no upper cover and symmetrically the same is true for $(y \wedge (x \vee z)) \vee (x \wedge z)$. Since these are the elements in the canonical meet representation of w, we can conclude that w has no upper cover.

8 Tschantz's theorem and semisingular elements

The last two sections present the basic theory of covers in free lattices. Using this theory, Nation and I were able to answer several fundamental questions about free lattices. For example, we were able to prove the following results.

Theorem 8.1 1. *A chain of covers in* $\mathbf{FL}(n)$ *not contained in the connected component of* 0 *or of* 1, *has length at most two.*

2. *A connected component not containing* 0 *or* 1 *does not have* $\mathbf{2} \times \mathbf{2}$ *as a cover preserving sublattice.*

3. *The finite interval sublattices of a free lattice are precisely the subintervals of the connected component of* 0 *in* $\mathbf{FL}(3)$, *which is diagrammed in Figure* 10, *and their duals.*

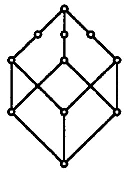


Figure 10

After proving these results, Nation and I sought to show that every infinite interval contained a copy of **FL**(3) and hence **FL**(ω). By Corollary 1.12, we only needed to show that every infinite interval had a set of three elements which is join and meet irredundant. We thought this would be easy to do, certainly much easier than our work on covers. We were quite surprised when we not only were unable to do this, we were not even able to prove that every infinite interval contained two incomparable elements. That is, we could not rule out the possibility that there was an infinite interval which was a chain!

We worked on trying to show that such an interval could not exist. We were able to draw several unlikely consequences from the existence of such an interval, but none of them lead to a direct contradiction. We showed the problem to my student, Tom Harrison. He was able to derive a series of much stronger consequences of the existence of such an interval, but still could not derive a contradiction. We also showed the problem to Steve Tschantz. In [30], he was able to show that such intervals cannot exist and that in fact every infinite interval contains **FL**(3).

Theorem 8.2 *Every infinite interval in a free lattice contains a sublattice isomorphic to* **FL**(ω).

His proof is quite involved. Most of the work goes into showing that an infinite interval cannot be a chain.

Let w be completely join irreducible. By Theorem 6.29 the canonical meetands of w_* are $\kappa(w)$ and those canonical meetands of w not above $\kappa(w)$. In most cases the canonical meetands of w_* include all of those of w. Elements where this is not so are called *semisingular*, i.e., a completely join irreducible element w is *semisingular* if $w_i \geq \kappa(w)$ for some canonical meetand w_i of w. A completely join irreducible element w is *singular* if $w_i \geq \kappa(w)$ for every canonical meetand of w, i.e., $w_* = \kappa(w)$. Singular elements are characterized in [20].

Although the concept of a semisingular element may seem technical, these elements turn out to be very important in the study of free lattices. Very recently Freese, Ježek, and Nation have characterized these elements. This characterization has several important consequences. First it gives a much shorter proof of Tschantz's Theorem. It also can be used to show that every interval of a free lattices which is neither atomic nor dually atomic has a maximal chain isomorphic to the rationals.

Theorem 8.3 $w \in$ **FL**(X) *is semisingular if and only if* $w_* = \kappa(w)$ *or* w *is the middle element of a three element interval.*

The proof of this theorem and the next corollary will first appear in [19]. Three element intervals are well understood, see section 10 of [18]. It is shown there that if w is meet reducible and the middle element of a three element interval $w_* \prec w \prec u$ and w_1 is the canonical meetand of w not above u, then $w = u \wedge w_1$ canonically, see Figure 11. Using this it is not hard to derive the following corollary.

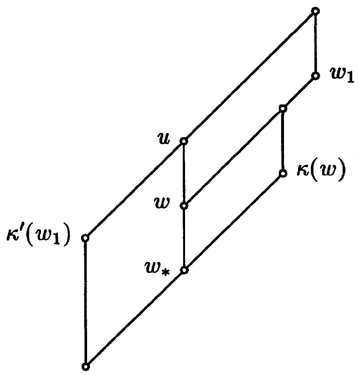


Figure 11

Corollary 8.4 *Let w be a completely join irreducible element of* $\mathbf{FL}(X)$ *with $\kappa(w) \neq w_*$. If $u \wedge \kappa(w) = w_*$, then either $u = w_*$, or $u = w$, or $u \succ w$. Moreover, if $u \succ w$ then w is semisingular.*

Using this corollary, it is not hard to show that no interval in a free lattice can be an infinite chain. For suppose that the interval c/d is an infinite chain in a free lattice. By Day's Theorem we can find a and b such that $d < b \prec a < c$. Moreover, since chains of covers are short, we can find such a and b so that c/a and b/d are infinite. Let q be a canonical joinand of a not below b. Then q is completely join irreducible with $q_* = q \wedge b$ and $\kappa(q) \geq b$. Since c/d is a chain, $c \wedge \kappa(q) = b \leq a = d \vee q$ and this contradicts (W) unless $q = a$ or $\kappa(q) = b$. By duality we may assume $q = a$ and hence $b = q_*$. But now $q_* = c \wedge \kappa(q)$ and since c/a is infinite, c cannot cover (or equal) $q = a$. Thus by Corollary 8.4, $q_* = \kappa(q)$, i.e., q is singular. It is shown in [20] that every such singular element is either in the connected component of 0 or of 1. But since $q = a$, this contradicts the fact that c/a and b/d are infinite.

9 Coverless elements

In this section we use Tschantz's Theorem to prove the following complement to Day's Theorem. This result will be used in the next section to construct maximal dense chains.

Theorem 9.1 *Every infinite interval of a free lattice contains a coverless element.*

Throughout this section $\sigma : \mathbf{FL}(X) \to \mathbf{FL}(Y)$ denotes a lattice embedding. Let

$$w = w_1 \vee \cdots \vee w_n$$

be the canonical form of an element of $\mathbf{FL}(X)$, and assume $n > 1$. The theorem will be proved with the aid of Theorem 8.2 and Theorem 6.3. In order to apply the latter theorem, we need to understand how σ affects the canonical form of w. While it is not true that every $\sigma(w_i)$ is a canonical joinand of $\sigma(w)$, we will show that this is true if w_i is not a generator.

Lemma 9.2 *If w_1 is a canonical joinand of w and $w_1 \notin X$ then $\sigma(w_1)$ is a canonical joinand of $\sigma(w)$.*

Proof Let the canonical form of w be given by (9) and let $w_1 = w_{11} \wedge \cdots \wedge w_{1k}$ canonically. Since $w_1 \notin X$, we have $k > 1$.

Let $\sigma(w) = u_1 \vee \cdots \vee u_r$ canonically. Since σ is one-one, $\sigma(w) = \sigma(w_1) \vee \cdots \vee \sigma(w_n)$, but $\sigma(w) > \sigma(w_2) \vee \cdots \vee \sigma(w_n)$. It follows that $\{u_1, \ldots, u_r\} \ll \{\sigma(w_1), \ldots, \sigma(w_n)\}$ but $\{u_1, \ldots, u_r\} \not\ll \{\sigma(w_2), \ldots, \sigma(w_n)\}$. Hence for some i we have

$$u_i \leq \sigma(w_1) = \sigma(w_{11}) \wedge \cdots \wedge \sigma(w_{1k}) \leq \sigma(w) = u_1 \vee \cdots \vee u_r.$$

For all j, we have $w_{1j} \not\leq w$, by one of the basic properties of canonical forms. Thus, since σ is one-one, $\sigma(w_{1j}) \not\leq \sigma(w)$ for all j. Hence, by (W), $u_i \leq \sigma(w_1) \leq u_j$, for some j. This forces $i = j$ and $u_i = \sigma(w_1)$, proving the lemma. □

Lemma 9.3 *If $w \in \mathbf{FL}(X)$, $w = w_1 \wedge \cdots \wedge w_n$ canonically with $n > 1$ and $w_1 = w_{11} \vee w_{12}$ canonically where $w_{1j} \notin X$ and $w_{1j} \not\leq w$, for $j = 1, 2$, then $\sigma(w)$ is not completely join irreducible in $\mathbf{FL}(Y)$.*

Proof Applying Lemma 9.2 to w_1, we see that both $\sigma(w_{11})$ and $\sigma(w_{12})$ are canonical joinands of $\sigma(w_1)$. Moreover, since σ is one-one, $\sigma(w_{1j}) \not\leq \sigma(w)$ for $j = 1, 2$. Now by Theorem 6.3, $\sigma(w)$ is not completely join irreducible. □

Now we are ready to complete the proof of Theorem 9.1. Take an infinite interval in a free lattice $\mathbf{FL}(X)$. By Tschantz's Theorem, there is an embedding σ of $\mathbf{FL}(\omega)$ into this interval. Let w be given by

$$w = ((x_0 \vee x_1) \wedge [(x_2 \wedge x_3) \vee (x_4 \wedge x_5)]) \vee (x_6 \wedge [(x_7 \wedge x_8) \vee (x_9 \wedge x_{10})]).$$

By the dual of Lemma 9.3 (with $w_1 = (x_0 \vee x_1) \wedge [(x_2 \wedge x_3) \vee (x_4 \wedge x_5)]$) we have that $\sigma(w)$ has no upper cover. By the same lemma, neither of its joinands has a lower cover. Thus, by Theorem 6.4, $\sigma(w)$ has no lower cover. □

10 Maximal chains

In this section we characterize those intervals in free lattices which contain a maximal dense chain, i.e., one with no covers. It follows from Theorem 6.4 that an element of a free lattice can have no more upper covers than the number of its canonical meetands. Thus it can have only finitely many upper (and lower) covers. This implies that if an interval is atomic, then every maximal chain in that interval must contain an atom and thus cannot be dense. Of course if it is dually atomic it also cannot have a dense chain. The Theorem 10.2 proves a converse to this.

First we require a lemma on connected components from [13]. The first two statements are the content of Lemma 1 of [13]. The third statement follows from the proof of that lemma.

Theorem 10.1 *Let $a \in \mathbf{FL}(n)$ and suppose that the connected component of a does not contain 0 or 1 and that it contains the elements indicated in Figure 12, where crosshatches indicate coverings. Then*

1. *a has no upper cover.*
2. *a has no lower cover except $b \wedge c$.*
3. *a is a proper join.*

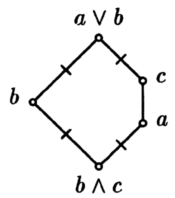


Figure 12

Theorem 10.2 *Suppose that $c > d$ in a free lattice and that c/d is neither atomic nor dually atomic. Then there is a maximal chain from c to d without any covers.*

Proof First we prove the theorem under the stronger hypothesis that c/d has no atom and no coatom and then we shall show how to derive the full result from this.

Let $a_i \succ b_i$, $i > 0$, be an enumeration of all the covers in c/d. Let $C_0 = \{c, d\}$. We build chains C_i's with the following properties.

1. Each C_i is finite.
2. Each element of $C_i - \{c, d\}$ is coverless.
3. $C_i \subseteq C_j$ if $i \leq j$.
4. C_i has an element which is incomparable with at least one of a_i and b_i.

Let $C = \bigcup C_i$. Clearly any maximal chain in c/d which contains C will have no cover.

Inductively, suppose that $C_0, \ldots, C_{i-1}$ have been constructed satisfying (1)–(4). Let $a = a_i$ and $b = b_i$. If C_{i-1} already has an element incomparable with either a or b, we let $C_i = C_{i-1}$. Otherwise, since C_{i-1} is finite, there are elements $e > f$ in C_{i-1} with $e \geq a \succ b \geq f$ and C_{i-1} has no element strictly between e and f. Since the elements of $C_{i-1} - \{c, d\}$

are coverless and c/d has no atoms and no coatoms, we must have $e > a \succ b > f$. Note that this implies that the connected component of a (which of course is the same as the connected component of b) does not contain 0 or 1.

Suppose that a is join reducible and let q be the canonical joinand of a which is not below b. Then q is completely join irreducible and $\kappa(q) \geq b$. If the interval $a/q \vee f$ is infinite, we can choose a coverless element r in this interval by Theorem 9.1. In this case we let $C_i = C_{i-1} \cup \{r\}$. Of course, r in incomparable to b.

Now suppose that $q \vee f < a$, but that $a/q \vee f$ is finite. By Theorem 8.1(1) any chain of covers containing a can length at most two. Note that

$$(q \vee f) \wedge \kappa(q) \prec q \vee f.$$

Indeed, $(q \vee f) \wedge \kappa(q)$ is above f and hence joins with q to $q \vee f$. Now the definition of κ shows that the above is a covering.

Thus we must have $a \succ q \vee f \succ (q \vee f) \wedge \kappa(q)$ and also $(q \vee f) \wedge \kappa(q) \leq b$. Let $v = [(q \vee f) \wedge \kappa(q)] \vee \kappa'(q \vee f)$. Then $v \succ (q \vee f) \wedge \kappa(q)$. This implies that $v < b$, since otherwise $\mathbf{2} \times \mathbf{2}$ is contained in the connected component of a, which would contradict Theorem 8.1. This situation is diagramed in Figure 13.

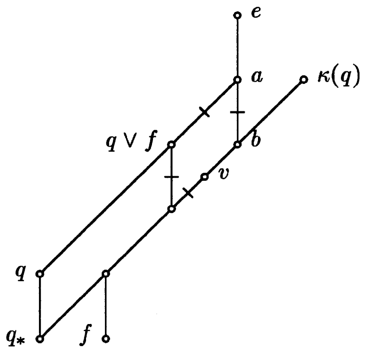


Figure 13

Assume now that b is meet reducible so that $b < \kappa(q)$. If $e \wedge \kappa(q)/b$ is infinite, we can again let $C_i = C_{i-1} \cup \{r\}$ for any coverless element r in this interval. If $e \wedge \kappa(q) = b$, then

$$e \wedge \kappa(q) = b \leq a = v \vee q$$

gives a violation of (W). Thus $e \wedge \kappa(q)/b$ is finite, and by the dual of the arguments above, $b \prec e \wedge \kappa(q)$. But by the dual of Theorem 10.1(2), the only cover of b is a.

Thus this situation cannot occur and we can conclude that either $q = a$ or $\kappa(q) = b$. By duality we may assume $q = a$ and so $q_* = b$. If $e \wedge \kappa(a) = b = a_*$, then, by Corollary 8.4, e either covers a or is equal to a. But this contradicts the fact that e has no lower cover. If $e \wedge \kappa(a)/b$ is infinite we can simply augment C_{i-1} with a coverless element from this interval. Thus we may assume that $e \wedge \kappa(a)/b$ is finite. Arguments as above show that we have the situation diagramed in Figure 14.

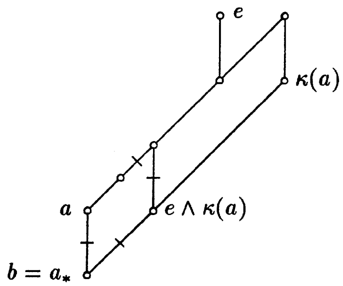


Figure 14

However, this contradicts Theorem 10.1(3), proving the theorem in the case c/d has no atom and no coatom.

Now suppose that c/d is a nontrivial interval which is neither atomic nor dually atomic. Then, by definition, we can find e and f in c/d such that e/d has no atom and c/f has no coatom.

Since e/d is atomless, we can find a descending chain $e_0 = e > e_1 > e_2 > \cdots > d$ with $\bigwedge e_n = d$. Suppose that for some n, $e_n \vee f < c$. Then since e_n/d is infinite, it contains a coverless element v such that $d < v < e_n$. Since c/f has no coatom, $c/(e_n \vee f)$ also has no coatom. In particular it is infinite. Thus we can find a coverless element $u \in c/(e_n \vee f)$ with $e_n \vee f < u < c$. Now all of the intervals v/d, u/v, and c/u are neither atomic nor dually atomic and hence all contain maximal dense chains by what was proved above. The union of these chains is a maximal dense chain in c/d.

Thus we may assume that $e_n \vee f = c$ for all n. But then since free lattices are continuous by Theorem 3.1,

$$f = f \vee d = f \vee \bigwedge e_n = \bigwedge f \vee e_n = c,$$

a contradiction. □

Despite all our knowledge about free lattices, many questions remain open. The following question was suggested by Tschantz.

Open Question 2 *Can a free lattice have an interval with $a/b = a/c \bigcup c/b$ and both infinite?*

Although it seems very unlikely such an interval could exist, we have not been able to prove that it cannot.

References

[1] Crawley, P., and Dilworth, R.P., *Algebraic Theory of Lattices*, Prentice-Hall, Englewood Cliffs, New Jersey, 1973.

[2] Czedli, G., and Freese, R., On congruence distributivity and modularity, *Algebra Universalis* **17** (1983), 216–219.

[3] Davis, A.C., A characterization of complete lattices, *Pacific J. Math.* **5** (1955), 311–319.

[4] Day, A., Doubling constructions in lattice theory, *Canad. J. Math.* **44** (1992), 252–269.

[5] Day, A., Lattice conditions implying congruence modularity, *Algebra Universalis* **6** (1976), 291–301.

[6] Day, A., Splitting lattices and congruence-modularity, in: *Contributions to Universal Algebra* (*Szeged 1975*) (B. Csákány and J. Schmidt, eds.), Colloq. Math. Soc. János Bolyai **17**, North-Holland, Amsterdam 1977, 57–71.

[7] Day, A., Splitting lattices generate all lattices, *Algebra Universalis* **7** (1977), 163–170.

[8] Day, A., Characterizations of finite lattices that are bounded-homomorphic images or sublattices of free lattices, *Canad. J. Math* **31** (1979), 69–78.

[9] Day, A., and Freese, R., A characterization of identities implying congruence modularity I., *Canad. J. Math.* **32** (1980), 1140–1167.

[10] Day, A., and Nation, J.B., A note on finite sublattices of free lattices, *Algebra Universalis* **15** (1982), 90–94.

[11] Dilworth, R.P., Lattices with unique complements, *Trans. Amer. Math. Soc.* **57** (1945), 123–154.

[12] Freese, R., Some order theoretic questions about free lattices and free modular lattices, in: *Ordered Sets* (I. Rival, ed.), Reidel (Kluwer), Dordrecht, 1982, NATO ASI Ser. C 83, 355–377.

[13] Freese, R., Connected components of the covering relation in free lattices, in: *Universal Algebra and Lattice Theory* (S. Comer, ed.), Lecture Notes in Math. **1149**. Springer-Verlag, Berlin-Heidelberg-New York, 1985, 82–93.

[14] Freese, R., Free lattice algorithms, *Order* **3** (1987), 331–344.

[15] Freese, R., Finitely presented lattices: continuity and semidistributivity, in: *Lattices, Semigroups, and Universal Algebra* (*Lisbon 1988*), Plenum Press, New York, 1990, 67–70.

[16] Freese, R., Kearnes, K., and Nation, J.B., Congruence lattices of semidistributive algebras, to appear in: *Proc. Conf. in Honor of Garrett Birkhoff* (*Darmstadt 1991*), Springer-Verlag, New York.

[17] Freese, R., and Nation, J.B., 3-3 lattice inclusions imply congruence modularity, *Algebra Universalis* **7** (1977), 191–194.

[18] Freese, R., and Nation, J.B., Covers in free lattices, *Trans. Amer. Math. Soc.* **288** (1985), 1–42.

[19] Freese, R., Ježek, J., and Nation, J.B., Free Lattices. Partially complete manuscript.

[20] Freese, R., Ježek, J., Nation, J.B., and Slavík, V., Singular covers in free lattices, *Order* **3** (1986), 39–46.

[21] Galvin, F., and Jónsson, B., Distributive sublattices of a free lattice, *Canad. J. Math.* **13** (1961), 265–272.

[22] Jónsson, B., Algebras whose congruence lattices are distributive, *Math. Scand.* **21** (1967), 110–121.

[23] Jónsson, B., Relatively free lattices, *Colloq. Math.* **21** (1970), 191–196.

[24] Jónsson, B., and Nation, J.B., A report on sublattices of a free lattice, in: *Contributions to Universal Algebra* (*Szeged 1975*) (B. Csákány and J. Schmidt, eds.), Colloq. Math. Soc. János Bolyai **17**, North-Holland, Amsterdam 1977, 233–257.

[25] McKenzie, R., Equational bases and non-modular lattice varieties, *Trans. Amer. Math. Soc.* **174** (1972), 1–43.

[26] McKenzie, R., McNulty, G., and Taylor, W., *Algebras, Lattices, Varieties,* Vol. I, Wadsworth and Brooks/Cole, Monterey, California, 1987.

[27] Nation, J.B., Bounded finite lattices, in: *Universal Algebra* (*Esztergom 1977*) (B. Csákány *et al.*, eds.), Colloq. Math. Soc. János Bolyai **29**, North-Holland, Amsterdam 1982, 531–533.

[28] Schweigert, D., Über endliche, ordnungspolynomvollständige Verbände, *Monatsh. Math.* **78** (1974), 68–76.

[29] Tarski, A., A lattice-theoretical fixpoint theorem and its applications, *Pacific J. Math.* **5** (1955), 285–309.

[30] Tschantz, S.T., Infinite intervals in free lattices, *Order* **6** (1990), 367–388.

[31] Whitman, Ph.M., Free lattices, *Ann. of Math. (2)* **42** (1941), 325–330.

[32] Wille, R., Eine Charakterisierung endlicher, ordnungspolynomvollständiger Verbände, *Arch. Math.* **28** (1977), 557–560.

A Survey of Boolean Algebras with Operators

Bjarni JÓNSSON
Department of Mathematics
Vanderbilt University
Box 1543, Sta. B
Nashville, TN 37240
U. S. A.

Abstract

The purpose of this survey is to call attention to the unifying role of the concept of a Boolean algebra with operators. The first chapter contains a brief history of this concept and a list of theorems from universal algebra that play an important role in the treatment of the subject. The next two chapters are concerned with the general theory, in particular canonical extensions and dualities, while each of the remaining chapters treats a particular class of BAO's. The presentation is highly incomplete. In each chapter some basic concepts are introduced and their properties are illustrated by stating a number of key theorems, mostly without proofs. The bibliography contains only a miniscule portion of the relevant literature, but should be sufficient to open up the subject to a reader who wants to pursue a particular topic further.

Acknowledgement

I wish to thank the organizers of the symposium for the opportunity to present this material to a large group of young mathematicians. Thanks are also due to my colleagues and students at Vanderbilt University who over the last several years have participated in seminars where most of these topics were discussed. Special thanks are due to Peter Jipsen, whose advice and assistance have improved this manuscript both in style and in substance, as well as in appearance.

1 Background and history

1.1 The origin of relation algebras

The theory of Boolean algebras with operators evolved from Tarski's work on relation algebras. The calculus of binary relations had been intensively investigated during the second half of the last century by DeMorgan, Peirce and Schröder, and a detailed account of their work can be found in Schröder [1895]. Tarski's aim was to present their work in an axiomatic framework. His 1941 paper in the Journal of Symbolic Logic set the stage for this, and the axioms that he eventually used were, with some minor modifications, the ones presented there. His primitive concepts were, in addition to the Boolean operations

I. G. Rosenberg and G. Sabidussi (eds.), Algebras and Orders, 239–286.

and constants, the binary operation of relative multiplication (;), the unary operation of conversion ($\smile$) and the unit, or identity element (1'). Thus he considered structures

$$\mathbf{A} = (\mathbf{A}_0, ;, 1', ^\smile)$$

such that

$\mathbf{A}_0 = (A, +, 0, \cdot, 1, ^-)$ *is a Boolean algebra*

$(A, ;, 1', ^\smile)$ *is a monoid with involution*

; *and* $^\smile$ *distribute over* $+$.

These axioms are of the most elementary nature, and they are quite weak. The final axiom, which gives his system its distinctive flavor, can be written

$$a; (a^\smile; b^-)^- \leq b.$$

This axiom is not as simple as the other ones. Modulo the other axioms, it is equivalent to the condition

$$a; x \leq y \quad \text{iff} \quad x \leq a \backslash y, \quad \text{where} \quad a \backslash y = (a^\smile; y^-)^-.$$

Using the fact that $^\smile$ is an involution, we can also write this in the form

$$x; a \leq y \quad \text{iff} \quad x \leq y/a, \quad \text{where} \quad y/a = (y^-; a^\smile)^-.$$

In other words, the axiom asserts that the operation ; is residuated, with the right and left residuals \ and / given by the indicated formulas. An equivalent property, extensively used by Tarski, is the condition

$$(x; y) \cdot z = 0 \quad \text{iff} \quad (x^\smile; z) \cdot y = 0, \quad \text{iff} \quad (z; y^\smile) \cdot x = 0.$$

In his terminology, this asserts that the operations $y \mapsto a; y$ and $z \mapsto a^\smile; z$ are conjugates of each other or, equivalently, that the operations $x \mapsto x; a$ and $z \mapsto z; a^\smile$ are conjugates of each other.

Tarski defined a proper relation algebra to be an algebra of subrelations of an equivalence relation, with the indicated operations. We shall refer to these as the primary models. It is obvious that the axioms hold in the primary models, and they are therefore suitable, but it is far less obvious whether they are adequate. This question could be interpreted in several ways. Minimally, in order for a set Σ of identities to be regarded as adequate, it would have to have the property that (1) many, or preferably most, of the important identities that hold in the primary models are consequences of Σ. This is of course a subjective criterion. Ideally it should be the case that (2) every identity that holds in the primary models is a consequence of Σ. An apparently stronger condition would be that (3) every model of Σ is isomorphic to one of the primary models, but Tarski showed later that (2) and (3) are equivalent. In his 1941 paper he did mention the representation problem, the question whether his axioms satisfy the criterion (3), but formulated no conjecture. He only expressed the opinion that he could prove from his axioms "the hundreds of theorems to be found in Schröder's 'Algebra der Logik'." This claim was largely confirmed in his seminars at Berkeley in the '40's, where he developed the arithmetic of relation algebras, with some contributions from his students, including Julia Robinson, Louise Chin and me.

1.2 The extension theorem for relation algebras

Although Tarski did not talk much about the representation problem, it was probably on his mind. Late in 1946, after I had left Berkeley, I received from him a communication where he showed that every relation algebra can be embedded in a complete and atomic relation algebra. He did not say so, but I assume that his motivation was that this would be a step towards proving a representation theorem. If this was the case, then he was mistaken, for in Lyndon [1950] it was shown that the representation problem has a negative solution. However, his result was obviously of independent interest, and as it turned out, his method could be applied in a much more general context. Let me therefore describe the basic ideas of his proof.

We consider a relation algebra

$$\mathbf{A} = (\mathbf{A}_0, ;, 1', ^{\smile}),$$

and we want to embed it in a complete and atomic relation algebra

$$\mathbf{A}^\sigma = (\mathbf{A}_0^\sigma, ;, 1', ^{\smile}).$$

The Boolean algebra $\mathbf{A}_0$ can be embedded in a complete and atomic Boolean algebra $\mathbf{A}_0^\sigma$, called the *canonical extension* of $\mathbf{A}_0$, with the following two properties:

(a) *For any distinct atoms p and q of $\mathbf{A}_0^\sigma$, there exists an element $a \in A$ with $p \leq a$ and $q \leq a^-$.*

(b) *Every subset of A whose supremum in $\mathbf{A}_0^\sigma$ is 1 has a finite subset whose supremum is also 1.*

This is an algebraic way of describing the extension that arises from the Stone Duality Theorem, which asserts that every Boolean algebra is isomorphic to the field of all clopen subsets of a Boolean space. To see this equivalence, observe that a field $\mathbf{F}$ of subsets of a set U is the field of all clopen subsets of a Boolean space iff the following two conditions hold:

(a') *For any distinct members p and q of U, there exists $X \in \mathbf{F}$ with $p \in X$ and $q \notin X$.*

(b') *Every sub-family $\mathbf{G}$ of $\mathbf{F}$ with $U = \bigcup \mathbf{G}$ has a finite sub-family $\mathbf{H}$ with $U = \bigcup \mathbf{H}$.*

A set field $\mathbf{F}$ is said to be *regular* if it has these properties. Thus $\mathbf{F}$ is regular iff the family of all subsets of U is a canonical extension of $\mathbf{F}$.

The advantage of the algebraic characterization is that it enables us to work more on one level. Nothing is lost. In particular, the topology is still there. The atoms of $\mathbf{A}_0^\sigma$ can be taken to be the points of the Stone space $\mathbf{X}$ of $\mathbf{A}_0$, and the map

$$a \mapsto \{p \in \mathbf{X} : p \leq a\}$$

is an isomorphism from $\mathbf{A}_0$ onto the set field of all clopen subsets of $\mathbf{X}$. The members of A are therefore referred to as *clopen elements* and the members of A^σ that are suprema or infima of subsets of A are referred to, respectively, as *open elements* and as *closed elements.*

With this extension of $\mathbf{A}_0$, Tarski defined the relative product of two atoms p and q, and the converse of an atom p, by the formulas

$$\begin{aligned} p;q &= \prod\{a;b : p \leq a \in A \text{ and } q \leq b \in A\}, \\ p^{\smile} &= \prod\{a^{\smile} : p \leq a \in A\}. \end{aligned}$$

The operations were then extended to arbitrary elements of A^σ by additivity. Of course it had to be checked that the new operations agree with the old ones on the original set, and the axioms had to be checked one by one in order to show that the new structure was a relation algebra.

1.3 The origin of Boolean algebras with operators

It is obvious that Tarski's construction can be carried out in a much more general setting. An operation f on a Boolean algebra $\mathbf{A}_0$ is called an *operator* if it is additive in each of its arguments. If, in addition, f takes on the value 0 whenever one of the arguments is 0, then f is said to be *normal.* An algebra $\mathbf{A} = (\mathbf{A}_0, f_i,\ i \in I)$ in which all the operations f_i are operators on the Boolean algebra $\mathbf{A}_0$ is called a *Boolean algebra with operators*, or briefly a BAO, and if all the f_i's are normal, then $\mathbf{A}$ is said to be *normal.* For an n-ary normal operator f, the *canonical extension* f^σ can be defined analogously to the extension of the relative product and the converse in a relation algebra, namely

$$f^\sigma(x) = \sum\{\prod\{f(y) : p \leq y \in A^n\} : x \geq p \in P^n\}, \text{ for all } x \in (A^\sigma)^n,$$

where P is the set of all atoms of $\mathbf{A}_0^\sigma$, but if f is not normal, then P must be replaced by the set $Q = P \cup \{0\}$. The operation f^σ is then a completely additive operator on $\mathbf{A}_0^\sigma$ that agrees with f on A, and if f is normal, then f^σ is obviously also normal. Finally, the *canonical extension* of a BAO $\mathbf{A} = (\mathbf{A}_0, f_i, i \in I)$ is defined to be $\mathbf{A}^\sigma = (\mathbf{A}_0^\sigma, f_i^\sigma, i \in I)$.

The concept of a Boolean algebra with operators was introduced in Jónsson, Tarski [1951], and the basic properties of canonical extensions of BAO's were developed there. The second part of that paper, Jónsson, Tarski [1952], is primarily devoted to the applications to relation algebras. The results listed in this section and in the next one are from the first part.

1.3.1 The Extension Theorem *The canonical extension $\mathbf{A}^\sigma$ of a BAO $\mathbf{A}$ is a complete and atomic BAO containing $\mathbf{A}$ as a subalgebra. If $\mathbf{A}$ is normal, then so is $\mathbf{A}^\sigma$.*

By saying that $\mathbf{A}^\sigma$ is atomic, we mean that its Boolean reduct $\mathbf{A}_0^\sigma$ is atomic; and by saying that $\mathbf{A}^\sigma$ is complete we mean that $\mathbf{A}_0^\sigma$ is complete, and that each of the operators f_i^σ is completely additive in all its arguments.

The notion of a BAO evolved out of the study of relation algebras, but an investigation of the new concept would not be justified unless there were other important examples. At the time, the only other BAO's that had been investigated to any extent were closure algebras, in McKinsey, Tarski [1944], [1946], [1948]. The concept of a projective algebra, introduced in Everett, Ulam [1946], served as a precursor to the notion of a cylindric algebra, which was just emerging. However, the Extension Theorem suggested a large supply of BAO's. In a normal BAO that is complete and atomic, each operator is completely determined by its action on the atoms. This action can be described by a relation on the set of atoms, and the operator can be reconstructed from this relation. To put this more precisely, we introduce the notion of the complex algebra of a relational structure. First, if R is a relation of rank $n+1$ on a set U, then we define R^+ to be the operation of rank n on the power set of U such that, for any subsets $X_0, X_1, \ldots, X_{n-1}$ of U,

$$\begin{aligned} R^+(X_0, X_1, \ldots, X_{n-1}) \; = \; & \{y \in U : (x_0, x_1, \ldots, x_{n-1}, y) \in R \\ & \text{for some } x_0 \in X_0, x_1 \in X_1, \ldots, x_{n-1} \in X_{n-1}\}. \end{aligned}$$

By the *complex algebra* of a relational structure $\mathbf{U} = (U, R_i, i \in I)$ we mean the BAO $\mathbf{U}^+ = (U^+, R_i^+, i \in I)$, where U^+ is the Boolean algebra of all subsets of U. Obviously, $\mathbf{U}^+$ is a complete, atomic and normal BAO. On the other hand, if $\mathbf{A}$ is a BAO that is complete, atomic and normal, then there is a unique relational structure $\mathbf{U}$ with $U = \mathrm{At}(\mathbf{A})$ such that $\phi : \mathbf{A} \cong \mathbf{U}^+$, where $\phi(a) = \{p \in U : p \leq a\}$ for all $a \in A$. Combined with the Extension Theorem, this yields a very general representation theorem: *Every normal BAO can be embedded in the complex algebra of some relational structure.* In order to be of interest, this result needs to be strengthened to record what is known about the relation between the given BAO and the relational structure used to represent it. A subalgebra $\mathbf{A}$ of a complete and atomic BAO $\mathbf{A}'$ is said to be a *regular subalgebra* of $\mathbf{A}'$ if the identity automorphism of $\mathbf{A}$ can be extended to an isomorphism from $\mathbf{A}^\sigma$ onto $\mathbf{A}'$.

1.3.2 The Representation Theorem *For every normal BAO* $\mathbf{A}$ *there exists, up to isomorphism, a unique relational structure* $\mathbf{U}$ *such that* $\mathbf{A}$ *is isomorphic to a regular subalgebra of* $\mathbf{U}^+$.

1.4 The Preservation Theorem

It is not of much value to know that a normal BAO $\mathbf{A}$ can be embedded in the complex algebra of a relational structure $\mathbf{U}$ unless we also know how the properties of $\mathbf{U}$ are related to properties of $\mathbf{A}$. In order to describe this relationship, we need to know what properties of BAO's are preserved by canonical extensions. In particular, we ask which identities are preserved. Thus we want to know whether, for terms s and t, $s^{\mathbf{A}} = t^{\mathbf{A}}$ implies $s^{\mathbf{A}^\sigma} = t^{\mathbf{A}^\sigma}$. A natural way to try to prove this would be to show that, in general, $t^{\mathbf{A}^\sigma} = (t^{\mathbf{A}})^\sigma$, but an obvious trouble with this approach is that, for an arbitrary term t, the operation $t^{\mathbf{A}}$ need not be additive, in which case its canonical extension is undefined. We therefore need to define canonical extensions for a larger class of operations. Assuming that f is an isotone operation of rank n on a Boolean algebra $\mathbf{A}_0$, we define

$$f^\sigma(x) = \sum\{\prod\{f(z) : y \leq z \in A^n\} : x \geq y \in K^n\} \text{ for } x \in A^n,$$

where K is the set of all closed elements of $\mathbf{A}_0^\sigma$. For operators, this agrees with the earlier notion. The isotone maps form a clone, and one might hope that the map $f \mapsto f^\sigma$ from the clone of isotone maps on $\mathbf{A}_0$ to the clone of isotone maps on $\mathbf{A}_0^\sigma$ preserved composition. This turned out not to be the case, and in Jónsson, Tarski [1951] we therefore had to consider a smaller clone, the clone generated by the operators. For operations in this clone, we proved that

(1) $$(f[g_0, g_1, \ldots, g_{n-1}])^\sigma = f^\sigma[g_0^\sigma, g_1^\sigma, \ldots, g_{n-1}^\sigma].$$

A term or an equation in the language of BAO's is said to be *strongly positive* if it does not contain the symbol for the complementation. An operation in the clone of a BAO $\mathbf{A}$ is said to be strongly positive if it can be represented by a strongly positive term, and the clone consisting of all such operations of $\mathbf{A}$ is referred to as the strongly positive clone of $\mathbf{A}$. Applied to the members of this clone, the formula (1) yields a preservation theorem for a large class of identities.

1.4.1 The Preservation Theorem *Every strongly positive identity that holds in a BAO* $\mathbf{A}$ *holds also in* $\mathbf{A}^\sigma$.

From this, other preservation theorems were obtained, in particular that if f and g are conjugate operators, then so are f^σ and g^σ. The original proof of the Preservation Theorem was rather indirect, largely due to the fact that there was no direct characterization of the functions involved. A simpler proof was given in Ribeiro [1952]. Ribeiro's key lemma was that (1) holds whenever f is an operator and all the g_i's are isotone. About twenty years later, a new proof was given in Henkin [1970]. Henkin proved (1) for a larger clone of operations, the so-called ω-additive operations. Although these operations are explicitly defined, his proof is rather long. In Sections 2.2-2.3 we shall describe a shorter proof using his concept.

The very general theorems stated in this and the preceding section set the stage for a more detailed study of the connection between particular classes of normal BAO's and classes of relational structures. Some results along these lines can be found in Jónsson, Tarski [1951], [1952]. Among other things, it is shown that a BAO $\mathbf{A} = (\mathbf{A}_0, f)$ with a single unary operation is a closure algebra iff $\mathbf{A}^\sigma$ is isomorphic to the complex algebra of a quasi-ordered set, and that the complex algebra of a generalized Brandt groupoid is a relation algebra. It should be noted in this context that in Duffin, Pate [1943] complex algebras of groups were investigated axiomatically, and that it is clear from their axioms that the structures they obtain are relation algebras. For a long time, little work was done on BAO's in general, although special classes of BAO's were intensively investigated, notably cylindric algebras by Henkin, Monk and Tarski, and polyadic algebras by Halmos.

1.5 Modal algebras

Although not much happened in the general theory of BAO's for a long while, there were certain related developments, notably in connection with modal logic. Modal logics are obtained from classical logic by adding a single unary predicate $\Diamond$. The axioms for (normal) modal logic are such that the Lindenbaum, Tarski algebra of the propositional modal calculus is a BAO with a normal unary operator. Such algebras are therefore referred to as modal algebras. The Lindenbaum, Tarski algebras yield a bijective correspondence between modal logics and varieties of modal algebras. In McKinsey [1941] this correspondence was used to prove that two important modal logics, S4 and S5, are decidable. The *finite modal property* introduced there has been extensively used to prove similar results for other modal logics. In McKinsey, Tarski [1948], closure algebras are used to study the logics S4 and S5. The notion of a *possible world*, introduced in Kripke [1963], has played a fundamental role in modal logic. *Kripke structures* are simply binary structures $\mathbf{U} = (U, R)$. In a footnote in his paper, Kripke mentions that while writing the paper he had become aware of overlapping results in Jónsson, Tarski [1951], [1952], but that he had not read these papers. In a series of papers by E. J. Lemmon, the techniques of McKinsey and Tarski are combined with those of Kripke, but unaware of the Jónsson, Tarski papers he rediscovered special cases of results from these papers. A fundamental question in modal logic was settled in Fine [1974] and Thomason [1974], where it was shown that some modal logics are incomplete. In a more algebraic terminology, this means that some varieties of modal algebras are not generated by complex algebras of Kripke frames.

1.6 Varieties

During the long period when BAO's were in hibernation, much happened in universal algebra that later played a role in the development of the subject. Birkhoff's Preservation Theorem came earlier, in 1935. This asserts that every class of algebras that is closed under the operations $\mathbb{H}$, $\mathbb{S}$ and $\mathbb{P}$ of forming homomorphic images, subalgebras and direct products is a variety, or an equational class. As noted by Tarski, this can be expressed compactly and conveniently by the formula $\mathbb{V}\mathrm{ar}(\mathcal{K}) = \mathbb{HSP}(\mathcal{K})$. Together with Birkhoff's Subdirect Product Theorem, which appeared almost a decade later, this laid the foundation for the theory of varieties. But for a detailed study of special varieties, stronger results were needed. Alfred Foster began his study of primal algebras in the early '50's. A primal algebra is a finite algebra in which every operation on the universe is a term operation. At the time, this appeared to me to be quite special, and even artificial. Obviously I am not a logician! However, I was intrigued by some of his results, in particular by the fact that the variety generated by a primal algebra $\mathbf{A}$ is simply the class of all subdirect powers of $\mathbf{A}$,— $\mathbb{V}\mathrm{ar}(\mathbf{A}) = \mathbb{P}\mathrm{s}(\mathbf{A})$,— and that the only subdirectly irreducible member of this variety is $\mathbf{A}$ itself. This is in sharp contrast to the general situation, for even in the variety generated by a single finite algebra there are often subdirectly irreducible members of arbitrarily large cardinality, and the subdirectly irreducibles need not form an elementary class. I finally found a weaker form of this phenomenon that applies to arbitrary congruence distributive varieties.

1.6.1 Theorem *If the class $\mathcal{K}$ of algebras generates a congruence distributive variety, then* $\mathbb{V}\mathrm{ar}(\mathcal{K}) = \mathbb{P}\mathrm{s}\mathbb{HS}\mathbb{P}\mathrm{u}(\mathcal{K})$.

Here $\mathbb{P}\mathrm{u}$ is the operation of taking ultraproducts. For me this result was right at the time, for I was interested in varieties of lattices. I did in fact realize that the discriminator played a central role in Foster's argument, but dismissed it because the discriminator operation on a non-trivial lattice is never a term operation. How wrong I was! Other people gradually realized the importance of this operation, and eventually the theory of discriminator varieties was born. Many people, besides Foster, contributed to this development, notably

Stan Burris, Stephen Comer, Alden Pixley, Robert Quackenbush and Heinrich Werner. For an excellent exposition, see Werner [1976].

We recall the definition of a discriminator variety, and the basic properties of these varieties.

1.6.2 Definition

(i) By the *ternary discriminator* on a set U we mean the operation f of rank 3 on U such that, for all $a, b, c \in U$,

$$f(a,b,c) = \begin{cases} a & \text{if } a \neq b \\ c & \text{if } a = b \end{cases}$$

(ii) A ternary term t is called a *discriminator term* on a class $\mathcal{K}$ of algebras if t represents the discriminator on each member of $\mathcal{K}$.

(iii) A variety $\mathcal{V}$ of algebras is called a *discriminator variety* if $\mathcal{V}$ is generated by a class $\mathcal{K}$ of algebras such that there exists a discriminator term on $\mathcal{K}$.

Why is the discriminator important? We first make the obvious observation that if a clone $\mathbf{C}$ of operations on a set U contains the discriminator operation on U, then $\mathbf{C}$ is closed under definition by cases. That is, if the n-ary operations h_0, h_1, h_2, h_3 are in $\mathbf{C}$, then so is the operation h with

$$h(x) = \begin{cases} h_0(x) & \text{if } h_2(x) = h_3(x) \\ h_1(x) & \text{otherwise} \end{cases}$$

Related to this is the less obvious fact, established in the next two theorems, that for a class of algebras possessing a discriminator term certain formulas that are not equations can be replaced by equations.

1.6.3 Theorem *Suppose $\mathcal{K}$ is a class of algebras for which a discriminator term exists. Let Φ be the smallest class of formulas that contains all the equations and is closed under the operations of forming conjunctions, disjunctions and implications. Then there is an effective way of associating with each member ϕ of Φ a term ϕ^* such that $\mathcal{K} \models \phi \leftrightarrow \phi^* \approx x$, where x is a variable that does not occur in ϕ.*

The set Φ in the above theorem includes all strong open Horn formulas. It is not closed under negation. Indeed, in an algebra $\mathbf{A} = (A, f)$ with f the discriminator operation on A, the condition $x \neq y$ cannot be expressed as an equation, for every equation will be satisfied in $\mathbf{A}$ whenever the same value is assigned to all the variables. However, for most of the classes considered here, Φ can be taken to be the set of all open formulas.

1.6.4 Theorem *Suppose $\mathcal{K}$ is a class of algebras for which a discriminator term exists, and suppose there exist terms 0 and 1 such that $\mathcal{K} \models \neg(0 \approx 1)$. Then there is an effective way of associating with each open formula ϕ a term ϕ^* such that $\mathcal{K} \models \phi \leftrightarrow \phi^* \approx 0$.*

The following omnibus theorem contains most of the basic facts about discriminator varieties.

1.6.5 Theorem *Suppose $\mathcal{V}$ is a discriminator variety generated by a class $\mathcal{K}$ of algebras, with t a discriminator term for $\mathcal{K}$. Then the following statements hold.*

I. *$\mathcal{V}$ is congruence permutable, congruence distributive, congruence extensile, congruence regular, congruence uniform, and semisimple.*

II. *For any non-trivial algebra $\mathbf{A} \in \mathcal{V}$, the following conditions are equivalent:*

  (i) $\mathbf{A}$ *is subdirectly irreducible.*

  (ii) $\mathbf{A}$ *is simple.*

  (iii) $\mathbf{A}$ *is directly indecomposable.*

  (iv) $\mathbf{A} \in \mathbb{S}\mathbb{P}\mathrm{u}(\mathcal{K})$.

  (v) t *represents the discriminator operation on* $\mathbf{A}$.

III. *For every $\mathbf{A} \in \mathcal{V}$,*

  (i) *For all $a, b \in A$, $\mathrm{Cg}(a,b) = \{(x,y) \in A^2 : t^{\mathbf{A}}(a,b,x) = t^{\mathbf{A}}(a,b,y)\}$.*

  (ii) *The principal congruences on $\mathbf{A}$ form a relatively complemented sublattice of* $\mathrm{Con}(\mathbf{A})$.

  (iii) *Every compact congruence on $\mathbf{A}$ is a principal congruence.*

  (iv) *Every principal congruence on $\mathbf{A}$ is a factor congruence.*

IV. *The class $\mathbb{H}\,\mathbb{S}\mathrm{i}(\mathcal{V})$, which consists of the members of $\mathbb{S}\mathrm{i}(\mathcal{V})$ and of the trivial algebra, is a universal class, and the map $\mathcal{M} \mapsto \mathbb{V}\mathrm{ar}(\mathcal{M})$ is an isomorphism from the lattice of all universal subclasses of $\mathbb{H}\,\mathbb{S}\mathrm{i}(\mathcal{V})$ onto the lattice of all subvarieties of $\mathcal{V}$. The inverse of this map is $\mathcal{U} \mapsto \mathbb{H}\,\mathbb{S}\mathrm{i}(\mathcal{U}) = \mathcal{U} \cap \mathbb{H}\,\mathbb{S}\mathrm{i}(\mathcal{V})$.*

Some of the terminology used in the above theorem may need an explanation. A variety $\mathcal{V}$ is said to be *congruence regular* if every congruence relation R on an algebra in $\mathcal{V}$ is determined by any one of its blocks, and $\mathcal{V}$ is said to be *congruence uniform* if all the blocks of R always have the same cardinality. We say that $\mathcal{V}$ is *congruence extensile* if every congruence relation on a subalgebra of an algebra $\mathbf{A} \in \mathcal{V}$ can be extended to a congruence relation on $\mathbf{A}$. To say that $\mathcal{V}$ is *semisimple* means that every subdirectly irreducible member of $\mathcal{V}$ is simple.

The property III(i) shows that, in a discriminator variety, principal congruences are equationally definable. The larger class of varieties characterized by this condition shares several of the important properties of discriminator varieties. This class was first investigated by P. Köhler and D. Pigozzi, whose work evolved out of earlier studies by E. Fried, G. A. Grätzer and R. Quackenbush of congruence schema. A series of papers by W. J. Blok and D. Pigozzi extends their results and applies them to modal algebras.

1.6.6 Definition A variety $\mathcal{V}$ is said to have *equationally definable principal congruences*, — briefly EDPC,— if for some $n \in \omega$ there exist ternary terms s_i, t_i, $0 \leq i \leq n$, such that, for all $\mathbf{A} \in \mathcal{V}$ and $a, b, c, d \in A$,

$$(c,d) \in \mathrm{Cg}(a,b) \quad \text{iff} \quad s_i^{\mathbf{A}}(a,b,c,d) = t_i^{\mathbf{A}}(a,b,c,d) \quad \text{for} \quad 0 \leq i \leq n.$$

The following concept will be used in the summary of the results about varieties with EDPC. Given two elements a and b in a join semilattice $\mathbf{L}$, if the inclusion $a \leq b + x$ has a smallest solution $x = x_0$ in $\mathbf{L}$, then we refer to x_0 as the *dual relative pseudocomplement of* b *in* a, and write $x_0 = b * a$. If $b * a$ exists for all $a, b \in L$, then $\mathbf{L}$ is said to be *dually Brouwerian*. The principal results from Köhler, Pigozzi [1980] are contained in the next three theorems.

1.6.7 Theorem *Suppose $\mathcal{V}$ is a variety with EDPC, and let s_i, t_i, $i \leq n$, be as in* 1.6.6. *Then, for any $\mathbf{A} \in \mathcal{V}$ and $a, b, c, d \in A$, the congruence relation*

$$\sum\{\mathrm{Cg}(s_i^{\mathbf{A}}(a,b,c,d), t_i^{\mathbf{A}}(a,b,c,d)) : 0 \leq i \leq n\}$$

is a dual relative pseudocomplement of $\mathrm{Cg}(c,d)$ *in* $\mathrm{Cg}(a,b)$ *in the join semilattice of all compact congruence relations on* $\mathbf{A}$.

1.6.8 Theorem *A variety $\mathcal{V}$ has EDPC iff, for all $\mathbf{A} \in \mathcal{V}$, the join semilattice of all compact congruence relations on $\mathbf{A}$ is dually Brouwerian.*

1.6.9 Theorem *Every variety with EDPC is congruence distributive and congruence extensile.*

Splitting algebras, introduced by R. N. McKenzie in his investigations of varieties of lattices have also played an important role in the study of varieties of BAO's. A subdirectly irreducible algebra $\mathbf{S}$ in a variety $\mathcal{V}$ is said to be *splitting* in $\mathcal{V}$ if there exists a largest subvariety $\mathcal{U}$ of $\mathcal{V}$ with $\mathbf{S} \notin \mathcal{U}$, and the variety $\mathcal{U}$ is then called the *conjugate variety of* $\mathcal{U}$. We denote by $\mathbb{H}_\omega(\mathbf{A})$ the class of all algebras that are isomorphic to $\mathbf{A}/\mathrm{R}$ for some compact congruence relation R on the algebra $\mathbf{A}$. The following result, Corollary 3.2 in Blok, Pigozzi [1982], is particularly useful.

1.6.10 Theorem *If $\mathcal{V}$ is a variety with EDPC, then every finitely presentable, subdirectly irreducible algebra $\mathbf{S} \in \mathcal{V}$ is splitting in $\mathcal{V}$, and the conjugate variety of $\mathbf{S}$ is*

$$\{\mathbf{A} \in \mathcal{V} : \mathbf{S} \notin \mathbb{S}\mathbb{H}_\omega(\mathbf{A})\}.$$

2 Canonical extensions

The canonical extension $\mathbf{A}^\sigma$ of a Boolean algebra $\mathbf{A}$, defined in Section 1.1, is characterized by the properties that $\mathbf{A}^\sigma$ is a complete and atomic Boolean algebra, containing $\mathbf{A}$ as a subalgebra, and that

(a) *For any distinct atoms p and q of $\mathbf{A}^\sigma$, there exists $a \in A$ with $p \leq a$ and $q \leq a^-$.*

(b) *Every subset of A whose supremum in $\mathbf{A}^\sigma$ is 1 has a finite subset whose supremum is also 1.*

We now define, for a map $f : \mathbf{A} \to \mathbf{B}$, where $\mathbf{A}$ and $\mathbf{B}$ are Boolean algebras, the canonical extension $f^\sigma : \mathbf{A}^\sigma \to \mathbf{B}^\sigma$ of f. In the general case, "extension" is a misnomer, but if f is isotone, then f^σ agrees with f on $\mathbf{A}$. The central result is that the map $f \mapsto f^\sigma$ is a contravariant functor on a certain category of maps between Boolean algebras, and from this the fundamental preservation theorems for canonical extensions are obtained. A more detailed account can be found in Jónsson [a].

2.1 Isotone maps

We assume that $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are Boolean algebras.

2.1.1 Definition For any map $f : \mathbf{A} \to \mathbf{B}$, we define

$$f^\sigma(x) = \sum\{\prod\{f(z) : y \leq z \in A\} : x \geq y \in K\} \text{ for all } x \in A^\sigma,$$

where K is the set of all closed elements of $\mathbf{A}^\sigma$.

Note that

$$\begin{aligned} f^\sigma(y) &= \prod\{f(z) : y \leq z \in A\} && \text{for all} \quad y \in K, \\ f^\sigma(x) &= \sum\{f^\sigma(y) : x \geq y \in K\} && \text{for all} \quad x \in A^\sigma. \end{aligned}$$

Clearly, f^σ is isotone, and if f is isotone, then f^σ agrees with f on A. Isotone maps between Boolean algebras form a category, and since $^\sigma$ sends objects in this category into objects, and morphisms into morphisms, one might hope that $^\sigma$ was a functor. This, however, is not the case. Example: Take $\mathbf{B}$ to be any infinite Boolean algebra, let $\mathbf{A} = \mathbf{B} \times \mathbf{B}$, and let $\mathbf{C}$ be the two-element Boolean algebra. Let $f : \mathbf{A} \to \mathbf{B}$ and $g : \mathbf{B} \to \mathbf{C}$ be the maps such that $f(x, y) = x + y$ for all $x, y \in B$, $g(x) = 0$ whenever $1 \neq x \in B$, and $g(1) = 1$. Then $(gf)^\sigma(x, y) = 1$ for all $x, y \in B^\sigma$ with $x + y = 1$, and in particular, $(gf)^\sigma(x, x^-) = 1$ for all $x \in B^\sigma$, but $g^\sigma f^\sigma(x, x^-) = 0$ whenever x is not clopen. (Observe that we are identifying $\mathbf{A}^\sigma$ with $\mathbf{B}^\sigma \times \mathbf{B}^\sigma$.) However, the following does hold.

2.1.2 Theorem (H. Ribeiro [1952]) *If $f : \mathbf{A} \to \mathbf{B}$ and $g : \mathbf{B} \to \mathbf{C}$ are isotone maps, then*

$$(gf)^\sigma(x) \leq g^\sigma f^\sigma(x) \quad \textit{for all} \quad x \in A^\sigma,$$

with equality holding whenever x is closed.

2.2 ω-additive operations

Having seen that the category of all isotone maps is too large for our purposes, we now turn to a smaller category. As before, we assume that $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are Boolean algebras. We denote by $S_m(U)$ the set of all subsets of U with at most m elements.

2.2.1 Definition A map f from $\mathbf{A}$ to $\mathbf{B}$ is said to be *m-additive*, where m is a positive integer, if for every finite subset U of A,

$$f(\sum U) = \sum\{f(\sum V) : V \in S_m(U)\}. \tag{1}$$

If $\mathbf{A}$ and $\mathbf{B}$ are complete, and if (1) holds for every subset U of A, then f is said to be *completely m-additive.* If f is m-additive, or completely m-additive, for some positive integer m, then f is said to be *ω-additive*, or *completely ω-additive*, respectively.

2.2.2 Lemma *For an isotone map $f : \mathbf{A} \to \mathbf{B}$, the following conditions are equivalent.*

(i) *f is m-additive.*

(ii) *For all $a \in A$ and $p \in \mathrm{At}(\mathbf{B}^\sigma)$ with $p \leq f(a)$, there exists $x \in A^\sigma$ of height at most m with $x \leq a$ and $p \leq f^\sigma(x)$.*

(iii) *For all $a \in A^\sigma$ and $p \in \mathrm{At}(\mathbf{B}^\sigma)$ with $p \leq f^\sigma(a)$, there exists $x \in A^\sigma$ of height at most m with $x \leq a$ and $p \leq f^\sigma(x)$.*

2.2.3 Lemma *If $f : \mathbf{A} \to \mathbf{B}$ is m-additive, then*

$$f^\sigma(x) = \sum\{\prod\{f(z) : y \leq z \in A\} : x \geq y \in K\} \text{ for all } x \in A^\sigma,$$

where K is the set of all elements of A^σ of height at most m.

2.2.4 Theorem *If the map $f : \mathbf{A} \to \mathbf{B}$ is m-additive, then the map $f^\sigma : \mathbf{A}^\sigma \to \mathbf{B}^\sigma$ is completely m-additive.*

2.2.5 Theorem *If the maps $f : \mathbf{A} \to \mathbf{B}$ and $g : \mathbf{B} \to \mathbf{C}$ are n-additive and m-additive, respectively, then the map gf is mn-additive.*

2.2.6 Theorem *If $f : \mathbf{A} \to \mathbf{B}$ is isotone and $g : \mathbf{B} \to \mathbf{C}$ is ω-additive, then*

$$(gf)^\sigma = g^\sigma f^\sigma.$$

2.2.7 Theorem *The map $f \mapsto f^\sigma$ is a covariant functor from the category of all ω-additive maps between Boolean algebras to the category of all completely ω-additive maps between complete and atomic Boolean algebras.*

2.3 Identities preserved by canonical extensions

The applications of the results in the preceding section to BAO's are based on three observations. The first observation is that an operator of rank n on a Boolean algebra $\mathbf{A}$, when regarded an a map $f : \mathbf{A}^n \to \mathbf{A}$, is ω-additive. A more general version of this is formulated in Lemma 2.3.2 below. The second observation is that the canonical extension $(\mathbf{A}^n)^\sigma$ of $\mathbf{A}^n$ may be identified with $(\mathbf{A}^\sigma)^n$, and the map $f^\sigma : (\mathbf{A}^n)^\sigma \to \mathbf{A}^\sigma$ can therefore be treated as an operation of rank n on $\mathbf{A}^\sigma$. The third observation is a triviality of a very general nature. When considering a composition $f[g_0, g_1, \ldots, g_{n-1}]$, where f and the g_i's are operations on $\mathbf{A}$ of ranks n and m, respectively, we can think of $[g_0, g_1, \ldots, g_{n-1}]$ as a map $h : \mathbf{A}^m \to \mathbf{A}^n$ with $h(x) = (g_0(x), g_1(x), \ldots, g_{n-1}(x))$. It is easy to check that

$$[g_0, g_1, \ldots, g_{n-1}]^\sigma = [g_0^\sigma, g_1^\sigma, \ldots, g_{n-1}^\sigma].$$

Although we are primarily concerned with BAO's, we formulate the preservation results for the larger class of algebras introduced by Henkin. By a *translate* of an operation f (of some positive rank n) we mean an operation of rank one obtained by assigning fixed values to all but one of the arguments in f. It is not in general true for an operation f on a Boolean algebra that if all the translates of f are ω-additive, then f is ω-additive. This is why the next definition must be formulated with some care. An algebra $\mathbf{A} = (\mathbf{A}_0, f_i, i \in I)$, consisting of a Boolean algebra $\mathbf{A}_0$ and operations f_i on the universe A of $\mathbf{A}_0$, will be referred to as an *expanded Boolean algebra.* The algebra $\mathbf{A}^\sigma = (\mathbf{A}_0^\sigma, f_i^\sigma, i \in I)$ is referred to as the *canonical extension* of $\mathbf{A}$.

2.3.1 Definition An expanded Boolean algebra $\mathbf{A} = (\mathbf{A}_0, f_i, i \in I)$ is called a *Henkin algebra* if, for each $i \in I$, there exists a positive integer n such that all the translates of f_i are n-additive.

Equivalently, $\mathbf{A}$ is a Henkin algebra if, for each $i \in I$, the map $f_i : \mathbf{A}^{\nu(i)} \to \mathbf{A}$ is ω-additive, where $\nu(i)$ is the rank of f_i.

2.3.2 Lemma *Suppose* $\mathbf{A}$ *is a Henkin algebra.*

(i) *Every member of the strongly positive clone of* $\mathbf{A}$ *is* ω*-additive.*

(ii) *If the n-ary operation f and the m-ary operations* $g_0, g_1, \ldots, g_{n-1}$ *are in the strongly positive clone of* $\mathbf{A}$, *then*

$$(f[g_0, g_1, \ldots, g_{n-1}])^\sigma = f^\sigma[g_0^\sigma, g_1^\sigma, \ldots, g_{n-1}^\sigma].$$

2.3.3 Theorem *If* $\mathbf{A}$ *is a Henkin algebra, then for any strongly positive term t in the language of* $\mathbf{A}$, $t^{\mathbf{A}^\sigma} = (t^{\mathbf{A}})^\sigma$.

2.3.4 Theorem *For any Henkin algebra* $\mathbf{A}$, *and for any strongly positive terms s and t in the language of* $\mathbf{A}$,

$$\mathbf{A} \models s \approx t \quad \textit{implies} \quad \mathbf{A}^\sigma \models s \approx t.$$

2.4 Other properties preserved by canonical extension

Quasi-identities are not in general preserved by canonical extensions. Example: Let $\mathbf{A} = (\mathbf{A}_0, f)$, where $\mathbf{A}_0$ is a complete and atomic Boolean algebra with countably many atoms $p_0, p_1, \ldots$ and f is the completely additive unary operator with $f(p_i) = p_{i+1}$ for all $i \in \omega$. Then the only fixed point of f is the zero element, but the meet in $\mathbf{A}^\sigma$ of the elements $f^n(1)$, $n = 0, 1, \ldots$, is a non-zero fixed point of f^σ. Therefore, if t is the term corresponding to the operation f, then the quasi-identity

$$t(x) \approx x \to x \approx 0$$

holds in $\mathbf{A}$, but not in $\mathbf{A}^\sigma$. However, certain special classes of quasi-identities and certain other first order properties are preserved.

2.4.1 Theorem *Suppose* $\mathbf{A}$ *is a Henkin algebra. If the formula* α *has one of the following forms, and if* $\mathbf{A} \models \alpha$, *then* $\mathbf{A}^\sigma \models \alpha$.

(i) α *is* $s \approx 0 \rightarrow t \approx u$, *with* s, t, u *strongly positive terms.*

(ii) α *is* $\neg(s \approx 0) \rightarrow t \approx u$, *with* s,t,u *strongly positive terms.*

(iii) α *is* $\beta \rightarrow t \approx u$, *with* t *and* u *strongly positive terms and* β *a conjunction of Boolean equations that is satisfied by the sequence* $(0, 0, \ldots, 0)$.

The proofs of (i) and (ii) are quite easy. We adjoin to $\mathbf{A}$ the unary operator g such that $g(x) = 1$ for $x \neq 0$ and $g(0) = 0$. If v is the term corresponding to the new operation, then (i) and (ii) are equivalent to the identities

$$v(s) + t \approx v(s) + u, \qquad v(s) \cdot t \approx v(s) \cdot u,$$

and each of these is preserved by Theorem 2.3.4. The proof of (iii) involves a somewhat more detailed analysis.

The results listed in the preceding section, and so far in this one, are obviously not in their most general form. Similar results apply to "operations" involving two or more Boolean algebras. The reason why they have not been stated in the more general form is that this would involve the more cumbersome language of heterogeneous algebras, but would require no new ideas. In most instances where such generalizations are needed, either there is a simple direct proof, or else the argument used in the homogeneous case carries over routinely. As a simple but important example, consider homomorphisms between Henkin algebras. The following notation will be useful both here and later. Given a map $h : U \rightarrow V$ and a positive integer n, we let $h^{[n]} : U^n \rightarrow V^n$ be the map such that, for all $p_0, p_1, \ldots, p_{n-1} \in U$,

$$h^{[n]}(p_0, p_1, \ldots, p_{n-1}) = (h(p_0), h(p_1), \ldots, h(p_{n-1})).$$

2.4.2 Lemma *Suppose* f *and* g *are* ω*-additive operations of rank* n *on the Boolean algebras* $\mathbf{A}$ *and* $\mathbf{B}$, *respectively, and suppose* $h : \mathbf{A} \rightarrow \mathbf{B}$ *is* ω*-additive. Then*

$$gh^{[n]} = hf \quad \textit{implies} \quad g^\sigma h^{\sigma[n]} = h^\sigma f^\sigma.$$

2.4.3 Theorem *The map* $f \mapsto f^\sigma$ *is a covariant functor from the category of all homomorphisms between Henkin algebras to the category of all complete homomorphisms between complete and atomic Henkin algebras.*

The twin concepts of residuation and conjugacy, discussed in Chapter 1, apply to maps between Boolean algebras.

2.4.4 Definition Suppose $\mathbf{A}$ and $\mathbf{B}$ are Boolean algebras and $f : \mathbf{A} \rightarrow \mathbf{B}$.

(i) By a *residual* of f we mean a map $g : \mathbf{B} \to \mathbf{A}$ such that, for all $x \in A$ and $y \in B$,

$$f(x) \leq y \quad \text{iff} \quad x \leq g(y).$$

If such a g exists, then we say that f is *residuated.*

(ii) By a *conjugate* of f we mean a map $h : \mathbf{B} \to \mathbf{A}$ such that, for all $x \in A$ and $y \in B$,

$$f(x) \cdot y = 0 \quad \text{iff} \quad x \cdot h(y) = 0.$$

The maps g and h, if they exist, are unique. If one exists, then so does the other, and they are related by the formulas

$$h(y) = g(y^-)^-, \qquad g(y) = h(y^-)^-.$$

The relation of conjugacy is symmetric: If h is a conjugate of f, then f is a conjugate of h. If f is residuated, then it is a normal operator. In fact, it preserves all existing joins:

$$x = \sum\{x_i : i \in I\} \quad \text{implies} \quad f(x) = \sum\{f(x_i) : i \in I\}.$$

Of course the conjugate h of f has the same properties. The residual g of f has the dual properties:

$$y = \prod\{y_i : i \in I\} \quad \text{implies} \quad g(y) = \prod\{g(y_i) : i \in I\}.$$

It is for this reason that, when considering canonical extensions, we usually treat the conjugate as the basic concept.

2.4.5 Theorem *Suppose* $\mathbf{A}$ *and* $\mathbf{B}$ *are Boolean algebras. If the maps* $f : \mathbf{A} \to \mathbf{B}$ *and* $g : \mathbf{B} \to \mathbf{A}$ *are conjugates of each other, then so are* $f^\sigma : \mathbf{A}^\sigma \to \mathbf{B}^\sigma$ *and* $g^\sigma : \mathbf{B} \to \mathbf{A}^\sigma$.

The notions of residuality can also be applied to operations of rank greater than one, or more generally, to maps from a direct product of two or more Boolean algebras into a Boolean algebra. We consider only the binary case.

2.4.6 Definition Suppose $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are Boolean algebras, and consider a map $\circ : \mathbf{A} \times \mathbf{B} \to \mathbf{C}$.

(i) By a *right conjugate* of $\circ$ we mean a map $\rhd : \mathbf{A} \times \mathbf{C} \to \mathbf{B}$ such that, for all $x \in A$, $y \in B$ and $z \in C$,

$$(x \circ y) \cdot z = 0 \quad \text{iff} \quad (x \rhd z) \cdot y = 0.$$

(ii) By a *left conjugate* of $\circ$ we mean a map $\lhd : \mathbf{C} \times \mathbf{B} \to \mathbf{A}$ such that, for all $x \in A$, $y \in B$ and $z \in C$,

$$(x \circ y) \cdot z = 0 \quad \text{iff} \quad (z \lhd y) \cdot x = 0.$$

(iii) If $\circ$ has a right and a left conjugate, then we say that $\circ$ is *residuated.*

Equivalently, $\circ$ is residuated if, for all $a \in A$ and $b \in B$, the maps $y \mapsto a \circ y$ and $x \mapsto x \circ b$ are residuated, the conjugates of these maps being $z \mapsto a \triangleright z$ and $z \mapsto z \triangleleft b$.

2.4.7 Theorem *Suppose* $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{C}$ *are Boolean algebras, and suppose the map* $\circ : \mathbf{A} \times \mathbf{B} \to \mathbf{C}$ *has a right conjugate* $\triangleright : \mathbf{A} \times \mathbf{C} \to \mathbf{B}$ *and a left conjugate* $\triangleleft : \mathbf{C} \times \mathbf{B} \to \mathbf{A}$. *Then the map* $\circ^\sigma : \mathbf{A}^\sigma \times \mathbf{B}^\sigma \to \mathbf{C}^\sigma$ *has* $\triangleright^\sigma : \mathbf{A}^\sigma \times \mathbf{C}^\sigma \to \mathbf{B}^\sigma$ *as its right conjugate and* $\triangleleft^\sigma : \mathbf{C}^\sigma \times \mathbf{B}^\sigma \to \mathbf{A}^\sigma$ *as its left conjugate.*

3 Dualities

The construction of the complex algebra of a relational structure gives rise to a duality between a category of suitably defined morphisms between structures and the category of all complete homomorphisms between normal, complete and atomic BAO's. The Stone duality between Boolean algebras and Boolean spaces can be extended to a duality between the category of all normal BAO's of a fixed type and a category whose objects are certain topological relational structures. These two dualities have been investigated in detail in Goldblatt [1989], and most of the results listed here can be found in that paper.

3.1 Complex algebras

For any set U there is, up to isomorphism, a unique complete and atomic Boolean algebra whose set of atoms is U. We denote this algebra by U^+. This departure from the notation in Section 1.3, where U^+ denoted the Boolean algebra of all subsets of U, results in a change in the definition of a complex algebra, although up to isomorphism the new concept agrees with the old one.

3.1.1 Definition

(i) If R is a relation of rank $n+1$ on a set U, then R^+ is defined to be the operation of rank n on U^+ such that, for all $x \in (U^+)^n$,

$$R^+(x) = \sum\{q \in U : (p,q) \in R \text{ for some } p \in U^n \text{ with } p \leq x\}.$$

(ii) For a relational structure $\mathbf{U} = (U, R_i, i \in I)$, we let

$$\mathbf{U}^+ = (U^+, R_i^+, i \in I),$$

and we refer to $\mathbf{U}^+$ as the *complex algebra of* $\mathbf{U}$.

If $\mathbf{A}$ is a complete and atomic Boolean algebra, then we let $\mathbf{A}_+$ be the set of all atoms of $\mathbf{A}$. Observe that, for any set U, $(U^+)_+ = U$. It is convenient, and usually harmless, to identify $(\mathbf{A}_+)^+$ with $\mathbf{A}$. This will make the description of the fundamental dualities a great deal simpler.

3.1.2 Definition

(i) Suppose $\mathbf{A}$ is a complete and atomic Boolean algebra, and let $U = \mathbf{A}_+$. If f is a complete normal operator of rank n on $\mathbf{A}$, then we let

$$f_+ = \{(p,q) : p \in U^n,\ q \in U \text{ and } q \leq f(p)\}.$$

(ii) For any normal, complete and atomic BAO $\mathbf{A} = (\mathbf{A}_0,\ f_i,\ i \in I))$, we let

$$\mathbf{A}_+ = ((\mathbf{A}_0)_+,\ (f_i)_+,\ i \in I)),$$

and we refer to $\mathbf{A}_+$ as the *atomic structure* of $\mathbf{A}$.

The involutive correspondence $\mathbf{A} \mapsto \mathbf{A}_+$ and $U \mapsto U^+$ between complete and atomic Boolean algebras and sets is part of a duality between two categories, the category of all complete homomorphisms between complete and atomic Boolean algebras and the category of all maps between sets. For morphisms $g : \mathbf{A} \to \mathbf{B}$ and $h : U \to V$ in these categories, the dual morphisms $g_+ : \mathbf{B}_+ \to \mathbf{A}_+$ and $h^+ : V^+ \to U^+$ are defined by

$$g_+(q) = \prod\{a \in A : q \leq g(a)\} \quad \text{for } q \in \mathbf{B}_+,$$

$$h^+(b) = \sum\{p \in U : h(p) \leq b\} \quad \text{for } b \in V^+.$$

This notation is ambiguous, for if g is a complete operator of rank n on a complete and atomic Boolean algebra, then according to an earlier definition, g_+ is a relation of rank $n+1$ on the set of all atoms. However, the intended meaning will always be clear from the context.

There is a similar duality between the category of all complete homomorphisms between normal, complete and atomic BAO's of a given type and a suitably defined category of morphisms between relational structures. Following Goldblatt, we refer to these morphisms as bounded.

3.1.3 Definition Suppose $\mathbf{U} = (U, R_i,\ i \in I)$ and $\mathbf{V} = (V, S_i,\ i \in I)$ are similar structures. A map $h : U \to V$ is called a *bounded morphism* from $\mathbf{U}$ to $\mathbf{V}$ if, for all $i \in I$, $p \in U$ and $y \in V^n$, where R_i and S_i are of rank $n+1$,

$$(y, h(p)) \in S_i \quad \text{iff} \quad (\exists x \in U^n)[(x,p) \in R_i \text{ and } h^{[n]}(x) = y].$$

3.1.4 Theorem *Suppose* $\mathbf{U}$ *and* $\mathbf{V}$ *are similar relational structures. A map* $h : U \to V$ *is a bounded morphism from* $\mathbf{U}$ *to* $\mathbf{V}$ *iff* $h^+ : \mathbf{V}^+ \to \mathbf{U}^+$ *is a complete homomorphism.*

3.1.5 Theorem *Let* $\mathbf{BAO}^{ca}$ *be the category of all complete homomorphisms between normal, complete and atomic BAO's, and let* $\mathbf{RS}$ *be the category of all bounded morphisms between relational structures. Then* $g \mapsto g_+$ *is a contravariant functor from* $\mathbf{BAO}^{ca}$ *to* $\mathbf{RS}$, $h \mapsto h^+$ *is a contravariant functor from* $\mathbf{RS}$ *to* $\mathbf{BAO}^{ca}$, *and the two functors are inverses of each other.*

It should be noted that, since the notion of bounded morphism differs from the usual notion of a homomorphism between structures, the notion of a substructure will also undergo a change.

3.1.6 Definition By an *inner substructure* of a structure $\mathbf{U} = (U, R_i, i \in I)$ we mean a substructure $\mathbf{V} = (V, S_i, i \in I)$ of $\mathbf{U}$ such that, for all $i \in I$ and for all $p \in V$ and $x \in U^n$, where the rank of R_i is $n+1$, the condition $(x, p) \in R_i$ implies that $x \in V^n$.

3.1.7 Theorem *Suppose* $\mathbf{V}$ *is a substructure of* $\mathbf{U}$. *Then* $\mathbf{V}$ *is an inner substructure of* $\mathbf{U}$ *iff the injection* $V \to U$ *is a bounded morphism.*

The duality correlates to each construction in one of the two classes a construction in the other. The following is a simple example.

3.1.8 Theorem *Suppose* $\mathbf{U}_i$, $i \in I$, *are similar relational structures. Then*

$$\prod\{\mathbf{U}_i^+ : i \in I\} \cong \mathbf{U}^+,$$

where $\mathbf{U}$ *is the disjoint union of the* $\mathbf{U}_i$*'s.*

3.2 Topological duality

The Stone space $\mathbf{A}_\delta = (X,\mathbf{T})$ of a Boolean algebra $\mathbf{A}$ can be taken to have the atoms of $\mathbf{A}^\sigma$ as its points, with the sets $F(a) = \{p \in X : p \leq a\}$, $a \in A$, as a basis for the topology $\mathbf{T}$. Conversely, the dual algebra $\mathbf{X}^\delta$ of a Boolean space $\mathbf{X} = (X, \mathbf{T})$ can be taken to be the regular subalgebra of X^+ whose universe consists of the elements $a \in X^+$ such that the set $F(a)$ is clopen. Thus $(\mathbf{A}_\delta)^\delta = \mathbf{A}$ and $(\mathbf{X}^\delta)_\delta = \mathbf{X}$. For a continuous map $g : \mathbf{X} \to \mathbf{Y}$, the dual $g^\delta : \mathbf{Y}^\delta \to \mathbf{X}^\delta$ is the restriction of g^+ to $\mathbf{Y}^\delta$, while the dual $h_\delta : \mathbf{B}_\delta \to \mathbf{A}_\delta$ of a homomorphism $h : \mathbf{A} \to \mathbf{B}$ is the map $(h^\sigma)_+$.

The topological dual $\mathbf{A}_\delta$ of a normal BAO $\mathbf{A} = (\mathbf{A}_0, f_i, i \in I)$ will be the topological structure $\mathbf{X} = (X,\mathbf{T}, R_i, i \in I)$, where $(X,\mathbf{T})$ is the dual space of the Boolean algebra $\mathbf{A}_0$ and $R_i = (f_i^\sigma)_+$ for $i \in I$. The correspondence $\mathbf{A} \mapsto \mathbf{A}_\delta$ is part of a duality between the category of all morphisms between normal BAO's and a category whose objects are certain topological relational structures. To describe this duality, we must characterize the objects and the morphisms in the latter category. It is clear that each object will consist of a Boolean space and a relational structure, both with the same universe, and the only question that remains is the connection of the relations to the topology. This question was settled in Halmos [1962] for the case of binary relations, and in Goldblatt [1989] for the general case. The term "Boolean relation" is due to Halmos.

3.2.1 Definition

(i) Suppose $\mathbf{X} = (X, \mathbf{T})$ is a Boolean space. A relation R of rank $n+1$ on X is said to be *Boolean* if, for every clopen subset U of X^n, the set

$$\{p \in X : (q, p) \in R \text{ for some } q \in U\}$$

is clopen, and for every $p \in U$, the set $\{q \in X^n : (q, p) \in R\}$ is closed.

(ii) By a *Boolean structure* we mean a structure $\mathbf{X} = (\mathbf{X}_0, R_i, i \in I)$ such that $\mathbf{X}_0 = (X, \mathbf{T})$ is a Boolean space and each R_i is a Boolean relation on X.

(iii) Suppose $\mathbf{X} = (\mathbf{X}_0, R_i, i \in I)$ and $\mathbf{Y} = (\mathbf{Y}_0, S_i, i \in I)$ are Boolean structures. By a *morphism* from $\mathbf{X}$ to $\mathbf{Y}$ we mean a continuous function from $\mathbf{X}_0$ to $\mathbf{Y}_0$ that is a bounded morphism from $(X, R_i, i \in I)$ to $(Y, S_i, i \in I)$.

The categories involved have now been specified, and we are ready to define the two functors.

3.2.2 Definition

(i) For a normal operator f on a Boolean algebra $\mathbf{A}$, we define $f_\delta = (f^\sigma)_+$.

(ii) For a normal BAO $\mathbf{A} = (\mathbf{A}_0, f_i, i \in I)$, we define $\mathbf{A}_\delta = ((\mathbf{A}_0)_\delta, (f_i)_\delta, i \in I)$.

(iii) For a homomorphism $g : \mathbf{A} \to \mathbf{B}$, where $\mathbf{A}$ and $\mathbf{B}$ are normal BAO's, we define $g_\delta = (g^\sigma)_+$.

3.2.3 Definition

(i) For a Boolean relation R on a Boolean space $\mathbf{X}$, we define R^δ to be the restriction of R^+ to $\mathbf{X}^\delta$.

(ii) For a Boolean structure $\mathbf{X} = (\mathbf{X}_0, R_i, i \in I)$, we define $\mathbf{X}^\delta = ((\mathbf{X}_0)^\delta, (R_i)^\delta, i \in I)$.

(iii) For a morphism $h : \mathbf{X} \to \mathbf{Y}$, where $\mathbf{X}$ and $\mathbf{Y}$ are Boolean structures, we define h^δ to be the restriction of h^+ to $\mathbf{Y}^\delta$.

3.2.4 Theorem *Let* **BAO** *be the category of all homomorphisms between normal BAO's, and let* **BRS** *be the category of all morphisms between Boolean structures. Then* $g \mapsto g_\delta$ *is a contravariant functor from* **BAO** *to* **BRS**, $h \mapsto h^\delta$ *is a contravariant functor form* **BRS** *to* **BAO**, *and the two functors are inverses of each other.*

The diagram in Figure 3.1 shows eight functors between four categories. The categories are

BAO *normal Boolean algebras with operators and homomorphisms.*

BAO^{ca} *complete and atomic normal BAOs and complete homomorphisms.*

BRS *Boolean structures and continuous bounded morphisms.*

RS *Relational structures and bounded morphisms.*

The functors represented by the arrows labelled $(\cdot)^\sigma, (\cdot)^+, (\cdot)_+, (\cdot)^\delta$ and $(\cdot)_\delta$ have been discussed above. The left arrow in the top line represents the injection function, and the right arrow in the bottom line is the functor that forgets the topology of the Boolean structure. The functor labelled $(\cdot)_\beta$ in the bottom line sends an object (a relational structure) $\mathbf{U}$ into

$$\begin{tikzcd}[column sep=huge, row sep=huge]
\mathbf{BAO} \arrow[r, shift left, "(\cdot)^{\sigma}"] \arrow[d, shift right, "(\cdot)_{\delta}"'] & \mathbf{BAO}^{ca} \arrow[l, shift left, "\text{injection}"] \arrow[d, shift right, "(\cdot)_{+}"'] \\
\mathbf{BSR} \arrow[r, shift left, "\text{forgetful}"] \arrow[u, shift right, "(\cdot)^{\delta}"'] & \mathbf{RS} \arrow[l, shift left, "(\cdot)_{\beta}"] \arrow[u, shift right, "(\cdot)^{+}"']
\end{tikzcd}$$

Figure 3.1

$\mathbf{U}_\beta = (\mathbf{U}^+)_\delta$, which can also be constructed as the Stone-Čech compactification of the discrete topology on U, with the relations on $\mathbf{U}_\beta$ given by the closures of the relations on $\mathbf{U}$.

Under the duality in Theorem 3.2.4, direct products of finitely many normal BAO's correspond to disjoint unions of Boolean structures, but the disjoint union of infinitely many compact topological spaces is not compact. The direct product operation on Boolean structures can be used to define an operation on normal BAO's that appears to be of some interest.

3.2.5 Definition If $\mathbf{A}$ and $\mathbf{B}$ are similar normal BAO's, then we let $\mathbf{A} \otimes \mathbf{B} = (\mathbf{A}_\delta \times \mathbf{B}_\delta)^\delta$, and we refer to $\mathbf{A} \otimes \mathbf{B}$ as the *tensor product* of $\mathbf{A}$ and $\mathbf{B}$.

The tensor product of two Boolean algebras coincides with their free product. The general notion of a tensor product has not been much investigated, but it appears to be of some interest. Example: If $\mathbf{A}$ is a relation algebra and $\mathbf{B}$ is the algebra of all binary relations on an n-element set, then $\mathbf{A} \otimes \mathbf{B}$ is the relation algebra of all n by n matrices over $\mathbf{A}$.

3.3 A Galois correspondence

The general representation theorem, which asserts that every normal BAO can be embedded in the complex algebra of a relational structure, raises a large number of questions. In order to facilitate the discussion of some of these, we introduce some terminology.

3.3.1 Definition

(i) For a class $\mathcal{H}$ of relational structures, we let

$$\mathcal{H}^+ = \{\mathbf{U}^+ : \mathbf{U} \in \mathcal{H}\}.$$

(ii) For a class $\mathcal{K}$ of normal, complete and atomic BAO's, we let

$$\mathcal{K}_+ = \{\mathbf{A}_+ : \mathbf{A} \in \mathcal{K}\}.$$

3.3.2 Definition

(i) For a relational structure $\mathbf{U}$, we let $\mathbf{U}_\sigma = ((\mathbf{U}^+)^\sigma)_+$, and we refer to $\mathbf{U}_\sigma$ as the *ultrafilter extension* of $\mathbf{U}$.

(ii) For a class $\mathcal{H}$ of relational structures, we let

$$\mathcal{H}_\sigma = \{\mathbf{U}_\sigma : \mathbf{U} \in \mathcal{H}\}.$$

3.3.3 Definition A variety $\mathcal{V}$ of normal BAO's is said to be

(i) *complete* if $\mathcal{V} = \mathbb{V}\text{ar}(\mathcal{H}^+)$ for some class $\mathcal{H}$ of structure;

(ii) *complex* if $\mathcal{V} = \mathbb{S}(\mathcal{H}^+)$ for some class $\mathcal{H}$ of structures;

(iii) *canonical* if $\mathbf{A}^\sigma \in \mathcal{V}$ for every $\mathbf{A} \in \mathcal{V}$.

3.3.4 Definition

(i) For a class $\mathcal{H}$ of relational structures, we let

$$\mathcal{H}^a = \mathbb{V}\text{ar}(\mathcal{H}^+).$$

(ii) For a variety $\mathcal{V}$ of normal BAO's, we let

$$\mathcal{V}^s = \{\mathbf{U} : \mathbf{U}^+ \in \mathcal{V}\}.$$

It is obvious that if $\mathcal{V}$ is canonical, then $\mathcal{V}$ is complex, and if $\mathcal{V}$ is complex, then $\mathcal{V}$ is complete. The fundamental preservation theorem, 2.3.4, assures us that every variety of BAO's that is defined by strongly positive identities is canonical. For a long time it was an open problem whether there exist varieties of modal algebras, — normal BAO's with a single normal unary operator, — that are not complete. This question was of fundamental importance because of the program, initiated by S. Kripke, of using complex algebras to investigate modal logics. In Fine [1974] and Thomason [1974] it was shown that the answer is negative, and in Blok [1980] it was shown that incomplete varieties are the rule rather than the exception. There exist complete varieties that are not complex, but it is still an open question whether every complex variety is canonical.

Under the Galois correspondence $\mathcal{H} \mapsto \mathcal{H}^a$, $\mathcal{V} \mapsto \mathcal{V}^s$ in Definition 3.3.4, a variety $\mathcal{V}$ is Galois closed iff it is complete. The next result, which is obviously of fundamental importance, is due to R. Goldblatt, but the proof is based in part on ideas due to K. Fine and J. A. F. K. van Bentham. For a detailed proof, see Goldblatt [1989] and [1991].

3.3.5 Theorem *If $\mathcal{H}$ is an elementary class of structures, then $\mathcal{H}^a$ is canonical.*

Actually the conclusion holds more generally whenever $\mathcal{H}$ is closed under ultraproducts. The following is a key lemma.

3.3.6 Lemma *For any structure* $\mathbf{U}$, $\mathbf{U}_\sigma$ *is the image of an ultrapower of* $\mathbf{U}$ *under a bounded morphism.*

To prove this lemma, we need a map from an ultrapower $U^J/\mathbf{F}$ of the universe U of $\mathbf{U}$ onto the universe U_σ of the ultrafilter extension $\mathbf{U}_\sigma$ of $\mathbf{U}$. The members of U_σ can be taken to be the ultrafilters on U. If $\mathbf{F}$ is any ultrafilter on a set J, then for $x \in U^J$ the family

$$\psi(x/\mathbf{F}) = \{X \subseteq U : \{j \in J : x(j) \in X\} \in \mathbf{F}\}. \tag{1}$$

is an ultrafilter on U, and does not depend on the choice of x. In other words, ψ is a map from $U/\mathbf{F}$ into U_σ. However, it is only for very special ultrafilters that this map is onto.

3.3.7 Lemma *If* $\mathbf{F}$ *is a good ultrafilter on* J*, then the map* ψ *in* (1) *is a bounded morphism from* $\mathbf{U}^J/\mathbf{F}$ *onto* $\mathbf{U}_\sigma$.

For the notion af a good ultrafilter, and for the proof of the existence of such filters, see e. g., Bell and Slomson [1969] or Chang and Keisler [1973]. One more lemma is needed. Denote by $\mathbb{U}\mathrm{n}$ and $\mathbb{H}\mathrm{b}$ the operations of taking disjoint unions and images under bounded morphisms. Of course, $\mathbb{P}\mathrm{u}$ is the operation of taking ultraproducts.

3.3.8 Lemma *For any class* $\mathcal{H}$ *of structures,*

$$\mathbb{P}\mathrm{u}\,\mathbb{U}\mathrm{n}(\mathcal{H}) \subseteq \mathbb{H}\mathrm{b}\,\mathbb{U}\mathrm{n}\,\mathbb{P}\mathrm{u}(\mathcal{H}).$$

To prove this, consider an ultraproduct $\mathbf{U} = \prod\{\mathbf{U}_j : j \in J\}/\mathbf{F}$, where each $\mathbf{U}_j$ is the union of disjoint structures $\mathbf{U}_{j,k}$, $k \in K_j$, in $\mathcal{H}$. Let $F = \prod\{K_j : j \in J\}$, and for $f \in F$, let $\mathbf{V}_f = \prod\{\mathbf{U}_{j,f(i)} : j \in J\}/\mathbf{F}$. The injection from $\prod\{\mathbf{U}_{j,f(j)} : j \in J\}$ to $\prod\{\mathbf{U}_j : j \in J\}$ induces an injection ϕ_f from $\mathbf{V}_f$ to $\mathbf{U}$, and the union of the maps ϕ_f is a bounded morphism from the disjoint union of the structures $\mathbf{V}_f$ onto $\mathbf{U}$. Hence $\mathbf{U} \in \mathbb{H}\mathrm{b}\,\mathbb{U}\mathrm{n}\,\mathbb{P}\mathrm{u}(\mathcal{H})$.

Theorem 3.3.5 now readily follows. Let

$$\mathcal{V} = \mathcal{H}^a = \mathbb{H}\mathbb{S}\mathbb{P}(\mathcal{H}^+).$$

It suffices to show that $\mathbf{A}^\sigma \in \mathcal{V}$ whenever $\mathbf{A} \in \mathbb{P}(\mathcal{H}^+)$. We have $\mathbf{A} \cong \mathbf{U}^+$, where $\mathbf{U}$ is the union of disjoint members of $\mathcal{H}$. Hence $\mathbf{A}^\sigma \cong (\mathbf{U}^+)^\sigma = (\mathbf{U}_\sigma)^+$, and

$$\mathbf{U}_\sigma \in \mathbb{H}\mathrm{b}\,\mathbb{P}\mathrm{u}\,\mathbb{U}\mathrm{n}(\mathcal{H}) \subseteq \mathbb{H}\mathrm{b}\,\mathbb{H}\mathrm{b}\,\mathbb{U}\mathrm{n}\,\mathbb{P}\mathrm{u}(\mathcal{H}) = \mathbb{H}\mathrm{b}\,\mathbb{U}\mathrm{n}(\mathcal{H}),$$

which implies that $(\mathbf{U}_\sigma)^+ \in \mathbb{S}\mathbb{P}(\mathcal{H}^+) \subseteq \mathcal{V}$.

4 Modal algebras and tense algebras

4.1 Modal logics and modal algebras

Modal propositional logic is obtained from the classical propositional logic by adjoining a single unary connective, the possibility operator. The resulting Lindenbaum, Tarski algebra is therefore a Boolean algebra with a single unary operation, indeed a normal unary operator. In order to conform with our choice of basic operations for Boolean algebras, we take $\vee$, $\wedge$ and $\neg$ as the basic logical connectives, together with the logical constants $\mathfrak{f}$ and $\mathfrak{t}$, and define $\rightarrow$ and $\leftrightarrow$ in the usual manner. The possibility operator will be written $\Diamond$, and the necessity operator is defined to be $\Box = \neg\Diamond\neg$. The propositional modal formulas therefore form an absolutely free algebra

$$\mathbf{PMF} = (PMF, \vee, \mathfrak{f}, \wedge, \mathfrak{t}, \neg, \Diamond)$$

generated by the infinite set P of propositional variables, and the modal logics are certain subsets of the set PMF.

4.1.1 Definition A *propositional modal logic* is a set Γ of propositional modal formulas with the following properties:

(i) All the classical tautologies belong to Γ.

(ii) For all $x, y \in PMF$, if $x \rightarrow y \in \Gamma$, then $\Diamond x \rightarrow \Diamond y \in \Gamma$.

(iii) $\Box\mathfrak{t} \in \Gamma$ and, for all $x, y \in PMF$, $\Diamond(x \vee y) \leftrightarrow (\Diamond x \vee \Diamond y) \in \Gamma$.

(iv) Γ is closed under modus ponens.

(v) Γ is closed under substitution.

The condition (iv) means that if x and $x \rightarrow y$ are in Γ, then so is y, and (v) means that every endomorphism of **PMF** takes Γ into itself.

4.1.2 Definition For $\Gamma \subseteq PMF$, we define

$$x \approx_\Gamma y \quad \text{iff} \quad x \leftrightarrow y \in \Gamma.$$

4.1.3 Theorem *A subset Γ of PMF is a modal logic iff $\approx_\Gamma$ is a fully invariant congruence relation on* **PMF** *such that* $\mathbf{PMF}/\approx_\Gamma$ *is a Boolean algebra with a normal operator, and Γ is a $\approx_\Gamma$-block.*

4.1.4 Definition By a *modal algebra* we mean a normal BAO $\mathbf{A} = (\mathbf{A}_0, f)$ with f a unary operator.

4.1.5 Theorem *For any modal logic Γ, let $\mathcal{V}(\Gamma)$ be the variety generated by the algebra* $\mathbf{PMF}/\approx_\Gamma$. *Then the map $\Gamma \mapsto \mathcal{V}(\Gamma)$ is a dual isomorphism from the lattice of all modal logics onto the lattice of all varieties of modal algebras.*

4.2 Elementary facts about modal algebras

A congruence relation R on a modal algebra $\mathbf{A} = (\mathbf{A}_0, f)$ is determined by any one of its blocks, in particular by the (Boolean) ideal $0/R$. We call an ideal I of $\mathbf{A}$ a congruence ideal if $I = 0/R$ for some congruence relation R on $\mathbf{A}$, and we call an element $a \in A$ a congruence element if the principal ideal $A \cdot a = \{x \in A : x \leq a\}$ is a congruence ideal. Clearly, an ideal is a congruence ideal iff it is closed under f, and an element u is a congruence element iff $f(u) \leq u$.

For any element $u \in A$, the principal ideal $A \cdot u$ is a Boolean algebra under the join and meet operations and the relative complementation

$$x^{-u} = u \cdot x^{-}.$$

The operation f can be relativized by taking

$$f_u(x) = f(x) \cdot u.$$

We refer to the algebra $\mathbf{A} \cdot u = (\mathbf{A}_0 \cdot u, f_u)$ as a *relative subalgebra* of $\mathbf{A}$. The map

$$\phi(x) = u \cdot x \quad \text{for} \quad x \in A$$

is a homomorphism from $\mathbf{A}_0$ onto $\mathbf{A}_0 \cdot u$. In order for ϕ to be a homomorphism from $\mathbf{A}$ onto $\mathbf{A} \cdot u$ it is necessary and sufficient that u^- be a congruence element. Relative subalgebras can also be used to describe the direct decompositions of $\mathbf{A}$. The factor relations on $\mathbf{A}_0$ are the Boolean congruence relations R for which the ideal $0/R$ is a principal ideal $A \cdot u$, and the complementary factor relation is then determined by the ideal $A \cdot u^-$. Hence the complementary pairs of factor relations on $\mathbf{A}$ are the kernels of homomorphisms $\mathbf{A} \to \mathbf{A} \cdot u$ and $\mathbf{A} \to \mathbf{A} \cdot u^-$ with u and u^- congruence elements. From this it is seen that the elements $u \in A$ with $f(u) \leq u$ and $f(u^-) \leq u^-$ form a complemented lattice, isomorphic to the lattice of all factor relations on $\mathbf{A}$.

A simple construction due to J. C. C. McKinsey has been extensively used in modal logic. It can be used to show that certain varieties of modal algebras are generated by their finite members, and hence that the corresponding modal logics are decidable. Suppose $\mathbf{A} = (\mathbf{A}_0, f)$ is a modal algebra and suppose $\mathbf{B}_0$ is a finite subalgebra of $\mathbf{A}_0$. For $x \in B$, let

$$g(x) = \prod\{f(y) : x \leq y \in B \text{ and } f(y) \in B\}.$$

Then $\mathbf{B} = (\mathbf{B}_0, g)$ is a modal algebra and $g(x) = f(x)$ whenever x and $f(x)$ are in B. The usefulness of this construction lies in the fact that any identity that fails in $\mathbf{A}$ also fails in one of the algebras $\mathbf{B}$. Consequently, the variety of all modal algebras is generated by its finite members, and so is every subvariety that is closed under McKinsey's construction.

4.3 Closure algebras

Closure algebras were introduced by J. C. C. McKinsey and A. Tarski to provide an algebraic treatment of the topological closure operation. With their subsequent use of these algebras in the study of Brouwerian logic and of the modal logic S4, they pioneered a new approach to the study of propositional logics.

4.3.1 Definition By a *closure algebra* we mean a modal algebra $\mathbf{A} = (\mathbf{A}_0, f)$ with

$$x \leq f(x) = ff(x) \quad \text{for all} \quad x \in A.$$

In view of the origin of the concept, the primary models are algebras consisting of all subsets of a topological space $\mathbf{X}$, with the topological closure as the modal operator. This is indeed a closure algebra; we call it the closure algebra of $\mathbf{X}$. The axioms are therefore suitable. That they are also adequate follows from the fact that, for any closure algebra $\mathbf{A}$, $\mathbf{A}^\sigma$ is isomorphic to the closure algebra $\mathbf{A}'$ of a topological space $\mathbf{X}$, and $\mathbf{A}$ is therefore embeddable in $\mathbf{A}'$. However, $\mathbf{X}$ is a rather unorthodox topology, for the closure operator is completely additive or, in other words, the intersection of a family of open sets is always open. Also, the topology will in general not be Hausdorff. It is much more difficult to prove that the closure algebras of certain familiar spaces generate the variety of all closure algebras. A very general result of this kind was proved in McKinsey, Tarski [1944]. Their result contains the following theorem as a special case.

4.3.2 Theorem *Suppose* $\mathbf{A}$ *is the closure algebra of a separable, metrizable topological space that is dense in itself. Then* $\mathbf{A}$ *generates the variety of all closure algebras.*

Some topological concepts carry over in a natural way to closure algebras. In particular, an element a is said to be closed if $f(a) = a$, open if a^- is closed, and clopen if both a and a^- are closed. These terms must not be confused with the notions of closed, open and clopen elements of the canonical extension of a closure algebra.

In spite of their historical origin, closure algebras have been mostly investigated because of their connection with modal logics. The modal logic that has been most investigated, the logic S4, can be characterized by the axioms

$$p \to \Diamond p, \qquad \Diamond\Diamond p \to \Diamond p.$$

From this we obtain the following theorem.

4.3.3 Theorem *In the notation of* Theorem 4.1.5, *if* Γ *is the logic* S4, *then* $\mathcal{V}(\Gamma)$ *is the variety of all closure algebras.*

The variety of all closure algebras is canonical, and it is therefore complete. The corresponding class of relational structures, under the Galois correspondence of Definition 3.3.4, is easy to determine.

4.3.4 Theorem *For a Kripke structure* $\mathbf{U} = (U, R)$, $\mathbf{U}^+$ *is a closure algebra iff* $\mathbf{U}$ *is quasiordered by* R.

4.3.5 Corollary *If* $\mathcal{V}$ *is the variety of all closure algebras, then* $\mathcal{V}^s$ *is the class of all quasiordered sets.*

4.3.6 Lemma *Every quasiordered set is the image of a poset under a bounded morphism.*

4.3.7 Theorem *If* $\mathcal{H}$ *is the class of all posets, then* $\mathcal{H}^a$ *is the variety of all closure algebras.*

4.3.8 Theorem *The variety $\mathcal{V}$ of all closure algebras is closed under the McKinsey construction, and is therefore generated by its finite members. On the other hand, $\mathcal{V}$ is not finitely generated. In fact, the $\mathcal{V}$-free algebra on one generator is infinite.*

To prove the second part of the theorem, it suffices to exhibit, in some closure algebra $\mathbf{A} = (\mathbf{A}_0, f)$, an element a that generates an infinite subalgebra. Let $\mathbf{A}$ be the complex algebra of the poset $(\omega, \leq)$, and take a to consist of the even natural numbers. Let $g(x) = f(x) \cdot x^-$. Then the elements $g^k(a)$, $k = 0, 1, \ldots$ are pairwise distinct.

4.3.9 Definition By a *Brouwerian algebra* we mean an algebra $(A, +, 0, \cdot, 1, \dot{-})$ such that $(A, +, 0, \cdot, 1)$ is a bounded distributive lattice and, for all $a, b, x \in A$,

$$a \leq b + x \qquad \text{iff} \qquad a \dot{-} x \leq b.$$

4.3.10 Theorem *Suppose $\mathbf{A} = (\mathbf{A}_0, f)$ is a non-trivial closure algebra.*

(i) *An element $a \in A$ is a congruence element iff a is closed.*

(ii) *The closed elements of $\mathbf{A}$ form a Brouwerian algebra under the join and meet operations inherited from $\mathbf{A}$, with $a \dot{-} b = a \cdot f(b^-)$ for all closed elements $a, b \in A$.*

(iii) *The compact elements of* $\operatorname{Con}(\mathbf{A})$ *form a Brouwerian algebra isomorphic to the algebra of all closed elements of $\mathbf{A}$ under the map $a \mapsto \operatorname{Cg}(a, 0)$.*

(iv) *$\mathbf{A}$ is subdirectly irreducible iff the set of all non-zero closed elements of $\mathbf{A}$ has a smallest member.*

(v) *$\mathbf{A}$ is finitely subdirectly irreducible iff the meet of two non-zero closed elements of $\mathbf{A}$ is never zero.*

(vi) *$\mathbf{A}$ is simple iff the only closed elements of $\mathbf{A}$ are* 0 *and* 1.

(vii) *$\mathbf{A}$ is directly indecomposable iff the only clopen elements of $\mathbf{A}$ are* 0 *and* 1.

4.3.11 Theorem *The variety of all closure algebras has EDPC. For any closure algebra $\mathbf{A} = (\mathbf{A}_0, f)$, and for all $a, b \in A$,*

$$(b, 0) \in \operatorname{Cg}(a, 0) \quad \textit{iff} \quad b \leq f(a).$$

4.4 Monadic algebras

A proper extension of **S4**, called **S5**, has been extensively investigated. This logic is obtained by adjoining the axiom $p \to \Box\Diamond p$ to the axioms for **S4**. The corresponding modal algebras have a very simple characterization.

4.4.1 Definition By a *monadic algebra* we mean a closure algebra with the property that, for all $x \in A$,

$$f(f(x)^-) = f(x)^-.$$

4.4.2 Theorem *For a modal logic* Γ, $\mathbf{S5} \subseteq \Gamma$ *iff* $\mathbf{PMF}/\approx_\Gamma$ *is a monadic algebra.*

Monadic algebras also arise in the algebraization of first order predicate logic. They are the one-dimensional cylindric algebras and the one-dimensional polyadic algebras. The Lindenbaum, Tarski algebra of first order predicate logic with a single variable is therefore a free monadic algebra, whence the term "monadic", which was coined by P. Halmos. Monadic algebras can be characterized by each of the following conditions:

The complement of a closed element is always closed.

The closed elements form a subuniverse.

An element is open iff it is closed.

4.4.3 Theorem *Suppose* $\mathbf{A} = (\mathbf{A}_0, f)$ *is a closure algebra, and let* $\mathbf{U} = (U, R)$ *be the atomic structure of* $\mathbf{A}^\sigma$*. Then the following conditions are equivalent.*

(i) $\mathbf{A}$ *is monadic.*

(ii) f *is self-conjugate.*

(iii) R *is an equivalence relation on* U.

4.4.4 Theorem *The variety of all monadic algebras is a discriminator variety.*

Consequently, for a monadic algebra, the properties of being simple, subdirectly irreducible and directly indecomposable are equivalent. Now a non-trivial Boolean algebra can be turned into a simple monadic algebra in a unique way, by defining

$$f(0) = 0 \quad \text{and} \quad f(x) = 1 \text{ for all } x \neq 0.$$

From this we easily obtain a description of the lattice of all varieties of monadic algebras.

4.4.5 Theorem *The varieties of monadic algebras form a chain of type* $\omega + 1$,

$$\mathcal{V}_0 \subseteq \mathcal{V}_1 \subseteq \ldots \subseteq \mathcal{V}_n \subseteq \ldots \subseteq \mathcal{V}_\omega,$$

where $\mathcal{V}_0$ *is the trivial variety and* $\mathcal{V}_n$*, with* $0 < n \in \omega$*, is generated by the subdirectly irreducible monadic algebra of order* 2^n*, while* $\mathcal{V}_\omega$ *is the variety of all monadic algebras. For* $0 < n \in \omega$*, an equational basis for* $\mathcal{V}_n$ *is obtained by adjoining the identity*

$$f(\prod\{x_i : i \leq n\}) = \prod\{f(\prod\{x_i : j \neq i \leq n\}) : j \leq n\}$$

to an equational basis for the variety of all monadic algebras.

4.5 Varieties of closure algebras

Varieties of closure algebras were investigated in Blok, Dwinger [1975] and in Blok [1980]. Their work provides a rather clear picture of the bottom part of the lattice of all varieties of closure algebras.

4.5.1 Lemma *If* $\mathbf{A}$ *is a closure algebra of order* 2^n, *with* $0 < n \in \omega$, *then* $\mathbf{A}$ *has a subalgebra of order* 2^{n-1}.

The proof is quite simple. For any distinct atoms p and q of $\mathbf{A}_0$, we obtain a subalgebra of $\mathbf{A}_0$ of order 2^{n-1} whose atoms are the remaining $n-2$ atoms of $\mathbf{A}_0$, together with the element $p+q$. The proof is completed by showing that p and q can be so chosen that whenever the closure of one of the other atoms contains one of them, then it contains both. If there exist two distinct atoms having the same closure, then p and q can be chosen to be such a pair. If distinct atoms always have distinct closures, then p and q can be so chosen that neither one of them is below the closure of one of the remaining atoms.

The preceding lemma gives much information about the subvarieties of a finitely generated variety $\mathcal{V}$. The next lemma tells us about the case when $\mathcal{V}$ is not finitely generated.

4.5.2 Lemma *Suppose* $\mathbf{A}$ *is a closure algebra of order* 2^n, *whose closed elements form a maximal chain. Then the conjugate variety of* $\mathbf{A}$ *is locally finite.*

4.5.3 Corollary *Every infinite closure algebra has arbitrarily large finite subalgebras.*

4.5.4 Theorem *Suppose* $\mathcal{V}$ *is a variety generated by a subdirectly irreducible closure algebra* $\mathbf{A}$. *If* $\mathbf{A}$ *is of finite order* 2^n, *then* $\mathcal{V}$ *is of height n or more in the lattice of all varieties of closure algebras, but if* $\mathbf{A}$ *is infinite, then the height of* $\mathcal{V}$ *is infinite.*

It is now easy to draw a picture of the bottom part of the lattice of all varieties of closure algebras. Figure 4.2 shows part of the partially ordered set of join irreducibles in this lattice, each variety labelled by its generating member listed in Figure 4.1 (black circles correspond to closed elements). There is one atom, the variety generated by the two element closure algebra. This variety has two covers generated by $\mathbf{A}_2$ and $\mathbf{A}_3$. Each of these is covered by their join, and the remaining covers are generated by closure algebras of order 8. Of the five varieties generated by subdirectly irreducible closure algebras of order 8, one covers $\mathbb{V}\mathrm{ar}(\mathbf{A}_2)$, three cover $\mathbb{V}\mathrm{ar}(\mathbf{A}_3)$, and one covers the join of these two varieties. Each variety of finite height has only finitely many covers. For any positive integer n, the number $\mu(n)$ of varieties at the n-th level is finite. It seems likely that the function μ is strictly increasing, but I am not aware of a proof of this.

4.6 Varieties of modal algebras

Some of the results about varieties of closure algebras extend to varieties of modal algebras $\mathbf{A} = (\mathbf{A}_0, f)$ that satisfy the inclusion $f^2(x) \leq f(x)$, as is shown in Blok [1980a], but very few extend to the lattice of all modal algebras. This is abundantly shown in Blok [1980], [1980a]. First, a positive result due to Makinson [1971].

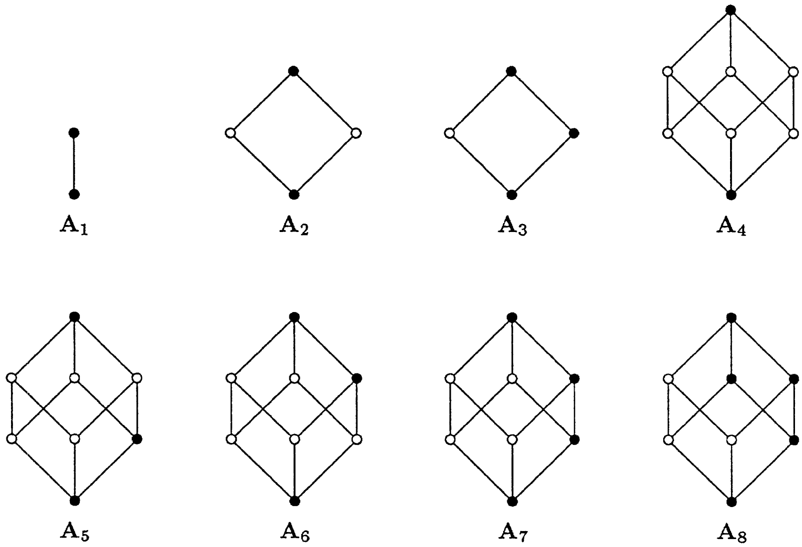


Figure 4.1

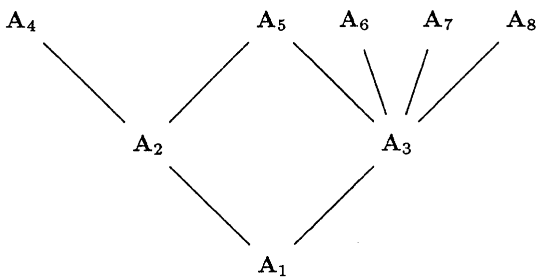


Figure 4.2

4.6.1 Theorem *The lattice of all varieties of modal algebras has just two atoms, the varieties generated by the two-element modal algebras.*

Already at the next level, things go spectacularly wrong.

4.6.2 Theorem *The variety generated by the two-element modal algebra with $f(1) = 1$ has infinitely many finitely generated covers, and it also has a cover that is not finitely generated.*

Call this variety $\mathcal{V}_0$. For each prime p, Blok constructs a structure $\mathbf{U} = (U, R)$ of order p whose complex algebra generates a cover of $\mathcal{V}_0$. The set U consists of the natural numbers $0, 1, \ldots, p-1$, and iRj iff $j = i$ or $j \equiv i+1 \bmod p$. The construction of the non-finitely generated cover uses the veiled recession frame. The *recession frame* is the frame (ω, R) with mRn iff $m \geq n-1$, and the *veiled recession frame* is the subalgebra $\mathbf{A}$ of $(\omega, R)^+$ whose elements are the finite and the co-finite subsets of ω. In the variety $\mathcal{V}$ generated by $\mathbf{A}$, the only finite subdirectly irreducible algebra is the two-element algebra, and every infinite subdirectly irreducible algebra contains an isomorphic copy of $\mathbf{A}$ as a subalgebra. Hence, $\mathcal{V}$ covers $\mathcal{V}_0$. This is also one of the simplest examples of an incomplete variety: The complex algebra of a Kripke structure belongs to $\mathcal{V}$ iff it belongs to $\mathcal{V}_0$. The variety of all modal algebras does not have EDPC, and finitely generated subdirectly irreducible members are not in general splitting. In Blok [1978], the splitting algebras are characterized.

4.6.3 Theorem *A finite subdirectly irreducible modal algebra $\mathbf{A} = (\mathbf{A}_0, f)$ is splitting in the variety of all modal algebras iff, for some $n \in \omega$, $f^n(1) = 0$.*

A conjugate equation for each splitting algebra is also found, and it is shown that every variety that is the intersection of conjugate varieties is generated by its finite members, and is therefore complete.

4.7 Tense algebras

If R is a binary relation on a set U, then the operator R^+ on U^+ is residuated, the conjugate being $R^{\smile +}$. In fact, the class of all BAO's $\mathbf{A} = (\mathbf{A}_0, f, g)$ with f and g conjugate unary operators on the Boolean algebra $\mathbf{A}_0$ is the variety generated by the complex algebras $(U, R, R^{\smile})^+$. The name "tense algebra", applied to these algebras, comes from tense logic. Tense algebras have the same relationship to tense logic as modal algebras have to modal logic. In particular, U is thought of as a set of states, and the relation R is thought of as representing a temporal relationship; pRq might be read as "if the state p prevailes at a certain time, then it is possible that q will prevail at a later time". A related interpretation is obtained by thinking of R as representing the action of a (possibly non-deterministic) computer program, and reading pRq as "if the program is executed with the initial state p, then q is a possible terminal state". The brief discussion that follows will not be concerned with these interpretations, but will treat tense algebras purely from an algebraic point of view.

4.7.1 Definition By a *tense algebra* we mean a BAO $\mathbf{A} = (\mathbf{A}_0, f, g)$ with f and g conjugate unary operators on the Boolean algebra $\mathbf{A}_0$.

In addition to the operators f and g, we shall have occasion to use the self-conjugate operator $h(x) = f(x) + g(x)$.

4.7.2 Lemma *Suppose* $\mathbf{A}$ *is a tense algebra. The congruence ideal generated by a set* $U \subseteq A$ *is equal to the Boolean ideal generated by the set* $\{h^n(x) : n \in \omega \text{ and } x \in U\}$.

4.7.3 Lemma *Suppose* $\mathbf{A}$ *is a tense algebra and* $u \in A$.

(i) u *is a congruence element iff* $f(u) \leq u$ *and* $g(u) \leq u$.

(ii) $f(u) \leq u$ *iff* $g(u^-) \leq u^-$.

(iii) *If* u *is a congruence element, then so is* u^-.

4.7.4 Corollary *A non-trivial tense algebra* $\mathbf{A}$ *is directly indecomposable iff the only congruence elements of* $\mathbf{A}$ *are* 0 *and* 1.

4.7.5 Theorem *For any variety* $\mathcal{V}$ *of tense algebras, the following conditions are equivalent:*

(i) $\mathcal{V}$ *is a discriminator variety.*

(ii) $\mathcal{V}$ *has EDPC.*

(iii) *For some* $n \in \omega$, $\mathcal{V} \models h^n(x) \approx h^{n+1}(x)$.

4.7.6 Corollary *Every finitely generated variety of tense algebras is a discriminator variety.*

4.7.7 Corollary *Every variety of tense algebras that satisfies the conditions* $ff(x) \leq f(x)$ *and* $fg(x) = gf(x)$ *is a discriminator variety.*

4.7.8 Theorem *Let* $\mathcal{H}$ *be the class of all structures* $(U, R, R^{\smile})$ *such that* R *totally orders* U. *Then*

(i) $\mathcal{H}^a$ *is the variety of all tense algebras that satisfy the conditions*

$$x \leq f(x) = ff(x) \quad \text{and} \quad fg(x) = gf(x) = f(x) + g(x).$$

(ii) $\mathcal{H}^a$ *is a discriminator variety.*

(iii) *A structure* $(U, R, R^{\smile})$ *belongs to* $\mathcal{H}^{as}$ *iff* R *quasiorders* U *and any two members of* U *that are in the same component are comparable.*

5 Residuated m-algebras

5.1 Definitions and basic properties

An operator of rank $n > 1$ is said to be residuated if each of its translates is residuated. Such an operator has n residuals and n conjugates, each of these being an operation of rank n. In particular, a residuated binary operator $\circ$ has left and a right residuals, $/$ and $\backslash$, and left and right conjugates, $\lhd$ and $\rhd$.

5.1.1 Definition

(i) A normal BAO $\mathbf{A} = (\mathbf{A}_0, \circ)$ with $\circ$ a binary operator is called a *Boolean groupoid*, and if the operation $\circ$ has a unit element e, then $\mathbf{A} = (\mathbf{A}_0, \circ, e)$ is called a *unital Boolean groupoid.*

(ii) A normal BAO $\mathbf{A} = (\mathbf{A}_0, \circ, \rhd, \lhd)$ with binary operators $\circ$, $\rhd$, and $\lhd$ is said to be *residuated* if $\rhd$ and $\lhd$ are, respectively, right and left conjugates of $\circ$. If $\circ$ has a unit e, then $\mathbf{A} = (\mathbf{A}_0, \circ, e, \rhd, \lhd)$ is said to be a *unital residuated algebra.*

(iii) If the operation $\circ$ is associative, then the algebras in (i) and (ii) are said to be *associative.*

(iv) An associative unital Boolean groupoid is referred to as a *monoidal algebra* or, briefly, an *m-algebra.* A residuated Boolean algebra is called an *r-algebra*, and a unital residuated Boolean algebra is called a *ur-algebra.* An associative unital residuated algebra is called a *residuated monoidal algebra* or, briefly, an *rm-algebra.*

Given an r-algebra $\mathbf{A} = (\mathbf{A}_0, \circ, \rhd, \lhd)$, the right and left residuals of $\circ$ are defined by

$$a\backslash b = (a \rhd b^-)^-, \qquad a/b = (a^- \lhd b)^-.$$

We list here some of the basic arithmetic properties of r-algebras.

5.1.2 Theorem *The following statements hold for arbitrary elements a, b, c, a', b', c' in an r-algebra $\mathbf{A}$.*

(i) $a \circ b \leq c$ *iff* $a \leq c/b$ *iff* $b \leq a\backslash c$.

(ii) $a \circ (a\backslash c) \leq c$ *and* $(c/b) \circ b \leq c$.

(iii) $(a \circ b)/b \geq a$ *and* $a\backslash(a \circ b) \geq b$.

(iv) $(c \cdot c')/b = (c/b) \cdot (c'/b)$ *and* $a\backslash(c \cdot c') = (a\backslash c) \cdot (a\backslash c')$.

(v) $c/(b + b') = (c/b) + (c/b')$ *and* $(a + a')\backslash c = (a\backslash c) + (a'\backslash c)$.

(vi) $a \rhd c = (a\backslash c^-)^-$ *and* $a\backslash c = (a \rhd c^-)^-$.

(vii) $c \lhd b = (c^-/b)^-$ *and* $c/b = (c^- \rhd b)^-$.

(viii) $(a \circ b) \cdot c = 0$ *iff* $(a \triangleright c) \cdot b = 0$ *iff* $(c \triangleleft b) \cdot a = 0$.

(ix) $a \circ 0 = 0 \circ b = 0$.

(x) $a \triangleright 0 = 0 \triangleright c = 0$ *and* $0 \backslash c = a \backslash 1 = 1$.

(xi) $0 \triangleleft b = c \triangleleft 0 = 0$ *and* $c/0 = 1/b = 1$.

(xii) $(a \circ b) \cdot c \leq a \circ (b \cdot (a \triangleright c)$ *and* $(a \circ b) \cdot c \leq (a \cdot (c \triangleleft b)) \circ b$.

If $\mathbf{A}$ *is associative, then the following statements hold.*

(xiii) $a \backslash (a' \backslash c) = (a' \circ a) \backslash c$ *and* $a \triangleright (a' \triangleright c) = (a' \circ a) \triangleright c$.

(xiv) $(c/b)/b' = c/(b' \circ b)$ *and* $(c \triangleleft b) \triangleleft b' = c \triangleleft (b' \circ b)$

(xv) $a \triangleright (c \triangleleft b) = (a \triangleright c) \triangleleft b$.

If $\mathbf{A}$ *is unital, then the following statements hold.*

(xvii) $e \backslash c = e \triangleright c = c/e = c \triangleleft e = c$.

5.2 Congruence ideals

The notions of a congruence ideal and of a congruence element, introduced in Section 4.2 for modal algebras, apply to arbitrary BAO's, and so does the notion of a relative subalgebra. By a *congruence ideal* of a BAO $\mathbf{A} = (\mathbf{A}_0, f_i, i \in I)$ we mean an ideal J of $\mathbf{A}_0$ such that $J = 0/R$ for some $R \in \mathrm{Con}(\mathbf{A})$, and by a *congruence element* of $\mathbf{A}$ we mean an element $u \in A$ such that $A \cdot u$ is a congruence ideal. Clearly J is a congruence ideal iff it is closed under all the translates of the operations f_i, and u is a congruence element iff $g(u) \leq u$ for every such translate g. An n-ary operation on A is relativized to $A \cdot u$ by letting

$$g_u(x) = g(x) \cdot u \text{ for all } x \in (A \cdot u)^n,$$

and we call the algebra $\mathbf{A} \cdot u = (\mathbf{A}_0 \cdot u, (f_i)_u, i \in I)$ a *relative subalgebra* of $\mathbf{A}$.

5.2.1 Lemma *Suppose* $\mathbf{A}$ *is an r-algebra.*

(i) *An ideal* J *of* $\mathbf{A}_0$ *is a congruence ideal of* $\mathbf{A}$ *iff, for all* $x \in J$*, the elements*

$$1 \circ x, \quad x \circ 1, \quad 1 \triangleright x, \quad x \triangleright 1, \quad 1 \triangleleft x, \quad x \triangleleft 1$$

are in J.

(ii) *An element* $u \in A$ *is a congruence element of* $\mathbf{A}$ *iff*

$$1 \circ u \leq u, \quad u \circ 1 \leq u, \quad 1 \triangleright u \leq u, \quad u \triangleright 1 \leq u, \quad 1 \triangleleft u \leq u, \quad u \triangleleft 1 \leq u.$$

5.2.2 Theorem *An element u in an r-algebra is a congruence element iff*

$$u \circ u \leq u, \quad u^- \circ u^- \leq u^-, \quad u \circ u^- = 0, \quad u^- \circ u = 0.$$

5.2.3 Corollary *An element u in an r-algebra is a congruence element iff u^- is a congruence element.*

5.2.4 Corollary *Suppose u is a congruence element in an r-algebra $\mathbf{A}$. Then, for all $x, y \in A$*

$$(x \circ y)u = xu \circ yu, \quad (x \rhd y)u = xu \rhd yu, \quad (x \lhd y)u = xu \lhd yu.$$

5.2.5 Corollary *For any element u of an r-algebra $\mathbf{A}$, the following conditions are equivalent.*

(i) *u is a congruence element of $\mathbf{A}$.*

(ii) *The map $x \mapsto xu$ is a homomorphism from $\mathbf{A}$ to $\mathbf{A} \cdot u$.*

(iii) *The map $x \mapsto (xu, xu^-)$ is an isomorphism from $\mathbf{A}$ onto $(\mathbf{A} \cdot u) \times (\mathbf{A} \cdot u^-)$.*

5.2.6 Theorem *Every finite, directly indecomposable r-algebra is simple.*

5.2.7 Theorem *Every finitely generated variety of r-algebras is a discriminator variety.*

5.3 Relation algebras

We are not going to present here the extensive theory of relation algebras; our primary goal is to show how it fits into the general theory of BAO's. This is the approach advocated in Birkhoff [1948]. We begin by rephrasing the definition of a relation algebras given in Chapter 1.

5.3.1 Definition A *relation algebra* is a BAO $\mathbf{A} = (\mathbf{A}_0, \circ, e, {}^\smile)$ such that

(i) $(A, \circ, e, {}^\smile)$ is a monoid with involution.

(ii) For all $a, b \in A$, $a \circ (a^\smile \circ b^-)^- \leq b$.

As noted in Section 1.1, (ii) means that the operation $\circ$ is right residuated, the right residual being $a \backslash x = (a^\smile \circ x^-)^-$. It follows that the right conjugate is $a \rhd x = a^\smile \circ x$, and using the involution ${}^\smile$, we find that $\circ$ is also left residuated, and that the left residual and the left conjugate are $x/b = (x^- \circ b^\smile)^-$ and $x \lhd b = x \circ b^\smile$. The operation ${}^\smile$ can be defined in terms of either one of the operations $\rhd$ and $\lhd$, and the variety of all relation algebras is therefore definitionally equivalent to a variety of rm-algebras. This variety was characterized in Hoare, Jifeng [1986]. Several other characterizations can be found in Jónsson, Tsinakis [a], including the three given in (ii)-(iv) of the next theorem.

5.3.2 Theorem *For any rm-algebra $\mathbf{A}$, the following conditions are equivalent.*

(i) *For some unary operation* $\smile$ *on* A, $(\mathbf{A}_0, \circ, e, ^\smile)$ *is a relation algebra.*

(ii) *For all* $a, b, x, y \in A$, $(a \circ x) \cdot (y \circ b) = 0$ *iff* $(a \rhd y) \cdot (x \lhd b) = 0$.

(iii) *For all* $a, b, x \in A$, $a \rhd (x \circ b) = (a \rhd x) \circ b$.

(iv) *For all* $a, b, x \in A$, $(a \circ x) \lhd b = a \circ (x \lhd b)$.

For any element x of a relation algebra $\mathbf{A}$, the element $u = 1 \circ x \circ 1$ is a congruence element, and hence by Corollary 5.2.2, so is u^-. From this it readily follows that the variety of all relation algebras is a discriminator variety. The congruence elements of $\mathbf{A}$ form a relation algebra $\mathbf{A}'$ that is "almost" a subalgebra of $\mathbf{A}$: The Boolean operations and the operations $\circ$ and $\smile$ agree with the corresponding operations in $\mathbf{A}$, but the unit element is $1 \circ e \circ 1 = 1$. The congruence lattice of $\mathbf{A}$ is isomorphic to the congruence lattice of $\mathbf{A}'$, which is isomorphic to the lattice of all (Boolean) ideals in $\mathbf{A}'$.

Every simple relation algebra satisfies one of the three identities

$$(1) \qquad e^- \circ e^- = 0, \qquad e^- \circ e^- = e, \qquad e^- \circ e^- = 1$$

and every relation algebra factors uniquely into a direct product of three algebras, corresponding to these three identities. The three cases are realized by the full relation algebra on a set U of order 1, of order 2, and of order three or more, respectively. The variety of all relation algebras is therefore the direct product of the three subvarieties defined by these identities. The first of these is essentially just the variety of all Boolean algebras, with the operators defined by $a \circ b = a \cdot b$, $a^\smile = a$, $e = 1$. Such relation algebras are called Boolean, or discrete. The second variety is generated by the full relation algebra on a two-element set, and it therefore has just one non-trivial proper subvariety. An equational basis for this subvariety is obtained by adding the identity $a^\smile = a$. The lattice of all subvarieties of the third variety is quite complicated, and offers many challenging problems. The unique atom is known to have a number of covers, a list of which can be found in Jipsen and Lukács [a], but it is not known if this list is complete. In fact, it is not known whether the number of covers is finite. It is also known that there are infinitely many co-atoms, a somewhat unusual situation. For any integer $n > 2$, the full relation algebra on an n element set is an absolute retract, and its conjugate variety is therefore a co-atom. These and related facts can be found in Jónsson [1982].

The variety of all relation algebras is canonical. In fact, as was discussed in Section 1.1, it was in the study of relation algebras that the notion of a canonical extension first arose. The relational structures whose complex algebras are relation algebras form a strictly elementary class. This class was characterized in Comer [1983]. In describing these structures it is convenient to treat a relation of rank $n + 1$ as a poly-operation of rank n, to use the complex algebra notation, and not to distinguish between a one-element set and its sole member.

5.3.3 Definition By a *poly-group* we mean a relational structure $\mathbf{U} = (U, \circ, E, ^\smile)$ with $\circ$ a binary poly-operation on U, E a subset of U and $\smile$ a unary operation on U such that, for all $p, q, r \in U$,

(i) $p \circ (q \circ r) = (p \circ q) \circ r$.

(ii) $p \circ E = E \circ p = p$.

(iii) $p \in q \circ r$ iff $q \in p \circ r^{\smile}$, iff $r \in q^{\smile} \circ p$.

5.3.4 Theorem *For any relational structure* $\mathbf{U}$, $\mathbf{U}^{+}$ *is a relation algebra iff* $\mathbf{U}$ *is a polygroup.*

5.4 Adjoining units to r-algebras

The principal question addressed here is: Which r-algebras can be embedded in r-algebras with a unit or, which r-algebras are subreducts of ur-algebras? Refining the question, we consider a variety $\mathcal{V}$ of ur-algebras, and ask for a characterization of the class (variety?) of r-algebras that are subreducts of members of $\mathcal{V}$. The results described here are from Jipsen, Jónsson, Rafter [a].

5.4.1 Theorem *For every r-algebra* $\mathbf{A}$, *the following conditions are equivalent:*

(i) $\mathbf{A}$ *is embeddable in an r-algebra with a unit.*

(ii) $\mathbf{A}^{\sigma}$ *has a unit.*

(iii) *For all* $x, y, z \in A$,

$$x \circ (y/y)(z/z) \geq x, \quad x \circ (y\backslash y) \geq x,$$
$$(y\backslash y)(z\backslash z) \circ x \geq x, \quad (y/y) \circ x \geq x.$$

Obviously (ii) implies (i), and to show that (i) implies (iii) we merely note that in an r-algebra with a unit the inclusions

$$y/y \geq e, \qquad y\backslash y \geq e$$

hold. The proof is completed by showing that if (iii) holds, then the element

$$e = \prod\{u \in A : u \geq u/u\}$$

is a unit of $\mathbf{A}^{\sigma}$.

5.4.2 Corollary *If* $\mathbf{A}$ *is a subalgebra of an r-algebra with a unit, then* $\mathbf{A}^{\sigma}$ *has a unit.*

5.4.3 Corollary *If* $\mathbf{A}$ *is a finite subalgebra of an r-algebra with a unit, then* $\mathbf{A}$ *has a unit.*

5.4.4 Theorem *Suppose* $\mathcal{U}$ *is a variety of ur-algebras. Let* $\mathcal{K}$ *be the class of all r-algebras that are reducts of members of* $\mathcal{U}$, *and let* $\mathcal{V} = \mathbb{V}\mathrm{ar}(\mathcal{K})$. *Then* $\mathcal{V} = \mathbb{S}(\mathcal{K})$. *If* $\mathcal{U}$ *is canonical, then* $\mathcal{V}$ *is canonical and, for every r-algebra* $\mathbf{A}$,

$$\mathbf{A} \in \mathcal{V} \textit{ iff } \mathbf{A}^{\sigma} \in \mathcal{K}.$$

Several decades ago, Tarski proposed the problem of finding an equational basis for the set of all identities that hold in every relation algebra and do not involve the unit element. This problem arose again recently because it appears that models of these axioms, "relation algebras without a unit element," are suitable for an abstract treatment of program specifications.

5.4.5 Definition By a specification algebra we mean an algebra $\mathbf{A} = (\mathbf{A}_0, \circ, ^{\smile})$ that satisfies all the identities that hold in every relation algebra and do not involve the constant denoting the unit.

5.4.6 Theorem *For any algebra* $\mathbf{A} = (\mathbf{A}_0, \circ, ^{\smile})$, *the following conditions are equivalent:*

(i) $\mathbf{A}$ *is a specification algebra.*

(ii) $\mathbf{A}$ *is a subreduct of a relation algebra.*

(iii) $\mathbf{A}^\sigma$ *is a reduct of a relation algebra.*

(iv) *The algebra* $(\mathbf{A}_0, \circ, \triangleright, \triangleleft)$ *with* $a \triangleright b = a^{\smile} \circ b$ *and* $a \triangleleft b = a \circ b^{\smile}$ *is an associative* $r-$*algebra and, for all* $a, b \in A$,

$$(a \circ b)^{\smile} = b^{\smile} \circ a^{\smile}, \qquad a^{-\smile} = a^{\smile -},$$

$$a \circ (b/b)(b^{-}/b^{-}) \geq a, \qquad a^{\smile\smile} = a.$$

5.5 Geometric structures

The ternary relations of betweenness and of collinearity are often used as basic concepts in axiomatic treatments of geometries. The complex algebras of these relations, especially of the betweenness relation, have been extensively investigated by Walter Prenowitz and James Jantosciak. A comprehensive account of their work can be found in Prenowitz and Jantosciak [1979]. They treat the betweenness relation as a poly-operation, and also make use of the conjugate operation.

5.5.1 Definition A *join structure* is a poly-algebra $\mathbf{U} = (U, \circ, \triangleright, \triangleleft)$ with binary poly-operations $\circ$, $\triangleright$, and $\triangleleft$ such that $\triangleright$ and $\triangleleft$ are, respectively, the right and the left conjugates of $\circ$, and for all $p, q, r, s \in U$,

(i) $p \circ q \neq \emptyset$.

(ii) $p \circ q = q \circ p$.

(iii) $p \circ (q \circ r) = (p \circ q) \circ r$.

(iv) $p \circ p = p$.

(v) $p \triangleright q \neq \emptyset$.

(vi) $(p \circ q)(r \circ s) = \emptyset$ implies $(p \triangleright s)(r \triangleright q) = \emptyset$.

(vii) $p \triangleright p = p$.

With the exception of (iv) and (vii), these axioms translate readily into conditions on the complex algebra of the poly-algebra, and the same is true of many of the theorems. Of course additional predicates and axioms must be introduced in order to completely axiomatize Euclidean geometry, and it seems doubtful that such axioms, say the ones in Tarski [1959], can be expressed in the language of BAO's.

In the case of projective geometries, the basic predicate is collinearity. This relation is symmetric, and the three poly-operations $\circ$, $\triangleright$ and $\triangleleft$ are therefore equal. This poly-operation is not associative, for if p and q are distinct points , then $p \circ (p \circ q)$ consists of all the points on the line through p and q except p, while $(p \circ p) \circ q$ omits both p and q. This can be remedied by adjoining a unit e to the structure, defining $p \circ p$ to be $\{p, e\}$ rather than p. This construction is used in Lyndon [1961] to obtain examples of non-representable relation algebras with very strong additional properties.

5.5.2 Lemma *Suppose* **G** *is a projective geometry with the property that each line of* **G** *contains at least four points. Choosing an element $e \notin G$, let $U = G \cup \{e\}$ and define $\circ$ to be the poly-operation on U such that, for distinct points p and q of* **G**, *$p \circ q$ is the set of all points r such that p,q,r are distinct but collinear. Also let $p \circ p = \{p, e\}$, $p \circ e = e \circ p = p$ and $e \circ e = e$. Then the structure $\mathbf{U} = (U, \circ, \{e\}, ^{\smile})$, where $p^{\smile} = p$ for all $p \in U$, is a commutative poly-group.*

By Theorem 5.3.4, the complex algebra of poly-group **U** is a relation algebra. In the present case, the complex algebra is symmetric ($a^{\smile} = a$), and therefore commutative. We are now ready to state Lyndon's result.

5.5.3 Theorem *For a poly-group* **U** *in the preceding lemma, the relation algebra* $\mathbf{U}^+$ *is representable iff* **G** *can be embedded as a hyperplane in a non-degenerate projective space.*

Thus, if **G** is a non-Arguesian projective plane, then $\mathbf{U}^+$ is not representable. Also, it is known that there are infinitely many positive integers n that cannot occur as the order of a line in a projective plane. If **G** is a projective line of such an order, then $\mathbf{U}^+$ is not representable.

6 Boolean modules and dynamic algebras

This final chapter is concerned with BAO's whose operators themselves form algebraic structures. Boolean modules over relation algebras and dynamic algebras are known examples of structures of this kind. We call attention to certain modifications of these concepts that appear to be worth investigating.

6.1 Boolean modules

As was noted earlier, the notion of residuals can be applied to maps $\circ : \mathbf{A} \times \mathbf{B} \to \mathbf{C}$ where $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are posets, and if these structures are Boolean algebras, then we can also speak of left and right conjugates of $\circ$. The right residual and the right conjugate of $\circ$, if they exist, are maps

$$\backslash : \mathbf{A} \times \mathbf{C} \to \mathbf{B} \quad \text{and} \quad \triangleright : \mathbf{A} \times \mathbf{C} \to \mathbf{B},$$

while the left residual and the left conjugate are maps

$$/ : \mathbf{C} \times \mathbf{B} \to \mathbf{A} \quad \text{and} \quad \triangleleft : \mathbf{C} \times \mathbf{B} \to \mathbf{A}.$$

The case when $\mathbf{A} = \mathbf{C}$ is particularly interesting. In this case we can think of the members of $\mathbf{B}$ as scalars acting on $\mathbf{A}$. A primary example of this is obtained by taking $\mathbf{A}$ and $\mathbf{C}$ to be the Boolean algebra of all subsets of a set U, and $\mathbf{B}$ the relation algebra of all binary relations on U, with

$$X \circ R = R^{+}(X) = \{y \in U : xRy \text{ for some } x \in X\}$$

for all $X \in A$ and $R \in B$. In this case,

$$\begin{aligned} Y / R &= \{x \in U : xRy \text{ for all } y \in Y\}, \\ Y \triangleleft R &= \{x \in U : xRy \text{ for some } y \in Y\}. \end{aligned}$$

This example inspired the notion of a Boolean module over a relation algebra introduced in Brink [1981]. We define here the more general notion of a Boolean module over an r-algebra.

6.1.1 Definition By a *module* over an r-algebra, or a *ur*-algebra, $\mathbf{A}$, we mean a Boolean algebra $\mathbf{M}$ with bilinear maps $\circ, \triangleleft : \mathbf{M} \times \mathbf{A}_0 \to \mathbf{M}$ such that

(i) $\triangleleft$ is a left conjugate of $\circ$.

(ii) For all $x \in M$ and $a, b \in A$, $(x \circ a) \circ b = x \circ (a \circ b)$.

If $\mathbf{A}$ is a *ur*-algebra, and if in addition $x \circ e = x$ for all $x \in M$, then $\mathbf{M}$ is said to be a *unital module* over $\mathbf{A}$.

Many of the properties listed in 5.1.2 apply to modules over r-algebras. Roughly, all the properties that are meaningful in the present setting are valid. In particular,

$$(x \triangleleft a) \triangleleft b = x \triangleleft (b \circ a)$$

for all $x \in M$ and $a, b \in A$.

In our set theoretic models, the scalar multiplication is also right residuated, with

$$X \backslash Y = (X \times Y) \cup (X^{-} \times Y^{-}),$$

$$X \triangleright Y = X \times Y$$

for all $X, Y \in A$. In these models, the right conjugation, the Cartesian multiplication, satisfies some very strong conditions which are described in the next definition.

6.1.2 Definition By a *strong module* over an r-algebra, or a ur-algebra, $\mathbf{A}$ we mean a Boolean algebra $\mathbf{M}$ with maps $\circ, \triangleleft : \mathbf{M} \times \mathbf{A} \to \mathbf{M}$ and $\triangleright : \mathbf{M} \times \mathbf{M} \to \mathbf{A}$ such that 6.1.1(i)(ii) hold and also

(iii) $\triangleright$ is a right conjugate of $\circ$.

(iv) $1_{\mathbf{M}} \triangleright 1_{\mathbf{M}} = 1_{\mathbf{A}}$.

(v) For all $x, y, z \in M$, $x \triangleright yz = (x \triangleright y)(x \triangleright z)$.

In the set theoretic models, the operation $\circ$ also satisfies the right distributive law

$$xy \triangleright z = (x \triangleright z)(y \triangleright z).$$

It is not known whether this follows from the above axioms, but in the modules considered below it does hold.

6.1.3 Definition By a *module* over a specification algebra $\mathbf{A} = (\mathbf{A_0}, \circ, ^{\smile})$ we mean a module $\mathbf{M}$ over the r-algebra $\mathbf{A}' = (\mathbf{A_0}, \circ, \triangleright, \triangleleft)$, where for all $a, b \in A$,

$$a \triangleright b = a^{\smile} \circ b, \qquad a \triangleleft b = a \circ b^{\smile},$$

with the property that $x \triangleleft a = x \circ a^{\smile}$ for all $x \in M$ and $a \in$A. If $\mathbf{M}$ is a strong module over $\mathbf{A}'$, then we say that $\mathbf{M}$ is a *strong module* over $\mathbf{A}$.

If $\mathbf{M}$ is a module over the specification algebra $\mathbf{A}$, then $\mathbf{M}^\sigma$ is a module over $\mathbf{A}^\sigma$, and the property of being a strong module is also preserved. Less obvious is the fact that if $\mathbf{M}$ is a strong module over the specification algebra $\mathbf{A}$, then $\mathbf{M}^\sigma$ is a unital module over the relation algebra $\mathbf{A}^\sigma$. The properties listed in the next two theorems constitute an outline of a proof of this.

6.1.4 Theorem *Suppose* $\mathbf{M}$ *is a module over a specification algebra* $\mathbf{A}$. *Then the following statements hold for all* $x, y, z, u \in M$ *and* $a \in A$.

(i) $x \leq x \circ 1_{\mathbf{A}}$.

(ii) $(x \circ a)y \leq x(y \circ a^{\smile}) \circ a$.

(iii) $x(1_{\mathbf{M}} \circ a) \leq x \circ a^{\smile} \circ a$.

(iv) *If a is an equivalence element (i. e., if $a = a^{\smile} = a \circ a$), and if $1_{\mathbf{M}} \circ a = 1_{\mathbf{M}}$, then $x \leq x \circ a$.*

(v) *If $a \circ 1_{\mathbf{A}} = 1_{\mathbf{A}}$, then $x \leq x \circ a \circ a^{\smile}$.*

If the scalar multiplication is right residuated, then the following hold.

(vi) $x \triangleright x = 0$ *iff* $x = 0$.

(vii) $x \circ (x \triangleright x) \geq x$.

(viii) $(x \backslash y)^{\smile} = y^{-} \backslash x^{-}$.

(ix) $(x \triangleright y)^{\smile} = y \triangleright x$.

(x) $x \rhd (y \circ a) = (x \rhd y) \circ a$.

(xi) $(x \rhd y) \circ (y \rhd z) = x(y \circ 1_{\mathbf{A}}) \circ z(y \rhd 1_{\mathbf{A}})$.

(xii) $(x \circ a)y \leq x \circ a(x \rhd y)$.

If $\mathbf{M}$ *is a strong module over* $\mathbf{A}$*, then the following statements hold.*

(xiii) $1_{\mathbf{M}} \circ a = 0$ *iff* $a = 0$.

(xiv) $xy \rhd zu = (x \rhd z)(y \rhd u)$.

(xv) $x \circ (x \rhd x) = x$.

(xvi) $x \circ (y \rhd y) = 0$ *iff* $xy = 0$.

In the set theoretic case, the element $\pi(x) = (x\backslash x)(x^-\backslash x^-)$ is an equivalence relation with two blocks (one if $x = 0$ or $x = 1$) and the intersection of arbitrarily many elements of this form is therefore an equivalence relation. In fact, the intersection of all the elements $\pi(x)$ is the identity relation. We try to imitate this for a strong module over a specification algebra $\mathbf{A}$, in order to show that the unit of $\mathbf{A}^\sigma$ is also a unit for the scalar multiplication on $\mathbf{M}^\sigma$.

6.1.5 Theorem *Suppose* $\mathbf{M}$ *is a strong module over a specification algebra* $\mathbf{A}$*, and for* $x \in M$ *let* $\pi(x) = (x\backslash x)(x^-\backslash x^-)$*. Then the following statements hold for all* $x, y \in M$*.*

(i) $\pi(x)$ *is an equivalence element.*

(ii) $\pi(x) = (x \rhd x) + (x^- \rhd x^-)$.

(iii) $x \circ \pi(y) \geq x$.

(iv) $x \circ \pi(x) = x$.

In $\mathbf{A}^\sigma$*, let* $u = \prod\{\pi(x) : x \in A\}$*, and let* e *be the unit element of* $\mathbf{A}^\sigma$*. Then the following statements hold.*

(v) u *is an equivalence element with* $u \geq e$.

(vi) $x \circ u = x \circ e = x$ *for all* $x \in M$.

Although u and e are both right units for the scalar multiplication, they need not be equal. Example: Let $\mathbf{A}$ be a relation algebra with $e \neq 1$ and let $\mathbf{M}$ be a two-element Boolean algebra. For $x \in M$ and $a \in A$, let $x \circ a = 0$ if $x = 0$ or $a = 0$, but $x \circ a = x$ otherwise. Then $u = 1$.

6.2 Program specifications

Specification algebras constitute a natural generalization of relation algebras, and contrary to our expectations they have a relatively simple characterization. Aesthetically this is satisfying, but as the name suggest, we have an ulterior motive in introducing these algebras: a suspected connection with computer programming. The speculations that follow, trying to justify this suspicion, are of a very tentative nature; they undoubtedly contain many misconceptions and errors, perhaps even a fatal flaw. At best, they represent a rather narrow view of the concepts of program and specification.

At the operational level, a specification is a requisition, a request from a customer to a supplier to provide a program that will perform a certain task. The specification may include certain considerations such as compatibility and efficiency, but the only aspect that will be considered here is the nature of the task to be performed, the required relationship between the input state and the output state. Our first abstraction is to think of these states as relational structures, consisting of one or more sets together with certain relations and operations on these sets, including nullary operations, or distinguished elements. We assume that, when a program is executed, the only difference between the input and the output states is in the values of some of these distinguished elements. Thus, if the input is a structure $\mathbf{U} = (\mathbf{U}_0, u)$ with u a string of elements, then the output will be a structure $\mathbf{U}' = (\mathbf{U}_0, u')$. The specification will therefore be a statement about the acceptable structures $(\mathbf{U}_0, u, u')$. We refer to these as specification structures. Our next assumption is that every program specification can be expressed as a first order sentence in the language of the specification structures. Modulo the usual equivalence relation $\equiv$, these sentences form a Boolean algebra, the Lindenbaum, Tarski algebra of the specification language. This algebra is isomorphic to the Boolean algebra of all strictly elementary classes of specification structures.

It is possible to define on these algebras two operations, one binary, the other unary. If we think of the specification structures as ordered pair $((\mathbf{U}_0, u), (\mathbf{U}_0, u'))$, then we can define in an obvious way the relative product of two classes of specification structures, and the converse of a class of specification structures. Next, corresponding operations on sentences in the specification language are defined. Let x and x' be sequences of constants in the specification language that denote the distinguished elements in the strings u and u'. Then the converse of a sentence $\phi(x, x')$ will be $\phi(x', x)$ and the relative product of $\phi(x, x')$ and $\psi(x, x')$ will be $(\exists\xi)(\phi(x, \xi) \wedge \psi(\xi, x'))$, where ξ is a string of variables. Of course, we only have to quantify over those variables that actually occur in the formulas. Modulo the equivalence relation $\equiv$, this determines the operations on the Lindenbaum, Tarski algebra. The algebra so obtained, call it $\mathbf{Q}$, will be a specification algebra in the sense of Definition 5.4.5.

If we assume that the number of distinguished elements is finite, then $\mathbf{Q}$ will have a unit, the specification that requires the output to be identical with the input. But this seems to be a rather stringent condition, for it puts a fixed upper bound on the number of elements that can be involved in any calculation.

Since we view program specifications as relations on the class of all states, we can define the scalar product $X \circ R$ of a class X of states and a specification R, and this operation will have a left conjugate $X \triangleleft R = X \circ R^{\smile}$. Furthermore, if X is strictly elementary, then so are $X \circ R$ and $X \triangleleft R$. Thus the Boolean algebra $\mathbf{M}$ of all classes of states that are strictly elementary relatively to the class of all states is a module over the specification algebra $\mathbf{Q}$. It is in fact a strong module, with

$$X \triangleright Y = \{(\mathbf{U}_0, u, u') : (\mathbf{U}_0, u) \in X \text{ and } (\mathbf{U}_0, u') \in Y\}$$

It follows that $\mathbf{M}^\sigma$ is a strong module over $\mathbf{Q}^\sigma$. We can take $\mathbf{M}^\sigma$ to be the Boolean algebra of all abstract classes of states (i. e., classes closed under elementary equivalence), and $\mathbf{Q}^\sigma$ can be taken to consist of all abstract classes of specification structures.

The abstract notion of a strong module over a specification algebra appears to capture many of the important properties of the concrete models, but something seems to be missing, for in the concrete models the element u defined in Theorem 6.1.5 is always equal to the unit element e, but this does not follow from the axioms. We therefore introduce an additional property.

6.2.1 Definition A module $\mathbf{M}$ over a specification algebra $\mathbf{A}$ is said to be *faithful* if, for any distinct elements $a, b \in A$, there exists $x \in M$ with $x \circ a \neq x \circ b$.

It is not known whether, given a strong module $\mathbf{M}$ over a specification algebra $\mathbf{A}$, it is always the case that $\mathbf{M}^\sigma$ is faithful over $\mathbf{A}^\sigma$. The next theorem shows that an affirmative answer to this question would be of some interest.

6.2.2 Theorem *Suppose* $\mathbf{M}$ *is a strong module over a specification algebra* $\mathbf{A}$*, and assume that* $\mathbf{M}^\sigma$ *is faithful over* $\mathbf{A}^\sigma$*. Let* $X = \mathrm{At}(\mathbf{M}^\sigma)$ *and*

$$E = \{(p, q) \in X^2 : p \circ 1 = q \circ 1\}.$$

Let $\mathbf{A}'$ *be the specification algebra of all subrelations of* E*, and let* $\mathbf{M}'$ *be the* $\mathbf{A}'$*-module of all subsets of* X*. Then* $(\mathbf{M}^\sigma, \mathbf{A}^\sigma)$ *is isomorphic to* $(\mathbf{M}', \mathbf{A}')$ *under the correspondence*

$$\begin{array}{lll} x \mapsto \{p \in X : p \leq x\} & & \textit{for all } x \in M^\sigma, \\ a \mapsto \{(p, q) \in E : p \triangleright q \leq a\} & & \textit{for all } a \in A^\sigma. \end{array}$$

6.3 Algebras of programs

Even the most complex computer program can be obtained from simple programs by a repeated application of simple constructions that can be regarded as operations on the set of all programs. In investigating the resulting algebra of programs, one of the objectives is to find a small set of basic operations that generates the clone of all operations. The solution of this problem will depend on the definition of a program, for that will determine what constructions are admissible. Since there exists no precise definition, the choice of operations is based on pragmatic rather than theoretical considerations. The operations most often taken as basic are the join operation, the composition, and the star operation, denoted $+$, $\circ$ and *, together with $p?$, where p is a property, or a strictly elementary class, of states. The suggested readings are

$$\begin{aligned} \alpha+\beta &: \quad \textit{Do either } \alpha \textit{ or } \beta. \\ \alpha\circ\beta &: \quad \textit{First do } \alpha\textit{, then } \beta. \\ \alpha^* &: \quad \textit{Do } \alpha \textit{ a finite number of times.} \\ p? &: \quad \textit{Test } p. \end{aligned}$$

The programs *skip* and *abort* are also treated as nullary operations, or distinguished elements. These will be written ι and 0.

In order for $+$ and * to be admissible, we have to admit non-deterministic programs. We could also allow the meet operation and the complementation, but this is rarely done. The reasons for this appear to be that they are not needed and that the complementation is somehow "more non-deterministic" than the other constructions. It is not clear how one can give a conclusive proof that these operations suffice, but the all-important operations *if – then – else* and *while – do* are in the clone generated:

$$(\textit{if } p \textit{ then } \alpha \textit{ else } \beta) = (p?)\circ\alpha + (\neg p)?\circ\beta,$$

$$(\textit{while } p \textit{ do } \alpha) = ((p?)\circ\alpha)^*\circ(\neg p)?.$$

For an algebraic treatment of programs, we therefore consider algebras

$$\mathbf{P} = (P, +, 0, \circ, \iota, ^*).$$

What should the axioms be? If we treat programs as syntactic entities, as formal expressions, then it is natural to assume that $\mathbf{P}$ is an absolutely free algebra. Alternatively, we can identify two programs if, whenever they are executed with the same input state, the set of possible output states is the same. It is this semantic approach that will be considered here, although for some purposes the syntactic concept is the appropriate one.

If we interpret programs as binary relations on the class of all states, then the operations $+$, $\circ$ and * should be interpreted as the union, relative multiplication and reflexive-transitive closure, and the programs *skip* and *abort* as the identity relation and the null relation, respectively. This provides us with more concrete models, algebras

$$\mathbf{A} = (A, +, 0, \circ, \iota, ^*)$$

consisting of a set A of binary relations, closed under the indicated operations. We shall refer to these as *concrete Kleene algebras,* and by a *representable Kleene algebra* we shall mean an algebra that is isomorphic to a concrete Kleene algebra.

We are now faced with an axiomatization problem similar to the one considered in Section 1.1 for the classical calculus of binary relations. There the basic facts were that the representable relation algebras form a variety that is not finitely based, but there is a relatively simple set of identities from which most of the important identities that hold in the representable algebras can be derived. Here the situation is much less satisfying. In the first place, the representable Kleene algebras do not form an elementary class. This, of course, is to be expected, in view of the non-finitary character of the operation *. Secondly, not only is the variety generated by the concrete Kleene algebras non-finitely based, but there is no known finite set of identities comparable to Tarski's axioms for relation algebras. For these reasons, there is no general agreement on how the concept of a Kleene algebra should be defined. In fact, in Kozen [1990], ten different notions of a Kleene algebra are listed. We refer to that paper and to Pratt [1987], together with the references listed in these papers, for a more detailed account of these concepts. Two of these concepts are defined below.

6.3.1 Definition By a *Kleene algebra* we mean an algebra $\mathbf{P} = (P, +, 0, \circ, \iota, ^*)$ such that

(i) $(P, +, 0)$ is a lower bounded join semilattice.

(ii) $(P, \circ, \iota)$ is a monoid.

(iii) $\circ$ distributes over $+$ and, for all $a \in A$, $0 \circ a = a \circ 0 = 0$.

(iv) For all $a, b \in A$,

$$\iota \leq a^*, \quad a \leq a^* \quad \text{and} \quad a^* \circ a^* = a^*,$$
$$a \circ b \leq b \quad \text{implies} \quad a^* \circ b \leq b,$$
$$b \circ a \leq b \quad \text{implies} \quad b \circ a^* \leq b.$$

6.3.2 Definition A Kleene algebra $\mathbf{P}$ is said to be *continuous* if, for all $a \in P$,

$$a^* = \sum(a^n : n \in \omega\}.$$

Since programs are represented by their graphs, they can be treated as scalars acting on the Boolean algebra of all classes of states, but unlike specifications, they do not in general map strictly elementary classes onto strictly elementary classes. The simplest programs, the assignments and the programs *skip* and *abort*, have this property, but the property is not preserved under the operation *. However, * preserves a weaker property: If α sends open classes (complements of elementary classes) into open classes, then so does α^*. In order to imitate this in the abstract situation, we consider a Kleene algebra $\mathbf{P}$ and a Boolean algebra $\mathbf{M}$, and postulate that the members of P act as scalars on $\mathbf{M}^\sigma$ rather than on $\mathbf{M}$. This results in a modified version of the notion of a dynamic algebra. This concept will be introduced in two stages.

6.3.3 Definition By a *module over a Kleene algebra* $\mathbf{P}$ we mean a Boolean algebra $\mathbf{M}$ with bilinear maps $\circ, \triangleleft : \mathbf{M} \times \mathbf{P} \to \mathbf{M}$ such that $\triangleleft$ is a left conjugate of $\circ$ and, for all $x \in M$ and $\alpha, \beta \in P$,

$$x \circ \iota = x, \quad x \circ (\alpha \circ \beta) = (x \circ \alpha) \circ \beta, \quad x \circ \alpha^* = sup\{x \circ \alpha^n : n \in \omega\}.$$

6.3.4 Definition By a *dynamic algebra* over a Kleene algebra $\mathbf{P}$ we mean a Boolean algebra $\mathbf{M}$ with maps $\circ, \triangleleft : \mathbf{M}^\sigma \times \mathbf{P} \to \mathbf{M}^\sigma$ and $? : \mathbf{M} \to \mathbf{P}$ such that

(i) $\mathbf{M}^\sigma$ is a module over $\mathbf{P}$.

(ii) For all $\alpha \in P$, the maps $x \mapsto x \circ \alpha$ and $x \mapsto x \triangleleft \alpha$ take open elements into open elements.

(iii) For all $p \in M$ and $x \in M^\sigma$, $x \circ p? = x \cdot p$.

The two features of this definition that appear to be new are that the scalar multiplication is required to be residuated, and that it is not required to map members of M to members of M. The existence of residuals makes for a much richer arithmetic, but the importance of the other modification, and of the axiom (ii) above, is not clear.

Bibliography

Bell, J.L., and Slomson, A.B.

[1969] *Models and Ultraproducts: An Introduction*, North-Holland Publishing Co.

Birkhoff, G.

[1935] On the structure of abstract algebras, *Proc. Cambridge Phil. Soc.* **31**, 433-454.

[1944] Subdirect unions of universal algebras, *Bull. Amer. Math. Soc.* **50**, 764-768.

[1948] *Lattice Theory,* Amer. Math. Soc. Colloq. Publ. vol. 25, 2nd ed. Amer. Math. Soc., Providence, RI.

Blok, W.J.

[1978] On the degree of incompleteness of modal logics and the covering relation in the lattice of modal logics, Report.

[1980] The lattice of varieties of modal algebras: An algebraic investigation, *J. Symbolic Logic* **45**, 221-238.

[1980a] Pretabular varieties of modal algebras, Trends in Logic, *Studia Logica* **39**, 101-124.

Blok, W.J., and Dwinger, Ph.

[1975] Equational classes of closure algebras, *Indag. Math.* **37**, 189-198.

Blok, W.J., and Pigozzi, D.

[1982] On the structure of varieties with equationally definable principal congruences I, *Algebra Universalis* **15**, 195-227.

Brink, C.

[1981] Boolean modules, *J. Algebra* **71**, 291-313.

Chang, C.C., and Keisler, J.H.

[1973] *Model Theory,* North-Holland Publishing Co, Amsterdam.

Comer, S.D.

[1983] A new foundation for the theory of relations, *Notre Dame J. Formal Logic* **24**, 181-187.

Duffin, R.J., and Pate, R.S.

[1943] An abstract theory of Jordan-Hölder composition series, *Duke Math. J.* **10**, 743-750.

Everett, C.J., and Ulam, S.M.

[1946] Projective algebras I, *Amer. J. Math.* **68**, 77-88.

Fine, K.

[1974] An incomplete logic containing S4, *Theoria* **40**, 23-29.

Goldblatt, R.

[1989] Varieties of complex algebras, *Ann. Pure Appl. Logic* **44**, 173-242.

[1991] On closure under canonical embedding algebras, in: *Algebraic Logic* (*Proc. Conf. Budapest 1988*) (H. Andréka *et al.*, eds.), Colloq. Math. Soc. János Bolyai **54**, North-Holland, Amsterdam, 217–231.

Halmos, P.

[1962] *Algebraic Logic,* Chelsea Publ. Co., New York.

Henkin, L.

[1970] Extending Boolean operations, *Fund. Math.* **22**, 723-752.

Hoare, C.A.R., and Jefing, H.

[1986] The weakest prespecification, II, *Fund. Inform.* **9**, 217-252.

Jipsen, P., Jónsson, B., and Rafter, J.

[a] Adjoining units to residuated Boolean algebras, preprint.

Jipsen, P., and Lukács, E.

[a] Minimal relation algebras, *Algebra Universalis*, to appear.

Jónsson, B.

[1967] Algebras whose congruence lattices are distributive, *Math. Scand.* **21**, 110-121.

[1982] Varieties of relation algebras, *Algebra Universalis* **15**, 273-298.

[a] The preservation theorem for canonical extensions of Boolean algebras with operators, to appear.

Jónsson, B., and Tarski, A.

[1951] Boolean algebras with operators I, *Amer. J. Math.* **73**, 891-939.

[1952] Boolean algebras with operators II, *Amer. J. Math.* **74**, 127-162.

Jónsson, B., and Tsinakis, C.

[a] Relation algebras as residuated Boolean algebras, to appear in *Algebra Universalis.*

Köhler, P., and Pigozzi, D.

[1980] Varieties with equationally definable principal congruences, *Algebra Universalis* **11**, 213-219.

Kozen, D.

[1990] On Kleene algebras and closed semirings, in: *Mathematical Foundations of Computer Science*, Lecture Notes in Computer Science **452**, Springer-Verlag, Berlin-Heidelberg-New York, 26-47.

Kripke, S.

[1963] Semantic analysis of modal logic I, *Grundlagen der Math.* **9**, 67-96.

Lemmon, E.J.
[1977] *An Introduction to Modal Logic,* Amer. Phil. Quarterly Monograph Ser., Blackwell, Oxford.

Lyndon, R.C.
[1950] The representation of relation algebras, *Ann. of Math.* **51**, 707-729.
[1961] Relation algebras and projective geometries, *Michigan Math. J.* **8**, 21-28.

McKinsey, J.C.C.
[1941] A solution of the decision problem for the Lewis systems S2 and S4 with an application to topology, *J. Symbolic Logic* **6**, 117-134.

McKinsey, J.C.C., and Tarski, A.
[1944] The algebra of topology, *Ann. of Math.* **45**, 141-191.
[1946] On closed elements in closure algebras, *Ann. of Math.* **47**, 122-162.
[1948] Some theorems about the sentential calculi of Lewis and Heyting, *J. Symbolic Logic* **13**, 1-15.

Makinson, D.C.
[1971] Some embedding theorems for modal logic, *Notre Dame J. Formal Logic* **12**, 252-254.

Pratt, V.
[1988] Dynamic algebras as a well-behaved fragment of relation algebras, in: *Algebraic Logic and Universal Algebra in Computer Science*, Lecture Notes in Computer Science **425**, Springer-Verlag, Berlin-Heidelberg-New York, 77-110.

Prenowitz, W., and Jantosciak, J.
[1979] *Join Geometries,* Springer-Verlag, Berlin-Heidelberg-New York.

Ribeiro, H.
[1952] A remark on Boolean algebras with operators, *Amer. J. Math.* **74**, 162-167.

Schröder, E.
[1895] *Algebra der Logik,* Band 3, Teubner, Leipzig.

Tarski, A.
[1941] On the calculus of relations, *J. Symbolic Logic* **18**, 73-89.
[1959] What is elementary geometry?, in: *The Axiomatic Method*, North-Holland, Amsterdam.

Thomason, S.K.
[1974] An incompleteness theorem for modal logic, *Theoria* **40**, 30-34.

Werner, H.
[1978] *Discriminator Algebras,* Studien der Algebra und ihre Anwendungen 6, Berlin.

Essentially Minimal Groupoids

Hajime MACHIDA
Department of Mathematics
Hitotsubashi University
2-1 Naka, Kunitachi, Tokyo 186
Japan

Ivo G. ROSENBERG
Département de mathématiques et de statistique
Université de Montéral
C.P. 6128, Succ. A
Montréal, Qué., H3C 3J7
Canada

Abstract

A clone C is essentially minimal if it contains an essential nonidempotent operation and every proper subclone of C is essentially unary. For a finite universe we determine all essentially minimal clones generated by groupoids of a certain type by means of four varieties and a family of varieties. We narrow the essentially minimal clones generated by groupoids of another type into three families.

1 Introduction

1.1 Let A be a finite universe with at least three elements. An n-ary operation on A is a map $f : A^n \to A$. Denote by $O_A^{(n)}$ the set of all n-ary operations on A and put

$$O_A := \bigcup_{n=1}^{\infty} O_A^{(n)}$$

(we use the symbol := for definitions; i.e., in the same sense as $\triangleq$ or $=_{\text{def}}$). A *clone* on A is a composition-closed subset of O_A containing all projections or, equivalently, the set of all term operations of a universal algebra on A. The set L_A of all clones on A, ordered by containment, forms an algebraic and dually algebraic lattice of cardinality $2^{\aleph_0}$ which remains largely unknown even for $|A| = 3$. Naturally the question arises about determining the extreme cases of either very large or very small clones. Although the co-atoms (= dual atoms) of L_A, called *maximal clones*, are known for $|A| \geq 3$, even the submaximal clones (i.e. those of distance 2 from the top) are not yet fully known. The atoms, called *minimal clones*, are fully known only for $|A| \leq 3$ [Cs 83]. There are many of them, and they seem

I. G. Rosenberg and G. Sabidussi (eds.), Algebras and Orders, 287–316.

to be hard to determine; cf. Quackenbush's recent survey [Qu 91]. The applications of minimal clones so far seem to be limited (e.g. they determine minimal p_n-sequences), hence there are few compelling reasons for their study which perhaps explains why so little is known.

In 1981 the first author introduced a hopefully more tractable variant of a minimal clone [Ma 81]. For its definition we need the following terminology and notations. Let $1 \leq i \leq n$. We say that an n-ary operation f on A *depends* on its i-th variable (or that its i-th variable is *essential*) if

$$f(a_1, \ldots, a_n) \neq f(a_1, \ldots, a_{i-1}, b_i, a_{i+1}, \ldots, a_n)$$

for some $a_1, \ldots a_n \in A$; else the i-th variable is nonessential (also fictitious or dummy). The operation f is *essential* if it depends on at least two variables. A clone is *essential* if it contains an essential operation and is *essentially unary* otherwise. Notice that an essentially unary clone C is just a "blown-up" monoid of selfmaps of A in the sense that C consists of all $f \in O_A$ of the form $f(x_1, \ldots, x_n) \approx g(x_i)$, where $1 \leq i \leq n$ and $g \in C \cap O_A^{(1)}$ (the symbol $\approx$ is reserved for identities whose variables range over A). Denote by E_A the set of essential clones. In [Ma81, Ma82a,b] a clone C is called essentially minimal if it is a minimal element of E_A; in other words, if $C \in E_A$ and each proper subclone is essentially unary. Obviously, every minimal clone which is essential is essentially minimal in this sense. Given the difficulties with minimal clones, we prefer to restrict the concept to the nonminimal clones. Thus in this paper a clone on A is *essentially minimal* if (i) C is nonminimal and (ii) a minimal element of E_A. Put differently, C is essentially minimal if (a) C is essential, (b) every proper subclone of C is essentially unary, and (c) C is not idempotent (where, as usual, C is idempotent if it consists of idempotent operations, i.e., f satisfying $f(x, \ldots, x) \approx x$). For $f \in O_A$ the clone *generated* by f is the least clone $[f]$ containing f and f *generates* every $h \in [f]$. Clearly, a clone C is essentially minimal if and only if (1) C is generated by an essential and nonidempotent operation f, and (2) if an essential h is generated by f, then h generates back f. Call an operation f *essentially minimal* if $[f]$ is essentially minimal.

1.2 For a selfmap h of A denote by $\langle h \rangle$ the subset of A permuted by h, that is, $\langle h \rangle$ consists of all elements on the cycles of the graph (or diagram) of h. Call h *reflective* if $h^2 = h \circ h = h$ (i.e. $h(x) = x$ for all $x \in$ im h; the usual term is idempotent which we already use in the universal algebra sense; we have tried to call it absorptive [MR84] or retractive [MR85]). In [MR84] we proved:

Let $f \in O_A^{(n)}$. Define $g \in O_A^{(1)}$ by $g(x) :\approx f(x, \ldots, x)$ and put $B := \langle g \rangle$. Let the restriction h of f to B be essential. Then f is essentially minimal if and only if

(i) *g is reflective and $f(x_1, \ldots, x_n) \approx f(g(x_1), \ldots g(x_n))$, and*

(ii) *either* (a) *$f(B^n) \subseteq B$ and h is an essentially minimal operation on B or* (b) *$g(f(x_1, \ldots, x_n))$ is nonessential.*

This theorem leaves open the case that the restriction of f to B is nonessential. This case seems to be much harder and so in this paper we study only the binary operations of this type. For notational simplicity we assume that $A = \boldsymbol{k} := \{0, \ldots, k-1\}$ and $B = \boldsymbol{l} = \{0, \ldots, l-1\}$. As usual, we write $f(x, y)$ multiplicatively as xy or $x \cdot y$, and refer to $G = \langle \boldsymbol{k}\,; \cdot \rangle$ as a *groupoid*. For $i > 0$ put

$$x *_i y :\approx g^i(xy).$$

Call G *vanishing* if some operation $*_i$ is nonessential. In Chapter 2 we find all nonvanishing essentially minimal finite groupoids (of our type and up to isomorphism). This is done through a rather tedious elimination process heavily dependent on the finiteness of the universe. However, the resulting groupoids are described by identities (i.e., axioms or laws), and the proof of their essential minimality applies to every essential member of their variety which can be studied separately. (We have some preliminary results for each of the three groups). Some of the results of this paper were announced in [MR85].

1.3 Groupoids have been studied from many points of view. It is likely but by no means necessary that essentially minimal groupoids will have relatively few or manageable n-ary terms depending on all their variables (semilattices are extreme cases of minimal groupoids with this property), and so essentially minimal groupoids may provide examples of interesting varieties of groupoids.

It is hoped that in contrast to the minimal groupoids the selfmap g will yield a tool for the determination of the essentially minimal groupoids.

The essentially minimal nonvanishing groupoids are given as the essential members of the varieties $V_1 - V_4$ and the family of varieties V_{5p} (p prime) given by the following identities. As usual xy^2 stands for $x(yy)$.

(1) $(xy)z \approx x(xy) \approx xy^2 \approx xy$

(2) $(xy)z \approx x(y(xx^2)) \approx xy,\ x(xy) \approx xx^2$

(3) $(xy)z \approx x^2,\ (xy)^2 \approx (xy)y \approx x^2y \approx xy$

(4) $(xy)z \approx xy,\ x(yz) \approx xx^2$

(5) There exists p prime and an element 0 such that

$(xy)z \approx (xz)y,\ (xy)0 \approx xy$

$x(yz) \approx x0 \approx (\ldots((xy)y)\ldots)y$ (with y repeated p times).

In Chapter 3 we narrow the vanishing essentially minimal groupoids into three groups.

We denote by id the identity selfmap of $\boldsymbol{k}$ (i.e. $\mathrm{id}(x) \approx x$). It is well known that to every selfmap h of $\boldsymbol{k}$ there exists $j \geq 0$ such that $f := h^j$ is reflective. The least j with this property is obtained as follows. Consider the relation $h^\circ := \{(x, h(x)) : x \in \boldsymbol{k}\}$ called the *graph* (or diagram) of h. It consists of vertex disjoint cycles (including the loops, i.e., the fixed points of h) and from each element of $\boldsymbol{k}$ not on a cycle there is a unique path to an element on a cycle. Denote by m the least common multiple of the cycle lengths, and by p the maximum of the above path lengths. Then $j = nm$, where n is the least integer such that $nm \geq p$. Moreover, im h^j is the set $\langle h \rangle$ of the elements on the cycles of h as well as the set of fixed points of h^j. Recall that we assume that $\langle g \rangle = \boldsymbol{l} = \{0, \ldots, l-1\}$ and that the restriction of G to $\boldsymbol{l}$ is nonessential. Since we can replace the groupoid by its reverse groupoid, we may assume without loss of generality that

$$xy = x^2 \quad \text{for all} \quad x, y \in \boldsymbol{l}. \tag{1}$$

For all $i \geq 0$ put $x *_i y :\approx g^i(xy)$ (where $x *_0 y \approx xy$). We say that G is *vanishing* if some $\langle \boldsymbol{k}\,; *_i \rangle$ is nonessential.

The sentence "without loss of generality we may assume that G has the property α" will indicate that G is term equivalent to a groupoid isomorphic to a groupoid satisfying α.

2 Nonvanishing groupoids

2.1 In this section $\langle \boldsymbol{k}\,; \cdot \rangle$ is an essentially minimal groupoid such that $*_{2j-1}$ is essential (where $g(x) \approx x^2$, j is the least positive integer such that $h := g^j$ is reflective and $\langle g \rangle = \boldsymbol{l}$ for some $0 < l < k$). We start with the following:

2.2 Lemma *Without loss of generality we may assume that $l > 1$. Then*

$$xy \in \boldsymbol{l} \quad \textit{for all} \quad x, y \in \boldsymbol{k}, \tag{2}$$

$$xy = x \quad \textit{for all} \quad x, y \in \boldsymbol{l}. \tag{3}$$

Proof Put $h := g^j$ and denote $*_{2j-1}$ by $\Box$. First notice that

$$x \Box y \approx h(x *_{j-1} y).$$

According to (1) for all $x, y \in \boldsymbol{l}$

$$x *_{j-1} y = g^{j-1}(xy) = g^{j-1}(x^2) = g^{j-1}(g(x)) = h(x) = x$$

and so

$$x \Box y = h(x *_{j-1} y) = h(x) = x$$

proving (3) for $\Box$. Clearly the range of $\Box$ is a subset of $\boldsymbol{l} =$ im h but, in view of (3), it actually equals $\boldsymbol{l}$ and so $\Box$ satisfies (2). Finally, $\Box$ is essential by the basic assumption of this chapter. Clearly $\square$ is a derived groupoid of $\langle \boldsymbol{k}; \cdot \rangle$ and so it suffices to replace $\langle \boldsymbol{k}; \cdot \rangle$ by $\langle \boldsymbol{k}; \square \rangle$. $\square$

2.3 We distinguish two cases according to whether G is essential or nonessential on $\boldsymbol{k} \times \boldsymbol{l}$:

$$\text{(A)} \qquad x_0 y_0 \neq x_0 0 \quad \text{for some} \quad k > x_0 \geq l > y_0 > 0, \tag{4}$$

$$\text{(B)} \qquad x \in \boldsymbol{k}, y \in \boldsymbol{l} \Rightarrow xy = x0. \tag{5}$$

In 2.4 – 2.14 we study the case (A) and in 2.15 the case (B).

2.4 Lemma *Let G satisfy* (2)–(4). *Then without loss of generality we may assume that G also satisfies*

$$x \in \boldsymbol{l}, \quad y \in \boldsymbol{k} \Rightarrow xy = x. \tag{6}$$

Proof Put $x * y :\approx xy^2$. Observe that according to (3) for all $x \in \boldsymbol{k}$ and $y \in \boldsymbol{l}$ we have

$$x * y = xy^2 = xy.$$

Combining this with (3) and (4) we obtain the essentiality of $*$. Observe that for all $x \in \boldsymbol{l}$ and $y \in \boldsymbol{k}$ we have $x * y = xy^2 = x$ on account of (3) and $y^2 \in \boldsymbol{l}$. Thus $*$ satisfies (6). By its definition the range of $*$ is a subset of $\boldsymbol{l}$. Since $*$ satisfies (6), its range is $\boldsymbol{l}$ and so $*$ satisfies (2). Now it suffices to replace G by $\langle \boldsymbol{k}; * \rangle$. $\square$

2.5 For each $z \in \boldsymbol{k}$ define a selfmap r_z of $\boldsymbol{k}$ by setting

$$r_z(x) :\approx zx.$$

From (6) we see that r_z is the constant map γ_z with value z for every $z \in \boldsymbol{l}$. Recall that $\langle r_z \rangle$ is the set of all elements on the cycles of $\{(x, r_z(x)) : x \in \boldsymbol{k}\}$. We have two cases:

$$(\alpha) \qquad |\langle r_t \rangle| > 1 \quad \text{for some} \quad l \leq t < k, \tag{7}$$

$$(\beta) \qquad |\langle r_t \rangle| = 1 \quad \text{for all} \quad l \leq t < k. \tag{8}$$

The case (α) is solved in 2.6–2.10 and the case (β) in 2.11–2.14. Note that (7) implies (4). We need the following well-known result:

2.6 Fact *There exists $q > 0$ such that all $r_0^q, \ldots, r_{k-1}^q$ are reflective.*

Proof Denote by m the least common multiple of the lengths of all the cycles of $r_0, \ldots, r_{k-1}$, by p_i the maximum path length joining an element not on a cycle of r_i to a cycle of r_i $(i = 0, \ldots, k-1)$, and put $p := \max(p_0, \ldots, p_{k-1})$. Then q is the least multiple of m not less than p. $\square$

2.7 Lemma *Let G satisfy* (2), (6) *and* (7). *Then without loss of generality we may assume that G also satisfies*

$$x(xy) \approx xy. \tag{9}$$

Proof Put

$$x * y :\approx r_x^q(y) \approx x(x(\dots(xy)\dots)).$$

Let $x \in \boldsymbol{l}$. Applying q times (6) we obtain $x * y = x$ for all $y \in \boldsymbol{k}$ proving that $*$ satisfies (6). In view of (7) and $\langle r_t \rangle \subseteq \boldsymbol{l}$ the reflective map r_t^q satisfies $r_t^q(0) \neq r_t^q(y_0)$ for some $y_0 \in \langle r_t \rangle \subseteq \boldsymbol{l}$, hence

$$t * y_0 = r_t^q(y_0) \neq r_t^q(0) = t * 0$$

proving (4) for $*$. Next, $*$ satisfies (2) because it satisfies (6), and its range is a subset of $\boldsymbol{l}$. Next $r_t'(x) :\approx t * x$ satisfies $r_t' = r_t^q$; hence $\langle r_t' \rangle = \langle r_t \rangle$ and therefore (7) holds for $*$. Finally

$$x * (x * y) \approx r_x^q(r_x^q(y)) \approx r_x^q(y) \approx x * y$$

by the reflectivity of r_x^q and so $*$ satisfies (9). Now it suffices to replace G by $\langle \boldsymbol{k}; * \rangle$. □

2.8 Lemma *Let G satisfy* (2), (6), (7) *and* (9). *Then without loss of generality we may assume that G also satisfies*

$$xy^2 \approx xy. \tag{10}$$

Proof Put $x * y :\approx xy^2$. First by (6) for each $x \in \boldsymbol{l}$ we have $x * y \approx x \approx xy$ for all $y \in \boldsymbol{k}$, and similarly for $y \in \boldsymbol{l}$ we have $x * y = x(y^2) = xy$. Thus $*$ and $\cdot$ agree everywhere except possibly on $(\boldsymbol{k}\backslash\boldsymbol{l})^2$. In particular, $*$ satisfies (2) and (6). For every $t \in \boldsymbol{k}$ put $r_t'(x) :\approx t * x$. By what has been shown above, r_t and r_t' agree on $\boldsymbol{l}$. Moreover, $\langle r_t \rangle \subseteq$ im $r_t \subseteq \boldsymbol{l}$, thus $\langle r_t \rangle \subseteq \langle r_t' \rangle$ and so $*$ satisfies (7). By the definition, (6) and (9) we have

$$x * (x * y) \approx x * (xy^2) \approx x((xy^2)(xy^2)) \approx xy^2 \approx x * y$$

and so $*$ satisfies (9). From (9) we have $y(y^2) \approx y^2$ and using $y^2 \in \boldsymbol{l}$ and (6) we get

$$x * (y * y) \approx x * (y(y^2)) \approx x * (y^2) \approx x(y^2y^2) \approx xy^2 \approx x * y$$

proving (10) for $*$. Now it suffices to replace G by $\langle \boldsymbol{k}; * \rangle$. □

From (6) and (2) we get

$$(xy)z \approx xy. \tag{11}$$

We show that the laws (9)–(11) imply the essential minimality for any essential groupoid.

2.9 Theorem *Let $G = \langle K, \cdot \rangle$ be an essential groupoid. If G satisfies*

$$(xy)z \approx x(xy) \approx xy^2 \approx xy, \tag{12}$$

then G is an essentially minimal groupoid.

Proof Let h be an n-ary essential operation on K derived from G. We need the following facts.

Claim 1. There exist $p > 1$ and $1 \leq i_1, \ldots, i_p \leq n$ such that $i_1 \neq i_2 \neq \ldots \neq i_p$ and

$$h(x_1, \ldots, x_n) \approx x_{i_1}(x_{i_2}(\ldots(x_{i_{p-1}}x_{i_p})\ldots)). \tag{13}$$

Proof. The operation h is given by a term (or formula) w built from $x_1, \ldots, x_n$ and parentheses. Choose w so that it contains the least possible total number of symbols from $\{x_1, \ldots, x_n\}$. From $(xy)z \approx xy$ we see that w has the form (13) for some $1 \leq i_1, \ldots, i_h \leq n$. From $x(xy) \approx xy$ it is evident that $i_1 \neq i_2 \neq \ldots \neq i_{p-1}$ and finally from $x(yy) \approx xy$ we obtain $i_{p-1} \neq i_p$.

Put $x \circ_1 y :\approx x$, denote $\cdot$ by $\circ_2$ and for $r > 2$ set

$$x \circ_r y :\approx x_1(x_2(\ldots(x_{r-1}x_r)\ldots)), \tag{14}$$

where $x = x_1 = x_3 = \ldots$ and $y = x_2 = x_4 = \ldots$ (e.g. $x \circ_3 y \approx x(yx), x \circ_4 y \approx x(y(xy))$, etc.).

Claim 2. The operation h generates $\circ_r$ for some $r > 1$.

Proof. Rename the variables so that $i_1 = 1$ in (13). As h is essential, we have $p > 1$. Consider $x * y :\approx h(x, y, \ldots y)$. Using $x(xz) \approx x(zz) \approx xz$, the expression (13) for $h(x, y, \ldots, y)$ can be simplified to (14) for some $r > 1$. Thus $\circ_r = *$ is generated by h.

Claim 3. The operation h generates $\circ$.

Proof. There is nothing to prove if $r = 2$ in Claim 2. Thus let $r > 2$. Form $x * y :\approx x \circ_r (y \circ_r y)$. From $x(y^2) \approx xy$ we obtain $y(y^2) \approx y^2$. Now $y \circ_r y \approx y^2$ and therefore

$$x * y \approx x \circ_r (y \circ_r y) \approx x \circ_r (y^2).$$

Put $z := x \circ_{r-2} (y^2)$. Then $x \circ_r (y^2) \approx x(y^2 z)$. From (12) we have $y^2 z \approx y^2$, $xy^2 \approx xy$ and consequently

$$x * y \approx x(y^2 z) \approx xy^2 \approx xy.$$

To complete the proof of the theorem we must show that G is not minimal.

Claim 4. An idempotent G such that $(xy)z \approx xy$ satisfies $xy \approx x$.

Proof. $xy \approx x^2 y \approx x^2 \approx x$. □

2.10 Remarks (1) Let G satisfy the assumptions of Theorem 2.9. In general, for $r > 2$ the groupoids $\langle K; \circ_r \rangle$ may form a large family providing an example of an essentially minimal clone with many binary terms. Claim 1 gives a canonical form for the term operations of G.

(2) The way we arrived at (12) indicates a construction of groupoids satisfying (12). Let L be a non-singleton proper subset of K. Let g be a reflective map from K onto L. If we set

$$xy = \begin{cases} x & \text{for all } x \in L \text{ and } y \in K \\ g(y) & \text{otherwise} \end{cases}$$

we obtain a groupoid satisfying (12).

(3) As a small example consider $k = 3, l = 2$ and $r_2(0) := 0$, $r_2(1) := 1$, $r_2(2) := a \in \mathbf{2} = \{0, 1\}$. Clearly r_2 is reflective, and r_0, r_1 and r_2 satisfy (i)–(iii) in (2) (with $L = \mathbf{2}$). The multiplication table of the corresponding groupoid is

	0	1	2
0	0	0	0
1	1	1	1
2	0	1	a

Having solved the case (α) from 2.5 we turn to the case (β).

2.11 Lemma *Let G be a groupoid with range $\mathbf{l}$ satisfying*

$$x_0 y_0 \neq x_0 0 \quad \textit{for some} \quad k > x_0 \geq l > y_0 > 0, \tag{4}$$

$$xy = x \quad \textit{for all} \quad x \in \mathbf{l} \quad \textit{and} \quad y \in \mathbf{k}, \tag{6}$$

$$|\langle r_z \rangle| = 1 \quad \textit{for all} \quad z \in \mathbf{k}. \tag{8}$$

Then without loss of generality we may assume that G also satisfies

$$x(xy) \approx x(x^2). \tag{15}$$

Proof Let $z \in \mathbf{k}$. By (8) we have $\langle r_z \rangle = \{z_0\}$ for some $z_0 \in \mathbf{k}$, and therefore there exists $j > 0$ such that $r_z^j(x) = z_0$ for all $x \in \mathbf{k}$. Denote by j_z the least j with this property, and put $q := \max(j_0, \ldots, j_{k-1})$. Here $q > 1$ because G is essential. Put $x * y :\approx r_x^{q-1}(y)$. From (6) we see that $*$ also satisfies (6). For every $z \in \mathbf{k}$ put $r'_z(x) :\approx z * x$. From $r'_z = r_z^{q-1}$ we see that $\langle r'_z \rangle = \langle r_z \rangle$ and so $*$ also satisfies (8). By the definition of q, there exist $a, b \in \mathbf{k}$ such that $r_a^{q-1}(b) \neq r_a^q(b)$, and consequently for $b' := r_a(b)$ we have $a * b \neq a * b'$. Combining this with (6) we obtain that $*$ is essential. Next,

$$x * (x * y) \approx r_x^{q-1}(r_x^{q-1}(y) \approx r_x^{2q-2}(y) \approx r_x^{2q-2}(x) \approx x * (x * x)$$

because $2q - 2 \geq q$ and so r_x^{2q-2} is constant. Thus $\langle \mathbf{k}; * \rangle$ satisfies (15) and we can replace G by $\langle \mathbf{k}; * \rangle$. □

Lemma 2.12 *Let G satisfy* (2), (4), (6) *and* (15). *Then without loss of generality we may assume that G also satisfies*

$$x(y(x(x^2))) \approx xy. \tag{16}$$

Proof Put $x*y :\approx x(y(x(x^2)))$. Let $x \in \boldsymbol{l}$ and $y \in \boldsymbol{k}$. Then from (6) we have $x*y = x = xy$ and $y*x = y(x(y(y^2)) = yx$, and therefore $*$ agrees with $\cdot$ everywhere except possibly on $(\boldsymbol{k}\backslash\boldsymbol{l})^2$. In particular, $*$ satisfies (4) and (6) and so $*$ is essential. We have

$$x * x \approx x(x(x(x^2))) \approx r_x^2(r_x^2(x)) \approx r_x^2(x) \approx x(x^2) \tag{17}$$

(because by (15) the map r_x^2 is constant).

We show that $*$ satisfies (15). Let $x, y \in \boldsymbol{k}$, $z := x * y$ and $t := x(x^2)$. Using $z \in \boldsymbol{l}$, (6), (15) and (17) we calculate

$$\begin{aligned} x * (x * y) &= x * z = x(z(x(x^2))) = x(zt) = xz \\ &= x(x(yt)) = x(x^2) = x * x, \end{aligned} \tag{18}$$

and therefore $*$ satisfies (15). We prove that $*$ satisfies (16). Notice that $y * t = yt \in \boldsymbol{l}$ due to $t \in \boldsymbol{l}$ and what has been shown at the beginning of the proof. By the same token, $x * (yt) = x(yt)$. Applying (18) we compute

$$\begin{aligned} x * (y * (x * (x * x))) &= x * (y * t) = x * (yt) = x(yt) \\ &= x(y(x(x^2)) = x * y \end{aligned}$$

and so $*$ satisfies (16) as well. □

As before (cf. the proof of 2.8), (6) and (2) imply

$$(xy)z \approx xy. \tag{11}$$

We show that (11), (15) and (16) are already sufficient for essential minimality.

2.13 Theorem *Every essential groupoid $G = \langle K; \cdot \rangle$ such that*

$$(xy)z \approx x(y(x(x^2))) \approx xy, \quad x(xy) \approx x(x^2) \tag{19}$$

is essentially minimal.

Proof Let h be an n-ary essential operation on K derived from G. To prove that h generates G we use the following claims.

Claim 1. There are $p \geq n$ and $1 \leq i_1, \ldots, i_p \leq n$ such that

$$i_1 \neq i_2 \neq \ldots \neq i_{p-2}; \quad i_{p-2} = i_{p-1} \Rightarrow i_{p-1} = i_p, \tag{20}$$

$$h(x_1, \ldots, x_n) \approx x_{i_1}(x_{i_2}(\ldots(x_{i_{p-1}} x_{i_p}) \ldots)). \tag{21}$$

Proof. Proceeding as in the proof of Claim 1 from the proof of Theorem 2.9 we obtain (21) for some $1 \le i_1, \ldots, i_p \le n$. From $x(xy) \approx x(x^2)$ we obtain the above restrictions (20).

For $m > 1$ and $j \in \{1,2\}$ put

$$x \circ_{mj} y :\approx x_1(x_2(\ldots(x_{m-1}x_m^j)\ldots)), \tag{22}$$

where $x = x_1 = x_3 = \ldots$ and $y = x_2 = x_4 = \ldots$ (and $x_m^1 := x_m$). We need:

Claim 2. The operation h generates a groupoid $\circ_{mj}$ for some $m > 1$ and $1 \le j \le 2$.

Proof. We may assume that $i_1 = 1$ and $i_2 = 2$ in (21) (if it is not, just rename the variables). Form $x * y :\approx h(x, y, \ldots, y)$.

(a) Suppose that both $i_{q-1} > 1$ and $i_q > 1$ for some q and denote by m the least value of such q. In view of $i_1 = 1$ and $i_2 = 2$ we have $m \ge 3$. If $m = p$, then by (20) clearly $*$ equals $\circ_{p2}$ and we are done. Thus let $m < p$. Applying the second and the last identities of (19) we obtain that $*$ equals $\circ_{mj}$ for some $j \in \{1,2\}$ concluding the proof in this case.

(b) Thus let $i_{j-1} = 1$ or $i_j = 1$ for all $j = 2, \ldots, p$. In view of (20) and $i_1 = 1, i_2 = 2$ we obtain $i_{2l-1} = 1$ for all $1 \le l \le \frac{1}{2}(p+1)$ and $i_{2l} > 1$ for all $1 \le l \le \frac{1}{2}p$. Now clearly $*$ equals $\circ_{pj}$ for $j = 1$ or $j = 2$.

Claim 3. If $1 \le j \le 2$ and $m > 1$ then $\circ_{mj}$ generates G.

Proof. Observe that $x \circ_{21} y \approx xy$ and so there is nothing to prove for $m = 2$ and $j = 1$. Thus let $n := m+j-1 \ge 3$. Write $\circ$ instead of $\circ_{mj}$. First note that $x \circ x \approx x(x(\ldots(x^2)\ldots))$ with n symbols x on the right hand side. Applying the last identity of (19) we obtain $x \circ x \approx x(x^2)$. Now from the first identity of (19)

$$y \circ (x \circ x) \approx y((x(x^2))(y((x(x^2))(\ldots))) \approx y(x(x^2)).$$

Finally by the same token and the second identity from (19)

$$x \circ (y \circ (x \circ x)) \approx x \circ (y(x(x^2))) \approx x(y(x(x^2))) \approx xy$$

proving the claim.

Combining Claims 1–3 we obtain that h generates G. To complete the proof it suffices to invoke Claim 4 from the proof of Theorem 2.9. □

2.14 Remark Lemmas 2.11 and 2.12 show a way of constructing groupoids satisfying (19). Let L be a proper subset of K and let the maps $r_x : K \to L$ $(x \in K)$ satisfy:

(i) $r_x(y) \approx x$ for all $x \in L$,

(ii) r_x^2 is constant for all $x \in K \backslash L$,

(iii) r_x is nonconstant on L for at least one $x \in K \backslash L$, and

(iv) $r_x(r_y(r_x^2(x))) = r_x(y)$ holds for all $x \in K \backslash L$ and $y \in K$.

Then $xy :\approx r_x(y)$ is essential and satisfies (19) and is therefore essentially minimal. Here is an example of such a family $\{r_x : x \in K\}$.

Example. Let $|K| \geq 4$, $a \in K$ and $\{L_1, L_2, b\}$ a partition of $L := K \backslash \{a\}$. Let r_a map $\{a, b\} \cup L_1$ onto b and L_2 into L_1, and let $r_x(y) := x$ for all $x \in L$ and $y \in K$.

It is easy to check (i)–(iii). We verify (iv). If $y \in L$ then $r_a(r_y(r_a^2(a))) = r_a(y)$ while for $y = a$ the condition (iv) reduces to $r_a^4(a) = b = r_a(a)$.

For $K := 4, a := 3, b := 0$ and $L_1 := \{1\}, L_2 := \{2\}$ the corresponding operation is given by the following table:

	0	1	2	3
0	0	0	0	0
1	1	1	1	1
2	2	2	2	2
3	0	0	1	0

.

□

We have solved the case (A) from 2.3 and so we turn to the case (B). For every $y \in \boldsymbol{k}$ put $c_y(x) :\approx xy$.

2.15 We assume that $\langle \boldsymbol{k}; \cdot \rangle$ is a groupoid with range $\boldsymbol{l}$ satisfying

$$x, y \in \boldsymbol{l} \Rightarrow xy = x, \tag{3}$$

$$x \in \boldsymbol{k}, y \in \boldsymbol{l} \Rightarrow xy = x0, \tag{5}$$

$$c_{y_0} \neq c_0 \text{ for some } y_0 \in \boldsymbol{k} \backslash \boldsymbol{l}. \tag{23}$$

2.16 For every $i > 0$ set

$$x \circ_i y :\approx c_y^i(x),$$

e.g. $x \circ_2 y \approx (xy)y$ and $x \circ_3 y \approx ((xy)y)y$. Let $y \in \boldsymbol{l}$ and $i > 1$. In view of $x \circ_{i-1} y \in \boldsymbol{l}$ and (3) we have $x \circ_i y = (x \circ_{i-1} y)y = x \circ_{i-1} y$. Combining this with (5) we obtain

$$x \circ_i y = x \circ_{i-1} y = \ldots = xy = x0. \tag{24}$$

If, moreover, $x \in \boldsymbol{l}$, then (24) yields $x \circ_i y = x0 = x$ and so $\circ_i$ satisfies (3). Since $0 \in \boldsymbol{l}$, the equation (24) gives $x \circ_i y \approx x0 \approx x \circ_i 0$ proving that $\circ_i$ satisfies (5).

2.17 According to Fact 2.6 there exists $q > 0$ such that all the maps $c_0^q, \ldots, c_{k-1}^q$, are reflective. Choose q to be the least positive integer with this property. Put $x * y :\approx x \circ_q y$. From the reflectivity we have

$$(x * y) * y \approx c_y^{2q}(x) \approx c_y^q(x) \approx x * y$$

and so $(x * y) * y \approx x * y$. With respect to the essentiality of $*$ we have three possible cases:

$$\text{(a)} \qquad x_0 * y_0 \neq x_0 \quad \text{for some} \quad 0 \leq x_0 < l \leq y_0 \leq k, \tag{25}$$

$$\text{(b)} \qquad x \in \boldsymbol{l}, y \in \boldsymbol{k} \Rightarrow x * y = x, \tag{26}$$

$$x_0 * y_0 \neq x_0 * 0 \quad \text{for some} \quad x_0, y_0 \in \boldsymbol{k} \backslash \boldsymbol{l}, \tag{27}$$

$$\text{(c)} \qquad x * y \approx x * 0. \tag{28}$$

First we consider the case (a).

2.18 Lemma *Let G have range $\boldsymbol{l}$ and satisfy* (3), (5) *and* (25). *Then without loss of generality we may assume that G also satisfies*

$$(xy)y \approx xy. \tag{29}$$

Proof We have shown in 2.16–2.17 that the groupoid $*$ satisfies (3), (5) and (29). Due to (3) and (25) it is also essential and so we can replace $\cdot$ by $*$. □

2.19 Lemma *Let G have range $\boldsymbol{l}$ and satisfy* (3), (5), (25) *and* (29). *Then without loss of generality we may assume that G also satisfies*

$$(x^2)y \approx xy, \tag{30}$$

$$x(yz) \approx x^2, \quad (xy)(xy) \approx xy. \tag{31}$$

Proof Put $x * y :\approx (x^2)y$. If $x \in \boldsymbol{l}$, then by (3) we have $x^2 = x$ and consequently $x * y = xy$ for all $y \in \boldsymbol{k}$. In particular, $*$ satisfies (3) and (25). Next, if $y \in \boldsymbol{l}$, then

$$x * y = (x^2)y = x^2 = (x^2)0 = x * 0$$

proving (5) for $*$.

Using $(x^2)y \in \boldsymbol{l}$, (3) and (29) we obtain

$$(x * y) * y \approx (((x^2)y) * y \approx ((x^2)y)((x^2)y))y \approx ((x^2)y)y \approx (x^2)y \approx x * y$$

and therefore $*$ satisfies (29). From (29) and (3) we get $(x^2)x \approx x^2 \approx (x^2)(x^2)$ and so

$$\begin{aligned}(x * x) * y &\approx ((x^2)x) * y \approx (x^2) * y \\ &\approx ((x^2)(x^2))y \approx (x^2)y \approx x * y\end{aligned}$$

proving (30) for $*$. Thus we may assume that already $\cdot$ satisfies (3), (5), (25), (29) and (30). It remains to show that $\cdot$ also satisfies (31). Setting $y = 0$ in (30) and taking into account that $0, x^2 \in \boldsymbol{l}$ we have $x0 \approx (x^2)0 \approx x^2$. Combining this with $yz \in \boldsymbol{l}$, from (5) we obtain $x(yz) \approx x0 \approx x^2$. Finally $(xy)^2 \approx xy$ follows from $xy \in \boldsymbol{l}$ and (3). □

The conditions (25) and (29)–(31) are already sufficient for essential minimality.

2.20 Theorem *Every essential groupoid $\langle K; \cdot\rangle$ satisfying*

$$x(yz) \approx x^2, \quad (xy)^2 \approx (xy)y \approx x^2y \approx xy \tag{32}$$

is essentially minimal.

Proof Let h be an n-ary essential term of G. We need:

Claim 1. There are $p > 1$ and $1 \leq i_1, \ldots, i_p \leq n$ such that $i_1 \neq i_2 \neq \ldots \neq i_p$ and

$$h(x_1, \ldots, x_n) \approx (\ldots((x_{i_1}x_{i_2})x_{i_3})\ldots)x_{i_p}. \tag{33}$$

Proof. Proceeding as in the proof of Claim 1 from the proof of Theorem 2.9, we obtain that h is of the form (33) for some $p > 1$ and $1 \leq i_1, \ldots, i_p \leq n$. From $(xy)y \approx (x^2)y \approx xy$ we see that indeed $i_1 \neq i_2 \neq \ldots \neq i_p$.

For $m > 1$ put

$$x \cdot_m y :\approx ((\ldots((x_1x_2)x_3)\ldots)x_{m-1})x_m \tag{34}$$

with $x_1 = x_3 = \ldots = x$ and $x_2 = x_4 = \ldots = y$.

Claim 2. The operation h generates $\cdot_m$ for some $m > 1$.

Proof. We may assume $i_1 = 1$ in (33). Put $x\Box y :\approx h(x, y, \ldots, y)$. Applying $(zy)y \approx zy$ and $x^2y \approx xy$ it is easy to verify that $\Box$ equals $\circ_m$ for some $m > 1$.

Claim 3. If $m > 1$, then $\cdot_m$ generates G.

Proof. From $(x^2)y \approx xy$ we have $(x^2)x \approx x^2$ and so $x \cdot_m x \approx x^2$. For $n > 1$ form

$$x *_n y :\approx (x \cdot_n x) \cdot_n y \approx (x^2) \cdot_n y. \tag{35}$$

By induction on $n = 2, 3, \ldots, m$ we prove that $*_n$ equals $\cdot$. From (35) and (32) we have

$$x *_2 y \approx x^2 \cdot_2 y \approx x^2y \approx xy$$

and so the statement holds for $n = 2$. Suppose it holds for some $n \geq 2$.

(a) Let n be odd. Then (34), the inductive assumption and (32) yield the required $x *_{n+1} y \approx (x *_n y)y \approx (xy)y \approx xy$.

(b) Let n be even. By the same argument $x *_{n+1} y \approx (x *_n y)x^2 \approx (xy)x^2$. For $z := xy$ and $t = u := x$ the identities of (31) yield

$$(xy)x^2 \approx z(tu) \approx z^2 \approx (xy)^2 \approx xy$$

proving $x *_{n+1} y \approx xy$ in this case.

Combining Claims 1–3 we obtain that h generates G. The fact that G is not minimal follows from the final claim:

Claim 4. $\langle K; \cdot \rangle$ *is not idempotent.*

Proof. Suppose it is idempotent. Then by (32) we have $xy \approx x(y^2) \approx x^2 \approx x$ in contradiction to the essentiality of $\langle K; \cdot \rangle$. □

2.21 Remark The sections 2.15–19 yield the following construction of groupoids satisfying (32). Let L be a proper subset of K, let $0 \in L$ and let $\{c_y : y \in K\}$ be a family of reflective maps from K into L such that

(i) $c_0(x) = x$ for all $x \in L$ and $c_y \circ c_0 = c_y$ for all $y \in K$,

(ii) $c_y = c_0$ for all $y \in L$,

(iii) $c_{y_0}(x_0) \neq x_0$ for some $y_0 \in K \backslash L$ and $x_0 \in L$, and

(iv) $c_x(x) = c_0(x)$ for all $x \in K$.

Then $\langle K; \cdot \rangle$ defined by $xy :\approx c_y(x)$ is essential and satisfies (32) and hence is essentially minimal.

Example. Let $L := \mathbf{2} \subset K$ and let $\{c_y : y \in K\}$ be a family of maps from K into L such that (i) $c_0 = c_1$ is nonconstant; (ii) if c_y is not constant then $c_y(0) = 0$ and $c_y(1) = 1$, and $c_y = c_0$; and (iii) at least one c_y is constant. For $K = \mathbf{3}$ such groupoids are

	0	1	2
0	0	0	0
1	1	1	0
2	0	0	0

	0	1	2
0	0	0	1
1	1	1	1
2	1	1	1

.

2.22 We turn to the case (b) from 2.17. By (26) and (27) the groupoid $*$ is essential. As shown in 2.17 it satisfies (5). Without loss of generality we may assume that already $\cdot$ is an essential groupoid with range $\boldsymbol{l}$ satisfying

$$x \in \boldsymbol{k}, y \in \boldsymbol{l} \Rightarrow xy = x0, \tag{5}$$

$$x \in \boldsymbol{l}, y \in \boldsymbol{k} \Rightarrow xy = x. \tag{26}$$

Now (26), (5) and $0, xy, yz, x^2 \in \boldsymbol{l}$ yield

$$(xy)z \approx xy, \quad x(yz) \approx x0 \approx x(x^2). \tag{36}$$

Now these identities are already sufficient for essential minimality:

2.23 Theorem *Every essential groupoid $G = \langle K; \cdot \rangle$ satisfying*

$$(xy)z \approx xy, \quad x(yz) \approx x(x^2) \tag{37}$$

is essentially minimal.

Proof It is easy to see that xy and yx are the only essential operations generated by G. Moreover, $\cdot$ is not idempotent because otherwise $xy \approx x(y^2) \approx x(x^2) \approx x$ in contradiction to the essentiality of $\cdot$. □

2.24 Example For $k = 3$ and $l = 2$ such groupoids are

	0	1	2
0	0	0	0
1	1	1	1
2	0	0	1

	0	1	2
0	0	0	0
1	1	1	1
2	1	1	0

.

2.25 Finally we consider the case (c) from 2.17 which turns out to be much more complex than the previous cases. In the remainder of Chapter 2 we assume that $\langle \boldsymbol{k}; \cdot \rangle$ is an essential groupoid with range $\boldsymbol{l}$ satisfying

$$c_0(x) = x \quad \text{for all} \quad x \in \boldsymbol{l}, \tag{3}$$

$$c_y = c_0 \quad \text{for all} \quad y \in \boldsymbol{l}, \tag{5}$$

$$c_0^q = \ldots = c_{k-1}^q \tag{38}$$

(where q is the least integer such that all $c_0^q, \ldots, c_{k-1}^q$ are reflective). Observe that c_0 maps $\boldsymbol{k}$ onto $\boldsymbol{l}$, and so from (3) we have $c_0^q = c_0$.

2.26 Lemma *Let $\langle \boldsymbol{k}; \cdot \rangle$ satisfy the conditions from* 2.25. *Then without loss of generality we may assume that there is a prime p such that*

$$c_0^p = \ldots = c_{k-1}^p = c_0, \tag{39}$$

$$c_z^a \neq c_0 \text{ for } \textit{some } z \in \boldsymbol{k} \text{ and } \textit{all } a = 1, \ldots, p-1. \tag{40}$$

Proof Let $y \in \boldsymbol{k} \backslash \boldsymbol{l}$. Taking into account $c_y^q = c_0$ and $c_0(x) = x$ for all $x \in \boldsymbol{l}$, we see that $\langle c_y \rangle = \boldsymbol{l}$ (i.e. $\boldsymbol{l}$ is the set of all vertices on the cycles of c_y). It is easy to see that q is the least common multiple of the lengths of the cycles of $c_l, \ldots, c_{k-1}$. Clearly $q > 1$ since otherwise $c_0 = \ldots = c_{k-1}$ and $\cdot$ would be nonessential. Choose a prime divisor p of q and put $r := q/p$. Now for each $y \in \boldsymbol{k}$ the map c_y^r has only cycles of length 1 (its fixed points) or p, and there is $l \leq z < k$ such that c_z^r has a cycle of length p. Put $x * y :\approx c_y^r(x)$. It is immediate that $*$ is essential and satisfies (3), (5), (39) and (40). Now it suffices to replace $\cdot$ by $*$. □

2.27 Lemma *If G satisfies* (3), (5), (39) *and* (40), *then $c_y = c_y \circ c_0$ for all $y \in \boldsymbol{k}$.*

Proof Let $x, y \in \boldsymbol{k}$. Put $t := c_y(x)$. Denote by u the unique element of $\boldsymbol{l}$ such that $c_y(u) = t$. Now by (39)

$$c_0(x) = c_y^p(x) = c_y^{p-1}(t) = c_y^p(u) = c_0(u) = u$$

and so

$$c_y(c_0(x)) = c_y(u) = t = c_y(x).$$

□

2.28 Remark For every $y \in \boldsymbol{k}$ denote by $\tilde{c}_y$ the restriction of c_y to $\boldsymbol{l}$. We know that $\tilde{c}_y^p$ is the identity map of $\boldsymbol{l}$, and hence $\tilde{c}_y$ is a permutation of $\boldsymbol{l}$ whose cycles are of length 1 or p.

By Lemma 2.27 the groupoid $\langle \boldsymbol{k}; \cdot \rangle$ is completely determined by the reflective map $c_0 : \boldsymbol{k} \to \boldsymbol{l}$ and the sequence $(\tilde{c}_l, \ldots, \tilde{c}_{k-1})$ of permutations of $\boldsymbol{l}$ (whose nontrivial members are of order p).

2.29 For all $n > 1$ we abbreviate $(\ldots((x_1x_2)x_3)\ldots)x_n$ by $x_1 \ldots x_n$. Here we also allow positive powers, e.g. xy^2x^2y stands for $((((xy)y)x)x)y$. Moreover, for $1 < i \leq m$, $x_1 \ldots x_{i-1}x_i^0x_{i+1} \ldots x_m$ denotes $x_1 \ldots x_{i-1}x_{i+1} \ldots x_m$. The following fact will often be invoked:

2.30 Fact *If $1 < i \leq n > 2$, $a_1, \ldots, a_n \in \boldsymbol{k}$ and $a_i \in \boldsymbol{l}$, then*

$$a_1 \ldots a_n = a_1 \ldots a_{i-1}a_{i+1} \ldots a_n. \tag{41}$$

Proof Put $b := a_1 \ldots a_{i-1}$. If $i > 2$ then $b \in \boldsymbol{l}$ and $ba_i = b$ by (3) and (5) proving (41). Thus let $i = 2$. From (5) and Lemma 2.27 we see that $a_1a_2a_3 = a_10a_3 = a_1a_3$ and so again we obtain (41). □

In 2.31–2.34 we restrict G to a groupoid with a simple shape of its term groupoids. For $m > 1$ and positive integers $\alpha_1, \dots, \alpha_m$ put $\alpha := (\alpha_1, \dots, \alpha_m)$ and

$$x *_\alpha y :\approx x_1^{\alpha_1} \dots x_m^{\alpha_m},$$

where $x = x_1 = x_3 = \dots$ and $y = x_2 = x_4 = \dots$ (e.g. for $\alpha = (1,2,2,1)$ we have $x *_\alpha y \approx xy^2x^2y \approx ((((xy)y)x)x)y)$. Put

$$o := \alpha_1 + \alpha_3 + \dots, \qquad e := \alpha_2 + \alpha_4 + \dots .$$

From Fact 2.30 we obtain:

2.31 Fact *Let $\alpha = (\alpha_1, \dots, \alpha_m)$ satisfy $m > 2$ or $\alpha_1 + \alpha_2 > 2$. If $y \in \boldsymbol{l}$, then*

$$x *_\alpha y = x^o$$

for all $x \in \boldsymbol{k}$.

2.32 Fact *Let $e = ph + e'$, where $h \geq 0$ and $0 \leq e' < p$. Then $x *_\alpha y = xy^{e'}$ for all $x \in \boldsymbol{l}$ and $y \in \boldsymbol{k}$. If p divides e, then $x *_\alpha y = x$ for all $x \in \boldsymbol{l}$ and $y \in \boldsymbol{k}$.*

Proof By Fact 2.30, (39) and (3) we have

$$\begin{aligned} x *_\alpha y &\approx xx^{\alpha_1 - 1}y^{\alpha_2}x^{\alpha_3} \dots \approx xy^{\alpha_2}y^{\alpha_4} \dots \\ &\approx xy^e \approx xy^p \dots y^p y^{e'} \approx x0^h y^{e'} \approx xy^{e'}. \end{aligned}$$

□

2.33 Lemma *Let G satisfy (3), (5), (39) and (40). Then without loss of generality we may assume that G also has the following property: If $m > 1$ and $\alpha_1, \dots, \alpha_m$ are positive integers such that $e := \alpha_2 + \alpha_4 + \dots = p$, then*

$$x^{\alpha_1}y^{\alpha_2}x^{\alpha_3}y^{\alpha_4} \dots \approx x^o \tag{42}$$

where $o := \alpha_1 + \alpha_3 + \dots$.

Proof By Facts 2.31 and 2.32 we have $x *_\alpha y = x$ for all $(x, y) \in \boldsymbol{l} \times \boldsymbol{k}$ and $x *_\alpha y = x^o$ for all $(x, y) \in \boldsymbol{k} \times \boldsymbol{l}$. Suppose that $*_\alpha$ is essential. Then it depends on its second variable and so $x_0 *_\alpha y_0 \neq x_0 *_\alpha 0$ for some $x_0, y_0 \in \boldsymbol{k} \backslash \boldsymbol{l}$. This means that $*_\alpha$ is of the type (b) from 2.17 which has been completely solved in 2.22-2.23. Thus we may assume that $*_\alpha$ depends on its first variable only. Notice that for $x \in \boldsymbol{l}$ by (3) and (5) we have $x = x^2 = \dots = x^o$ and so $x *_\alpha y \approx x^o$. □

2.34 Remark *Since $\boldsymbol{l}$ is the range of G and every c_y permutes $\boldsymbol{l}$, we have the following cancellation law:*

$$aby_1 \dots y_m = cdy_1 \dots y_m \Rightarrow ab = cd.$$

We have:

2.35 Lemma *Let G satisfy* (3), (5), (39), (40) *and* (42). *If $m > 1$ and $\alpha_1, \ldots, \alpha_m$ are positive integers, then*

$$x^{\alpha_1} y^{\alpha_2} x^{\alpha_3} y^{\alpha_4} \ldots \approx x^o y^e, \tag{43}$$

where $0 < o \le p$ and $0 < e \le p$ are determined by

$$o \equiv \alpha_1 + \alpha_3 + \ldots (\text{mod } p), \quad e \equiv \alpha_2 + \alpha_4 + \ldots (\text{mod } p). \tag{44}$$

Proof We need:

Claim. If $i > 0$ and $0 < j \le p$, then

$$x^i y^j x \approx x^i y^{j-1} x y. \tag{45}$$

Proof. First consider $1 < j \le p$. Applying Lemma 2.33 twice we obtain

$$x^i y^j x y^{p-j} \approx x^{i+1} \approx x^i y^{j-1} x y^{p-j+1}$$

and (45) follows by Remark 2.34. Finally let $i = 1$. By the same token $x^i y x y^{p-1} \approx x^{i+1} \approx x^{i+1} y^p$ and $x^i y x \approx x^{i+1} y$.

We prove (43) by induction on $n := \alpha_1 + \alpha_2 + \ldots$.

(1) Let $n = 2$. Then $\alpha_1 = \alpha_2 = o = e = 1$ and (43) reduces to $xy = xy$.

(2) Suppose that (43) holds for some $n \ge 2$ and let $\alpha_1, \ldots, \alpha_m$ be positive integers summing up to $n+1$. Let $e :\equiv \alpha_2 + \alpha_4 + \ldots (\text{mod } p)$ satisfy $0 < e \le p$. Put $z := x_1^{\alpha_1} \ldots x_m^{\alpha_m}$, where $x = x_1 = x_3 = \ldots$ and $y = x_2 = x_4 = \ldots$. We have $z \approx x_1^{\alpha_1} \ldots x_m^{\alpha_m - 1} x_m$. Applying the induction hypothesis to $x_1^{\alpha_1} \ldots x_m^{\alpha_m - 1}$, we obtain $z \approx x^o y^{e-1} y$ for m even and $z \approx x^{o-1} y^e x$ for m odd. In the former case we are done. In the latter case by the claim and again by the induction hypothesis

$$z \approx x^{o-1} y^e x \approx x^{o-1} y^{e-1} x y \approx x^o y^{e-1} y \approx x^o y^e.$$

This concludes the induction step and the proof. □

For all $y, z \in \boldsymbol{k}$ define a selfmap ψ_{yz} of $\boldsymbol{k}$ by

$$\psi_{yz}(x) :\approx xyzy^{p-1}z^{p-1}. \tag{46}$$

We have:

2.36 Lemma *If $x, y, z \in \boldsymbol{k}$ are either not pairwise distinct or $y \in \boldsymbol{l}$ or $z \in \boldsymbol{l}$, then*

$$\psi_{yz}(x) = x0. \tag{47}$$

Proof By Lemmas 2.33 and 2.26 we have:

$$\begin{aligned}\psi_{xz}(x) &\approx x^2zx^{p-1}z^{p-1} \approx x^{p+1} \approx x0,\\ \psi_{yx}(x) &\approx xyxy^{p-1}x^{p-1} \approx x^{p+1} \approx x0,\\ \psi_{yy}(x) &\approx xy^2y^{p-1}y^{p-1} \approx xy^{2p} \approx x0.\end{aligned}$$

If $y \in \boldsymbol{l}$ then by Fact 2.30 and (39) we have $\psi_{yz}(x) = xz^p = x0$. The case $z \in \boldsymbol{l}$ is analoguous. □

For all $y, z \in \boldsymbol{k}$ denote by φ_{yz} the restriction of ψ_{yz} to $\boldsymbol{l}$. This means that to every $x \in \boldsymbol{l}$ the map φ_{yz} assigns the value $xyzy^{p-1}z^{p-1}$. As noted in Remark 2.34, the map φ_{yz} is a permutation of $\boldsymbol{l}$.

For each $d \geq 0$ define a ternary operation f_d on $\boldsymbol{k}$ by setting

$$f_d(x,y,z) :\approx \psi^d_{yz}(x) \tag{48}$$

(i.e., $f_d(x,y,z) :\approx xyzy^{p-1}z^{p-1}\dots yzy^{p-1}z^{p-1}$, where $yzy^{p-1}z^{p-1}$ is iterated d times). Now we arrive at

2.37 Proposition *Let G satisfy* (3), (5), (39), (40) *and* (42). *Then without loss of generality we may assume that*

$$xyz \approx xzy. \tag{49}$$

Proof Our starting point is the following claim.

Claim 1. If f_1 is nonessential, then (49) *holds.*

Proof. Let f_1 be nonessential. Then by (48), (39), $0 \in \boldsymbol{l}$ and Fact 2.30,

$$\begin{aligned}xyzy^{p-1}z^{p-1} &\approx f_1(x,y,z) \approx f_1(x,x,x) \approx x^{2p+1}\\ &\approx x0^2 \approx x0 \approx xz^p \approx xz0z^{p-1}\\ &\approx xzy^pz^{p-1}.\end{aligned}$$

Applying Remark 2.34 we obtain (49).

In view of the claim we assume that f_1 is essential. Denote by m the least common multiple of the cycle lengths of the permutations φ_{yz} of $\boldsymbol{l}$ (for all $y, z \in \boldsymbol{k}$). Clearly $m > 1$ and $\varphi^m_{yz} = \mathrm{id}_{\boldsymbol{l}}$ for all $y, z \in \boldsymbol{k}$. Denote by q the least prime divisor of m and put $d := m/q$. A direct check shows that:

$$\varphi^d_{vw} \neq \mathrm{id}_{\boldsymbol{l}} \quad \textit{for some } v, w \in \boldsymbol{k}, \tag{50}$$

$$\varphi^m_{yz} = \mathrm{id}_{\boldsymbol{l}} \quad \textit{for all } y, z \in \boldsymbol{k}. \tag{51}$$

Put $h := f_d$. We need

Claim 2. The ternary operation h is essential, and

$$h(y,x,x) \approx h(x,y,x) \approx h(x,x,y) \approx x0, \tag{52}$$

$$h(x0,y,z) \approx h(x,y0,z) \approx (x,y,z0) \approx h(x,y,z). \tag{53}$$

Proof. The identity (52) follows from the definition and Lemma 2.36. Similarly, (53) is a consequence of Fact 2.30. By Lemma 2.36 for $y \in \boldsymbol{l}$ we have $h(x,y,z) = x0$ for all $x, z \in \boldsymbol{k}$, which shows that h depends on its 1st variable. By (50) there exist $u \in \boldsymbol{l}$ and $v, w \in \boldsymbol{k}$ such that $\varphi^d_{vw}(u) \neq u$. By Lemma 2.36 and (3) we have $\varphi_{0w}(u) = u$ and hence $h(u,0,w) = \varphi^d_{0w}(u) = u$ while $h(u,v,w) = \varphi^d_{vw}(u) \neq u$. This shows that h depends on its 2nd variable as well and so h is essential.

Claim 3. Every binary operation generated by f is nonessential.

Proof. Let b be a binary operation generated by h. Denote by w a word built correctly from the symbols h, x_1, x_2 and the parentheses such that $b(x_1,x_2) \approx w(x_1,x_2)$ and containing the least possible number of symbols x_1 and x_2. We show that w has no proper subword. Suppose to the contrary that it has a proper subword. Then w has a subword of the form $h(x_p,x_q,x_r)$ for some $p,q,r \in \{1,2\}$. In view of (52) we have $h(x_p,x_q,x_r) \approx x_l0$ for some $l \in \{1,2\}$. Now applying (53) we can shorten w. This contradiction shows that w has no proper subword. Thus $w(x_1,x_2) \approx h(x_p,x_q,x_r)$ for some $p,q,r \in \{1,2\}$, and by the same argument as above $w(x_1,x_2) \approx x_10$ proving that b is nonessential. □

By Claim 2 the operation h is an essential term of G. According to Claim 3 the operation h does not generate any essential binary operation, and so G is not essentially minimal and we are left with the case (49). (Remark: it is likely that h – which resembles a majority operation – is essentially minimal but in this paper we study only essentially minimal groupoids). □

The identity (49) already determines an essentially minimal groupoid. Recall that $x_1 \ldots x_n$ stands for $(\ldots(x_1x_2)x_3)\ldots)x_n$.

2.38 Proposition *Let $G = \langle K; \cdot \rangle$ be an essential groupoid. If there exist a prime p and an element 0 in K such that*

$$xyz \approx xzy \tag{54}$$

$$x(yz) \approx x0 \approx xy^p \tag{55}$$

$$xy0 \approx xy \tag{56}$$

then G is essentially minimal.

Proof Let h be an essential n-ary term operation of G.

Claim 1. *There are* $r > 1$, $1 \leq i_2 < \ldots < i_r \leq n$, $i_1 \in \{1,\ldots,n\}\backslash\{i_2,\ldots,i_r\}$ *and* $0 < \alpha_1 \leq p$, $0 < \alpha_2,\ldots,\alpha_r < p$, *such that*

$$h(x_1,\ldots,x_n) \approx x_{i_1}^{\alpha_1} \ldots x_{i_r}^{\alpha_r}. \tag{57}$$

Proof. The operation h is determined by a term (or formula) w correctly built from $x_1,\ldots,x_n$ and parentheses. Denote by $l(w)$ the number of symbols from $\{x_1,\ldots,x_n\}$ in w. Choose w so that $l(w)$ is the least possible. We show:

Fact 1. *The term* w *contains no subterm* v *of the form* $v_1(v_2v_3)$ *with* $l(v_i) > 0$ $(i = 1,2,3)$.

Proof. Suppose to the contrary that w contains such a subterm v. We show that $l(v_1) = 1$. Suppose to the contrary that $l(v_1) > 1$. Then $v_1 = zt$ where $l(z) > 0$ and $l(t) > 0$. Applying (55) and (56) we obtain

$$v \approx v_1 0 = (zt)0 \approx zt = v_1,$$

where $l(v) > l(v_1)$ in contradiction to the minimality of $l(w)$. Thus $l(v_1) = 1$, i.e. $v_1 = x_i$ for some $1 \leq i \leq n$. From (55) we obtain

$$v \approx x_i(v_2v_3) \approx x_i 0 \approx x_i(x_i^2)$$

proving that the term operation determined by v is essentially unary. Since h is essential, it follows that $w \neq v$ and so either w contains a subterm uv or a subterm vu with $l(u) > 0$. Consider the first case. Applying (55) we obtain $uv \approx u0 \approx u(x_1^2)$, where $l(uv) > l(ux_1^2)$, again in contradiction to the minimality of $l(w)$. Thus w contains a subterm vu with $l(u) > 0$. On account of (54) and (55) we have

$$vu \approx (x_1 0)u \approx (x_1 u)0 \approx x_1 u;$$

here again $l(vu) > l(x_1 u)$ provides the final contradiction.

Fact 1 yields immediately that

$$h(x_1,\ldots,x_n) \approx x_{j_1} \ldots x_{j_m}$$

for some $1 \leq j_1,\ldots,j_m \leq n$. From (54) we see that

$$h(x_1,\ldots,x_n) \approx x_{i_1}^{\beta_1} \ldots x_{i_r}^{\beta_r}$$

for some $\beta_1,\ldots,\beta_r > 0$ and $i_1,\ldots,i_r \in N := \{1,\ldots,n\}$ such that $i_2 < \ldots < i_r$ and $i_1 \in N\backslash\{i_2,\ldots,i_r\}$. Applying (54)–(56) we can reduce β_j to $0 < \alpha_j < p$ $(j = 2,\ldots,r)$ and β_1 to $0 < \alpha_1 \leq p$.

Claim 2. *The operation* h *generates the groupoid* $x^i y^j$ *for some* $0 < i \leq p$ *and* $0 < j < p$.

Proof. We can exchange the variables of h so that $i_m = m$ for all $m = 1, \ldots, r$. In view of (54)-(56) we have

$$\begin{aligned} h(x, y, x, \ldots, x) &\approx x^{\alpha_1} y^{\alpha_2} x^{\alpha_3 + \ldots + \alpha_m} \\ &\approx x^{\alpha_1 + \alpha_3 + \ldots + \alpha_m} y^{\alpha_2} \approx x^i y^j, \end{aligned}$$

where $0 < i \leq p$, $i \equiv \alpha_1 + \alpha_3 + \ldots + \alpha_m \pmod p$, and $j = \alpha_2$.

Claim 3. Let $0 < i \leq p$ and $0 < j < p$. The groupoid $x^i y$ is a term groupoid of $x^i y^j$.

Proof. Put $x \cdot_1 y :\approx x^i y^j$ and for $m = 2, 3, \ldots$ define $x \cdot_m y$ inductively by setting

$$x \cdot_m y :\approx (x \cdot_1 y) \cdot_{m-1} y.$$

By induction on $m = 1, 2, \ldots$ we prove that $x \cdot_m y \approx x^i y^{mj}$. Clearly this holds for $m = 1$. Suppose $m > 1$ and the statement holds for $m - 1$. Now by the definition and induction hypothesis

$$x \cdot_m y \approx (x^i y^j) y^{(m-1)j} \approx x^i y^j y^{(m-1)j} \approx x^i y^{mj},$$

thus completing the induction step. Now it suffices to choose $m > 0$ so that $mj \equiv 1 \pmod p$. Then $x^i y \approx x^i y^{mj} \approx x \cdot_m y$ where clearly $x \cdot_m y$ is a term groupoid of $x^i y^j$.

Claim 4. Let $0 < i \leq p$. Then G is a term groupoid of the groupoid $x^i y$.

Proof. There is nothing to prove if $i = 1$ and so let $1 < i \leq p$. The groupoid $x^i y$ generates the unary operation $h(x) :\approx x^i x \approx x^{i+1}$. Form $x \bullet y :\approx h(x)^i y$. We show that $x \cdot y \approx x^{i+1} y$. Denoting x^{i+1} by z and applying (55), (54) and (56) we get

$$x \bullet y \approx z^i y \approx z^{i-1} z y \approx z^{i-1} 0 y \approx z^{i-1} y 0 \approx z^{i-1} y.$$

Continuing in this fashion we obtain $x \bullet y \approx zy \approx x^{i+1} y$. This shows that $x^{i+1} y$ is a term groupoid of $x^i y$. By the same token $x^{i+2} y$ is a term groupoid of $x^i y$, etc., and finally $x^{p+1} y$ is a term groupoid of $x^i y$. However,

$$x^{p+1} y \approx x x^p y \approx x 0 y \approx x y 0 \approx xy$$

proving the claim.

In view of Claims 1–4 it remains to show that G is not idempotent. If $x^2 \approx x$, then $xy \approx x(y^2) \approx x0$, contrary to the essentiality of G. □

3 A classification of vanishing essentially minimal groupoids

3.1 In this chapter we classify the vanishing essentially minimal groupoids into three basic groups. We start by recalling our notation. Let $G = \langle \boldsymbol{k}; \cdot \rangle$ be an essential groupoid.

The selfmap g^G (or shortly g) of $\boldsymbol{k}$ is defined by $g^G(x) :\approx x^2$. Denote by j the least positive integer such that g^j is reflective (cf. 1.1) and put $\langle g\rangle :=$ im g^j. We assume that $\langle g\rangle = \boldsymbol{l} = \{0,\ldots,l-1\}$ for some $0 < l < k$. For every $a \in \boldsymbol{k}$ define the (row-map) r_a^G (or shortly r_a) by $r_a^G(x) :\approx ax$. Put

$$\Pi_G := \{r_a^G : a \in \boldsymbol{k}\}, \quad \sigma_G := |\text{im } r_0^G| + \ldots + |\text{im } r_{k-1}^G|. \tag{58}$$

Call $a \in \boldsymbol{k}$ *regular* if r_a^G is constant and *singular* otherwise (i.e., if $|\text{im } r_a^G| > 1$). Next denote by R_G and S_G (or shortly R and S) the sets of the regular and singular elements, and notice that S_G is nonvoid due to the essentiality of G. As in 1.1 for $i > 0$ put

$$x *_i y :\approx g^i(xy),$$

and call G *vanishing* if some $\langle \boldsymbol{k}; \cdot_i\rangle$ is nonessential. Finally, G is *taut* if

$$g(xy) \approx g^2(x). \tag{59}$$

Observe that (59) is equivalent to the validity of

$$\text{im } r_a \subseteq g^{-1}(g^2(a)) \tag{60}$$

for all $a \in \boldsymbol{k}$ (where $g^{-1}(b) := \{x \in \boldsymbol{k} : g(x) = b\}$).

We have:

3.2 Lemma *Let G be vanishing. If G is not taut, then*

$$|\Pi_H| \leq |\Pi_G|, \quad |R_H| \geq |R_G|, \quad \sigma_H < \sigma_G \tag{61}$$

for some taut $H = \langle \boldsymbol{k}; *_i\rangle$.

Proof By i denote the greatest integer such that the groupoid $*_i$ is essential. Then clearly $*_{i+1}$ is nonessential and hence without loss of generality we may assume it does not depend on its 2nd variable.

Put $x \circ y :\approx x *_i y$, $H := \langle \boldsymbol{k}; \circ\rangle$ and $h := g^H$. Observe that

$$h(x) \approx x * x \approx g^i(x^2) \approx g^{i+1}(x).$$

Now

$$\begin{aligned} h(x * y) &= g^{i+1}(g^i(xy)) \approx g^i(x *_{i+1} y) \approx g^i(x *_{i+1} x) \\ &\approx g^{2i+1}(x^2) \approx g^{2i+2}(x) \approx h^2(x), \end{aligned}$$

and so by (59) the groupoid H is taut. Since $r_a^H = g^i \circ r_a$, the first two inequalities in (61) are immediate and $\sigma_H \leq \sigma_G$. Suppose to the contrary that $\sigma_H = \sigma_G$. Then im $(g^i \circ r_a) =$ im r_a holds for all $a \in \boldsymbol{k}$ and so g^i permutes the set im r_a. This means that g is injective on every im r_a and so $g^{i+1} = g \circ g^i$ is injective on im r_a. Since $g^{i+1} \circ r_a$ is constant, this implies r_a constant for all $a \in \boldsymbol{k}$, in contradiction to the essentiality of G. □

3.3 The basic idea of our proof is the successive elimination of groupoids that are not essentially minimal until the only groupoids left are the essentially minimal ones. The inherent risk of moving on cycles will be avoided if each time we eliminate some groupoids certain parameters of the remaining groupoids improve. This can be formulated in such a way that we study only those groupoids for which the parameters are already optimal and eliminate all with non-optimal parameters. The parameters which will serve us are those from (61). Denote by E_G the set of essential term groupoids of G.

3.4 An essential groupoid G is *optimal* if

(i) $|\Pi_G| = \min\{|\Pi_K| : K \in E_G\}$;

(ii) $|R_G| = \max\{|R_K| : K \in E_G, |\Pi_K| = |\Pi_G|\}$; and

(iii) $\sigma_G = \min\{\sigma_K : K \in E_G, |\Pi_K| = |\Pi_G|, |R_K| = |R_G|\}$.

In other words, G is optimal if first within E_G it has the least number ρ of row-maps, secondly amongst groupoids from E_G with exactly ρ row-maps it has the greatest number γ of constant rows, and finally among those groupoids from E_G with exactly ρ row-maps and exactly γ constant rows it has the least σ.

Without loss of generality we may choose our essentially minimal groupoid to be optimal. From now on, G is always optimal and so according to Lemma 3.2 also taut. In the next theorem we separate the optimal groupoids into three basic groups.

3.5 Theorem *If G is optimal then it satisfies one of the following conditions*:

(i) $g^{-1}(S)$ *is empty*;

(ii) $\operatorname{im} g^2 = \boldsymbol{l}$ *and the equivalence*

$$\Theta := \{(x, y) \in \boldsymbol{k}^2 : r_x = r_y\} \tag{62}$$

has blocks $B_0, \ldots, B_{l-1}$ *such that for all* $i = 0, \ldots, l-1$ *we have* (1) $B_i \cap \boldsymbol{l} = \{i\}$, *and* (2) r_x *maps* B_i *into* $B_{g(i)}$ *for every* $x \in B_i$.

(iii) $\operatorname{im} g^3 = \{0\}$, $0 \in S$, *and* $r_a = r_0$ *for all* $a \in \operatorname{im} g$ *while* $r_b \neq r_0$ *for some* $b \in \boldsymbol{k} \setminus \operatorname{im} g$.

Proof For $i \geq 0$ put $x \circ_i y :\approx g^i(x)y$. We start with

Case 1 Let $S \cap \langle g \rangle = \emptyset$. Denote by d the largest integer such that $g^d(c) \in S$ for some $c \in \boldsymbol{k}$. If $d = 0$, then $g^{-1}(S)$ is void and we have the case (i). Thus let $d > 0$. Put $x \Box y :\approx x \circ_d y$, $H := \langle \boldsymbol{k}; \Box \rangle$ and $h := g^H$. Denote the row-maps of $\Box$ by r'_a, and put

$$S' := \{a \in \boldsymbol{k} : |\operatorname{im} r'_a| > 1\}.$$

Observe that for $v := g^d(a)$ we have

$$r'_a(x) \approx vx \approx r_v(x). \tag{63}$$

There exists $c \in \boldsymbol{k}$ such that $b := g^d(c) \in S$. By (63) we have $r'_c = r_b$ and so $|\text{im } r'_c| = |\text{im } r_b| > 1$ proving $c \in S'$ and $S' \neq \emptyset$. We need:

Claim 1. The groupoid $H := \langle \boldsymbol{k}; \square \rangle$ is essential.

Proof. Since S' is nonempty, clearly $\square$ depends on its 2nd variable. We show that $R' := \{a \in \boldsymbol{k} : |\text{im } r'_a| = 1\}$ is nonempty. From the assumption $\langle g \rangle \cap S = \emptyset$ we see that $\boldsymbol{l} = \langle g \rangle \subseteq R$. Put $v := g^d(0)$. Taking into account that g permutes $\langle g \rangle$, clearly $v \in \langle g \rangle$ and so from (63) we obtain that $r'_0 = r_v$ is constant, proving $R' \neq \emptyset$. Since both R' and S' are nonvoid, clearly $\square$ depends on its first variable as well.

Put $\Pi' := \Pi_H$ and $\sigma' := \sigma_H$. From (63) we see that $\{r'_a : a \in \boldsymbol{k}\} \subseteq \{r_a : a \in \boldsymbol{k}\}$ and so $\Pi' \subseteq \Pi$. From $|\Pi'| \leq |\Pi|$ and the optimality of G we have $|\Pi'| = |\Pi|$ and hence $\Pi' = \Pi$. We need the next claim (where $g^{-d}(S) := \{x \in \boldsymbol{k} : g^d(x) \in S\}$).

Claim 2. 1) *There is a map $\varphi : \boldsymbol{k} \to$ im g^d such that $r_x = r_{\varphi(x)}$ for all $x \in \boldsymbol{k}$;* 2) $g^{-d}(S) \subseteq S$; 3) *if $s \in S$ and $p \in \boldsymbol{k}$ satisfy $r_s = r_p$, then $g^2(s) = g^2(p)$, and* 4) $d = 1$.

Proof. 1) Let $x \in \boldsymbol{k}$. From $r_x \in \Pi = \Pi'$ we see that $r_x = r'_q$ for some $q \in \boldsymbol{k}$. Setting $p := g^d(q)$ by (63) we have $r'_q = r_p$ and so it suffices to set $\varphi(x) := p$.

2) Let $c \in g^{-d}(S)$ and suppose to the contrary that c is regular. Applying 1) we have $r_c = r_p$, where $p = \varphi(c) = g^d(q)$ for some $q \in \boldsymbol{k}$. From $r_c = r_p$ and the fact that they are constant we see that

$$g(c) = r_c(c) = r_p(c) = r_p(p) = g(p).$$

Since $c \in g^{-d}(S)$, we have $t := g^d(c) \in S$. Finally $t \in S$,

$$t = g^d(c) = g^{d-1}(g(c)) = g^{d-1}(g(p)) = g^d(p) = g^{2d}(q)$$

and $d > 0$ contradict the maximality of d (postulated at the beginning of the proof).

3) Let $r_s = r_p$ for some $s \in S$ and $p \in \boldsymbol{k}$. Then clearly $I :=$ im $r_s =$ im r_p, and from the tautness of G (cf. (60)) we have

$$I \subseteq g^{-1}(g^2(s)) \cap g^{-1}(g^2(p)).$$

It is immediate that for $a, b \in \boldsymbol{k}$ the sets $g^{-1}(a)$ and $g^{-1}(b)$ intersect if and only if $a = b$ and therefore $g^2(s) = g^2(p)$.

4) Suppose to the contrary that $d \geq 2$. Choose $c \in g^{-d}(S)$. Then c is singular by 2), and $r_c = r_p$ for some $p = g^d(q)$ by 1). Clearly $p \in S$ and from $r_c = r_p$ and by 3) we obtain $g^2(c) = g^2(p)$. Finally $g^d(c) \in S, d \geq 2$, and

$$g^d(c) = g^{d-2}(g^2(c)) = g^{d-2}(g^2(p)) = g^d(p) = g^{2d}(q)$$

which again contradicts the maximality of d. □

For every $x \in \boldsymbol{k}$ denote by $\tau(x)$ the least integer m such that $g^m(x) \in \langle g \rangle$, that is, the "distance" of x from its cycle. Further put $t := \max\{\tau(s) : s \in S\}$ and $V_i := \{x \in \boldsymbol{k} : \tau(x) = i\}$ for all $i = 0, \dots, t$. We need:

Claim 3. (i) $t = 2$, *and* (ii) $V_2 \subseteq S$.

Proof. Let $s \in S$ satisfy $\tau(s) = t$. Put $v := g(s), u := g^2(s)$ and $p := \varphi(s)$ (where φ is the map from Claim 2). Applying Claim 2,3) we obtain $g^2(p) = g^2(s) = u$. By Claim 2,4) we have $d = 1$ and so by Claim 2,1) we have $p = \varphi(s) \in$ im g, and hence $p = g(q)$ for some $q \in \boldsymbol{k}$. In view of $s \in S$ and $r_p = r_s$ clearly $p \in S$ and thus $q \in g^{-1}(S) \subseteq S$ by Claim 2,2) and 3). We assume that $\langle g \rangle$ and S are disjoint, and so $s, p \notin \langle g \rangle$. We have $t \geq \tau(q) \geq 2$ and therefore $v \notin \langle g \rangle$. The assumption $u \notin \langle g \rangle$ leads to the contradiction $\tau(q) = t + 1$ and so $u \in \langle g \rangle$.

Finally, we prove (ii). Suppose V_2 contains a regular element x_0. Put $x_1 := \varphi(x_0)$. By Claim 2,1) and 4) we have $x_1 \in R \cap$ im g. Moreover, $r_{x_0} = r_{x_1}$ and r_{x_0} is a constant map; consequently

$$g(x_0) = r_{x_0}(x_0) = r_{x_1}(x_0) = r_{x_1}(x_1) = g(x_1).$$

This and $x_0 \in V_2$ show that $x_1 \in V_2$. By induction on $n = 1, 2, \dots$ we construct $x_n \in R \cap V_{n+1} \cap \text{im } \varphi$ such that $g^n(x_n) = g(x_1)$. Clearly x_1 meets the condition. Suppose $n \geq 1$ and we have found $x_n \in R \cap V_{n+1} \cap \text{im } \varphi$ such that $g^n(x_n) = g(x_1)$. Due to $x_n \in \text{im } \varphi \subseteq \text{im } g$ we have $x_n = g(a)$ for some $a \in \boldsymbol{k}$. Clearly $a \in V_{n+2}$ and so a is regular since $a \in S$ would contradict $t = 2$. Put $x_{n+1} := \varphi(a)$. As above, we have $x_{n+1} \in R$ and $g(x_{n+1}) = g(a) = x_n$, and therefore

$$g^{n+1}(x_{n+1}) = g^n(g(x_{n+1})) = g^n(x_n) = g(x_1).$$

Clearly $x_{n+1} \in V_{n+2} \cap$ im φ. This concludes the induction step. Since $x_n \in V_{n+1}$ for all $n = 1, 2, \dots$, the elements $x_1, x_2, \dots$ are pairwise distinct elements of $\boldsymbol{k}$. This contradiction shows that $V_2 \subseteq S$, proving (ii). □

Claim 3, (i) and (ii) shows that $\emptyset \neq V_2 \subseteq S$. By Claim 3 (i) we have $V_n \subseteq R$ for all $n > 2$. In view of Claim 2,2) and 4) we have $g^{-1}(S) \subseteq S$ and therefore V_n is void for all $n > 2$. It follows that $\langle g \rangle \subseteq R \subseteq \langle g \rangle \cup V_1$. As $d = 1$ the set $V_1 \cap S$ in nonempty. We show that $R' \supseteq \langle g \rangle \cup V_1$. Indeed, let $x \in \langle g \rangle \cup V_1$. Then $r'_x = r_{g(x)}$, where $g(x) \in \langle g \rangle \subseteq R$. We obtain

$$|R'| \geq l + |V_1| > |R|$$

in contradiction to the optimality of G. Thus $S \cap \langle g \rangle = \emptyset$ does not hold for $d > 0$.

Now we consider:

Case 2 Let $T := S \cap \langle g \rangle \neq \emptyset$. We distinguish two subcases.

Subcase 2.a. Let $l > 1$. Denote by j the least integer such that g^j is reflective and put $G' := \langle \boldsymbol{k}; \square \rangle$, where $x \square y :\approx x \circ_j y \approx g^j(x)y$. Denote by r'_a the row-maps of $\square$ and put $g'(x) :\approx x \square x$. We show:

Claim 4. The groupoid G' is essential and the equivalence $\Theta := \ker g^j$ is nontrivial and satisfies

$$r'_a = r'_b \Leftrightarrow (a, b) \in \Theta. \tag{64}$$

Proof. We prove (64). According to (63) we have $r'_a = r'_b$ if and only if $r_c = r_d$ where $c := g^j(a)$ and $d := g^j(b)$. The implication $\Leftarrow$ is trivial. To prove $\Rightarrow$ consider $(a, b) \in \boldsymbol{k}^2 \backslash \Theta$ and put $c := g^j(a)$ and $d := g^j(b)$. We must prove that $r_c \neq r_d$. Notice that $c, d \in \boldsymbol{l}$. We have 2 cases.

1) Let $\{c, d\} \cap T$ be nonempty. By symmetry we may assume $c \in T$. From (60) we obtain im $r_c \subseteq g^{-1}(g^2(c))$. Now c being singular, clearly im $r_c \supset \{g(c)\}$. Taking into account that $\boldsymbol{l}$ consists of the cycles of g, we can choose $u \in$ im $r \backslash \boldsymbol{l}$. Clearly $g(u) = g^2(c)$. From $(a, b) \notin \Theta$ we have $c \neq d$. As $d \in \boldsymbol{l}$ and im $r_d \subseteq g^{-1}(g^2(d))$, it is easy to see that $u \notin$ im r_d and therefore $r_c \neq r_d$.

2) Let $c, d \in R$. Now r_c and r_d are constant with values $g(c)$ and $g(d)$. Here c and d are distinct elements of $\boldsymbol{l}$, and g permutes $\boldsymbol{l}$; consequently $g(c) \neq g(d)$ and $r_c \neq r_d$. We have proved (64).

We show that G' is essential. The restriction of g^j to $\boldsymbol{l}$ is the identity and so from (64) the maps $\{r'_a : a \in \boldsymbol{l}\}$ are all distinct. In view of $l > 1$ the groupoid G' has at least 2 distinct rows and so it depends on its 1st variable. Moreover, $T \neq \emptyset$ and $|\text{im } r_t| > 1$ for all $t \in T$, and so G' depends on its second variable as well.

Put $\Pi' := \{r'_x : x \in \boldsymbol{k}\}$. Clearly $\Pi' \subseteq \Pi$. From optimality we obtain $|\Pi'| = |\Pi|$ and so $\Pi' = \Pi$. This, Claim 4 and the fact that $h := g^j$ is the identity on $\boldsymbol{l}$ shows that

$$\Pi = \{r_x : x \in \boldsymbol{l}\}. \tag{65}$$

We need:

Claim 5. (i) im $g^2 = \boldsymbol{l}$, *and* (ii) *if $x \in \boldsymbol{l}$ and $y, z \in g^{-2}(x)$, then $r_y = r_z$.*

Proof. Let $y \in \boldsymbol{k}$. According to (63) we have $r_y = r_t$ for some $t \in \boldsymbol{l}$. As G is taut, we have

$$\{g^2(y)\} = g(\text{im } r_y) = g(\text{im } r_t) = \{g^2(t)\},$$

proving $g^2(y) = g^2(t)$. Thus $g^2(y) \in \boldsymbol{l}$ proving (i). In the situation of (ii) we have $r_y = r_t = v_z$.

Now it is easy to see that the condition (ii) of the theorem holds.

Subcase 2.b. We turn to the case $l = 1$. Now $\langle g \rangle = \mathbf{1} = \{0\}$. Since $T = \langle g \rangle \cap S$ is nonempty, we have $T = \{0\}$ and so 0 is singular. For $i \geq 0$ put

$$L_i := g^{-i}(0) \backslash g^{1-i}(0),$$

and denote by d the greatest integer h such that L_h is nonempty. Further, denote by j the greatest integer such that $r_b \neq r_0$ for some $b \in L_{d-j}$. Such j exists because G depends on its 1st variable. Put $x \Box y :\approx x \circ_j y$. The row-maps r'_x of $\Box$ satisfy $r'_0 = r_0 \neq r_b = r'_a$ for some $a \in L_d$, and so in view of $0 \in S$ the groupoid $\langle \boldsymbol{k}; \Box \rangle$ is essential. As before we have $\Pi = \Pi'$. Consider $b \in L_d$. We have $r_b = r'_e = r_c$ for some $e \in \boldsymbol{k}$ and $c = g^j(e)$. By (60) we have

$$\text{im } r_b \subseteq g^{-1}(g^2(b)), \quad \text{im } r_c \subseteq g^{-1}(g^{2+j}(e));$$

hence $e \in L_{d+j}$ and so $j = 0$.

Let $x \in L_{d-1}$. We have $r_x = r_0$ and from (60)

$$L_0 \cup L_1 \supseteq \text{ im } r_0 = \text{ im } r_x \subseteq L_{d-2}$$

which implies $d - 2 \leq 1$ and $d \leq 3$. This completes the proof that G satisfies (iii). □

We conclude with a proposition which further restricts our groupoids.

3.6 Proposition *Let G be optimal and satisfy one of the conditions of Proposition 3.5. Then without loss of generality we may assume that moreover either (α) r_a^2 is constant for every $a \in \boldsymbol{k}$ or (β) r_a is reflective for all $a \in \boldsymbol{k}$.*

Proof For every $i > 0$ put $x \cdot_i y :\approx r_x^i(y)$ and $H_i := \langle \boldsymbol{k}; \cdot_i \rangle$. We need:

Claim 1. If $H := H_i$ is essential, then H is taut and

$$|\Pi_H| = |\Pi_G|, \quad |R_H| = |R_G|, \quad \sigma_H = \sigma_G. \tag{66}$$

Proof. For $a \in \boldsymbol{k}$ put $r'_a := r_a^H$. From $r'_a = r_a^i$ it follows that $|\Pi_H| \leq |\Pi_G|$. If a is regular, then $r_a^i = r_a$ and therefore $|R_H| \geq |R_G|$. Next $\sigma_H \leq \sigma_G$ because im $r_a^i \subseteq$ im r_a for all $a \in \boldsymbol{k}$. Now (66) follows from the optimality of G. Moreover, H is taut by Lemma 3.2.

We have two cases:

1) H_j is nonessential for some $j > 1$, and 2) H_i is essential for all $i = 2, 3, \ldots$. We start with the case 1. Denote by i the greatest integer such that H_i is essential, denote by $H = \langle \boldsymbol{k}; \Box \rangle$ the groupoid H_i and abbreviate r_a^H and g^H by r'_a and g. We need:

Claim 2. r_a^{i+1} is constant for all $a \in \boldsymbol{k}$.

Proof. We must show that H_{i+1} does not depend on its second variable. Suppose to the contrary that it does. As H_{i+1} is nonessential, it does not depend on its first variable;

consequently there exists a selfmap r of $\boldsymbol{k}$ such that $r_a^{i+1} = r$ for all $a \in \boldsymbol{k}$. By (60) clearly

$$\operatorname{im} r = \operatorname{im} r_a^{i+1} \subseteq \operatorname{im} r_a \subseteq g^{-1}(g^2(a)) \tag{67}$$

holds for all $a \in \boldsymbol{k}$. Suppose that R is nonempty and choose $a \in R$. Then r_a^{i+1} is the constant selfmap γ_t of $\boldsymbol{k}$ with the value $t := g(a)$ and so

$$r = \gamma_t = r_x^{i+1}$$

for all $x \in \boldsymbol{k}$ contrary to our assumption that H_{i+1} depends on its second variable. Thus $R = \emptyset$. Then G is of type (ii) or (iii) from Proposition 3.5.

Suppose G is of type (ii). Choose distinct $a_0, a_1 \in \boldsymbol{l}$. Then for $j = 0, 1$ from (67) we have

$$\operatorname{im} r \subseteq g^{-1}(g^2(a_j))$$

leading to the contradiction

$$\operatorname{im} r \subseteq g^{-1}(g^2(a_0)) \cap g^{-1}(g^2(a_1)) = \emptyset.$$

Thus let G be of type (iii). From (67) we see that the corresponding number d is at most 2. In view of $g(L_0 \cup L_1) = \{0\}$ and (67) we have $r^2 = \gamma_0$ and so $r^{i+1} = \gamma_0$, again in contradiction to our assumption that H_{i+1} depends on its second variable.

Now for all $a \in \boldsymbol{k}$ the map r'_a satisfies $r'^2_a = r_a^{2i}$. As $2i > i+1$, from Claim 1 we see that r'^2_a is constant for all $a \in \boldsymbol{k}$. A direct check shows that H satisfies the same condition from Proposition 3.5 as G.

If we replace G by H we obtain the case (α).

2) Thus assume that H_i is essential for all $i = 1, 2, \ldots$. Denote by i the least integer such that $r_0^i, \ldots, r_{b-1}^i$ are all reflective. By Claim 1 the groupoid H_i is optimal, and as above satisfies the same condition from Proposition 3.5 as G. If we replace G by H_i we obtain the case (β). □

References

[Cs83] Csákány, B., All minimal clones on the three-element set, *Acta Cybernet.* **6** (1983), 227–238.

[Ma 81] Machida, H., Essentially minimal closed sets in multiple-valued logic, *Trans. IECE Japan E* **64** (1981) no. 4, 243–245.

[Ma 82a] Machida, H., A theorem on essential minimality in k-valued logic, *Trans. IECE Japan E* **65** (1982), 123–124.

[Ma 82b] Machida, H., Toward a classification of minimal closed sets in 3-valued logics, in: *Proc. 12th Internat. Symposium on Multiple-Valued Logic, Paris 1982*, IEEE Computer Soc. Publ. Office, 1982, 313–317.

[MR 84] Machida, H., Rosenberg, I.G., Classifying essentially minimal clones, in: *Proc. 14th Internat. Symposium on Multiple-Valued Logic, Winnipeg 1984*, IEEE Computer Soc. Publ. Office, 1984, 4–7.

[MR 85] Machida, H., Rosenberg, I.G., Essentially minimal groupoids, in *Proc. 15th Internat. Symposium on Multiple-Valued Logic, Kingston, Ont., May 1985*, IEEE Computer Soc. Publ. Office 1985, 338–344.

[MR 92] Machida, H., Rosenberg, I.G., A "large" essentially minimal clone over an infinite set, *Contemp. Math.* **131** (1992) part 3, 159–167.

[Qu 91] Quackenbush, R.W., A survey of minimal clones, Preprint, University of Manitoba 1991, 15 pp. To appear in *Aequationes Math.*

Functional and Affine Completeness and Arithmetical Varieties

Alden F. PIXLEY
Department of Mathematics
Harvey Mudd College
Claremont, CA 91711
U.S.A.

Abstract

Our understanding of affine complete and functionally complete algebras and varieties of affine complete algebras is closely related to our understanding of arithmetical algebras and varieties. Because of this we first survey basic results about arithmetical algebras and varieties, emphasizing finite algebras and finitely generated varieties. The main outlines of the theory of congruence distributive affine complete varieties are discussed and then attention is focused on arithmetical affine complete varieties, emphasizing some recent results describing their structure. In particular we shall examine properties of finite, arithmetical, affine complete algebras having no proper subalgebras (FACS algebras) since any affine complete arithmetical variety of finite type is generated by such an algebra. We shall present some interesting sufficient conditions for a FACS algebra to generate an arithmetical (and hence affine complete) variety.

1 Introduction

The cluster of ideas we shall discuss in these lectures have their origin in the study of Boolean algebras. Ever since universal algebra first evolved as an independent branch of algebra the variety of Boolean algebras has somehow always played a conspicuous role. This is partly because Boolean algebras (because of their connection with set theory and logic) often play a role in seemingly unrelated areas; the ultraproduct construction and the generally pervasive role of the prime ideal theorem are among the more obvious illustrations of this way in which Boolean algebras influence universal algebra. On the other hand Boolean algebras have also played an important role because the two element Boolean algebra, the generator of the variety of Boolean algebras, is the simplest example of a primal algebra. Primality, partly because of its interest to both logicians and computer scientists, but also, and not unimportantly, because of its concrete character, was an early popular area of study

I. G. Rosenberg and G. Sabidussi (eds.), Algebras and Orders, 317–357.

in universal algebra. The same can be said about the various algebraic concepts which generalize primality: functional completeness, affine completeness, quasi-primality, etc.

Now the defining property of primal algebras—all functions are term functions—while it has the advantage of concreteness, bears little connection with the ordinary basic concepts of abstract algebra. For this reason an early important discovery [3] was that finite primal algebras could also be characterized in terms of algebraic concepts which made obvious connections with other areas of study in universal algebra. This characterization is the following: a finite algebra is primal iff it is simple, has no proper subalgebras, has no proper automorphisms (i.e.: is rigid), and generates an arithmetical variety. Since arithmeticity is the conjunction of congruence distributivity and congruence permutability, and all of these are Mal'cev characterizable, the concept of primality is thus placed at the intersection of several mainstream concepts of universal algebra. Also it is interesting that congruence distributivity and permutability each makes its contribution to the theory in a conspicuous way: congruence distributivity implies that the variety generated by a primal algebra **A** consists of subdirect powers of **A**. Permutability implies that the finite members of the variety are direct powers of **A**. This is somewhat analogous to the facts that every distributive lattice is a subdirect power of **2** (from the fact that lattices are congruence distributive) and a distributive lattice is relatively complemented iff it is congruence permutable. In the generalizations of primality which we shall discuss, congruence distributivity and permutability make their contributions to the theory in similarly conspicuous ways.

In these lectures, for the reasons just mentioned, we begin in Section 2 by developing certain aspects of arithmetical lattices of equivalence relations, arithmetical algebras, and arithmetical varieties. The most noteworthy fact here, Theorem 2.2, and the real underlying determinant of the most remarkable features of the whole theory, is that (with suitable finiteness conditions) all three of these classes can be described, formally, in the same way. In Section 3 we introduce affine complete algebras and make important and remarkable links between affine completeness and arithmeticity. In Section 4 we develop the main features (to the extent that they are known) of the general theory of affine completeness. In particular we observe that at least for finitely generated affine complete varieties, the arithmetical case is much more transparent than the congruence distributive but non-permutable case. In Section 5 we present some recent results concerning arithmetical affine complete varieties; and in particular describe an interesting, but by no means exhaustive, class of affine complete arithmetical varieties. Section 6 presents two illuminating counter-examples.

It is a pleasure to acknowledge the help of Nikolai Weaver in preparing these lecture notes.

2 Arithmeticity

A sublattice **L** of the lattice **Eqv**A of all equivalence relations on the set A is *arithmetical* if it is distributive and permutable, the latter condition meaning that the join of two relations is given by their relation product. These two conditions can be expressed jointly by the single equality:

$$\theta \wedge (\phi \circ \psi) = (\theta \wedge \psi) \circ (\theta \wedge \phi)$$

for all $\theta, \phi, \psi \in L$. ($\circ$ = product.) Arithmeticity can also be expressed by just the inequality

$$\theta \wedge (\phi \circ \psi) \leq (\theta \wedge \psi) \circ (\theta \wedge \phi),$$

and we shall usually make use of this fact. To see that this is so substitute $\phi \circ \psi$ for θ in the inequality and infer $\phi \circ \psi \leq \psi \circ \phi$, and thus permutability. Hence $\circ = \vee$. From this it follows that the reverse inequality holds, and hence the equality.

The equation describing arithmeticity is an example of a *congruence equality*, by which we mean, in general, an equality $e = f$ where e and f are terms in variables and binary operation symbols, $\vee$, $\wedge$, and $\circ$. $e = f$ holds in a sublattice $\mathbf{L}$ of $\mathbf{Eqv}A$ if $e = f$ is true for every interpretation of the variables by members of $\mathbf{L}$ and where the operations are interpreted as join, meet, and relation product respectively. $e = f$ holds in an algebra $\mathbf{A}$ if it holds in Con$\mathbf{A}$. Arithmeticity is significant among congruence equations in that it is stronger than (i.e.: logically implies) every other non-trivial congruence equation. This is clear since if $e = f$ is a non-trivial congruence equation and $\mathbf{L}$ is arithmetical then $\mathbf{L}$ is distributive and hence every non-trivial lattice identity holds in $\mathbf{L}$ and, in particular the lattice identity obtained from $e = f$ by replacing $\circ$ by $\vee$. But then since $\circ = \vee$ in $\mathbf{L}$, it follows that $e = f$ in $\mathbf{L}$. Because of the strength of arithmeticity as a congruence equation we may anticipate some of the remarkable properties of arithmeticity which we shall describe below.

An alternative familiar characterization of arithmeticity is the following: $\mathbf{L}$ is arithmetical iff it satisfies the *Chinese remainder condition*:

For each finite set $\theta_1, \ldots, \theta_n$ of equivalence relations in L, and elements $a_1, \ldots, a_n$ in A, the system

$$x \equiv a_i \ (\theta_i) \quad i = 1, \ldots, n$$

is solvable iff for all $1 \leq i \leq j \leq n$,

$$a_i \equiv a_j \ (\theta_i \vee \theta_j).$$

The proof that arithmeticity is characterized by the Chinese remainder condition is an easy exercise.

An algebra $\mathbf{A}$ is arithmetical if its congruence lattice $\mathbf{ConA}$ is arithmetical, while a variety is arithmetical if each of its members is arithmetical. Arithmeticity of varieties is a "strong" Mal'cev condition and is characterized as follows:

Theorem 2.1 ([15]) *A variety V is arithmetical iff for some ternary term t(x,y,z) the Mal'cev type equations*

$$t(x, x, z) = z, \ t(x, y, x) = x, \ t(x, z, z) = x, \tag{1}$$

hold in V.

Proof If V is arithmetical let $\mathbf{F}$ be the V–free algebra with free generators x, y, z. Let $\theta(x,z), \theta(x,y), \theta(y,z)$ be the principal congruences of $\mathbf{F}$ which collapse the three pairs of generators. Then

$$(x,z) \in \theta(x,z) \wedge (\theta(x,y) \circ \theta(y,z))$$

so for some element (term) $t(x,y,z)$ of F, we must have

$$(x, t(x,y,z)) \in \theta(x,z) \wedge \theta(y,z) \text{ and } (t(x,y,z), z) \in \theta(x,z) \wedge \theta(x,y)$$

from which we infer that the equations (1) hold on the generators of $\mathbf{F}$ and thus are equations of V.

Conversely, if $t(x,y,z)$ is a term of V satisfying (1), let $\mathbf{A} \in V$. Suppose $\theta, \phi, \psi \in \mathrm{Con}\mathbf{A}$ and $(x,z) \in \theta \wedge (\phi \circ \psi)$ so that $(x,z) \in \theta$ and for some $y \in A$, $(x,y) \in \phi$ and $(y,z) \in \psi$. Then, applying equations (1), we have

$$x = t(x,y,x)\, \theta\, t(x,y,z)\, \theta\, t(z,y,z) = z, \text{ and}$$

$$x = t(x,y,y)\, \psi\, t(x,y,z)\, \phi\, t(y,y,z) = z,$$

so that $(x,z) \in (\theta \wedge \psi) \circ (\theta \wedge \phi)$. Hence $\mathbf{A}$ is arithmetical.

An example of Theorem 2.1 is the variety of Boolean algebras with

$$t(x,y,z) = (x \vee y') \wedge (x \vee z) \wedge (y' \vee z)$$

($'$ = complement). Another is the variety generated by the Galois field $GF(p^k)$ with operations $+$ and $\times$ and

$$t(x,y,z) = z + (x + (p-1)z)(x + (p-1)y)^{p^k - 1}.$$

Finally let V be the variety generated by a finite set of Galois fields. Let $p_1, \dots, p_n$ be the distinct characteristics of the fields and let $k_1, \dots, k_n$ be the least integers such that for each $i = 1, \dots, n$, each of the fields of characteristic p_i is contained in $GF(p_i^{k_i})$. Also let $t_i(x,y,z)$ be constructed as above for this field. Let a_i be a solution of the congruence

$$p_1 \cdots p_{i-1} p_{i+1} \cdots p_n x \equiv 1 \ \ (p_i).$$

Then

$$t(x,y,z) = \sum_{i=1}^{n} p_1 \cdots p_{i-1} p_{i+1} \cdots p_n a_i t_i(x,y,z)$$

satisfies (1) in V. This proves the "if" half of the following important theorem due to Michler and Wille [13]: *A variety of rings is arithmetical iff it is generated by a finite number of finite fields.* The other half of this result apparently requires quite deep but well known facts about ring theory.

One of the most interesting facts about arithmetical lattices of equivalence relations is that such lattices, if complete and if the underlying set A is countable, can be characterized by the property that they always support a compatible function satisfying the equations (1) of Theorem 2.1, that is by the same equations which characterize arithmetical *varieties*! In these lectures we will need this fact only for the finite case, which we establish below (Theorem 2.2).

A function is *L–compatible* if it has the substitution property with respect to members of L. It is convenient to describe this in terms of principal equivalence relations in L (if they exist; for example if **L** is complete.) Denote elements of A^k by boldface; $\mathbf{x} = (x_1, \ldots, x_k)$. If $\mathbf{x}, \mathbf{y} \in A^k$ let $\theta(\mathbf{x}, \mathbf{y})$ denote $\theta(x_1, y_1) \vee \cdots \vee \theta(x_k, y_k)$. Then for $X \subset A^k$, $f{:}X \to A$ is L–compatible means

$$\mathbf{x}, \mathbf{y} \in X \Longrightarrow (f(\mathbf{x}), f(\mathbf{y})) \in \theta(\mathbf{x}, \mathbf{y}).$$

Recall that by definition

$$(f(\mathbf{x}), f(\mathbf{y})) = f((x_1, y_1), \ldots, (x_k, y_k)),$$

the "componentwise extension" of f from a function mapping a subset $X \subset A^k$ to A to a function mapping the subset $\{(x_1, y_1), \ldots, (x_k, y_k) : \mathbf{x}, \mathbf{y} \in X\} \subset (A \times A)^k$ to $A \times A$. Sometimes this extension is denoted by $f \times f$. With this understanding, another way to describe L–compatibility of f is simply to require that each element in L be closed under the componentwise extension of f.

Theorem 2.2 ([16]) *For any nonempty set A, if* **L** *is a finite* 0-1 *sublattice of* **Eqv**A*, then* **L** *is arithmetical iff there is a function $f{:}A^3 \to A$ which is L–compatible and which satisfies equations* (1) *of Theorem* 2.1.

Proof $\Rightarrow$ For $\phi \in L$ let $h[\phi, 1]$ be the height of the interval $[\phi, 1]$. For $n \geq 1$ let $P(n)$ be the statement:

> *For each $\phi \in L$ with $h[\phi, 1] \leq n$ there exists a function $f_\phi{:}(A/\phi)^3 \to (A/\phi)$ satisfying* (1) *on A/ϕ and for which*
>
> $$\phi \leq \theta \Longrightarrow f_\phi(x/\phi, y/\phi, z/\phi) \subset f_\theta(x/\theta, y/\theta, z/\theta)$$
>
> *for all $x, y, z \in A$.*

For $n = 1$ (meaning ϕ is maximal) define f_ϕ as the discriminator:

$$\begin{aligned} f_\phi(x/\phi, y/\phi, z/\phi) &= z/\phi \text{ if } x/\phi = y/\phi \\ &= x/\phi \text{ otherwise.} \end{aligned}$$

This establishes $P(1)$.

Assume $P(n)$ and let $h[\phi, 1] = n + 1$. In case any pair of $x/\phi, y/\phi, z/\phi$ are equal define $f_\phi(x/\phi, y/\phi, z/\phi)$ to satisfy equations (1). Otherwise suppose $\phi_1, \ldots, \phi_k$ are all of the elements of L which cover ϕ. If $k = 1$ define $f_\phi(x/\phi, y/\phi, z/\phi)$ to be any ϕ-class contained in $f_{\phi_1}(x/\phi_1, y/\phi_1, z/\phi_1)$. In case $k > 1$, for each $x, y, z \in A$ pick $w_i \in f_{\phi_i}(x/\phi_i, y/\phi_i, z/\phi_i)$.

From $P(n)$ it follows that $w_i \equiv w_j \quad (\phi_i \vee \phi_j)$ for $1 \leq i < j \leq k$. Hence by the Chinese remainder condition there is a $w \in A$ such that $w \equiv w_i \quad (\phi_i)$, for $i = 1, \dots, k$, and thus

$$w/\phi \subset w/\phi_i = f_{\phi_i}(x/\phi_i, y/\phi_i, z/\phi_i)$$

for $i = 1, \dots, k$. Since the ϕ_i cover ϕ and $k > 1$, w/ϕ is uniquely determined so we define $f_\phi(x/\phi, y/\phi, z/\phi) = w/\phi$. Repeating this construction for all ϕ with $h[\phi, 1] = n+1$ we establish $P(n+1)$ and, by induction, $P(n)$ for all $n \geq 1$. Then define $f(x, y, z)$ to be the sole element in $f_0(x/0, y/0, z/0)$. Obviously f satisfies equations (1).

To check the compatibility of f, suppose $\phi \in L$ and the pairs $(x, x\prime), (y, y\prime), (z, z\prime) \in \phi$. Then $f_\phi(x/\phi, y/\phi, z/\phi) = f_\phi(x\prime/\phi, y\prime/\phi, z\prime/\phi)$ and by the inclusion condition of $P(h[0,1])$, each of $f(x, y, z)$ and $f(x\prime, y\prime, z\prime)$ is in this ϕ–class.

$\Leftarrow$ Suppose $f{:}A^3 \to A$ is L–compatible and satisfies equations (1). Suppose $\theta, \phi, \psi \in L$ and $(x, z) \in \theta \wedge (\phi \circ \psi)$ so that $(x, z) \in \theta$ and for some $y \in A$, $(x, y) \in \phi$ and $(y, z) \in \psi$. Then, applying equations (1), we have

$$x = f(x, y, x)\, \theta\, f(x, y, z)\, \theta\, f(z, y, z) = z, \text{ and}$$

$$x = f(x, y, y)\, \psi\, f(x, y, z)\, \phi\, f(y, y, z) = z,$$

so that $(x, z) \in (\theta \wedge \psi) \circ (\theta \wedge \phi)$. Hence **L** is arithmetical. Notice that we have formally copied the second half of the proof of Theorem 2.1 (with f replacing t).

For algebras the theorem above has the following immediate consequence. It can be extended to countable algebras but in these lectures we shall only need the finite case.

Theorem 2.3 *A finite algebra* **A** *is arithmetical iff there exists a Con***A***–compatible function $f{:}A^3 \to A$ satisfying equations* (1).

Another useful characterization of arithmetical equivalence lattices is the following due to Kaarli [8]. See [4] for this and related results in a more general setting.

Theorem 2.4 *For any nonempty set A, if* **L** *is a complete sublattice of* **Eqv***A, then* **L** *is arithmetical iff the following condition* (*the compatible function extension property*) *holds:*

> *For any positive integer k and finite subsets $X, Y \subset A^k$, with $X \subset Y$, any L-compatible function $f{:}X \to A$ has an L-compatible extension from Y to A.*

Proof $\Leftarrow$ Suppose, as in the $\Leftarrow$ direction of the proof of Theorem 2.2, that the pair $(x, z) \in \theta \wedge (\phi \circ \psi)$ so that $(x, z) \in \theta$, $(x, y) \in \phi$, $(y, z) \in \psi$. Let

$$X = \{(x, y, x), (x, y, y), (z, y, z), (y, y, z)\} \subset A^3$$

and define $f{:}X{\to}A$ by assigning value x for the first two triples in X and z for the last two. To check the compatibility of f first notice that if $\mathbf{a} = (x,y,y)$ and $\mathbf{b} = (y,y,z)$, then transitivity implies that $(f(\mathbf{a}), f(\mathbf{b})) = (x,z)$ is an element of any congruence which contains the pairs $(x,y), (y,y), (y,z)$. The other five cases are immediate since in each of these the value of $(f(\mathbf{a}), f(\mathbf{b})) = (f((a_1,b_1),(a_2,b_2),(a_3,b_3))$ is one of (a_i,b_i). Hence f can be extended compatibly to $X \cup \{(x,y,z)\}$. Then the same argument as in the second half of Theorem 2.1 and again in the second half of Theorem 2.2 applies to show that $(x,z) \in (\theta\wedge\psi)\circ(\theta\wedge\phi)$. For use below it is worth noticing that we have actually proved the following

Lemma 2.1 *If* $\mathbf{L}$ *is any sublattice of* $\mathbf{Eqv}A$ *and if for any set* $X \subset A^3$ *of the form* $X = \{(x,y,x),(x,y,y),(z,y,z),(y,y,z)\} \subset A^3$ *the (*$\mathbf{L}$*-compatible) function* $f{:}X{\to}A$, $f(x,y,x) = f(x,y,y) = x$, *and* $f(z,y,z) = f(y,y,z) = z$, *can be* $\mathbf{L}$*-compatibly extended to* $X \cup \{(x,y,z)\}$, *then* $\mathbf{L}$ *is arithmetical.*

$\Rightarrow$ Obviously we need only consider the case $X = \{\mathbf{x}^1,\ldots,\mathbf{x}^n\}$ and $Y = X \cup \{\mathbf{y}\}$, i.e.: addition of a single element to the domain of f. If $n = 1$ then defining $f(\mathbf{y}) = f(\mathbf{x}^1)$ is clearly a compatible extension. For $n \geq 2$ we can extend f compatibly by defining $f(\mathbf{y}) = w$ where w is a solution to the system

$$w \equiv f(\mathbf{x}^i)\ \ (\theta(\mathbf{x}^i,\mathbf{y}))\ \ i = 1,\ldots,n.$$

By the Chinese remainder condition the system is solvable if

$$f(\mathbf{x}^i) \equiv f(\mathbf{x}^j)\ \ (\theta(\mathbf{x}^i,\mathbf{y}) \vee \theta(\mathbf{x}^j,\mathbf{y})).$$

But compatibility implies $f(\mathbf{x}^i) \equiv f(\mathbf{x}^j)\ (\theta(\mathbf{x}^i,\mathbf{x}^j))$ and

$$\theta(\mathbf{x}^i,\mathbf{x}^j) \leq \theta(\mathbf{x}^i,\mathbf{y}) \vee \theta(\mathbf{x}^j,\mathbf{y})$$

so the extension is possible. Notice that only in this direction of the proof do we require $\mathbf{L}$ to be complete.

Remarks The observant reader may notice that we can obtain Theorem 2.2, at least for A finite, directly from Theorem 2.4. Indeed, if $\mathbf{L}$ is arithmetical, let $X \subset A^3$ be the set of all triples of the forms $(x,y,x), (x,x,y)$, or (x,y,y), and define $f{:}X \to A$ to satisfy equations (1) of Theorem 2.1. As in the proof of Theorem 2.4 it is easy to check that f is L-compatible. Hence f can be L-compatibly extended to $Y = A^3$ by Theorem 2.4. Notice that this proof is not only simpler than that of Theorem 2.2 but is also different in approach: here we partially define f and then extend L-compatibly to all of A^3. On the other hand, in the first proof of Theorem 2.2 we define each of the f_ϕ individually so as to insure compatibility. Also notice that in the first proof of Theorem 2.2 we satisfy (1) for the maximal ϕ by defining f_ϕ to be the discriminator, and then for successively lower ϕ in L, the equations (1) are inherited by f_ϕ from above, if any pair of $x/\phi, y/\phi, z/\phi$ are equal.

An advantage of this (first) approach is that we are able to conclude not only that equations (1) are satisfied, but also that principal equivalence relations are definable by f.

Specifically, for any $\phi \in L$, let $\mathbf{L}_\phi$ be the sublattice of $\mathbf{Eqv}A$ which is naturally isomorphic to the interval $[\phi, 1]$ of $\mathbf{L}$. Then by a slight elaboration of the proof of Theorem 2.2 we can prove the following statement:

For any $\phi \in L$ and $u, v, x, y \in A$,

$$x/\phi \equiv y/\phi \;\; \theta(u/\phi, v/\phi) \text{ in } \mathbf{L}_\phi \iff f_\phi(u/\phi, v/\phi, x/\phi) = f_\phi(u/\phi, v/\phi, y/\phi).$$

In particular, for $\phi = 0$ we have

$$x \equiv y \;\; \theta(u, v) \iff f(u, v, x) = f(u, v, y).$$

For arithmetical varieties this result can be applied to obtain principal congruence formulas in a simple and uniform way. Details can be found in [18].

For future reference (Theorem 2.8) we also notice another feature of the (first) proof of Theorem 2.2: we could have defined $f_\phi(x/\phi, y/\phi, z/\phi)$, starting with each maximal ϕ and ending with $\phi = 0$, for a *single* triple $(x, y, z) \in A^3$, and then repeating the process for another triple, etc., until A^3 is exhausted. Moreover each $f_\phi(x/\phi, y/\phi, z/\phi)$ is—except when a non-maximal $\phi \in L$ is covered by only a single element—completely prescribed in the proof. If we use this "triple by triple" approach, notice that the L-compatibility of f does not depend on making choices for the values of $f_\phi(x/\phi, y/\phi, z/\phi)$ which depend on the previous choices made for the preceding triples in A^3.

INTERPOLATION AND ARITHMETICITY

We begin this discussion by extending the concept of compatibility. First, as in Theorem 2.4 above, we may speak of compatible *partial* functions, i.e.: functions $f{:}X \to A$, where X is a subset of A^k. In particular if X is finite we call such a function a *finite partial function* on A. Second, we note that congruence relations of an algebra $\mathbf{A}$ are just special subuniverses in $A \times A$. Now suppose that K is any collection of subuniverses in $A \times A$ which is closed under intersection (and hence is a complete lattice of subuniverses in $A \times A$.) To say that a partial function f on A is *K-compatible* (where defined) means that each subuniverse in K is closed under f (where defined). This means that if $((a_1, b_1), \ldots, (a_k, b_k))$ is any k-tuple in $A \times A$ such that both $\mathbf{a} = (a_1, \ldots, a_k)$ and $\mathbf{b} = (b_1, \ldots, b_k)$ are in X, then $f((a_1, b_1), \ldots, (a_k, b_k))$, which is defined componentwise to be $(f(a_1, \ldots, a_k), f(b_1, \ldots, b_k)) = (f(\mathbf{a}), f(\mathbf{b}))$, is in the subuniverse in K generated by $((a_1, b_1), \ldots, (a_k, b_k))$, (i.e.: is in the intersection of all members of K containing these k pairs).

We shall be concerned with a particular lattice K of subuniverses in $A \times A$, namely those which we shall call rectangular: a subuniverse S in $A \times A$ is *rectangular* if

$$(x, y), (x, v), (u, v) \in S \implies (u, y) \in S,$$

i.e.: if three vertices of a rectangle are in S then so is the fourth. Notice that the family of rectangular subuniverses is closed under intersections. Also we have the following connection between rectangular subuniverses and congruences.

Lemma 2.2 *A subuniverse S in $A \times A$ is a congruence relation of* $\mathbf{A}$ *iff S is both rectangular and diagonal (i.e.: contains $\Delta = \{(x,x) : x \in A\}$.)*

Proof Let S be both rectangular and diagonal. Then for $(x,y) \in S$, we also have $(x,x),(y,y) \in S$ so rectangularity implies $(y,x) \in S$ (symmetry). S is transitive since $(x,y),(y,z) \in S \Rightarrow (x,y),(z,y),(z,z) \in S \Rightarrow (x,z) \in S$ using symmetry and rectangularity. Conversely, if S is a congruence of $\mathbf{A}$, then S is diagonal and if $((x,y),(x,v),(u,v) \in S$, then symmetry and transitivity imply $(u,y) \in S$.

Finally, if f is a finite partial function on A then f has an *interpolating* term function t (respectively, polynomial function p) if t (respectively p) agrees with f on $dom\, f = X$. Using these concepts we can prove the following important theorem due to Hagemann and Herrmann [6]. The proof below is somewhat more direct than their original version.

Theorem 2.5 *An algebra* $\mathbf{A}$ *is arithmetical if* $\mathbf{A}$ *satisfies the condition:*
a) *For any finite partial function f on A, f has an interpolating polynomial iff f is Con*$\mathbf{A}$*–compatible.*
An algebra $\mathbf{A}$ *has all of its subalgebras arithmetical if* $\mathbf{A}$ *satisfies the condition:*
b) *For any finite partial function f on A, f has an interpolating term function iff f is compatible with all rectangular subuniverses in $A \times A$.*

Proof The proof of the first statement is immediate from Lemma 2.1 since an interpolating polynomial for the function of the lemma will certainly be a Con$\mathbf{A}$–compatible extension of f to $X \cup \{(x,y,z)\}$.

For the proof of the second statement let $\mathbf{B} < \mathbf{A}$ and let $X \subset B^3$ be the same as in Lemma 2.1. We verify that the (Con$\mathbf{B}$–compatible) function f defined as in Lemma 2.1 is compatible with all rectangular subuniverses in $A \times A$. This verification is just a slight variation of the proof of Lemma 2.1: notice that in that proof to verify the $\mathbf{L}$–compatibility of f only the case $\mathbf{a} = (x,y,y)$ and $\mathbf{b} = (y,y,z)$ required any particular attention, namely we used the transitive property of an equivalence relation to conclude that $(x,y),(y,y),(y,z) \in L \Rightarrow (x,z) \in L$. But also in the present situation only this case requires attention and here any rectangular subuniverse in $A \times A$ which contains $(x,y),(y,y),(y,z)$ must also contain (x,z). Hence f is compatible with all rectangular subuniverses in $A \times A$ and therefore has an interpolating term function. This term function is a Con$\mathbf{B}$–compatible extension of f to $X \cup \{x,y,z\}$ so, as in the proof of Lemma 2.1, $\mathbf{B}$ is arithmetical.

It is important to notice that the converse of Theorem 2.5 is obviously false: let $\mathbf{A}$ be the two element set with no operations, which is arithmetical, but both of the conditions *a)* and *b)* are false in $\mathbf{A}$. On the other hand the next theorem asserts that if a variety V is arithmetical then then each of these interpolation properties holds in every member of V and hence each characterizes arithmeticity of a variety. These properties result from the fact that an arithmetical variety has both a ternary majority term and a Mal'cev term. We examine the effects of these separately.

Majority term. A ternary term $m(x, y, z)$ is a majority term for a variety V if

$$m(x, y, x) = m(x, x, y) = m(y, x, x) = x$$

are equations of V. In particular, if t satisfies (1), then

$$m(x, y, z) := t(x, t(x, y, z), z)$$

is a majority term.

Suppose $\mathbf{A}$ is an algebra in a variety with such a majority term. Suppose $X \subset A^k$ and that $f{:}X \to A$ is any function which is compatible with each subuniverse of $\mathbf{A} \times \mathbf{A}$. As noted earlier this means that for each pair of k-tuples $\mathbf{a} = (a^1, \ldots, a^k)$ and $\mathbf{b} = (b_1, \ldots, b_k)$ in X, $(f(\mathbf{a}), f(\mathbf{b}))$ is in the subalgebra of $\mathbf{A} \times \mathbf{A}$ generated by $(a_1, b_1), \ldots, (a_k, b_k)$. Accordingly, this means that there is a k–ary term function t such that

$$\begin{aligned} t(a_1, \ldots, a_k) &= f(a_1, \ldots, a_k) \\ \text{and } \; t(b_1, \ldots, b_k) &= f(b_1, \ldots, b_k), \end{aligned}$$

i.e.: for each pair of elements in its domain, f can be *interpolated* by a term function. Now if the domain of f is a finite subset of A^k, say

$$dom\, f = \{\mathbf{a}^1, \ldots, \mathbf{a}^m : \mathbf{a}^i = (a^i_1, \ldots, a^i_k)\},$$

then for each pair $\{\mathbf{a}^i, \mathbf{a}^j\}$ there will be a term t_{ij} interpolating f. Thus for the triple $\{\mathbf{a}^i, \mathbf{a}^j, \mathbf{a}^k\} \subset dom\, f$,

$$m(t_{ij}(\mathbf{a}^s), t_{ik}(\mathbf{a}^s), t_{jk}(\mathbf{a}^s)) = f(\mathbf{a}^s)$$

for $s \in \{i, j, k\}$. Hence the terms $t_{ijk} = m(t_{ij}, t_{ik}, t_{jk})$ interpolate f for each triple of points in its domain. Continuing, we see that the terms $t_{ijkl} = m(t_{ijk}, t_{ijl}, t_{ikl})$ interpolate four points at a time, etc., so that by induction f can be interpolated by a term function on all of *dom f*. We have proven half of

Proposition 2.1 ([1]) *A variety V has a ternary majority term iff it has the property:*

> *For each $\mathbf{A}$ in V if f is a finite partial function on A which is compatible with each subuniverse of $\mathbf{A} \times \mathbf{A}$ (where it is defined), then f can be interpolated by a term function.*

For the converse suppose V has this property and let $\mathbf{F}$ be the V–free algebra with free generators x, y. Let f in $F^3 \to F$ have domain

$$dom\, f = \{(x, x, y), (x, y, x), (y, x, x)\}$$

with values $f(x, x, y) = x,\ f(x, y, x) = x,\ f(y, x, x) = x$. Observe that, for example, for the first pair in *dom f*,

$$f((x, x), (x, y), (y, x)) = (f(x, x, y), f(x, y, x)) = (x, x)$$

which is of course in any subalgebra of $\mathbf{F} \times \mathbf{F}$ containing $(x,x),(x,y),(y,x)$. The same is obvious for the other two pairs in *dom* f. Hence f is compatible with all subuniverses in $F \times F$ (where defined) and hence f can be interpolated by a term function m. Then since

$$m(x,x,y) = x, \; m(x,y,x) = x, \; m(y,x,x) = x$$

hold and x, y are free generators, the above are equations of V so m is a majority term for V.

Proposition 2.1 applies, for example, to the variety V of all lattices since the lattice "median" term $m(x,y,z) = (x \vee y) \wedge (x \vee z) \wedge (y \vee z)$ is a majority term for V.

Mal'cev term. $p(x,y,z)$ is a Mal'cev term for V (characterizing congruence permutability of V) if $p(x,x,z) = z, \; p(x,z,z) = x$ are equations of V.

Proposition 2.2 *A variety V has a Mal'cev term iff for each pair of algebras $\mathbf{A}, \mathbf{B} \in V$, each subuniverse S in $A \times B$ is rectangular, meaning (as before) that*

$$(x,y),(x,v),(u,v) \in S \Longrightarrow (u,y) \in S,$$

(i.e.: if three vertices of a rectangle in $A \times B$ are in S then so is the fourth).

Proof $\Rightarrow p((x,y),(x,v),(u,v)) = (p(x,x,u), p(y,v,v)) = (u,y)$.
$\Leftarrow$ Consider $\mathbf{F}$ freely generated in V by x, y. For the subalgebra of $\mathbf{F} \times \mathbf{F}$ generated by $(x,x),(x,y),(y,y)$ to be rectangular requires that it contain (y,x). Hence there is a term p such that $p((x,x),(x,y),(y,y)) = (y,x)$, so $p(x,x,y) = y$ and $p(x,y,y) = x$ and again these must be identities of V.

We have proved most of the following key result:

Theorem 2.6 ([17]) *For a variety V the following conditions are equivalent:*

(1) *V is arithmetical.*

(2) *For any algebra $\mathbf{A} \in V$, each finite partial function f on A has an interpolating term function iff f is compatible with each rectangular subuniverse of $\mathbf{A} \times \mathbf{A}$ (where defined).*

(3) *For any algebra $\mathbf{A} \in V$, each finite partial function f on A has an interpolating polynomial iff f is compatible with Con$\mathbf{A}$ (where defined).*

Proof Each of the implications (2) $\Rightarrow$ (1) and (3) $\Rightarrow$ (1) follows directly from Theorem 2.5. Propositions 2.1 and 2.2 prove (1) $\Rightarrow$ (2).

(1) $\Rightarrow$ (3) Consider the algebra $\mathbf{A}^+$ obtained from $\mathbf{A}$ by adding all elements in A as new nullary operations, and observe that (1), and hence (2) hold in $V(\mathbf{A}^+)$. Further $\mathbf{Con}\mathbf{A}^+ = \mathbf{Con}\mathbf{A}$ and by Lemma 2.2 the rectangular subuniverses of $\mathbf{A}^+ \times \mathbf{A}^+$ are exactly the congruences of $\mathbf{A}$. Hence if f is a finite partial function on A and f is compatible with

Con A (where defined), we conclude from (2) that it can be interpolated by a term function of $\mathbf{A}^+$ and thus by a polynomial of $\mathbf{A}$. This proves (3).

GRAPH SUBALGEBRAS AND RECTANGULAR HULLS

For algebras $\mathbf{A}$ and $\mathbf{B}$, let $\mathbf{A}_1 < \mathbf{A}, \mathbf{A}_2 < \mathbf{B}$, $\phi_i \in \mathrm{Con}\mathbf{A_i}$, $i = 1,2$, and suppose

$$\sigma : \mathbf{A}_1/\phi_1 \to \mathbf{A}_2/\phi_2$$

is a (surjective) isomorphism. Then

$$S = \{(a_1, a_2) \in A_1 \times A_2 : \sigma(a_1/\phi_1) = a_2/\phi_2\}$$

is a subuniverse of $\mathbf{A} \times \mathbf{B}$. Since it is the union of the elements (blocks $a_1/\phi_1 \times a_2/\phi_2$) of the graph of σ we call the subalgebra $\mathbf{S}$ a *graph subalgebra* of $\mathbf{A} \times \mathbf{B}$ (defined by σ).

Proposition 2.3 *A subalgebra of* $\mathbf{A} \times \mathbf{B}$ *is a graph subalgebra iff it is rectangular.*

Proof $\Rightarrow$ If $\mathbf{S}$ is a graph subalgebra and $(x,y), (x,v), (u,v) \in S$, then

$$\sigma(x/\phi_1) = y/\phi_2, \ \ \sigma(x/\phi_1) = v/\phi_2, \ \ \sigma(u/\phi_1) = v/\phi_2.$$

The second two equations imply $x/\phi_1 = u/\phi_1$ and from this and the first equation we obtain $\sigma(u/\phi_1) = y/\phi_2$ so $(u,y) \in S$.

$\Leftarrow$ If $\mathbf{S}$ is rectangular with first and second projections $A_1 \subset A, A_2 \subset B$, then for $a, b \in A_1$ define ϕ_1 by

$$(a,b) \in \phi_1 \iff \exists y \in A_2 \text{ with } (a,y), (b,y) \in S.$$

ϕ_1 is reflexive and symmetric and has the substitution property directly from the definition. Finally, if $(x,y), (y,z) \in \phi_1$, then so is the triple $(y,x), (y,z), (z,z)$, so that by rectangularity so is (z,x), verifying transitivity. Hence ϕ_1 is a congruence of $\mathbf{A}_1$. Analogously define $\phi_2 \in \mathrm{Con}\mathbf{A}_2$ by

$$(a,b) \in \phi_2 \iff \exists x \in A_1 \text{ with } (x,a), (x,b) \in S.$$

By the definitions of ϕ_i it is immediate that for $(a_1, a_2) \in S$, $\sigma(a_1/\phi_1) = a_2/\phi_2$ defines a isomorphism and $\mathbf{S}$ is contained in the graph subalgebra defined by σ. To show that $\mathbf{S}$ equals the graph subalgebra, suppose $(a,b) \in S$ and $u \in A_1, v \in A_2$ are such that $(u,v) \in a/\phi_1 \times b/\phi_2$. We must show that $(u,v) \in S$. But we have $(u,a) \in \phi_1$ and $(v,b) \in \phi_2$ so that for some $x \in A_1, y \in A_2$,

$$(x,v), (x,b) \in S \ \text{ and } \ (u,y), (a,y) \in S.$$

Then rectangularity applied to $(x,b), (a,b), (a,y) \in S$ implies $(x,y) \in S$, and then applied again to $(u,y), (x,y), (x,v) \in S$ implies $(u,v) \in S$.

If **S** is any subalgebra in $\mathbf{A} \times \mathbf{B}$ then it is certainly in the rectangular subalgebra $\mathbf{A}_1 \times \mathbf{A}_2$ (the product of its projections). Also the intersection of all rectangular subalgebras containing **S** is rectangular; since it is the least rectangular subalgebra of $\mathbf{A} \times \mathbf{B}$ containing **S**, we call it the *rectangular hull* of **S** (in $\mathbf{A} \times \mathbf{B}$). We can also construct the rectangular hull of **S** as a graph subalgebra. Define $\phi_i \in \mathrm{Con}\mathbf{A}_i$ by

$$\phi_1 = \bigvee\{\theta(a,b) : \exists y \in A_2 \text{ with } (a,y) \text{ and } (b,y) \text{ in } S.\}$$

$$\phi_2 = \bigvee\{\theta(c,d) : \exists x \in A_1 \text{ with } (x,c) \text{ and } (x,d) \text{ in } S.\}$$

For each $a_1/\phi_1 \in A_1/\phi_1$, choose $a_2 \in A_2$ such that $(a_1, a_2) \in S$ and define $\sigma(a_1/\phi_1) = a_2/\phi_2$. From the definitions of ϕ_1 and ϕ_2, it is easy to check that $\sigma{:}A_1/\phi_1 \to A_2/\phi_2$ is an isomorphism and that the graph subalgebra defined by σ is the rectangular hull of **S**.

Propositions 2.2 and 2.3 together prove the well known fact (due essentially to I. Fleischer [2]) that a variety V is congruence permutable iff all subalgebras (in the product of pairs of algebras in V) are graph subalgebras. (The latter property has been known for groups for over a hundred years.) In [6] and elsewhere, rectangular subdirect products of $\mathbf{A} \times \mathbf{B}$ are called *pullbacks.*

NEAR UNANIMITY FUNCTIONS

Proposition 2.1 above has an unexpected consequence:

Corollary 2.1 ([11]) *If a clone on a finite set contains a ternary majority function, then it is finitely generated.*

Proof If C is such a clone on a finite set A, then C is the set of term functions of the algebra $\mathbf{A} = \langle A, C\rangle$ and hence (by Proposition 2.1) $f \in C$ iff f is compatible with each subuniverse of $\mathbf{A} \times \mathbf{A}$, and there are only finitely many such subuniverses. But also there are only finitely many subsets of $A \times A$ which are not subuniverses. For each such non-subuniverse B pick an $f \in C$ such that f is not compatible with B. These f together with the majority function will then generate C (by another application of Theorem 2.1).

The following are more general versions of Proposition 2.1 and Corollary 2.1. We establish them now for later use. An $(n+1)$–ary *near unanimity function* u is any (n+1)–ary function satisfying

$$u(x, ..., x, y, x, ...x) = x$$

where the single occurrence of y can occur in any of the $n+1$ argument positions.

Proposition 2.4 ([1]) *A variety V has an n–ary near unanimity term iff it has the property:*

> *For each $\mathbf{A} \in V$, if f is a finite partial function on A which is compatible with each subuniverse of $\mathbf{A}^n$ (where defined), then f can be interpolated by a term function.*

Corollary 2.2 *If a clone on a finite set contains any n–ary near unanimity function then it is finitely generated.*

Corollary 2.2 is proved from Proposition 2.4 by replacing $\mathbf{A} \times \mathbf{A}$ by $\mathbf{A}^n$ in the proof of Corollary 2.1.

Proof of Proposition 2.4 Let $\mathbf{A}$ be an algebra in a variety V having an $(n+1)$–near unanimity term u, $n \geq 2$. Let $X \subset A^k$, X finite, and suppose that $f{:}X \to A$ is any function which is *compatible* with all of the subuniverses in A^n, where defined. This means, extending the earlier definition, that each subuniverse of $\mathbf{A}^n$ must be closed under the componentwise extension of f from a function defined on X to a function defined on a certain subset of all k–tuples of elements of A^n. More specifically this means that for any k–tuple

$$\mathbf{a}_{1*} = (a_{11}, \ldots, a_{1n}), \ldots, \mathbf{a}_{k*} = (a_{k1}, \ldots, a_{kn})$$

of elements of A^n for which

$$\mathbf{a}_{*1} = (a_{11}, \ldots, a_{k1}), \ldots, \mathbf{a}_{*n} = (a_{1n}, \ldots, a_{kn})$$

are in X, so that $f(\mathbf{a}_{1*}, \ldots, \mathbf{a}_{k*})$, (which is defined componentwise to be $(f(\mathbf{a}_{*1}), \ldots, f(\mathbf{a}_{*n}))$ exists, then this function value is in the subalgebra of $\mathbf{A}^n$ generated by $\mathbf{a}_{1*}, \ldots, \mathbf{a}_{k*}$. The best way to vizualize this condition is as follows: for any $k \times n$ matrix M whose rows are elements of A^n, the new row vector obtained by applying f to the columns of M (provided they are in X) is in the subalgebra of A^n generated by the rows of M. Accordingly, this means that there is a k–ary term function t such that

$$\begin{aligned} t(\mathbf{a}_{*1}) &= f(\mathbf{a}_{*1}) \\ &\cdots \\ t(\mathbf{a}_{*n}) &= f(\mathbf{a}_{*n}), \end{aligned}$$

i.e.: that f can be interpolated by a term function at each subset of n elements of its domain X. Then if X has m elements, $X = \{\mathbf{a}^1, \ldots, \mathbf{a}^m\}$, $\mathbf{a}^i = (a_1^i, \ldots, a_k^i)$, for any subset Y of $n+1$ elements of X, we will have $n+1$ (n+1)–ary term functions $t_1, \ldots, t_{n+1}$ respectively interpolating f on the $n+1$ different n element subsets of Y. But then just as in the case for $n = 2$, it is easy to see that the term function $u(t_1, \ldots, t_{n+1})$ interpolates f on Y. Continuing as before we see, by induction, that f can be interpolated on all of X.

Regarding this proof we should note several features. First notice that if $|X| = m \leq n$ then the compatibility of f implies the existence of an interpolating term without use of u.

Also notice that if $Y \subset X$ and $n+1 < |Y| = r \leq |X|$, and if we already, by induction, have r terms $s_1, \ldots, s_r$, which interpolate f on the $r-1$ element subsets of Y, then we will only need the first $n+1$ of these to form a term $u(s_1, \ldots, s_{n+1})$ which will interpolate f on Y; i.e.: in general, there are many choices for the interpolating terms.

Finally notice that the final interpolating term is a composition whose innermost terms are interpolating terms for n out of the m domain elements at a time, encased in $(m-n)$

layers of the $(n+1)$–ary term u. A count shows that there will be $[(n+1)^{m-n}-1]/n$ uses of u in the final interpolation. Thus the final interpolating term may be highly complex. For example, let $\mathbf{A}$ be a 10 element lattice (so u is the ternary majority term for lattices, i.e.: $n=2$). If f is a ternary function which is compatible with all sublattices of $\mathbf{A}\times\mathbf{A}$, then $|dom\ f| = 1000$ and the interpolation represents f as a 998–layered composition involving $(3^{998}-1)/2$ uses of u!

For the proof of the converse of Proposition 2.4 suppose that V has the given property. Let $\mathbf{F}$ be the V–free algebra with free generators x and y. Let X be the following set of $n+1$ $(n+1)$–tuples:

$$X = \{(x,\ldots,x,y),\ (x,\ldots,y,x),\ldots,(y,x,\ldots,x)\}.$$

Let $f{:}X\rightarrow A$ have the value x at each $(n+1)$–tuple in X. Now if we delete any one of the elements of X and transpose the remaining into columns forming an $(n+1)\times n$ matrix, then since there will be n y's, each occurring in a different one of the $n+1$ rows, some row will consist only of x's. Since this is the value obtained by applying f to each of the n columns, this shows that f is compatible with all subuniverses of $\mathbf{F}^n$, and hence can be interpolated by an $(n+1)$–ary term function u. As before, u must then be a near unanimity term for V.

An important property of a variety having a near unanimity term is that it is necessarily congruence distributive. This was first proved by Mitschke [14] by explicitly constructing Jónsson terms from any given near unanimity term. The following Proposition yields the same result and the proof (from [4]) is somewhat more transparent than that given by Mitschke. We will need this result and Corollary 2.2 at the end of Section 4.

Proposition 2.5 *If an algebra* $\mathbf{A}$ *has a* $(n+1)$*–ary* Con$\mathbf{A}$*–compatible near unanimity function, then* $\mathbf{A}$ *is congruence distributive.*

Proof For $\theta,\phi,\psi\in \mathrm{Con}\mathbf{A}$, let $(a,b)\in\theta\wedge(\phi\vee\psi)$ so that $(a,b)\in\theta$ and for some elements $x_0=a,x_1,\ldots,x_n=b\in A$, $(x_i,x_{i+1})\in\phi\cup\psi$ for all $i<n$. Let u be a near unanimity function satisfying the required hypotheses. Then for each i we have

$$(u(a,...,a,x_i,b,...,b),u(a,...,a,x_{i+1},b,...,b))\in\phi\cup\psi,$$

and also

$$\begin{aligned} u(a,...,a,x_i,b,...,b) &\equiv_\theta u(a,...,a,x_i,a,...,a)\\ &= u(a,...,a,x_{i+1},a,...,a)\\ &\equiv_\theta u(a,...,a,x_{i+1},b,...,b). \end{aligned}$$

Then for each i we have

$$(u(a,...a,x_i,b,...b),u(a,...,a,x_{i+1},b,...,b))\in\theta\wedge(\phi\cup\psi).$$

But $\theta \wedge (\phi \cup \psi) = (\theta \wedge \phi) \cup (\theta \wedge \psi) \subset (\theta \wedge \phi) \vee (\theta \wedge \psi)$. Consequently, by transitivity we have

$$(u(a, ..., \underline{a}, b, ..., b), u(a, ..., a, \underline{b}, ..., b)) \in (\theta \wedge \phi) \vee (\theta \wedge \psi)$$

where for each argument position (underlined), the right side is obtained from the left by replacing the last a by a b. Hence we obtain

$$a = u(a, ..., a, b) \equiv u(a, ..., a, b, b) \equiv \cdots \equiv u(a, b, ...b) = b$$

mod $(\theta \wedge \phi) \vee (\theta \wedge \psi)$. Hence

$$\theta \wedge (\phi \vee \psi) \leq (\theta \wedge \phi) \vee (\theta \wedge \psi)$$

so Con$\mathbf{A}$ is distributive.

LOCALLY FINITE AND FINITELY GENERATED VARIETIES

For locally finite arithmetical varieties condition (2) of Theorem 2.6 extends to total functions:

Theorem 2.7 *If V is a locally finite arithmetical variety, then for any $\mathbf{A}$ in V, $f{:}A^k \to A$ is a term function of $\mathbf{A}$ iff f is compatible with all (rectangular) subuniverses of $\mathbf{A} \times \mathbf{A}$.*

Proof Let $\mathbf{F}$ be the free term algebra, in the subvariety $V(\mathbf{A}) \subset V$ generated by $\mathbf{A}$, with free generators $x_1, \ldots, x_k$. Since F is finite there is a set G of k-tuples of elements of A, say

$$G = \{\mathbf{a}^1, \ldots, \mathbf{a}^m : \mathbf{a}^i = (a_1^i, \ldots, a_k^i)\},$$

such that

$$t(x_1, \ldots, x_k) \mapsto (t(a_1^1, \ldots, a_k^1), \ldots, t(a_1^m, \ldots, a_k^m))$$

is an isomorphism of $\mathbf{F}$ onto a subalgebra of $\mathbf{A}^m$. Hence if two k-ary terms in F agree when evaluated at each element of G, then they agree everywhere in A^k. Since G is finite, by Theorem 2.6, f is interpolated by some term t on G. If $\mathbf{b}$ is any other element in A^k then f is also interpolated by some term s on $G \cup \{\mathbf{b}\}$. But then s and t , agreeing on G must also agree at $\mathbf{b}$, so $f(\mathbf{b}) = s(\mathbf{b}) = t(\mathbf{b})$. Since $\mathbf{b}$ is arbitrary, we conclude that f is a term function.

The argument above actually shows that for *any* locally finite variety, finite interpolation by term functions implies global interpolation by term functions. For example, if $\mathbf{A}$ is any distributive lattice and f is compatible with all of the sublattices of $\mathbf{A} \times \mathbf{A}$, then (since the "median" term is a majority term for all lattices) f can be finitely interpolated by lattice term functions. Hence f is a lattice term function.

What about the converse of Theorem 2.7? For the finitely generated case this raises the question:

If $V = V(\mathbf{A})$ is generated by a finite algebra $\mathbf{A}$ and for each $\mathbf{B} \in V$ the terms of $\mathbf{B}$ are precisely the functions compatible with all rectangular subalgebras of $\mathbf{B} \times \mathbf{B}$, is V arithmetical?

It does not seem to be known if this is true and in view of the strong requirement of statement (2) of Theorem 2.6 (certain *partial* functions must be interpolated by terms), it seems unlikely. For this reason Theorem 2.9 below seems all the more remarkable, since it requires that the term function test be met only by the single generating algebra of the variety. The only hypothesis is the reasonable requirement that the generating algebra have only arithmetical subalgebras. We can think of Theorem 2.9 as a kind of single algebra analog of the equivalence of (1) and (2) of Theorem 2.6. To prove it we need the following generalization of Theorem 2.3.

Theorem 2.8 *A finite algebra $\mathbf{A}$ has only arithmetical subalgebras iff there exists a function $f{:}A^3 \to A$ which is compatible with each rectangular subalgebra of $\mathbf{A} \times \mathbf{A}$ and which satisfies equations* (1) *of Theorem* 2.1.

Proof $\Leftarrow$ Add f as a new operation to $\mathbf{A}$ without changing the congruence lattices of any of the subalgebras. The resulting algebra generates an arithmetical variety so each subalgebra of $\mathbf{A}$ is arithmetical.

$\Rightarrow$ This can be accomplished by an elaboration of the scheme used to prove Theorem 2.2. We sketch the argument. (Complete details can be found in [16]). By Proposition 2.3 it is enough to exhibit an $f{:}A^3 \to A$ which is compatible with each graph subalgebra of $\mathbf{A} \times \mathbf{A}$ and which satisfies equations (1). To do this it will suffice to exhibit an f which satisfies the following:

1) For each $\mathbf{B} < \mathbf{A}$, B is closed under f.

2) f is $\mathrm{Con}\mathbf{B}$-compatible.

3) for each $\mathbf{B}, \mathbf{C} < \mathbf{A}$, $\phi \in \mathrm{Con}\mathbf{B}$, $\psi \in \mathrm{Con}\mathbf{C}$, and isomorphism $\sigma{:}\mathbf{B}/\phi \to \mathbf{C}/\psi$,

$$\sigma f_\phi(x/\phi, y/\phi, z/\phi) = f_\psi(\sigma(x/\phi), \sigma(y/\phi), \sigma(z/\phi))$$

for all $x, y, z \in B$. (f_ϕ, f_ψ are the functions on B/ϕ and C/ψ induced by f by virtue of 1) and 2), e.g.: $f_\phi(x/\phi, y/\phi, z/\phi) = f(x, y, z)/\phi$.)

4) All f_ϕ satisfy equations (1).

Now to construct f satisfying 1)–4) above, it is easy to see that we need only consider 3–generated subalgebras $\mathbf{B}, \mathbf{C} < \mathbf{A}$. Let $\langle x, y, z\rangle$ denote the subalgebra of $\mathbf{A}$ generated by the triple $(x, y, z) \in A^3$. We follow the plan of the proof of Theorem 2.2. As was observed in the last of the remarks made after the proof of Theorem 2.4, for this particular (x, y, z) we can define $f_\phi(x/\phi, y/\phi, z/\phi)$ for all $\phi \in \mathrm{Con}\langle x, y, z\rangle$, so that the resulting f is

$\mathrm{Con}\langle x,y,z\rangle$-compatible, without reference to the other elements in $\langle x,y,z\rangle^3$. The definition is completely prescriptive except when some non-maximal ϕ has only a single cover; only in this case do we need to make a choice. Thus to construct f we begin by letting $\prec$ be an arbitrary fixed total ordering of all of the triples in A^3. To satisfy 1)–4) above we must then define $f_\phi(x/\phi, y/\phi, z/\phi)$ for each $(x,y,z) \in A^3$ and $\phi \in \mathrm{Con}\langle x,y,z\rangle$ so that the following conditions are met:

1′) $f_\phi(x/\phi, y/\phi, z/\phi) \in \langle x,y,z\rangle/\phi$.

2′) If $\phi \leq \theta$ in $\mathrm{Con}\langle x,y,z\rangle$, then

$$f_\phi(x/\phi, y/\phi, z/\phi) \subset f_\theta(x/\theta, y/\theta, z/\theta).$$

3′) If $(u,v,w) \prec (x,y,z)$, $\theta \in \mathrm{Con}\langle x,y,z\rangle$, and $\sigma:\langle u,v,w\rangle/\theta \to \langle x,y,z\rangle/\phi$ is an isomorphism such that $\sigma(u/\theta) = x/\phi$, $\sigma(v/\theta) = y/\phi$, $\sigma(w/\theta) = z/\phi$, then

$$\sigma f_\theta(u/\theta, v/\theta, w/\theta) = f_\phi(x/\phi, y/\phi, z/\phi).$$

4′) All $f_\phi(x/\phi, y/\phi, z/\phi)$ satisfy equations *(1)* if any pair of $x/\phi, y/\phi, z/\phi$ are equal.

To accomplish this goal we define the $f_\phi(x/\phi, y/\phi, z/\phi)$ by induction according to the following scheme:

For the first triple (x,y,z) (in the order $\prec$) define $f_\phi(x/\phi, y/\phi, z/\phi) \in \langle x,y,z\rangle/\phi$ for all $\phi \in \mathrm{Con}\langle x,y,z\rangle$ as in the proof of Theorem 2.2. Suppose now that $f_\theta(u/\theta, v/\theta, w/\theta)$ has been defined for all $(u,v,w) \prec (x,y,z)$ and $\theta \in \mathrm{Con}\langle u,v,w\rangle$. Define $f_\phi(x/\phi, y/\phi, z/\phi)$ for this single triple and all ϕ as before except if there is a $(u,v,w) \prec (x,y,z)$ and θ in $\mathrm{Con}\langle u,v,w\rangle$ satisfying the hypothesis of 3′) above with θ (and hence ϕ) having a single cover. In this case define $f_\phi(x/\phi, y/\phi, z/\phi)$ to satisfy the conclusion of 3′), i.e.: make a choice for $f_\phi(x/\phi, y/\phi, z/\phi)$ which corresponds, under σ, to the choice made earlier for $f_\theta(u/\theta, v/\theta, w/\theta)$. Then continue in the same fashion with the next triple in order until A^3 is exhausted. It is not difficult to verify that the resulting f satisfies conditions 1′)—4′).

As a direct consequence of Theorems 2.6 and 2.8 we have

Theorem 2.9 *If* $\mathbf{A}$ *is finite and all subalgebras of* $\mathbf{A}$ *are arithmetical, then* $V(\mathbf{A})$ *is arithmetical iff the term functions of* $\mathbf{A}$ *are precisely the functions* $f{:}A^k \to A$ *which are compatible with all rectangular subalgebras of* $\mathbf{A} \times \mathbf{A}$.

Finally we have the following general result.

Theorem 2.10 *For a finite algebra* $\mathbf{A}$ *the following are equivalent:*

i) $V(\mathbf{A})$ *is arithmetical.*

ii) *All subalgebras of* $\mathbf{A}$ *are arithmetical and the term functions of* $\mathbf{A}$ *are precisely the total functions* $f{:}A^k \to A$ *which are compatible with all rectangular subalgebras of* $\mathbf{A} \times \mathbf{A}$.

iii) *Each partial function f on A has an interpolating term function iff f is compatible with all rectangular subuniverses in $A \times A$.*

Proof ii) $\Rightarrow$ i) by Theorem 2.9. i) $\Rightarrow$ iii) by (1) $\Rightarrow$ (2) of Theorem 2.6. iii) $\Rightarrow$ ii) by Theorem 2.5 and finiteness.

A special case of this theorem is the following which was actually the first one of them to be discovered ([15]) and which began the study of discriminator varieties and quasi-primal algebras.

Theorem 2.11 *For a finite algebra $\mathbf{A}$ the following are equivalent:*

i) *The discriminator is a term function of $\mathbf{A}$.*

ii) *All subalgebras of $\mathbf{A}$ are simple and $V(\mathbf{A})$ is arithmetical.*

iii) *The term functions of $\mathbf{A}$ are the functions $f: A^k \to A$ which are compatible with all graphs of isomorphisms of subalgebras of $\mathbf{A}$.*

Statement iii) of this theorem is the original definition of a *quasi-primal* algebra. If $\mathbf{A}$ has no proper subalgebras and is rigid then the equivalence of ii) and iii) is the primal algebra characterization referred to in Section 1. Today statement i) is also commonly used as the definition of a quasi-primal algebra. Notice that iii) $\Rightarrow$ i) since the discriminator is a *pattern* function, i.e.: its values are always one of its arguments, depending on the pattern of equalities among the arguments, and this pattern is obviously preserved by any isomorphism of subalgebras.

The last three theorems might suggest that there are also analogous equivalences for single finite arithmetical algebras, but stemming from the equivalence (1) $\Leftrightarrow$ (3) of Theorem 2.6 (instead of (1) $\Leftrightarrow$ (2)). Remarkably enough there are, although, reasonably, they do not involve varieties since (3) deals with polynomials instead of terms. Theorems 3.2, 3.3, and 3.4 of the next section are the analogous equivalences.

3 Affine completeness

In contrast to Theorem 2.7, condition (3) of Theorem 2.6 does not extend so directly as the following example shows. Let $\mathbf{A} = \langle\{0,1\}, +, \cdot\rangle$ be the two element field. Since $t(x,y,z) = z + (x+z)(x+y)$ satisfies equations (1), $V(\mathbf{A})$ is arithmetical. Let $\mathbf{B}$ be the subalgebra of $\mathbf{A}^\omega$ consisting of all sequences $x = \{x_n\}$ which are 0 for all but finitely many n. Define the function f componentwise by

$$\begin{aligned} f(x)_n &= x_n \text{ if } n \text{ is odd,} \\ &= 0 \text{ if } n \text{ is even.} \end{aligned}$$

Using the fact (from congruence distributivity) that congruences on finite subdirect products are "skew-free", one can check that f is compatible with Con**B**. (See the proof of Theorem 4.3 for complete details.) But if f were a polynomial then for some m, $(m+1)$-ary term t, and elements $a^1, \ldots, a^m \in B$, we would have $f(x) = t(x, a^1, \ldots, a^m)$ and hence for all sufficiently large n, say $n \geq N$, we would have the term

$$\begin{aligned} t(x_n, 0, \ldots, 0) &= x_n \text{ if } n \text{ is odd,} \\ &= 0 \text{ if } n \text{ is even,} \end{aligned}$$

Supposing $n \geq N$ is odd and taking $x \in B$ with $x_n = x_{n+1} = 1$, we obtain the contradiction $1 = 0$.

Affine Completeness. An algebra **A** is *affine complete* if for any *total* function $f: A^k \to A$, f is a polynomial iff f is Con**A**–compatible. A simple affine complete algebra is usually called *functionally complete.* Thus an algebra is functionally complete if *every* function on it is a polynomial. It is usual (though, as we shall see below, not necessary) to also require a functionally complete algebra to be finite. A variety is affine complete if each of its algebras is affine complete.

Add all Con**A**–compatible functions to any **A** and the new algebra is obviously affine complete with the congruence lattice unchanged. Hence every algebra is a reduct of some affine complete algebra. Thus it appears that little can be said about single affine complete algebras as a class. (But see Kaarli [8] for some tantalizing partial results.) For affine complete varieties, our principal interest from here on, the situation is much different and for varieties there is a strong connection between arithmeticity and affine completeness. On the other hand the ring of integers is arithmetical but not affine complete. An easy way to see this (observed by P. Palfy) is to note that $f(x) = (x^2/2)(x^2+1)$ is integer valued on the integers and is compatible with all of the congruences, (since $x - y$ divides $f(x) - f(y)$), but f is not an integer polynomial.

Theorem 2.6, (3) asserts that each finite member of an arithmetical variety V is affine complete and the example at the beginning of this section shows that the infinite members need not be. Algebras satisfying the condition (3) of Theorem 2.6 are usually called *locally affine complete* and thus a variety is arithmetical iff it is locally affine complete.

The most important example of an affine complete variety is the variety of Boolean algebras, an old result due to Grätzer ([5]). Hu ([7]) extended Grätzer's result by showing that varieties generated by finite sets of independent primal algebras are also affine complete. The key observation is that each of these finitely generated arithmetical varieties is generated by a finite algebra having no proper subalgebras. (In the example above, the two element field, with addition and multiplication as the only operations, has a proper subalgebra.) The following result is then a correct analog of Theorem 2.7, and implies both Grätzer's and Hu's results.

Theorem 3.1 *If V is arithmetical and is generated by a finite algebra having no subalgebras, then V is affine complete.*

Proof Suppose $V = V(\mathbf{A})$ where $\mathbf{A}$ is finite having no subalgebras. By congruence distributivity $V = \mathbf{IP}_S\mathbf{H}(\mathbf{A})$ so that if $\mathbf{B} \in V$, we may suppose

$$\mathbf{B} < \Pi\{\mathbf{A}_i : i \in I\}$$

where each $\mathbf{A}_i$ is $\mathbf{A}/\theta$ for some $\theta \in \mathrm{Con}\mathbf{A}$. Let $f{:}B^k \to B$ be compatible with Con$\mathbf{B}$. Then if $\mathbf{B} = \langle B, F\rangle$ it follows that the algebra $\mathbf{B}^* = \langle B, F \cup \{f\}\rangle$ has the same universe and congruence lattice as $\mathbf{B}$. Also, by compatibility, f naturally induces various functions on the subdirect factors of $\mathbf{B}$, say f_i is the function induced on A_i. It follows from this that $\mathbf{B}^*$ is a subdirect product

$$\mathbf{B}^* < \Pi\{\mathbf{A}_i^* : i \in I\},$$

where $\mathbf{A}_i^*$ is obtained by adding f_i to the operations of $\mathbf{A}_i$. Since A is finite there are only finitely many non-isomorphic $\mathbf{A}_i^*$. Hence if we pick any $a \in B$, then the subalgebra $\mathbf{B}^*(a)$ of $\mathbf{B}^*$ generated by a is isomorphic to a subdirect product in $\Pi\{\mathbf{A}_i^* : i \in I_0\}$ for some finite subset $I_0 \subset I$, and hence $\mathbf{B}^*(a)$ is finite. Therefore the reduct $\bar{\mathbf{B}}$ obtained from it by discarding the operation f is a finite subalgebra of $\mathbf{B}$. Also f (restricted) is in $\bar{B}^k \to \bar{B}$ and is compatible with Con $\bar{\mathbf{B}}$. Hence, by Theorem 2.6, (3), f (restricted) is a polynomial of $\bar{\mathbf{B}}$. Hence for some $(k+m)$–ary term t and $\mathbf{a} \in \bar{B}^m$, $f(\mathbf{x}) = t(\mathbf{x}, \mathbf{a})$ for all $\mathbf{x} \in \bar{B}^k$. But since $\mathbf{A}$ has no subalgebras, the projections of $\bar{B}$ are onto each of the factors A_i, for *all* $i \in I$. Hence for any $\mathbf{x} \in B^k$ and $i \in I$, choose $\bar{\mathbf{x}} \in \bar{B}^k$ with $\bar{\mathbf{x}}_i = \mathbf{x}_i$. Then

$$f(\mathbf{x})_i = f_i(\mathbf{x}_i) = f_i(\bar{\mathbf{x}}_i) = f(\bar{\mathbf{x}})_i = t(\bar{\mathbf{x}}, \mathbf{a})_i = t(\bar{\mathbf{x}}_i, \mathbf{a}_i) = t(\mathbf{x}_i, \mathbf{a}_i) = t(\mathbf{x}, \mathbf{a})_i$$

and hence f is a polynomial of $\mathbf{B}$. We conclude that V is affine complete.

In establishing the theorem above we have used congruence permutability only to conclude that any finite member of V is affine complete. Hence we have actually proved the following somewhat somewhat stronger statement:

> *If V is a congruence distributive variety generated by a finite algebra having no proper subalgebras, and if each finite algebra in V is affine complete, then V is affine complete.*

The following characterization of finite *arithmetical* affine complete algebras is a simple application of the results presented so far. It is intriguing to note the formal parallel between this result and the equivalence of (1) and (3) of Theorem 2.6. It is the analog (announced at the end of Section 2) of the first of the sequence of three theorems beginning with Theorem 2.9.

Theorem 3.2 *If $\mathbf{A}$ is a finite arithmetical algebra then $\mathbf{A}$ is affine complete iff there is a polynomial $p(x, y, z)$ of $\mathbf{A}$ such that*

$$p(x, x, z) = z,\ p(x, y, x) = x,\ p(x, z, z) = x,$$

for all $x, y, z \in A$ (i.e.: p satisfies equations (1) *of Theorem* 2.1).

Proof If $\mathbf{A}$ is affine complete then by Theorem 2.3 there is a Con$\mathbf{A}$ – compatible function (hence a polynomial) $p: A^3 \to A$ satisfying equations (1) of Theorem 2.1. Conversely, if $\mathbf{A}$ has a polynomial $p(x, y, z)$ satisfying the identities (1) of Theorem 2.1, then the algebra $\mathbf{A}^+$ (obtained by adding the elements of A as new nullary operations) generates an arithmetical variety and by Theorem 2.6, (3), (or Theorem 3.1) $\mathbf{A}^+$ is affine complete; hence so is $\mathbf{A}$. (We could also prove the theorem by applying Theorem 2.9 directly to $\mathbf{A}^+$.)

Next we have the following analog of Theorem 2.10:

Theorem 3.3 *For a finite algebra* $\mathbf{A}$ *the following are equivalent:*

i) *Some ternary polynomial of* $\mathbf{A}$ *satisfies equations* (1) *of Theorem* 2.1.

ii) $\mathbf{A}$ *is arithmetical and affine complete.*

iii) $\mathbf{A}$ *is locally affine complete.*

Proof i) $\Rightarrow$ ii) by Theorems 3.2 and 2.3. ii) $\Rightarrow$ iii) by Theorem 2.4 (the compatible function extension property). iii) $\Rightarrow$ i) by Theorem 2.5 and finiteness and Theorem 3.2.

Finally we have the following special case when $\mathbf{A}$ is simple, and which is the analog of Theorem 2.11:

Theorem 3.4 *For a finite algebra* $\mathbf{A}$ *the following are equivalent:*

i) *The discriminator of A is a polynomial of* $\mathbf{A}$.

ii) $\mathbf{A}$ *is simple and some ternary polynomial of* $\mathbf{A}$ *satisfies equations* (1) *of Theorem* 2.1.

iii) $\mathbf{A}$ *is functionally complete.*

(The equivalence of i) and iii) is Werner's characterization of functional completeness [20].)

4 Affine complete varieties; general theory

In this section we outline some of the main features of the theory of affine complete varieties. The material in Sections 4 and 5 is largely joint work of Kaarli and Pixley, ([11], [12]).

Theorem 4.1 *If V is affine complete then V is residually finite* (*i.e.: each subdirectly irreducible member of V is finite*).

Proof First an approach which doesn't quite work. Let $\mathbf{A}$ be subdirectly irreducible with monolith μ and suppose a, b are distinct elements in the same μ–block. Then *any* $f \in \{a,b\}^A$ is Con$\mathbf{A}$–compatible. Let $\mathbf{A} = \langle A, F\rangle$ and $Pol_{\mathbf{A}}$ be the set of polynomials of $\mathbf{A}$. Then

$$|Pol_{\mathbf{A}}| \leq |A| + |F| + \aleph_0.$$

Now if $\mathbf{A}$ has at most countable type ($|F| \leq \aleph_0$),

$$|A| < |\{a,b\}^A| \leq |Pol_{\mathbf{A}}| \leq |A| + \aleph_0$$

so $\aleph_0 \leq |A|$ is impossible. This shows that *any* subdirectly irreducible affine complete algebra *of at most countable type* is finite. This is both more and less than the claim of the theorem but suggests how to proceed in general. (Also this explains why a functionally complete algebra of at most countable type must be finite, and hence why finiteness is usually included in the definition of functional completeness.)

Suppose now that $|F|$ is arbitrary and $\aleph_0 \leq |A|$, so that

$$|Pol_{\mathbf{A}}| \leq |A| + |F|.$$

As above let a, b be distinct but congruent mod μ. Define $f{:}A^3 \to A$ for all $u, v, x \in A$ by

$$\begin{aligned} f(u,u,a) &= a \\ f(u,v,x) &= b \text{ if } v \neq u \text{ or } x \neq a. \end{aligned}$$

Then f is Con$\mathbf{A}$–compatible. Hence for some integer k and $(k+3)$–ary term $t(x,y,z,\mathbf{w})$,

$$f(x,y,z) = t(x,y,z,\mathbf{c})$$

for some $\mathbf{c} \in A^k$. Consider the first order sentence

$$\Phi := (\exists x,y)\{x \neq y \wedge (\exists \mathbf{c})(\forall u,v)[t(u,u,x,\mathbf{c}) = x \wedge t(u,u,y,\mathbf{c}) = y$$

$$\wedge (u \neq v \to t(u,v,x,\mathbf{c}) = t(u,v,y,\mathbf{c}))]\}.$$

Taking $x = a$ and $y = b$ we see that $\mathbf{A} \models \Phi$. Since $\aleph_0 \leq |A|$, by the upwards Löwenheim–Skolem Theorem, there is $\mathbf{B} \in V$ with

$$|F| + \aleph_0 \leq |B| \tag{2}$$

and such that $\mathbf{B} \models \Phi$. Accordingly, there are distinct $x, y \in B$ and $\mathbf{c} \in B^k$ such that for any distinct $u, v \in B$,

$$x = t(u,u,x,\mathbf{c})\theta(u,v)t(u,v,x,\mathbf{c}) = t(u,v,y,\mathbf{c})\theta(u,v)t(u,u,y,\mathbf{c}) = y.$$

Hence the pair (x,y) is in any non–zero congruence of $\mathbf{B}$. Therefore $\mathbf{B}$ is subdirectly irreducible. But then by (2) we have

$$|Pol_{\mathbf{B}}| \leq |B| + |F| + \aleph_0 = |B| < |\{x,y\}^B| \leq |Pol_{\mathbf{B}}|,$$

a contradiction. Hence A must be finite.

Notice that the proof does not say that an infinite subdirectly irreducible affine complete algebra does not exist (unless the type is at most countable); rather, it can be interpreted to say that in a variety containing an infinite subdirectly irreducible affine complete algebra there is always another algebra which is subdirectly irreducible but not affine complete.

CONGRUENCE DISTRIBUTIVE AFFINE COMPLETE VARIETIES

The following is a consequence of tame congruence theory.

Theorem 4.2 (*R. McKenzie, unpublished*) *If V is a locally finite affine complete variety then V is congruence distributive.*

Problem 4.1 *Is there a non congruence distributive affine complete variety?*

It is not at all clear how to attack Problem 4.1. In any case the remainder of our discussion will focus on congruence distributive affine complete varieties. In this setting we can say a great deal about affine complete varieties.

If $\mathbf{B}$ is an affine complete subalgebra of some arbitrary algebra $\mathbf{A}$, then each Con$\mathbf{B}$–compatible function $f{:}B^k \to B$ has at least one Con$\mathbf{A}$–compatible extension to $A^k \to A$, namely the polynomial which agrees with f on B^k. In a congruence distributive setting this extension is unique:

Theorem 4.3 *If V is a congruence distributive affine complete variety, $\mathbf{B}$ a subalgebra of $\mathbf{A} \in V$, and if $f{:}B^k \to B$ is Con$\mathbf{B}$-compatible, then there is exactly one Con$\mathbf{A}$-compatible extension of f to $A^k \to A$.*

Proof For simplicity of notation suppose $k = 1$. We suppose the theorem is false, namely that f has extensions g and h and that $g(u) \neq h(u)$ for some $u \in A \setminus B$. Let

$$D = \{\mathbf{a} = (a_1, a_2, ...) \in A^\omega : a_i \text{ are eventually constant and in } B\}.$$

Certainly $\mathbf{D} \in V$. Define $c{:}D \to D$ by

$$\begin{aligned} c(\mathbf{x})_n &= g(x_n) \;\; n \text{ odd}, \\ &= h(x_n) \;\; n \text{ even}. \end{aligned}$$

Since $\mathbf{x}$ is eventually constant in B, so is $c(\mathbf{x})$ and thus $c{:}D \to D$.

Next we show that c is Con$\mathbf{D}$–compatible. This requires that we show that

$$(c(\mathbf{x}), c(\mathbf{y})) \in \theta(\mathbf{x}, \mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in D$. Now for some integer N and elements $a, b \in B$, we have

$$\mathbf{x} = (x_1, x_2, \ldots, x_N, \mathbf{a})$$

and

$$\mathbf{y} = (y_1, y_2, \ldots, y_N, \mathbf{b})$$

where $\mathbf{a} = (a, a, a, \ldots)$ and $\mathbf{b} = (b, b, b, \ldots)$ are constants. But

$$\mathbf{D} = \mathbf{A}^N \times \mathbf{D}_N, \quad \mathbf{D}_N \cong \mathbf{D},$$

where D_N is the result of erasing everything before the $(N+1)$–st components of the members of D, and congruence distributivity implies that

$$\theta(\mathbf{x}, \mathbf{y}) = \theta(x_1, y_1) \times \cdots \times \theta(x_N, y_N) \times \theta(\mathbf{a}, \mathbf{b})$$

in Con$\mathbf{D}$. Also $(c(\mathbf{x}), c(\mathbf{y})) =$

$$((g(x_1), g(y_2)), (h(x_2), h(y_2)), \ldots, (g(x_N), g(y_N)), (f(\mathbf{a}), f(\mathbf{b})))$$

(assuming without loss of generality that N is odd). Consequently c is Con$\mathbf{D}$–compatible. Therefore for some, say $(m+1)$–ary term, t,

$$c(\mathbf{x}) = t(\mathbf{x}, \mathbf{d}^1, \ldots, \mathbf{d}^m)$$

where each $\mathbf{d}^i = (d_1^i, d_2^i, \ldots)$ is eventually constant, say d^i in B. Hence for n_0 sufficiently large and $n \geq n_0$

$$c(\mathbf{x})_n = t(x_n, d^1, \ldots, d^m).$$

For such an n select $\mathbf{x} \in D$ with $x_n = x_{n+1} = u$. Then

$$g(u) = c(x_n) = t(x_n, d^1, \ldots, d^m) = t(x_{n+1}, d^1, \ldots, d^m) = c(x_{n+1}) = h(u),$$

a contradiction.

Theorem 4.3 has several important consequences:

Corollary 4.1 *If $V(\mathbf{A})$ is a congruence distributive affine complete variety and $\mathbf{B}$ is a subalgebra of $\mathbf{A}$, then $V(\mathbf{B}) = V(\mathbf{A})$.*

Proof For if p and q are terms and $p \approx q$ is one of the defining equations of $V(\mathbf{B})$, then by Theorem 4.3 $p \approx q$ is also an equation of $\mathbf{A}$. Hence $V(\mathbf{B}) = V(\mathbf{A})$.

Notice that in Theorem 4.3 it is possible for the subalgebra $\mathbf{B}$ of $\mathbf{A}$ to have only a single element. Hence we have

Corollary 4.2 *If $\mathbf{A}$ is a non-trivial algebra in a congruence distributive affine complete variety, then $\mathbf{A}$ has no trivial subalgebras.*

Corollary 4.3 *If V is a congruence distributive affine complete variety, then the subdirectly irreducible algebras in V have no proper subalgebras.*

Proof Suppose $\mathbf{B} < \mathbf{A}$ where $\mathbf{A}$ is subdirectly irreducible and $V(\mathbf{A})$ is congruence distributive affine complete. The Corollary above asserts that $\mathbf{A} \in V(\mathbf{B}) = IP_SHS(\mathbf{B})$ by the finiteness of B. Hence the subdirect irreducibility of $\mathbf{A}$ asserts that $\mathbf{A} \in HS(\mathbf{B})$ so $B = A$.

The following theorem is the best result we presently have regarding the (finite) sizes of the subdirectly irreducibles in an affine complete variety.

Theorem 4.4 *If V is congruence distributive affine complete and has at most countable type, then for some integer N, $|A| \leq N$ for all subdirectly irreducible members of V, (i.e.: V is residually $\leq N$ for some integer N).*

Proof We know that each subdirectly irreducible is finite and has no proper subalgebras. If we suppose their sizes are unbounded then we can choose a sequence of them $\mathbf{A}_1, \mathbf{A}_2, \ldots$ with $|A_i| < |A_{i+1}|$ for all i. As in the proof of Theorem 4.3 we again use a combinatorial argument to obtain a contradiction. For each $\mathbf{A}_i$ let μ_i be the monolith and choose $\mathbf{a} = (a_1, a_2, \ldots) \in \prod A_i$ such that $|a_i/\mu_i| \geq 2$ for all i. Define $\mathbf{B}$ by taking

$$B = \{\mathbf{b} \in \prod A_i : \text{ for some integer } k \text{ and unary term } t,\ i > k \Rightarrow b_i = t(a_i)\},$$

(i.e.: B is the set of all elements in $\prod A_i$ which are eventually in the subalgebra generated by $\mathbf{a}$.) Then $\mathbf{B}$ is a subalgebra of $\prod \mathbf{A}_i$ and is subdirect since the $\mathbf{A}_i$ have no proper subalgebras. $\mathbf{B}$ is countable because V has at most countable type. Hence $|Pol_{\mathbf{B}}| \leq \aleph_0$. We obtain a contradiction by showing how to construct $2^{\aleph_0}$ Con$\mathbf{B}$–compatible functions $f{:}B \to B$.

First, by the countability of B, enumerate its elements in some fashion:

$$B = \{\mathbf{b}^1, \mathbf{b}^2, \ldots\}.$$

Define f inductively by:

i) $f(\mathbf{b}^1) = \mathbf{a}^1 = \mathbf{a}$

ii) if $f(\mathbf{b}^1), \ldots, f(\mathbf{b}^{k-1})$ are defined, take $f(\mathbf{b}^k) =$ any $\mathbf{a}^k \in B$ satisfying

  a) $\mathbf{a}^k$ is eventually equal to $\mathbf{a}$,

  b) $a_i^k \in a_i/\mu_i$ for all i, and,

  c) $(\forall j)(\forall m < k)(b_j^k = b_j^m \Rightarrow a_j^k = a_j^m)$.

Obviously at least one such $\mathbf{a}^k \in B$, namely $\mathbf{a}^k = \mathbf{a}$, always exists. Further, using the fact that the $|A_i|$ grow larger, it is not hard to see that for arbitrarily large m there is always $\mathbf{b}^m \in B$ such that for some index j, $b_j^m \notin \{b_j^1, \ldots, b_j^{m-1}\}$, and this means that c) imposes no restriction so that that we can choose a_j^m to be an arbitrary element in a_j/μ_j. But this means that for infinitely many values of k the choice of $\mathbf{a}^k$ satisfying $a)$, $b)$ and $c)$ is not unique and hence there are $2^{\aleph_0}$ choices for f.

To complete the proof it is necessary to show that f is Con**B**–compatible, which means

$$(\mathbf{a}^i, \mathbf{a}^j) \in \theta(\mathbf{b}^i, \mathbf{b}^j) \text{ for all } i, j. \tag{3}$$

But by $a)$, $a_k^i = a_k^j = a_k$ for all but finitely many k; wlog we can rearrange the factors so that for some N

$$a_1^i \neq a_1^j, \ldots, a_N^i \neq a_N^j \tag{4}$$

and

$$a_k^i = a_k^j \text{ for all } k > N. \tag{5}$$

For this N we also have

$$\mathbf{B} = \mathbf{A}_1 \times \cdots \times \mathbf{A}_N \times \mathbf{B}'_N$$

where the elements of B'_N are obtained by deleting the first N components of the members of B. Writing $\mathbf{b} \in B$ as

$$\mathbf{b} = (b_1, \ldots, b_N, \mathbf{b}_N), \ \ b_i \in A_i, \ \mathbf{b}_N \in \mathbf{B}'_N,$$

congruence distributivity then implies

$$\theta(\mathbf{b}^i, \mathbf{b}^j) = \theta(b_1^i, b_1^j) \times \cdots \times \theta(b_N^i, b_N^j) \times \theta(\mathbf{b}_N^i, \mathbf{b}_N^j).$$

Now condition $c)$ together with (4) implies $\mu_k \leq \theta(b_k^i, b_k^j)$, $k < N$, so from condition $b)$ we have

$$(a_k^i, a_k^j) \in \mu_k \leq \theta(b_k^i, b_k^j), \ \ k < N;$$

On the other hand (5) implies $\mathbf{a}_N^i = \mathbf{a}_N^j$ so this pair is in $\theta(\mathbf{b}_N^i, \mathbf{b}_N^j)$. Hence (3) holds.

Problem 4.2 *Is Theorem* 4.4 *true for varieties of uncountable type?*

It seems likely that this problem has an affirmative answer. In any case if V has finite type then from Theorem 4.4 we have that the number of non-isomorphic subdirectly irreducible members of V is finite. Their direct product generates V and by Corollary 4.1 so does a minimal subalgebra of the product. Hence we have

Theorem 4.5 *If V is a congruence distributive affine complete variety of finite type then V is generated by a finite algebra* $\mathbf{A}$ *having no proper subalgebras:* $V = IP_SH(\mathbf{A})$.

Combining Theorem 4.5 with Theorem 3.1 we obtain

Theorem 4.6 *An arithmetical variety of finite type is affine complete iff it is generated by a finite algebra having no proper subalgebras.*

We remark that while there are general methods for constructing congruence distributive affine complete varieties which are not congruence permutable (i.e.: not arithmetical)—one such method is described in detail below—the central problem concerning affine complete varieties remains open:

Problem 4.3 *Describe the finite algebras* **A** *in a congruence distributive variety which generate affine complete subvarieties.*

The variety of bounded distributive lattices shows that it is not enough for **A** to have no proper subalgebras; in this example the generating algebra itself is not affine complete. Hence one might ask: does a finite affine complete algebra (in a congruence distributive variety) having no proper subalgebras generate an affine complete variety. Unfortunately the answer to this question is also negative: K. Kaarli [10] has given an example of such an algebra which has a non-affine complete quotient algebra. In fact, from the remark following Theorem 3.1, we can restate Problem 4.3 more precisely as

> *For a congruence distributive variety V generated by a finite algebra having no proper subalgebras, when is each finite algebra in V affine complete?*

For an arithmetical variety we know, of course, that all of the finite members are affine complete.

NON-ARITHMETICAL AFFINE COMPLETE VARIETIES

We conclude this section by describing one instructive general method for constructing finitely generated affine complete varieties which are congruence distributive but not congruence permutable. This is the simplest of two methods described by Kaarli in [10] and generalizes (and simplifies) the construction given in [11].

For an integer m let $A_1, \ldots, A_m$ be finite sets each having at least two elements. Let $\mathbf{a} = (a_1, \ldots, a_m) \in A_1 \times \cdots \times A_m$ be fixed and let C be any subset of $A_1 \times \cdots \times A_m$ with the properties:

i) C projects onto each A_i, and

ii) for all $\mathbf{b} = (b_1, \ldots, b_m) \in C$, $\mathbf{b}\prime = (b_1\prime, \ldots, b_m\prime)$ is also in C provided that for each i, $b_i\prime$ equals either b_i or a_i.

Observe that there are many choices for C and, in particular, C can be chosen in many ways so that the kernels of the projections are not all pairwise permutable. For example, the smallest choice possible is

$$C = \bigcup_{i=1}^{m} \{a_1\} \times \cdots \times \{a_{i-1}\} \times A_i \times \{a_{i+1}\} \times \cdots \times \{a_m\}.$$

A largest possible (non–permutable) choice can be obtained by choosing $(a_{1'}, a_{2'}) \in A_1 \times A_2$ with $a_{1'} \neq a_1$ and $a_{2'} \neq a_2$, and then taking

$$C = \{(x_1, \ldots, x_m) \in A_1 \times \cdots \times A_m : x_1 \neq a_{1'} \text{ or } x_2 \neq a_{2'}\}.$$

For an integer n and any functions $g^i : A_i^n \to A_i$, $i = 1, \ldots, m$, let $g = g^1 \times \cdots \times g^m$ denote the combined componentwise actions of the g^i from $(A_1 \times \cdots \times A_m)^n$ to $A_1 \times \cdots \times A_m$. Let F be the collection of all such g, for all non–negative integers n, and such that C is closed under g. Specifically, this means that for $\mathbf{z}^1, \ldots, \mathbf{z}^n \in A_1 \times \cdots \times A_m$, if

$$g(\mathbf{z}^1, \ldots, \mathbf{z}^n) = (g^1(z_1^1, \ldots, z_1^n), \ldots, g^m(z_m^1, \ldots, z_m^n)) \notin C$$

then some $\mathbf{z}^k \notin C$. (We shall find this contrapositive form most useful in the sequel.) Another way to vizualize this closure condition is the following: for every $n \times m$ matrix M whose rows are elements of C, application of the component functions of $g = g^1 \times \cdots \times g^m$ to the respective columns of M should yield a row vector which is also in C, i.e.: if application of the g^i to the columns of M yields an m–tuple not in C then one of the rows of M is not in C.

With F defined as above let $\mathbf{C} = \langle C; F \rangle$. Also let $\mathbf{A}_i = \langle A_i; F_i \rangle$ where $F_i = \{g^i : g = g^1 \times \cdots \times g^i \times \cdots \times g^m \in F\}$, i.e.: the $\mathbf{A}_i$ are the componentwise projections of $\mathbf{C}$. Notice that each of the $\mathbf{A}_i$ is primal; this is clear since for an arbitrary n–ary g^i, $g = a_1 \times \cdots \times g^i \times \cdots \times a^m$ is in F (where the a_i here denote n–ary constant functions). Also notice that $\mathbf{C}$ has no proper subalgebras since each of the nullary constant functions is in F.

More important, for each $i = 1, \ldots, m$, let u^i be the $(m+1)$–ary near unanimity function on A_i which has value a_i if it is not the case that m of its arguments agree. Then $u = u^1 \times \cdots \times u^m$ is a near unanimity function on C. We must show that $u \in F$. To do this consider an $((m+1) \times m)$ matrix M (with i–th column entries from A_i) and suppose for some $k \leq m$, M has an $((m+1) \times k)$–ary submatrix in which each of the columns has m identical entries. Then by the pigeon–hole principle, application of the u^i to the columns of this submatrix will yield one of its $m+1$ rows while application of the appropriate u^i to the other columns of M will yield a_i. Hence $u \in F$. From this it follows, by Proposition 2.5 (or Mitschke's theorem [14]), that the variety generated by $\mathbf{C}$ is congruence distributive and also, by Corollary 2.2, that the clone F is finitely generated, so that we may take the algebra $\mathbf{C}$ to be of finite type. With the appropriate choice of C, $V(\mathbf{C})$ is, of course, not generally arithmetical.

Now we show that $V = V(\mathbf{C})$ is affine complete. By congruence distributivity, if $\mathbf{B} \in V$, then with no loss of generality we may take

$$\mathbf{B} < \mathbf{A}_1^{I_1} \times \cdots \times \mathbf{A}_m^{I_m}$$

for non-empty disjoint index sets I_i with $\bigcup I_i = I$. Also, since every element of C is a nullary operation, $\mathbf{B}$ contains a copy $\mathbf{C}'$ of $\mathbf{C}$ which we henceforth identify with $\mathbf{C}$. Let $f : B^n \to B$ be Con$\mathbf{B}$–compatible. Then for each $j \in I$, if $j \in I_i$, f induces a function $f_j : A_i^n \to A_i$. Let

$$\{d^1, \ldots, d^p\} = f(C') = \{f(z^1, \ldots, z^n) : z^1, \ldots, z^n \in C'\}$$

and note that since $\mathbf{C}\prime < \mathbf{B}$, $f(C\prime) \subset B$. Define $h = h^1 \times \cdots \times h^m$, $h^i: A_i^{n+p} \to A_i$, by

$$\begin{aligned} h^i(x_1,\ldots,x_n,u_1,\ldots,u_p) &= f_j(x_1,\ldots,x_n) \text{ if for some } j \in I_i, \\ &\quad (u_1,\ldots,u_p) = (d_j^1,\ldots,d_j^p), \\ &= a_i \text{ otherwise.} \end{aligned}$$

We must show that, i) the h^i are well defined, and ii) C is closed under h. For having shown i) and ii), then for any $j \in I$, say $j \in I_i$, we have

$$\begin{aligned} h^{\mathbf{B}}(x^1,\ldots,x^n,d^1,\ldots,d^p)_j &= h^{\mathbf{A}_i}(x_j^1,\ldots,x_j^n,d_j^1,\ldots,d_j^p) \\ &= h^i(x_j^1,\ldots,x_j^n,d_j^1,\ldots,d_j^p) \\ &= f_j(x_j^1,\ldots,x_j^n) = f(x^1,\ldots,x^n)_j. \end{aligned}$$

(The superscripts indicate the algebra in which the operation is interpreted.) Therefore the polynomial

$$h^{\mathbf{B}}(x^1,\ldots,x^n,d^1,\ldots,d^p) = f(x^1,\ldots,x^n)$$

for all $x^1,\ldots,x^n \in B$, so that V is affine complete.

To verify i) (h^i is well defined), suppose that for some $j_1, j_2 \in I_i$,

$$(d_{j_1}^1,\ldots,d_{j_1}^p) = (d_{j_2}^1,\ldots,d_{j_2}^p).$$

Let $x_1,\ldots,x_n \in A_i$. Also for $k = 1,\ldots,n$, let $c_k^{\mathbf{C}} \in F$ be nullary functions such that $c_k^{\mathbf{A}_i} = x_k$. (This can be done by taking $c_k^{\mathbf{A}_r} = a_r$ if $r \neq i$.) Then $f(c_1^{\mathbf{C}},\ldots,c_n^{\mathbf{C}}) \in f(C)$ and hence equals some $d^k \in \{d^1,\ldots,d^p\}$. Thus

$$\begin{aligned} d_{j_1}^k = f(c_1^{\mathbf{C}},\ldots,c_n^{\mathbf{C}})_{j_1} &= f_{j_1}(c_1^{\mathbf{A}_i},\ldots,c_n^{\mathbf{A}_i}) \\ &= f_{j_1}(x_1,\ldots,x_n) \end{aligned}$$

and likewise $d_{j_2}^k = f_{j_2}(x_1,\ldots,x_n)$. But $d_{j_1}^k = d_{j_2}^k$ so $f_{j_1}(x_1,\ldots,x_n) = f_{j_2}(x_1,\ldots,x_n)$. Hence h^i is well defined.

Finally, to show ii) (C is closed under h), suppose that

$$\begin{aligned} \mathbf{v} &= h(\mathbf{z}^1,\ldots,\mathbf{z}^n,\mathbf{w}^1,\ldots,\mathbf{w}^p) \\ &= (h^1(z_1^1,\ldots,z_1^n,w_1^1,\ldots,w_1^p),\ldots,h^m(z_m^1,\ldots,z_m^n,w_m^1,\ldots,w_m^p)) \notin C. \end{aligned}$$

Then this function value $\mathbf{v}$ is an element of $A_1 \times \cdots \times A_m$ which differs from $a = (a_1,\ldots,a_m)$ on at least two components. For ease of notation and wlog let us suppose that this function value is

$$\mathbf{v} = (a_1\prime, a_2\prime, a_3, a_4, \ldots, a_m)$$

where $a_i\prime \in A_i$, $a_i\prime \neq a_i$, $i = 1,2$. We remark that it follows that for any choices of $b_i \in A_i$,

$$\mathbf{v}\prime = (a_1\prime, a_2\prime, b_3, b_4, \ldots, b_m) \notin C.$$

We need to show that either some $\mathbf{z}^k \notin C$ or some $\mathbf{w}^k \notin C$. Our assumption about $\mathbf{v}$ implies that for some $j_1 \in I_1, j_2 \in I_2$, we have

$$\begin{aligned} (w_1^1, \ldots, w_1^p) &= (d_{j_1}^1, \ldots, d_{j_1}^p), \\ \text{and } (w_2^1, \ldots, w_2^p) &= (d_{j_2}^1, \ldots, d_{j_2}^p), \end{aligned}$$

and consequently

$$\begin{aligned} f_{j_1}(z_1^1, \ldots, z_1^n) &= a_{1'}, \\ \text{and } f_{j_2}(z_2^1, \ldots, z_2^n) &= a_{2'}. \end{aligned}$$

Now if some $\mathbf{z}^k = (z_1^k, z_2^k, z_3^k, \ldots, z_m^k) \notin C$ then we are done; otherwise all $\mathbf{z}^1, \ldots, \mathbf{z}^n \in C$ so we may take them to be nulllary operations such that the value of $\mathbf{z}^k$ in A_i is z_i^k. Then it follows that $f(\mathbf{z}^1, \ldots, \mathbf{z}^n) = d^k$ for some $d^k \in \{d^1, \ldots, d^p\}$. Hence we have, for this k,

$$\begin{aligned} w_1^k = d_{j_1}^k = f(\mathbf{z}^{1\,\mathbf{C}}, \ldots, \mathbf{z}^{n\,\mathbf{C}})_{j_1} &= f_{j_1}(\mathbf{z}^{1\,\mathbf{A}_1}, \ldots, \mathbf{z}^{n\,\mathbf{A}_1}) \\ &= f_{j_1}(z_1^1, \ldots, z_1^n) = a_{1'}. \end{aligned}$$

Likewise, $w_2^k = d_{j_2}^k = \cdots = a_{2'}$, and hence

$$\mathbf{w}^k = (a_{1'}, a_{2'}, \ldots)$$

and therefore, as remarked above, cannot be in C no matter what the remaining components are. This proves ii).

5 Arithmetical affine complete varieties

By a theorem of R. Magari every variety contains a simple algebra. By Theorem 4.1 an affine complete variety must even contain a *finite* simple, and hence functionally complete algebra $\mathbf{A}$. By Theorem 4.2, $\mathbf{A}$ must generate a congruence distributive variety and hence by Corollary 4.3 $\mathbf{A}$ has no proper subalgebras. Hence

Proposition 5.1 *A minimum affine complete variety is always generated by a finite functionally complete algebra having no proper subalgebras.*

It turns out that the converse of Proposition 5.1 is also true and in fact, a finite functionally complete algebra having no proper subalgebras even generates an arithmetical variety. (Theorem 5.2.) Hence the minimal affine complete varieties are arithmetical and are just of this type. But not only this is true: in the sequel we will exhibit an interesting class of finite, arithmetical (not just simple), affine complete algebras having no proper subalgebras which generate arithmetical (and hence affine complete) varieties. The functionally complete algebras will be a special case. The results of this section will appear in greater detail in [12]. They are motivated by results of Kaarli in [9].

Definition For brevity we shall say that an algebra is an FACS *algebra* if is

i) finite,
ii) arithmetical,
iii) affine complete, and
iv) has no proper subalgebras.

Notice that ii) and iii) can be replaced by the single requirement that the algebra be locally affine complete (by Theorem 3.3). From Theorem 4.6 we know that an affine complete arithmetical variety of finite type is necessarily generated by a FACS algebra. Hence we want to examine the question:

If $\mathbf{A}$ *is a* FACS *algebra, when is* $V(\mathbf{A})$ *arithmetical and hence affine complete?*

A technical (and not very interesting) characterization of those FACS algebras generating arithmetical varieties is available in [12]. Hence we look for interesting sufficient conditions assuring arithmeticity. We can obtain one quickly from Theorem 2.9:

Proposition 5.2 *If* $\mathbf{A}$ *is finite, arithmetical, and has no proper subalgebras, and if the term functions of* $\mathbf{A}$ *are the functions* $f{:}A^k \to A$ *which are both* Con$\mathbf{A}$*- and* Aut$\mathbf{A}$*-compatible, then* $V(\mathbf{A})$ *is arithmetical.*

Proof Applying Theorem 2.9, suppose the condition of the Proposition holds and let f be compatible with all rectangular (= graph) subalgebras of $\mathbf{A} \times \mathbf{A}$. Since the congruences and the automorphisms are rectangular subuniverses, f is a term function.

Since a function is Aut$\mathbf{A}$–compatible iff it commutes with each automorphism, we have

Corollary 5.1 *If* $\mathbf{A}$ *is a FACS algebra and the terms of* $\mathbf{A}$ *are precisely the polynomials which commute with each automorphism, then* $V(\mathbf{A})$ *is arithmetical.*

(The converses of both the Proposition and the Corollary are false.) While the Theorem and Corollary are interesting, at least one problem with the Corollary is that it gives no clue as to why a simple FACS algebra (i.e.: a functionally complete algebra having no proper subalgebras) should generate an arithmetical variety. In the remainder of this section we will give another interesting sufficient condition, which turns out to be a special case of the condition of the corollary, and which clearly implies $V(\mathbf{A})$ arithmetical for a simple FACS algebra.

The Cross Lemma. We remarked earlier in Section 2 that any subalgebra of the product of a pair of algebras is contained in its unique rectangular hull in the product. Suppose now that $\mathbf{B}$ is a subdirect product in $\mathbf{A} \times \mathbf{A}$ and consider the rectangular hull defined as a graph subalgebra $\sigma{:}\mathbf{A}/\phi_1 \to \mathbf{A}/\phi_2$. We want to observe a special property of this rectangular hull in case $\mathbf{A}$ is a FACS algebra. First, since A is finite, the congruences ϕ_i are defined by finite joins of principal congruences. For example, let

$$\begin{gathered}((a_1, b_1), (a_1, c_1)) \\ ((a_2, b_2), (a_2, c_2)) \\ \cdots \\ ((a_m, b_m), (a_m, c_m))\end{gathered}$$

be a listing of the set of all pairs $((x, y), (x\prime, y\prime))$ of elements of B with $x = x\prime$. Then

$$\phi_2 = \bigvee_{1 \le i \le m} \theta(b_i, c_i).$$

Now let N be any ϕ_2–block and choose the integer n so large that $2^n \ge |N|$.

Consider mn–tuples $\mathbf{x}$ of elements of A, thinking of such an $\mathbf{x}$ as consisting of a sequence of m constituent n–tuples, laid in a line and numbered from left to right (from 1 to m). Let K be the subset of all such mn–tuples with the special properties:

a) for each $i = 1, \ldots, m$ the elements of the i–th constituent n–tuple are either b_i or c_i;

b) the same pattern of b's and c's occurs in each of the m constituent n–tuples occurring in $\mathbf{x}$.

Because of these conditions, $|K| = 2^n$ and if $\mathbf{x}$,$\mathbf{y}$ are distinct elements of K, then

$$\theta(\mathbf{x}, \mathbf{y}) = \bigvee_{1 \le i \le mn} \theta(x_i, y_i) = \bigvee_{1 \le i \le m} \theta(b_i, c_i) = \phi_2.$$

Now choose $f{:}K \to N$ to be any function of K onto N, which is possible since $|K| = 2^n \ge |N|$. Then f is a partial Con$\mathbf{A}$–compatible function in $A^{mn} \to A$. If we further suppose that $\mathbf{A}$ is arithmetical, f has a Con$\mathbf{A}$–compatible extension with domain all of A^{mn} (and which we still denote by f). (We have used the compatible function extension property, Theorem 2.4, to make this extension.)

Next, if we also suppose that $\mathbf{A}$ is affine complete then f is a polynomial, i.e.:

$$f(\mathbf{x}) = t(\mathbf{x}, \mathbf{a})$$

for some $(mn + k)$–ary term t and $\mathbf{a} \in A^k$. Finally, if $\mathbf{A}$ has no proper subalgebras, then the unary terms act transitively on A, which implies that for any $a \in A$ there are terms $u_i(x)$ such that $a_i = u_i(a)$, $i = 1, \ldots, k$. Hence we have a $(mn + 1)$–ary term t and $a \in A$ such that

$$f(\mathbf{x}) = t(\mathbf{x}, a).$$

Since B is subdirect in $A \times A$, $(b, a) \in B$ for some $b \in A$.

Now consider the elements of B obtained as values of the polynomial

$$t((x_1, y_1), \ldots, (x_{mn}, y_{mn}), (b, a)) = (t(x_1, \ldots, x_{mn}, b), t(y_1, \ldots, y_{mn}, a))$$

where the mn–tuples $((x_1, y_1), \ldots, (x_{mn}, y_{mn}))$ are chosen so that the elements of the i–th constituent n–tuple are either (a_i, b_i) or (a_i, c_i) and so that $(y_1, \ldots, y_{mn})$ is in K. Then all of these mn–tuples are in B and since (b, a) is also in B, we obtain as function values the elements of B of the form

$$(t(a_1, \ldots, a_1, \ldots, a_m, \ldots, a_m, b), t(y_1, \ldots, y_{mn}, a)) = (c, f(y_1, \ldots, y_{mn}))$$

where c is the constant $t(a_1, \ldots, a_1, \ldots, a_m, \ldots, a_m, b)$ in A. Then since f maps K onto N, the elements of B so obtained are just $\{c\} \times N$.

In a completely analogous way, for any ϕ_1–block M we can find an element $d \in A$ with $M \times \{d\} \subset B$. Observe that for any ϕ_1–block M there is precisely one ϕ_2–block N such that $(M \times N) \cap B \neq \emptyset$. This is immediate since $\sigma: A/\phi_1 \rightarrow A/\phi_2$ is an isomorphism.

Summarizing, we have proved that B contains "crosses", one in each of the product blocks of the rectangular hull of $\mathbf{B}$. (Picture a cross as consisting of vertical and horizontal "line segments" inscribed in the product block.) We call this observation the "Cross Lemma".

Lemma 5.1 (*Cross Lemma*) *Let* $\mathbf{A}$ *be a FACS algebra,* $\mathbf{B} < \mathbf{A} \times \mathbf{A}$*, and suppose the rectangular hull of* $\mathbf{B}$ *is defined by* $\sigma: A/\phi_1 \rightarrow A/\phi_2$*. Then for any* $(a_1, a_2) \in A \times A$ *with* $\sigma(a_1/\phi_1) = a_2/\phi_2$*, there exists* $(c, d) \in a_1/\phi_1 \times a_2/\phi_2$ *such that*

$$(\{c\} \times a_2/\phi_2) \cup (a_1/\phi_1 \times \{d\}) \subset B.$$

The Cross Lemma has several applications to FACS algebras. We need only the following one.

Lemma 5.2 *If* $\mathbf{A}$ *is a FACS algebra,* θ *an atom in* $\mathbf{ConA}$*, and* B *is a subuniverse in* $A \times A$ *which is contained in* θ*, then* B *is either the graph of a non–trivial automorphism or* B *contains* Δ *(the graph of the trivial automorphism).*

Proof Let $\mathbf{H}$ be the rectangular hull of $\mathbf{B}$. Since $B \subset \theta$ and θ is a rectangular subuniverse in $\mathbf{A} \times \mathbf{A}$, it follows that $B \subset H \subset \theta$. Therefore if $\mathbf{H}$ is defined by the isomorphism $\sigma: A/\phi_1 \rightarrow A/\phi_2$, then $0 \leq \phi_1, \phi_2 \leq \theta$. Then since θ is an atom, by the finiteness of A, either

a) $\phi_1 = \phi_2 = 0$, and B is the graph of an automorphism, or

b) $\phi_1 = \phi_2 = \theta$, and σ is the identity automorphism of $\mathbf{A}/\theta$.

In case b), by the Cross Lemma $B \cap \Delta \neq \emptyset$. Since $\mathbf{A}$ has no proper subalgebras, Δ is a minimal subuniverse and hence we conclude $\Delta \subset B$.

Lemma 5.3 *Let* $\mathbf{A}$ *be a FACS algebra and let* θ *be the join of the atoms of* $\mathbf{ConA}$*. If there is a unary term* $t(x)$ *such that* $t(A)$ *is contained in some* θ*–block, then there is also a unary term* $u(x)$ *such that* $u(A)$ *is contained in some orbit of* $\mathbf{AutA}$*.*

Proof Among all unary terms $t(x)$ for which $t(A)$ lies in some θ–block, choose one, $u(x)$, such that $|u(A/\phi)|$ is minimal for all $\phi \in \mathrm{Con}\mathbf{A}$. This can be done by choosing, for each $\phi \in \mathrm{Con}\mathbf{A}$, a term u_ϕ such that $|u_\phi(A/\phi)|$ is minimal, and then taking u to be the composition of all such u_ϕ.

Let $b, c \in u(A)$. We show that the subalgebra $\mathbf{B} < \mathbf{A} \times \mathbf{A}$ generated by (b, c) is the graph of an automorphism of $\mathbf{A}$. This will show that $u(A)$ is contained in the c–orbit of $\mathbf{AutA}$. Now

$$B = \{(t(b), t(c)) : t \in T\}$$

where T is the set of unary terms. Hence we need to show that

$$t_1(b) = t_2(b) \;\Leftrightarrow\; t_1(c) = t_2(c) \tag{6}$$

for all $t_1, t_2 \in T$. Suppose, on the contrary, that for some t_1, t_2,

$$t_1(b) = t_2(b) \text{ and } t_1(c) \neq t_2(c).$$

Let Φ be the set of all congruences covered by θ. Since θ is the join of all of the atoms, $\bigwedge \Phi = 0$, and hence for some $\phi \in \Phi$

$$t_1(b/\phi) = t_2(b/\phi) \text{ and } t_1(c/\phi) \neq t_2(c/\phi). \tag{7}$$

Observe that $\mathbf{A}/\phi$ is a FACS algebra (by Theorem 3.2). Then apply Lemma 5.2 to $\mathbf{A}/\phi$ in which the congruence θ/ϕ covers 0 in $\mathrm{Con}(\mathbf{A}/\phi)$. By (7) b/ϕ and c/ϕ are distinct. Also $t(b/\phi) \neq t(c/\phi)$ for all $t \in T$, for otherwise $|t \circ u(A/\phi)| < |u(A/\phi)|$, contradicting the minimality of $|u(A/\phi)|$. Therefore the subalgebra of $\mathbf{A}/\phi \times \mathbf{A}/\phi$ generated by the pair $(b/\phi, c/\phi)$ does not contain the diagonal Δ/ϕ. Hence the subalgebra is the graph of an automorphism of $\mathbf{A}/\phi$, and this contradicts (7). Therefore (6) holds, so B is the graph of an automorphism.

Lemma 5.4 *If* $\mathbf{A}$ *is a FACS algebra with a unary term* $u(x)$ *such that* $u(A)$ *is contained in some orbit of* $\mathbf{AutA}$, *then* $V(\mathbf{A})$ *is arithmetical.*

Proof Since $\mathbf{A}$ is arithmetical there is a $\mathrm{Con}\mathbf{A}$–compatible function $f{:}A^3 \to A$ satisfying equations (1) of Theorem 2.1. Since $\mathbf{A}$ is affine complete, f is a polynomial. Since $\mathbf{A}$ has no proper subalgebras the unary terms act transitively on A so that for *any* $a \in A$ there is some 4-ary term $t(x, y, z, w)$ such that

$$f(x, y, z) = t(x, y, z, a).$$

But if σ is any automorphism then for example, $t(x, x, z, \sigma a) =$

$$t(\sigma\sigma^{-1}x, \sigma\sigma^{-1}x, \sigma\sigma^{-1}z, \sigma a) = \sigma t(\sigma^{-1}x, \sigma^{-1}x, \sigma^{-1}z, a) = \sigma\sigma^{-1}z = z,$$

that is, the equations (1) also hold for $t(x, y, z, \sigma a)$. Now suppose that for the term $u(x)$, $u(A)$ lies in the c–orbit of $\mathbf{AutA}$. Choosing the term $t(x, y, z, w)$ accordingly, it follows from the reasoning above that the term $t(x, y, z, u(x))$ also satisfies these equations. Hence $V(\mathbf{A})$ is arithmetical by Theorem 2.1.

Finally, by Lemmas 5.3 and 5.4 we have the following general result:

Theorem 5.1 *Let* $\mathbf{A}$ *be a FACS algebra and let* θ *be the join of the atoms of* $\mathbf{Con}\mathbf{A}$. *If there is some unary term* $t(x)$ *such that* $t(A)$ *is contained in some* θ*–block, then* $V(\mathbf{A})$ *is arithmetical.*

If $\mathbf{A}$ is functionally complete, then $1 \in \mathrm{Con}\mathbf{A}$ is an atom so certainly any $t(A)$ is contained in a 1–block. Hence as a special case of Theorem 5.1 we have Kaarli's theorem [9]: *If* $\mathbf{A}$ *is a finite functionally complete algebra having no proper subalgebras, then* $V(\mathbf{A})$ *is arithmetical.* It follows that $V(\mathbf{A})$ consists of subdirect powers of $\mathbf{A}$ and hence is minimal. (This result is also obtained by Szendrei [19] by different means.) Combining this with Proposition 5.1 we have

Theorem 5.2 *For an affine complete variety* V,

a) *If* V *is minimal then* V *is arithmetical, and*

b) V *is minimal iff* V *is generated by a finite functionally complete algebra having no proper subalgebras.*

WEAKLY DIAGONAL ALGEBRAS

Let us examine the significance of the condition "for some unary term $u(x)$, $u(A)$ is contained in an orbit of $\mathbf{Aut}\mathbf{A}$", which, by Lemma 5.4, is sufficient for $V(\mathbf{A})$ to be arithmetical. If $\mathbf{A}$ has no proper subalgebras and $u(A)$ is contained in the c–orbit of $\mathbf{Aut}\mathbf{A}$, let $\mathbf{S}$ be any subalgebra in $\mathbf{A}^I$ (I any index set). Pick any $\mathbf{b} \in S$; then $u(\mathbf{b}) \in S$ and for any $i \in I$, $u(\mathbf{b})_i = u(b_i) = \sigma_i c$ for some $\sigma_i \in \mathrm{Aut}\mathbf{A}$. Now if $a \in A$ is arbitrary, then for some term t, $a = t(c)$; hence $t(u(\mathbf{b})) \in S$ and for the same σ_i

$$t(u(\mathbf{b}))_i = t(\sigma_i c) = \sigma_i t(c) = \sigma_i a.$$

From this it is apparent that

For any power $\mathbf{A}^I$, *each* $\mathbf{S} < \mathbf{A}^I$ *is isomorphic to a subdirect product in* $\mathbf{A}^I$ *containing the diagonal.*

If $\mathbf{A}$ is a finite algebra having no proper subalgebras and $\mathbf{A}$ has this property, we say that $\mathbf{A}$ is *weakly diagonal.*

Conversely, if $\mathbf{A}$ is weakly diagonal in the above sense, then in particular it follows as a special case that each subuniverse in $A \times A$ contains the graph of an automorphism. Then choose $u(x)$ so that $|u(A)|$ is minimal. Then any term $t(x)$ is a bijection of $u(A)$ onto $t \circ u(A)$ (provided A is finite). Let $a, b \in u(A)$; then the subalgebra of $\mathbf{A} \times \mathbf{A}$ generated by (a, b) contains the graph of an automorphism σ. But this subalgebra consists of all $(t(a), t(b))$, where t is some unary term. Consequently for some t,

$$t(b) = \sigma t(a) = t(\sigma(a)). \tag{8}$$

But $a = u(a\prime)$ for some $a\prime \in A$, so

$$\sigma(a) = \sigma u(a\prime) = u(\sigma(a\prime)) \in u(A).$$

Then by (8), since t is bijective on $u(A)$, we have $b = \sigma(a)$ and therefore $u(A)$ is contained in the a–orbit of $\mathbf{AutA}$.

Notice that if $\mathbf{A}$ is weakly diagonal and rigid then it is *diagonal*, i.e.: every subalgebra of a power of $\mathbf{A}$ contains the diagonal. The two element Boolean algebra is diagonal and weakly diagonal FACS algebras capture many of its properties. See [12] for more details.

Finally observe that if $\mathbf{A}$ is weakly diagonal then $\mathbf{A}$ satisfies the hypothesis (of Corollary 5.1): if a polynomial commutes with all automorphisms, then it is a term. To see this suppose that for some $c \in A$, $u(A)$ is contained in the c–orbit of $\mathbf{AutA}$. Now if f is any polynomial, then (by the transitivity of the unary term functions) for some term and this c, $f(\mathbf{x}) = t(\mathbf{x}, c)$. But then the condition $\sigma f(\mathbf{x}) = f(\sigma \mathbf{x})$, all σ, is just equivalent to the condition: $t(\mathbf{x}, c) = t(\mathbf{x}, \sigma c)$, for all σ. But since $u(x)$ takes on only values σc, it follows that $f(\mathbf{x}) = t(\mathbf{x}, u(x))$, a term.

The converse of the implication just established is not true: there are examples of FACS algebras for which the terms are just the polynomials which commute with all automorphisms, but which are not weakly diagonal. (Again see [12] for details.)

CHARACTERIZING BOOLEAN ALGEBRAS

Any algebra on a two element set is necessarily simple and hence, if it generates an affine complete variety, by Theorem 4.2 this variety is congruence distributive, and hence consists of subdirect powers of the two element generator. Consequently this variety is minimal and Theorem 5.2 applies. Now if one inspects the Post classification of clones on a two element set it is not difficult to observe that there are precisely two clones B and C, such that the algebras $\mathbf{B} = \langle\{0,1\}; B\rangle$ and $\mathbf{C} = \langle\{0,1\}; C\rangle$ each generate affine complete varieties, i.e.: in view of Theorem 5.2, such that the algebras $\mathbf{B}$ and $\mathbf{C}$ have no proper subalgebras and generate arithmetical varieties. Letting t be the ternary discriminator on $\{0,1\}$, and $\prime$ be complementation, one can easily see from the Post classification that these two clones can be generated as follows:

$$B = \langle t, \prime, 0, 1\rangle, \text{ and } B = \langle t, \prime\rangle.$$

In the first instance, $V(\mathbf{B})$ is clearly term equivalent to the variety of Boolean algebras. In the second case $V(\mathbf{C})$ is again all subdirect powers of $\mathbf{C}$, but unlike $\mathbf{B}$, which is rigid, $\mathbf{C}$ has an automorphism (exchanging 0 and 1); $\mathbf{B}$ is diagonal while $\mathbf{C}$ is weakly diagonal. Also the variety $V(\mathbf{B})$ of Boolean algebras can be embedded in $V(\mathbf{C})$ in an obvious way. It would be interesting to know if this "extension" of the variety of Boolean algebras is interesting in other respects. In any case we have established the following characterization of Boolean algebras in terms of affine completeness:

Theorem 5.3 *The variety of Boolean algebras is the unique affine complete variety generated by a two element rigid algebra.*

6 Two counter-examples

We conclude these lectures with two instructive examples. Example 1 gives a FACS algebra which does not generate an arithmetical variety. Example 2 shows that for a finite functionally complete algebra having a one element subalgebra, the variety which it generates need satisfy no non-trivial congruence equation. Both of these examples appear in [9].

Example 1 A FACS algebra $\mathbf{A}$ with $V(\mathbf{A})$ not arithmetical.
Let $A = \{a, b, c, d\}$ and let θ be the equivalence relation on A with blocks $\{a, b\}$ and $\{c, d\}$. Define $f: A^4 \to A$ by the conditions:

i) f is compatible with θ and for any fixed $w \in A$, $f(x, y, z, w)$ induces the discriminator on A/θ;

ii) for all $x, y, z \in A$ and $w \in \{a, b\}$,

$$f(x, y, y, w) = f(x, y, x, w) = f(y, y, x, w) = x;$$

iii) for all $x, y, z, w \in A$, if $|\{x, y, z\}| = 3$ or $w \in \{c, d\}$, then

$$f(x, y, z, w) \in \{a, d\}.$$

It is easy to verify that f is uniquely defined by these conditions and that the polynomial $f(x, y, z, a)$ satisfies equations (1) (of Theorem 2.1). Next define two unary functions g and h by:

$$\begin{aligned} g(a) &= g(b) = b, & g(c) &= g(d) = d; \\ h(a) &= h(b) = c, & h(c) &= h(d) = a. \end{aligned}$$

Now consider the algebra $\mathbf{A} = \langle A; f, g, h\rangle$. $\mathbf{A}$ has no proper subalgebras since

$$g(a) = b,\ h(b) = c,\ g(c) = d,\ h(d) = a.$$

It is easy to see that $\theta \in \mathrm{Con}\mathbf{A}$. Let us prove that θ is the only non–trivial congruence. From $f(a, a, c, a) = c$, $f(a, b, c, a) = d$ it follows that $(c, d) \in \theta(a, b)$ and likewise from $f(b, c, c, a) = b$, $f(b, d, c, a) = a$ we have $(a, b) \in \theta(c, d)$. Hence $\theta(a, b) = \theta(c, d) = \theta$. Now we have to show that

$$\theta(a, c) = \theta(a, d) = \theta(b, c) = \theta(b, d) = A \times A$$

(the diversity congruence of $\mathbf{A}$). From the definitions of g and h it follows easily that

$$\theta(a, c) = \theta(b, d) \leq \theta(b, c),\ \theta(a, c) \leq \theta(a, d).$$

Hence it is sufficient to prove that $\theta(a,c) = A \times A$. But $f(b,a,d,a) = d$, $f(b,c,d,a) = a$, therefore $(a,d) \in \theta(a,c)$ so that (since $\theta(a,c) = \theta(b,d)$), collapsing a and c collapses all of a, b, c, d. Hence **ConA** is the three element chain consisting of $\{\Delta,\ \theta,\ A \times A\}$, so that $\mathbf{A}$ is arithmetical. It follows from Theorem 3.2 that $\mathbf{A}$ is affine complete and hence is a FACS algebra.

Now we show that $V(\mathbf{A})$ is not arithmetical. Consider the subset

$$B = \{(a,c), (a,d), (b,d), (c,a), (d,a), (d,b)\}$$

of $A \times A$. We shall show that this is a subuniverse of $\mathbf{A} \times \mathbf{A}$. By straightforward direct computation one can verify that for A/θ, interchanging $\{a,b\}$ and $\{c,d\}$ is an automorphism. Hence this automorphism induces a corresponding graph subalgebra of $\mathbf{A} \times \mathbf{A}$. This graph subalgebra clearly has universe equal to the (disjoint) union of B and $\{(b,c),(c,b)\}$. Thus to show that B is a subuniverse of $\mathbf{A} \times \mathbf{A}$ we must show that when we apply f, g or h to elements in B we never get either (b,c) or (c,b) as the result. For g and h this is obvious since $c \notin g(A)$ and $b \notin h(A)$. Suppose $(x_i, y_i) \in B$, $i = 1,2,3,4$ and $f((x_1,y_1),(x_2,y_2),(x_3,y_3),(x_4,y_4))$ equals either (b,c) or (c,b). Then by clause iii) of the definition of f, both x_4 and y_4 are in $\{a,b\}$ which contradicts $(x_4,y_4) \in B$. Hence B is a subuniverse of $\mathbf{A} \times \mathbf{A}$. Since B is clearly not rectangular, it follows (Proposition 2.2) that $V(\mathbf{A})$ is not arithmetical.

Example 2 A finite functionally complete algebra with a single proper subalgebra of one element and which generates a variety satisfying no non–trivial congruence equation.

Let $A = \{a,b,c\}$ and let $d{:}A^3 \to A$ be the discriminator. Define functions $f{:}A^4 \to A$, $g{:}A \to A$, and $h{:}A \to A$ by

$$\begin{aligned} f(x,y,z,u) &= d(x,y,z) \text{ if } u = b, \\ &= a \text{ else;} \end{aligned}$$

$$\begin{aligned} g(x) &= c \text{ if } x = b, \\ &= a \text{ else;} \end{aligned}$$

and

$$\begin{aligned} h(x) &= b \text{ if } x = c, \\ &= a \text{ else.} \end{aligned}$$

Let $\mathbf{A} = \langle A; f, g, h \rangle$. Since the discriminator is the polynomial $f(x,y,z,b)$, $\mathbf{A}$ is functionally complete. It is also evident that $\{a\}$ is the only proper subuniverse in A. Consider the two subsets

$$B = \{(b,c),(a,a),(a,b),(a,c),(b,a),(c,a)\}, \text{ and } C = B \setminus \{(b,c)\}$$

in $A \times A$. From the definitions of f, g, h it is not hard to verify that

$$f(B^4), g(B), h(B) \subset C.$$

Hence B is a subuniverse of $\mathbf{A} \times \mathbf{A}$ and also the equivalence relation θ which partitions B into the two disjoint sets C and $\{(b, c)\}$ is a congruence relation of the subalgebra $\mathbf{B}$ of $\mathbf{A} \times \mathbf{A}$. But also, because of the inclusions just noted, the operations on the two element algebra $\mathbf{B}/\theta$ all collapse to a single nullary operation with constant value C. Hence $\mathbf{B}/\theta$ generates the variety of pointed sets, which satisfies no non-trivial congruence equation.

References

[1] Baker, K. A., Pixley, A. F., Polynomial interpolation and the Chinese Remainder Theorem for algebraic systems, *Math. Z.* **143** (1975), 165–174.

[2] Fleischer, I., A note on subdirect products, *Acta Math. Acad. Sci. Hungar.* **6** (1955), 463–465.

[3] Foster, A. L., Pixley, A. F., Semi-categorical algebras II., *Math. Z.* **85** (1964), 169–184.

[4] Fried, E., Pixley, A. F., Compatible functions and the Chinese Remainder Theorem, to appear in *Algebra Universalis.*

[5] Grätzer, G., On Boolean functions (Notes on lattice theory II), *Rev. Roumaine Math. Pures Appl.* **7** (1962), 693–697.

[6] Hagemann, J., Herrmann, Ch., Arithmetical locally equational classes and representation of partial functions, in: *Universal Algebra*, Colloq. Math. Soc. János Bolyai **29**, North-Holland, Amsterdam 1982, 345–360.

[7] Hu, T. K., Characterization of algebraic functions in equational classes generated by independent primal algebras, *Algebra Universalis* **1** (1971), 187–191.

[8] Kaarli, K., Compatible function extension property, *Algebra Universalis* **17** (1983), 200–207.

[9] Kaarli, K., On varieties generated by functionally complete algebras, to appear in *Algebra Universalis.*

[10] Kaarli, K., On affine complete varieties generated by hemi-primal algebras with Boolean congruence lattices. Preprint.

[11] Kaarli, K., Pixley, A. F., Affine complete varieties, *Algebra Universalis* **24** (1987), 74–90.

[12] Kaarli, K., Pixley, A. F., Finite arithmetical affine complete algebras having no proper subalgebras. In preparation.

[13] Michler, G., Wille, R., Die primitiven Klassen arithmetischer Ringe, *Math. Z.* **113** (1970), 369–372.

[14] Mitschke, A., Near unanimity identities and congruence distributivity in equational classes, *Algebra Universalis* **4** (1974), 287–288.

[15] Pixley, A. F., The ternary discriminator function in universal algebra, *Math. Ann.* **191** (1971), 167–180.

[16] Pixley, A. F., Completeness in arithmetical algebras, *Algebra Universalis* **2** (1972), 179–196.

[17] Pixley, A. F., Characterizations of arithmetical varieties, *Algebra Universalis* **9** (1979), 87–98.

[18] Pixley, A. F., Principal congruence formulas in arithmetical varieties, in: *Universal Algebra and Lattice Theory* (S. Comer, ed.), Lecture Notes in Mathematics **1149**, Springer Verlag, New York, 1984, 238–254.

[19] Szendrei, Á., The primal algebra characterization theorem revisited, to appear in *Algebra Universalis.*

[20] Werner, H., Eine Charakterisierung funktional vollständiger Algebren, *Arch. Math. (Basel)* **21** (1970), 381–385.

Reading, Drawing, and Order

Ivan RIVAL
Department of Computer Science
The University of Ottawa
Ottawa, Ontario K1N 6N5
Canada

Abstract

The modern theoretical computer science literature is preoccupied with efficient data structures to code and store ordered sets. Among these data structures, graphical ones play a decisive role especially in decision-making problems. Choices must be made, from among alternatives ranked hierarchically according to precedence or preference. Loosely speaking, graphical data structures must be *drawn* in order that they may be easily *read.*

This is an introduction to, and survey of, the theory of *upward drawings* of ordered sets, highlighting the current directions of research.

1 The upward drawing

With the increasing interest in combinatorics and the rise of discrete mathematics, from the 1970s and onward, mathematicians and theoretical computer scientists have increasingly focussed attention on *ordered sets. Order* arises in, and has applications to, many branches of mathematics and it is, therefore, well-positioned in the mathematical landscape to react promptly to important developments. The most notable source of problems over the last decade surely springs from theoretical computer science.

The modern theoretical computer science literature attests to a preoccupation with efficient data structures to code and store ordered sets. Among these data structures, graphical ones are coming to play a decisive role about problems in which decisions must be made from among alternatives ranked according to precedence or preference. Loosely speaking, graphical data structures must be *drawn* in order that they may be easily *read.*

Chief among graphical data structures for ordered sets is the *upward drawing,* which we shorten to *drawing,* (alias *diagram, line diagram, Hasse diagram, directed covering graph*) according to which the elements of an ordered set P are drawn on a surface, usually the plane, as disjoint small circles, arranged in such a way that, for $a, b \in P$, the circle corresponding to a is higher than the circle corresponding to b whenever $a > b$ and a monotonic arc is drawn to join them just if *a covers b* (that is, for each $x \in P, a > x \geq b$ implies $x = b$). These arcs are drawn to avoid the incidence of any other circle on it, and thus avoid unwanted comparabilities. In symbols, we write $a \succ b$ and we also say that a is an

I. G. Rosenberg and G. Sabidussi (eds.), Algebras and Orders, 359–404.

upper cover of b or b is a *lower cover* of a. In this case, we draw a monotonic polygonal path consisting of line segments, from the vertex a to the vertex b. Wherever possible, and convenient, we use straight segments for the monotonic arcs. Moreover, as is customary, we identify an ordered set with its drawing. And, naturally, *covering graph* refers to the undirected companion of the upward drawing (see Figures 1, 2, and 3). See [Ri85a], [Ri89], and [Ri93a].

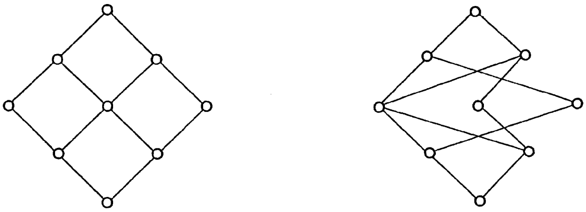


Figure 1: Upward drawings of a planar ordered set

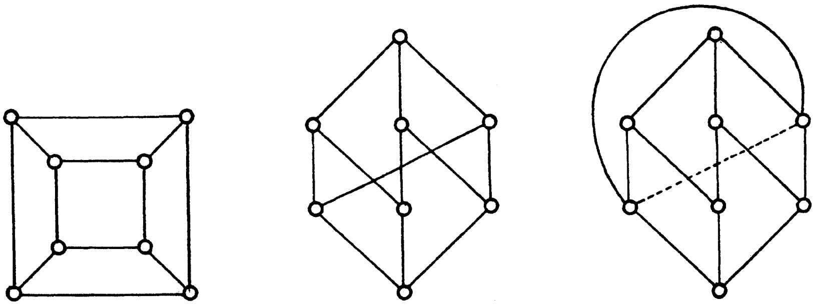


Figure 2: (i) The covering graph of $\mathbf{Q}_3$; (ii) upward drawing of $\mathbf{Q}_3$; (iii) not an upward drawing, although planar

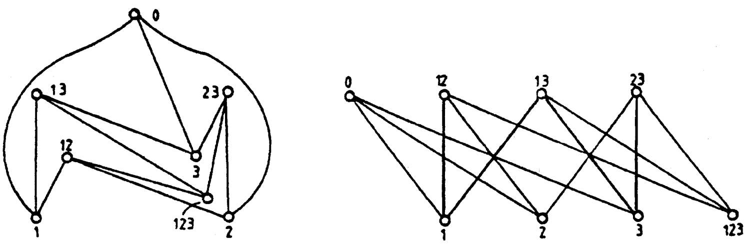


Figure 3: Upward drawings of a planar bipartite order with the same covering graph as $\mathbf{Q}_3$

1.1 Planarity

Planarity (or no crossing lines) is a classical theme in combinatorics and graph algorithm research. For undirected graphs, there are well-known and elegant combinatorial characterizations of the graphs with a planar representation and efficient algorithms for testing whether a graph has such a representation at all (cf. [HoTa74] and [LEC67]).

An ordered set is ***planar*** if it has an upward drawing in which no arcs cross. Thus, planarity is a property of the order and a planar ordered set may, indeed, have upward drawings which are not planar. Planarity can occasionally be recognized in an ordered set's upward drawing — even when drawn nonplanar. Thus, it is not too difficult to see that the ordered set $\mathbf{3} \times \mathbf{3}$, the direct product of the three-element chain $\mathbf{3} = \{0 < 1 < 2\}$ by itself, drawn in Figure 1, is planar, even when drawn with crossings. Apart from the last, each drawing has lines crossing, and the perforated line segments describe *transformations* to "unravel" the crossings to produce a planar upward drawing. On the other hand, the ordered set illustrated in Figure 3, which has a covering graph identical to the covering graph of $\mathbf{Q}_3$, the ordered set of all subsets of the three-element set ordered by set inclusion (Figure 2), is planar despite the fact that $\mathbf{Q}_3$ is not, and that $\mathbf{Q}_3$ would seem to have fewer crossings! Again the perforated line segments describe *transformations* to realize a planar representation of this planar, bipartite ordered set. Although the planar representation illustrated in this upward drawing, uses monotonic arcs which are not all straight segments, it is always possible to produce a planar drawing with the very same faces and in which all lines are straight segments [Ke87a].

At least for lattices, planarity is well understood [KeRi75]. Its application to a "dimension theory" of ordered sets is deep and surprising [KeTr82], [Ri78]. The *(order) dimension* of an ordered set P is the least number of linear extensions of P whose intersection is again P. The *completion* of P is the smallest lattice into which P can be order embedded; the completion of P is the bridge linking the dimension of orders to the geometry of lattices.

There are two decisive facts.

Theorem 1.1 [BFR71]

(i) *The dimension of an order equals the dimension of its completion.*

(ii) *A lattice is planar if and only if it has dimension at most two.*

A modern applied metaphor for planar lattices stems from a geometrical analogue of sorting: given a planar representation of a graph and a point located somewhere within one of its faces, derive an *effective* procedure to locate the point. An illuminating approach to this problem is to produce an upward drawing of a planar lattice, by adjoining a top and bottom, and (possibly) adding vertices to some edges (subdivision points). Indeed, planar lattices constitute the theoretical underpinnings to the solution of this problem (cf. [PrTa88]).

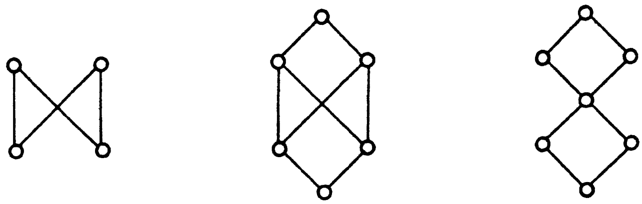


Figure 4: Dimension of an order equals dimension of its completion

The connection between planar lattices and planar graphs is striking.

Theorem 1.2 [Pl76] *A lattice is planar if and only if the graph obtained from its drawing by adjoining an edge from the top to the bottom is itself planar — as an (undirected) graph.*

This reduction, therefore, leads to an effective planarity testing procedure for lattices, based on graphs (cf. [HoTa74]). We may use this, in turn, to test whether an n-element ordered set P has dimension at most two. If it does, then its completion (which must, then, be an order sublattice of the direct product of two chains, each with at most n elements) must be a planar lattice with at most n^2 elements.

Moreover, it follows that planarity is easy to check if, for instance, it is required that all minimals and all maximals of the drawing appear on the exterior face, for, in this case, adjoining a top and bottom, necessarily produces a planar lattice (cf. [KeRi75]).

There is an effective procedure, too, if the vertical displacement of each element of P matches its relative position with respect to some arbitrary but fixed linear extension.

Theorem 1.3 [BaNa88] *Given an ordered set P with bottom and a linear extension L, there is a linear time procedure to test whether P has a planar upward drawing "consistent" with L.*

More generally, the problem for ordered sets with bottom is settled.

Theorem 1.4 [HuLu90], [Th89] *There is an efficient planarity testing procedure for an ordered set, provided that it has a bottom element.*

For ordered sets of width two there is an efficient planarity testing procedure based on a "forbidden" order characterization result.

Theorem 1.5 [CPR90a] *A finite ordered set P of width two is planar if and only if no member of the minimal list $\mathcal{L}$*(nonplanar, width two) *is a homeomorph of a subdrawing of P.*

1.2 Slope

What is a "good" drawing?

For the presentation of ordered data, the order among the elements must, of course, be readily apparent. Thus, for elements a and b represented by vertices on the plane with different y-coordinates, a is comparable to b if we find a monotonic path from the vertex a to the vertex b. A vertical path may be the easiest to discern.

Besides planarity, another natural criterion bears on the number of different slopes used in drawing the covering edges. This may be important in comparing drawings according to their "drawability". The "steepness" of the line segments has, for some time, remained a preoccupation of upward drawing schemes.

For $a \in P$, let *down degree* of a stand for the number of lower covers of a, that is, the number of $x \in P$ such that $a \succ x$; dually, let *up degree* of a mean the number of upper covers of a. Let *maximum degree* of P stand for the largest value from among $\{downdegree(a), updegree(a) | a \in P\}$. The number of different slopes required in a drawing of P is, evidently, at least the maximum degree of P, although this bound cannot always be attained (see Figure 5).

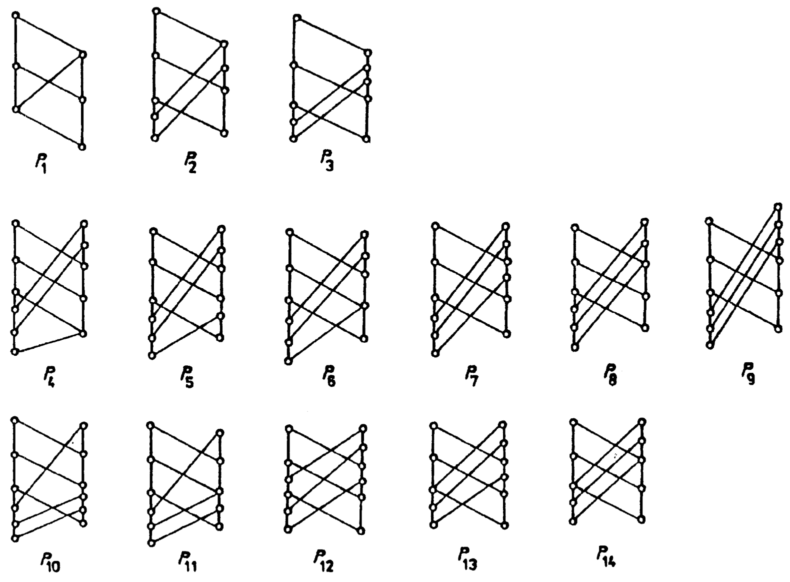


Figure 5: The minimal list of width two, nonplanar ordered sets

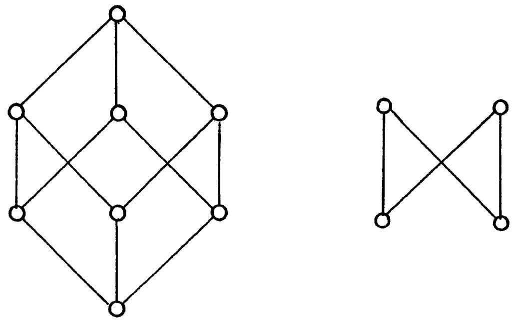


Figure 6: (i) Upward drawings requiring three slopes; (ii) upward drawing requiring three slopes although maximum degree is two

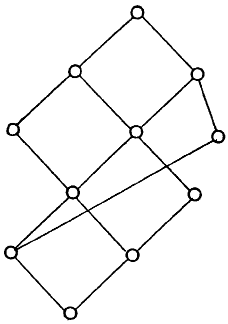


Figure 7: A lattice with maximum degree two requiring three slopes

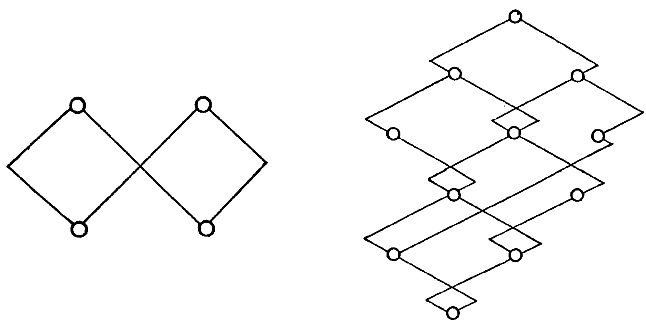


Figure 8: (i) A crooked two-slope upward drawing; (ii) a one-bend upward drawing using two slopes

The conjecture [Sa85] that the minimum number of slopes required to draw lattices is the maximum degree, is held up for distributive lattices, at least. However, it is not true even in the case that the maximum degree is two [CPR90b] (cf. [Cz91]) (see Figure 7).

These examples notwithstanding, the slope criterion seems to be in wide favour. And, as a matter of fact, with the simple artifice of "bends" on the line segments joining vertices in the covering relation of the drawing, it always becomes possible to produce a drawing with as few slopes as the maximum degree of the order [CPRU90]. For instance, the ordered set drawn in Figure 7 has an upward drawing using "crooked" edges, each with at most one bend, in which only two different slopes are ever used for the line segments (see Figure 8). Such an artifice requires relaxing the usual edge constraint, for the covering relation need no longer be represented by a line segment, although comparable vertices a, b will still be located at the ends of a monotonic polygonal path. (To avoid misinterpretation, bend points are all distinct in the drawing on the plane.)

Theorem 1.6 [CPRU90] *Any finite ordered set has a one-bend upward drawing using maximum degree many slopes.*

Thus, every covering edge is constructed using at most two line segments and every line segment is parallel to one of the maximum degree many lines. Moreover, starting from any given upward drawing, keeping its vertices in place, it is possible to join the appropriate pairs of vertices by crooked edges, at least if the maximum degree is even.

Theorem 1.7 [CPRU90] *For any upward drawing of a finite ordered set there is a two-bend upward drawing using the same vertex set location and using only maximum degree many slopes.*

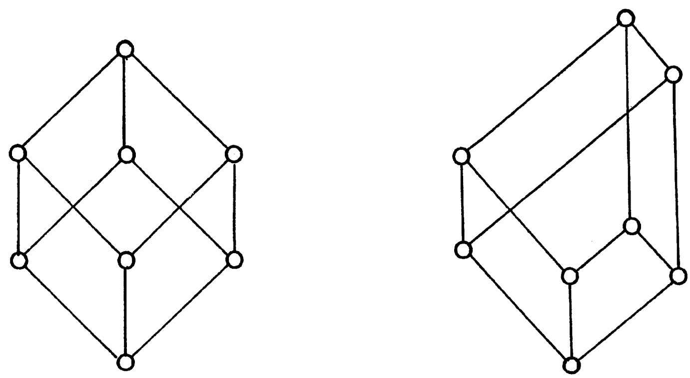


Figure 9: An "Escher-like", minimum-slope, upward drawing of $\mathbf{Q}_3$

Although parallelism seems to improve readability of drawings there are some notable exceptions. $\mathbf{Q}_3$ has another upward drawing with the minimum number of slopes (three) which, visually, is not at all as satisfying as either the other minimum slope drawing, or even a non-minimum slope drawing [Ste87] (cf. Figure 9).

There is already a precedent for the idea of "bends", for example, in VLSI circuit design in which a planar graph is presented on a given rectilinear grid [Sto80], [Ta87].

1.3 Levels

Among familiar data structures the tree (see Figure 10(i)) is surely the most common. One difficulty in drawing a tree is the crowding of circles at successive branchings. However, a common feature of essentially all tree drawings is that the successors of any circle are all drawn on the same horizontal level: this artifice produces drawings easier to read.

Does every ordered set have an upward drawing in which, for every element, all upper covers are on a horizontal level?

No.

The ordered set illustrated in Figure 10(ii) cannot be drawn so that, for every element, all upper covers lie on the same horizontal.

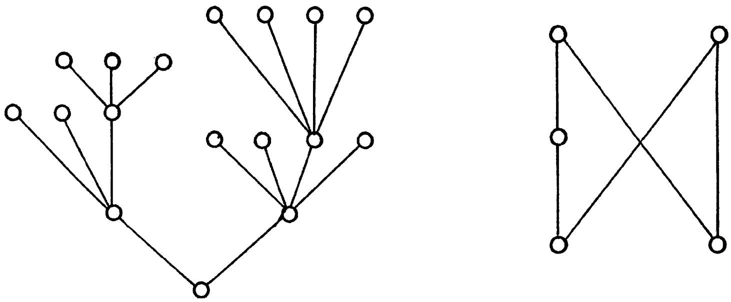


Figure 10: (i) A level drawing of a tree; (ii) no level drawing possible

Say that an ordered set is *upper levelled* if it has an upward drawing in which, for each element, all upper covers have the same y-coordinate. A subset $\{x_1, a_1, \ldots, a_n, c_1, \ldots, c_n\}$, $n \geq 2$, of an ordered set is an *alternating cover cycle* if $c_i \succ a_i$, for $i \leq n$, $c_{i+1} > a_i$, for $i \leq n-1$, and $c_1 > x > a_n$ are the only comparabilities (see Figure 11). These configurations arise, for example, in the study of *cutsets*, that is, subsets which intersect every maximal chain [RiZa85].

Theorem 1.8 [PeRi91] *A finite ordered set is upper levelled if and only if it contains no alternating cover cycle.*

Another characteristic result about alternating cover cycles is this.

Theorem 1.9 [RiZa85] *A finite ordered set is the union of antichain cutsets if and only if it contains no alternating cover cycle.*

These configurations arise as well in "orientation invariants" [BaRi89] and "single-machine scheduling" results [ElRi85] , [FaSc87].

The next result, too, follows from the characterization of upper levelled ordered sets.

Theorem 1.10 [PeRi91] *An ordered set is upper levelled if and only if it is lower levelled.*

An ordered set is *lower levelled* if it has an upward drawing in which, for each element, its lower covers all have the same y-coordinate.

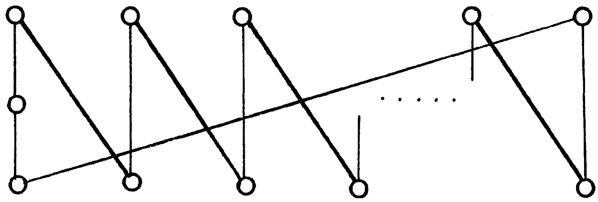


Figure 11: An alternating cover cycle

1.4 Bending and stretching

It is well known that, in a finite ordered set, the minimum number of chains whose union is all of the set equals the maximum size of an antichain [Di50]. In graphical terms it is convenient to visualize chains as "vertical paths", using vertical line segments for all of the edges. Of course, it is impossible to draw all chains vertically on the plane — unless, for every vertex both the down degree and the up degree is at most one. Sometimes it may be convenient to draw the disjoint chains of a given chain decomposition vertically (with non-vertical edges between these chains). In this case we call the drawing a *vertical drawing* — or a *k-vertical drawing* — if it has k disjoint, vertical chains.

For an ordered set and a positive integer k, a *k-channel drawing* is an upward drawing of it as a "subdrawing" — subgraph — of a k-vertical drawing. Imagine drawing an ordered set on an output tape of arbitrary length and bounded width, on which the circles for the elements, can be printed along a few vertical "channels", and straight line segments are drawn for the edges joining circles between these channels.

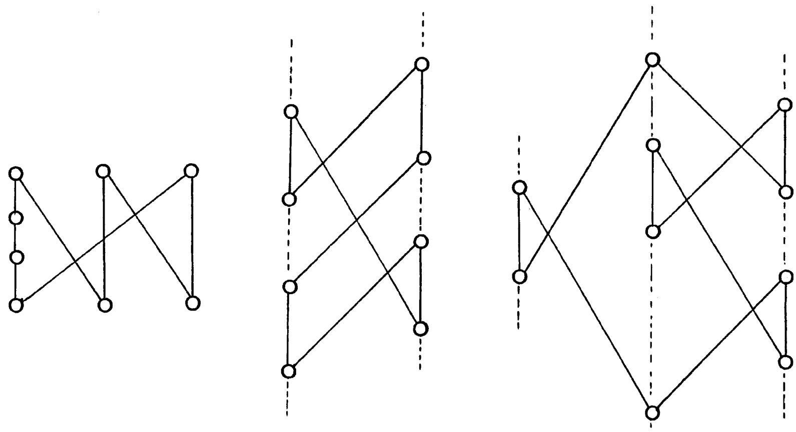


Figure 12: (i) An upward drawing; (ii) a two-channel drawing; (iii) not a two-channel drawing

The *channel number*, *channel*(P), of an ordered set P, is the smallest number k, such that P has a k-channel drawing. The channel number of P may well be smaller than *width*(P), the number of chains in a minimum chain decomposition. For instance, if P is the disjoint sum of chains then *channel*$(P) = 1$. Obviously, too, *channel*$(P) \leq$ *width*(P) and *channel*(P) may be strictly less than *width*(P).

For k-channel drawings we chart a course somewhere between monotonic arcs and straight segments. The circuit design metaphor is illustrative. In layout design a planar graph may be drawn on a grid in such a way that the edges of the graph follow the horizontal and vertical grid lines. For instance, if, for the graph edges of the complete graph $\mathbf{K}_4$, we use only straight segments, then such an embedding is impossible. With "bends" $\mathbf{K}_4$ has a grid embedding (cf. [Ta87]). In this sense, there are ordered sets P, for which there are ordered sets Q_P such that, using bends, P has a drawing as a subdrawing of a drawing of Q_P and *width*$(Q_P) <$ *channel*(P) (see Figure 14).

In a drawing using only straight segments for the edges, it may be impossible (even with bends) to construct a two-channel drawing, for instance, if there is a vertex with down degree greater than two. To this end, we use "stretching". We adopt the convention according to which straight edges rising or falling into a vertex of the drawing may be bent, and the bent edges, joining a vertex to its covers, may themselves meet at a bend, and continue along a common line until the common vertex is reached (see Figure 15).

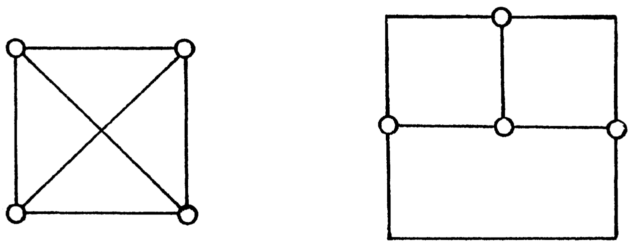


Figure 13: (i) $\mathbf{K}_4$; (ii) a grid layout of $\mathbf{K}_4$ using four bends

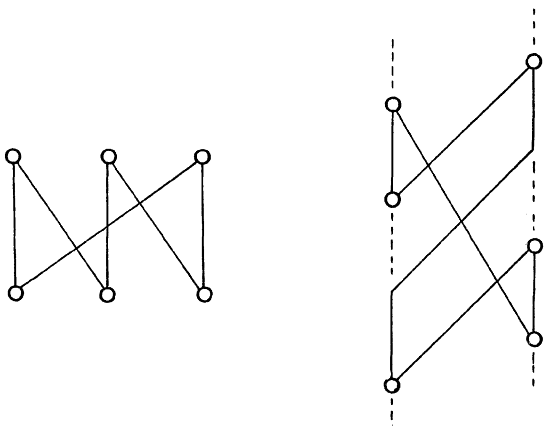


Figure 14: (i) An upward drawing; (ii) using bends, a two-channel drawing

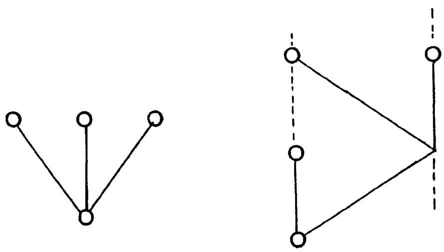


Figure 15: With stretching every element has down degree at most two and up degree at most two

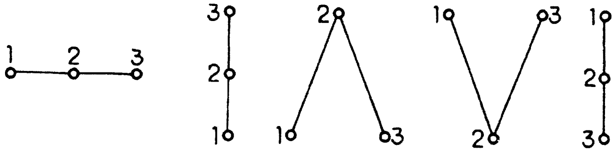


Figure 16: (i) A covering graph; (ii) all of its orientations

Here are the main results about channel drawings.

Theorem 1.11 [NR89a]

(i) *For any n-element ordered set there is a drawing which, with bending and stretching of the edges, transforms it into a three-channel drawing.*

(ii) *There are finite ordered sets with no two-channel drawing at all.*

(iii) *Moreover, there are finite ordered sets which cannot be embedded into any ordered set which, with bending and stretching, has a two-channel drawing.*

2 Transformations of upward drawings

2.1 Orientation and invariants

What is at stake for a better understanding of upward drawings?

An *orientation* of a graph is an order (or upward drawing) with this graph as covering graph.

We will need to search far before we come upon a nontrivial property — different, for example, from the number of elements, edges, etc. — invariant with respect to all orientations of a covering graph. For instance, none of the familiar parameters *height*, *width*, or *dimension*, is such an invariant, as we see from the four possible orientations of the covering graph consisting of the three-element path (see Figure 17).

Neither *planarity* nor the *fixed point property* is such an invariant either (see Figure 19).

A *doubly irreducible* element has at most one upper cover and at most one lower cover. The next result, while apparently a modest contribution, seems to involve considerable effort (see Figure 18).

Theorem 2.1 [JNR87] *In every lattice orientation of the covering graph of a planar lattice, there is always a doubly irreducible element.*

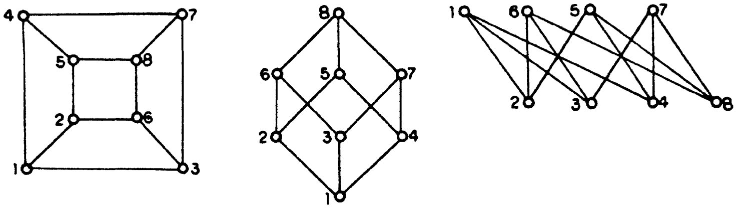


Figure 17: (i) A covering graph; (ii) some of its orientations

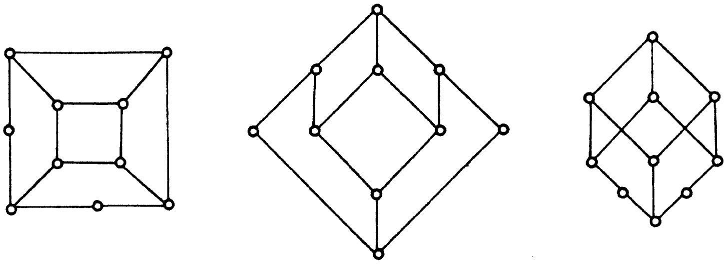


Figure 18: (i) A covering graph; (ii) lattice orientations of this graph

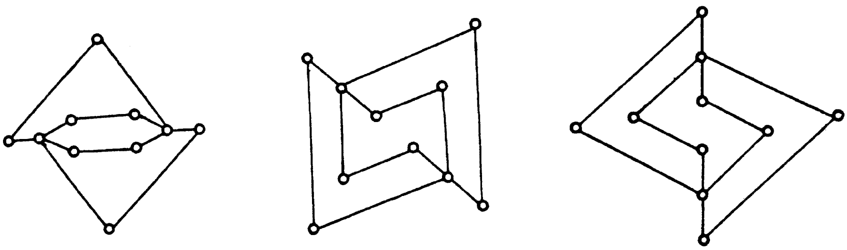


Figure 19: (i) A covering graph; (ii) a planar orientation of dimension four; (iii) a planar orientation of dimension two

Even the prospect of bounding the value of a combinatorial parameter in terms of its value for some orientation does not offer more consolation. For instance, the width, the height or the dimension can be small, and it can be large, for different orientations with fixed covering graph.

How to transform one orientation to another?

The best understood such transformation is the *pushdown* according to which a fixed, but arbitrary, maximal element becomes a minimal element, all of whose lower covers become its upper covers; the *pullup* transforms a fixed, but arbitrary, minimal element into a maximal element, all of whose upper covers become its upper covers (see Figure 20). This transformation is first recorded in [Mo72].

Theorem 2.2 [Mo72] *For a connected covering graph G and a vertex $a \in G$ there is an orientation whose top element is a.*

Intuitively, begin with any orientation and pushdown maximal elements successively until a remains the only maximal element. The reorientation produced by a pullup, however, can be reproduced by a sequence of pushdowns.

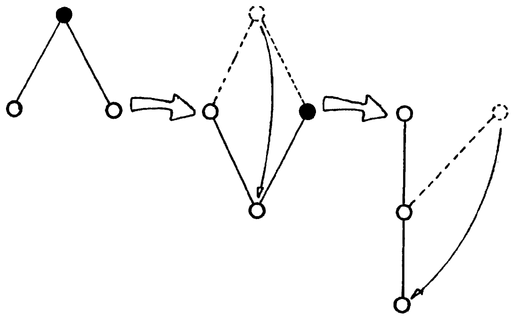


Figure 20: Pushdown\pullup

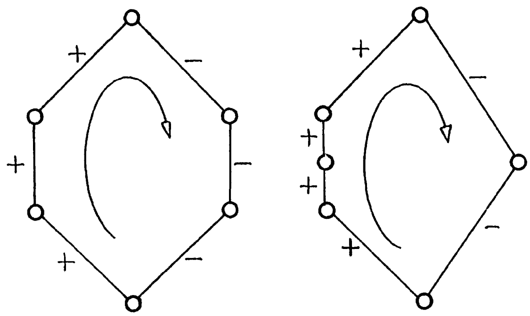


Figure 21: (i) $flow = 2$; (ii) $flow = 0$

Pushdowns or pullups cannot account for all transformations.

Consider, for instance, the six-element cycle as graph. Let P be the orientation consisting of two four-element chains with common top and bottom, and let Q be the orientation consisting of a six-element chain and a three-element chain with common top and bottom (see Figure 21). Then neither can be obtained from the other by any sequence of pushdowns or pullups. To see this, just associate a fixed clockwise rotation of the vertices of the underlying cyclic graph. In any orientation define the *flow* of this graph cycle as the difference between the number of "ups" and the number of "downs" with respect to the clockwise rotation. Thus, in P, this cycle has $flow = 0$ while in Q this cycle has $flow = 2$. On the other hand, it is easy to see that the flow must remain unchanged by any pushdown or pullup!

With this elementary observation in hand, a complete description is near, of just which orientations can be obtained by pushdowns or pullups.

Theorem 2.3 [Pr86] *Let P, Q be orientations of a (common) graph G. Then P can be obtained from Q by a sequence of pushdowns, if and only if, for every graph cycle C of G, $flow_P(C) = flow_Q(C)$.*

A striking consequence is this.

Theorem 2.4 [Pr86] *If P can be obtained from Q by a sequence of pushdowns and if, in addition, $top(P) = top(Q)$ then $P = Q$.*

To see this let (a, b) be a covering pair, nearest to the common top, with $a \succ b$ in P but $b \succ a$ in Q. Let A be a chain maximal between a and $top(P)$ in P and let B be a chain maximal between b and $top(Q)$ in Q. By hypothesis, $flow_P(C) = flow_Q(C)$ where $C = \{a, b, A, B\}$. Then, $1 + |A| - |B'| = -1 + |A| - |B|$, where B' is the orientation, in P of the chain B in Q. (Note that, in view of the minimality of the choice (a, b), the chain A in P is a chain A in Q.)

Here is a rather different description in terms of a distinguished subobject that, as we have already seen, is characteristic of upper levelled drawings [PeRi91], and of antichain cutset decompositions [RiZa85].

Theorem 2.5 [BaRi89] *Let P be any orientation of G. Then every orientation of G is obtained by a sequence of pushdowns if and only if there is no orientation of G which contains an alternating cover cycle.*

There is yet a third characterization of the pushdown or pullup transformations [LR91a], to which we turn below to estimate the number of orientations.

Even if the maximals of P coincide with the maximals of Q the orientations P, Q, obtained from each other by a sequence of pushdowns and pullups, need not be identical (see Figure 22).

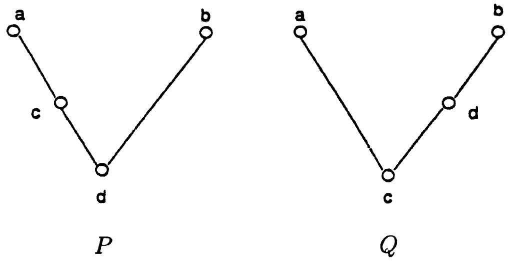


Figure 22: Q and P have identical maximals, Q is obtained from P by pushdowns, yet $P \neq Q$

2.2 Characterization

Let's clear the air from the start.

Theorem 2.6 [NeRö87] *The decision problem, whether an undirected graph is a covering graph, is NP-complete.*

The naive problem is, nevertheless, of considerable interest [Or62], [Ri85b]. The starting point is the *triangle* which, clearly, cannot be a covering graph — any attempt to orient it produces a "nonessential" edge. Apart from it, the smallest "noncovering" graph is the well known graph illustrated in Figure 23. A convenient proof can be fashioned in terms of the pushdown transformation.

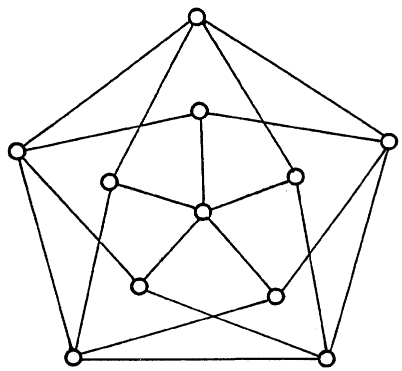


Figure 23: The smallest, triangle-free non-covering graph

Suppose it is orientable. Choose $top = a$. It has five lower covers and, then, below these must lie an orientation of the pentagon which contains a three-element chain. Every three-element path in this pentagon has endpoints among the lower covers of a and, therefore, once oriented, the three-element chain produces a "nonessential" edge.

The folklore of these problems is illuminating.

Theorem 2.7 *If, for a graph, the size of its smallest cycle exceeds the least number of colours needed to colour its vertices, then it is orientable.*

For, if the colour classes of the graph are denoted $C_1, C_2, \ldots, C_m$, define a relation $a < b$ if there is a positive integer $k \leq m$ and a path $a = v_0, v_1, v_2, \ldots, v_k = b$, in the graph, whose respective colours are strictly increasing. From the hypothesis, it follows that no such path can contain a chord — a nonessential edge — and, therefore, this relation produces an order.

Of course, an orientable graph is triangle-free and if, in addition, it is planar, then its vertices can be coloured with only three colours [Gr58]. From this, there follows a solution for planar graphs.

Theorem 2.8 *A planar graph is orientable if and only if the size of its smallest cycle exceeds the least number of colours needed to colour its vertices.*

A related recent result settles the characterization question for bipartite planar graphs. A *bipartite* orientation of a graph is an orientation of it in which every element is either minimal or maximal.

Theorem 2.9 [BLR90] *A bipartite graph has a bipartite planar orientation if and only if the underlying graph itself is planar.*

As a special case, the characterization of the covering graphs of distributive and modular lattices is accessible. Two quite different characterizations are known of covering graphs orientable as distributive lattices. The first is a concrete one cast in terms of the *distance* function in a graph, the least number of edges in a path from a to b. The *diameter* of the graph is the largest distance possible in the graph.

Theorem 2.10 [Al65] *A covering graph is orientable as a distributive lattice if and only if*

(i) *it is connected and has no odd cycles,*

(ii) *there are vertices $\top, \bot$ such that $diameter(G) = dist(\bot, \top)$ and, for vertices a, b, c, with $a \succ c$, $b \succ c$, $dist(a, \bot) = dist(b, \bot) = dist(c, \bot) + 1$, there is a unique vertex e such that $e \succ a, e \succ b$ and $dist(e, \bot) = dist(c, \bot) + 2$,*

(iii) *every subgraph, graph isomorphic to the covering graph of $\mathbf{Q}_3 \backslash \{q\}$, where $q \in \mathbf{Q}_3$, is contained in a "full" $\mathbf{Q}_3$, and*

(iv) *it contains no subgraph, graph isomorphic to* $\mathbf{K}_{2,3}$.

An alternate approach to orientability is based on the "retract" construction. A subgraph H of a graph G is a *retract* provided there is an edge-preserving map g — a *retraction* — of G to H such that $g(v) = v$ for every vertex $v \in H$. We write $H \trianglelefteq G$. Of course, any subgraph of an orientable graph is itself orientable. More striking, however, is this fact.

Theorem 2.11 [DuRi83] *Any graph with an orientable retract is itself orientable.*

For, if g is a retraction of G to an orientable subgraph H then, for distinct vertices $a, b \in G$ define $a < b$ just if there is a positive integer m and a path $a = x_0, x_1, x_2, \ldots, x_m = b$ such that $g(x_{i+1}) \succ g(x_i)$ in the orientation of H. In these terms the characterization of distributivity is particularly satisfying (see Figure 24).

Theorem 2.12 [DuRi83] *A graph is the covering graph of a distributive lattice of length n if and only if the graph is a retract of $\mathbf{Q}_n$ and it has diameter n.*

(This contrasts with the well known fact that any finite ordered set can be order embedded in some hypercube $\mathbf{Q}_n$ and, indeed, that every finite lattice is an "order retract" of some $\mathbf{Q}_n$, cf. [Ri82].)

Here is a suggestive reformulation of the last result.

Theorem 2.13 [DuRi83] *A graph G is orientable as a distributive lattice if and only if $G \trianglelefteq \prod_{i \in I} G_i$, where each $G_i \cong \mathbf{K}_2$.*

(Notice that the drawing of the direct product of ordered sets is precisely the Cartesian product of their respective drawings [Ri85a].)

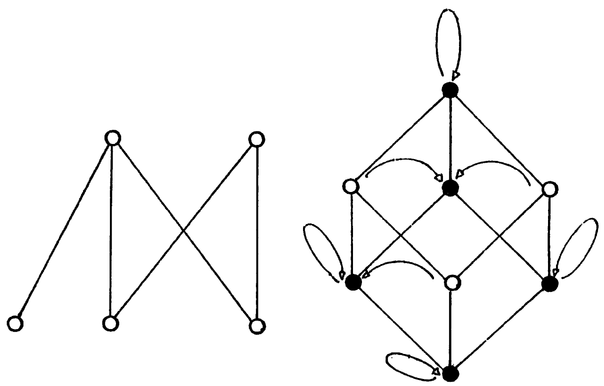


Figure 24: A retract of a cube

3 Enumeration and structure

How many orientations has an n-element covering graph?

This is different from the more elementary companion question: *how many n-element orders?* Our aim here is to enumerate the ways in which a covering graph can be drawn, that is, the number of its orientations.

3.1 Pushdowns

There are two easy cases. As an undirected graph an antichain corresponds to a set of independent vertices, so there is just one drawing possible. Trees, as graphs, are orientable, for they are planar and triangle-free. The number of drawings possible for an n-vertex tree is 2^{n-1} for it has $n-1$ edges and each part of a partition of the ordered pairs (a, b) of vertices can be independently oriented, $a < b$ or $b < a$. As a tree contains no cycles, all of its orientations are obtained by pushdowns, from any fixed orientation.

A naive approach to enumerate orientations distinguishes a subset of edges of a fixed orientation and "reverses" each of them. Call a subset E of edges of the drawing of P *reversible* if there is an orientation of the covering graph of P in which an edge has the same direction as in P, with the exception of those edges $a \succ b$ of E which become $a \prec b$ (see Figure 25). If reversible, the new edges (in the new drawing) are called *reversed.* This approach can be used to fashion an effective characterization of those orders obtained from a fixed one by a sequence of pushdowns.

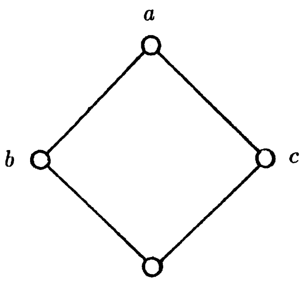


Figure 25: The edge $a \succ b$ is not reversible, although the pair $a \succ b$, $a \succ c$ of edges is reversible

Let E be a subset of the edges of P, say E consists of $a_1 \succ b_1, a_2 \succ b_2, \ldots$ and let $P \backslash E$ stand for the drawing obtained from P by removing all of the edges of E. Let U_E denote the subset of all vertices of P connected to some a_i in $P \backslash E$, that is, the vertices $a \in P$ for which there is a positive integer m and a sequence $a = x_0, x_1, x_2, \ldots, x_m = a_i$, for some i, such that $x_j \succ x_{j+1}$ or $x_{j+1} \succ x_j$ in $P \backslash E$. Let D_E denote the subset of all vertices of P connected to some b_i in $P \backslash E$. We call E a *cut* of P if $D_E \cap U_E = \emptyset$ (see Figure 26).

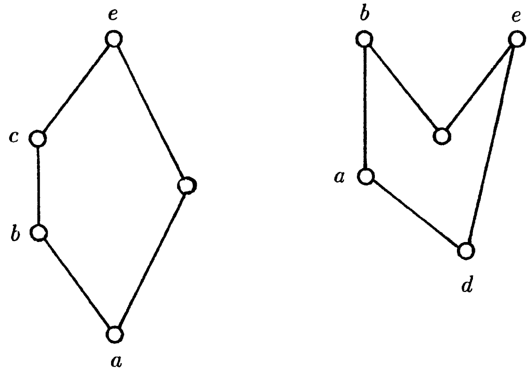


Figure 26: (i) $E = \{c \succ b, d \succ a\}$ is a cut; (ii) the orientation with the reversed edges

Theorem 3.1 [LR91a] *Let P be a finite ordered set. An ordered set can be obtained from P by a sequence of pushdowns or pullups if and only if the reversed edges can be partitioned into cuts of P.*

By partitioning the subset of reversed edges into minimal cuts and testing for connectivity, we derive an efficient procedure to test whether an orientation Q of the covering graph of P can actually be obtained from P by a sequence of pushdowns and pullups.

Theorem 3.2 [LR91a] *There are at least $\frac{n^2+2n}{2} - n\log_2 n$ distinct upward drawings obtained from any connected, n-element upward drawing.*

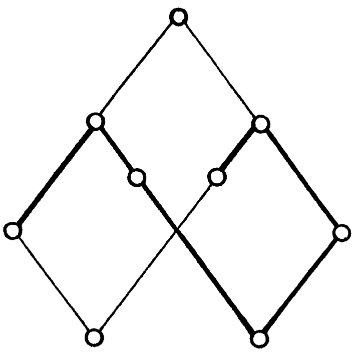


Figure 27: (i) A reversible subset of edges

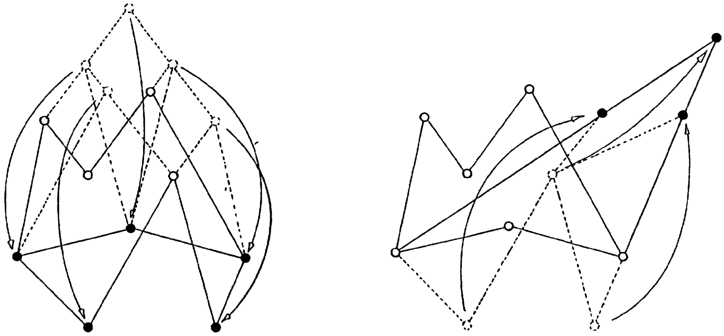


Figure 27: (ii) a sequence of pushdowns and pullups

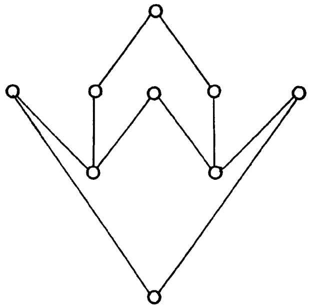


Figure 27: (iii) the orientation obtained by reversing these edges

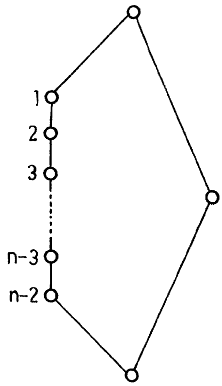


Figure 28: An upward drawing with $\frac{n^2-n}{2}$ distinct upward drawings obtained by pushdowns

3.2 Orientations

Apart from the few isolated classes of orders (e.g. trees), there seem to be far more orientations of a covering graph than can be produced, from a fixed orientation, by pushdowns.

There is an easily derivable lower bound and, for some apparently exceptional examples, a more difficult upper bound.

Theorem 3.3 (i) *Almost every n-element connected covering graph has at least $2^{\frac{n}{2}}$ orientations* [LR91b].

(ii) *There is an n-element covering graph with at most $2^{\frac{cn \log\log\log n}{\log\log n}}$ orientations* [BrNe91].

The key to the proof of (i) is this well known and important theorem from which we can piece it together.

Theorem 3.4 [KlRo75] *Almost every covering graph can be coloured with three colours.*

Thus, asymptotically speaking, almost every ordered set has three levels L_1 = *minimals*, L_3 = *maximals*, each with about $\frac{n}{4}$ elements, and the "middle level" L_2, with about $\frac{n}{2}$ elements, such that every element of L_1 is below about half of the elements of L_2 and every element of L_2 is below about half of the elements of L_3.

Next, observe that every antichain A (of some fixed orientation P) can become the subset of maximals for, just push down, successively, maximal elements of P until only the elements of A remain maximal. Then, for any $S \subseteq A$, there is an orientation P_S in which the maximals are precisely the elements of S. It follows that there are at least $2^{|A|}$ orientations (each obtained from P by pushdowns).

Finally, in view of the asymptotic estimate [KlRo75], there is, in almost every n-element ordered set, an antichain A with $\frac{n}{2}$ elements — the middle level.

One attack on the conjectured lower bound of $2^{\sqrt{n}}$ turns out to highlight several interesting [LR91b]. The approach consists of these two conjectures about a covering graph G and some fixed, but arbitrary orientation P of it.

(i) *Any independent subset of G is an antichain in some orientation.*

(ii) *For any matching M in P and any $M' \subseteq M$, there is an orientation of G that reverses all of the edges of M' and none of the edges of $M \backslash M'$, (while other edges may, or may not, be reversed).*

If true of an n-vertex covering graph and an orientation of it, it would follow that this covering graph has at least $2^{\frac{n}{3}}$ orientations. Here is the reason. First,

$$(\textit{maximum independent subset of } G) + 2(\textit{maximum matching}) \geq n,$$

for, if M is a maximum matching in P then the complement of M in G must be an independent subset. Thus, either

$$\textit{maximum independent subset} \geq \frac{n}{3}$$

or

$$\textit{maximum matching} \geq \frac{n}{3}.$$

Then, if $I \subseteq G$ is an independent subset and $|I| \geq \frac{n}{3}$ then an orientation Q of G whose maximals are precisely I, produces, by the pushdown transformation, at least $2^{\frac{n}{3}}$ further orientations, one for each subset of I. Otherwise, P has a matching M with at least $\frac{n}{3}$ edges and, as any $M' \subseteq M$ may be reversed, this, too, produces $2^{\frac{n}{3}}$ orientations.

Such a neat approach!

Yet, both (i) and (ii) are false! (See Figure 29 and Figure 30.) Nevertheless, both the general ideas and the particular examples are certainly of independent interest.

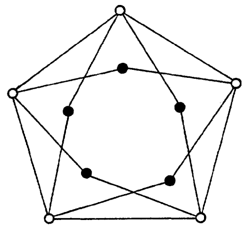


Figure 29: The covering graph G with independent subset I (shaded vertices), cannot be an antichain in any orientation, else, choose one with I as maximals and adjoin a new top element. This produces an orientation of the eleven-vertex non-covering graph, cf. Figure 22

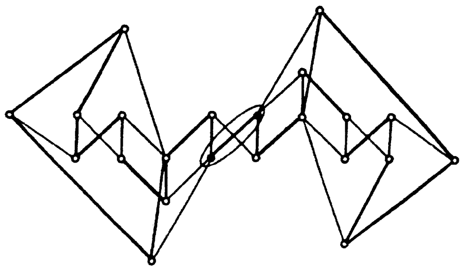


Figure 30: The single edge of $\{a \succ b\} = M' \subseteq M$ (distinguished edges) cannot be reversed, else either $a_1 \succ b_1$ or $a_2 \succ b_2$ of M must be reversed

3.3 Structure theory

A success story in the theory of ordered sets is about "comparability graphs". From a "structural" perspective comparability graphs are fully understood. For an ordered set P, its *comparability graph* is the graph whose vertices are the elements of P and whose edges are precisely those pairs (a, b) of elements such that either $a < b$ or $b < a$ in P.

The key idea is the "lexicographic decomposition". Let Q be an ordered set and let $(P_q \mid q \in Q)$ be a family of disjoint ordered sets indexed by Q. The *lexicographic sum* $\sum_{q\in Q} P_q$ of the ordered sets $P_q, (q \in Q)$ is an ordered set defined on the disjoint union $\bigcup_{q\in Q} P_q$ with $a < b$, if $a < b$ in some P_q, or, $a \in P_{q_1}, b \in P_{q_2}$ and $q_1 < q_2$ in Q. Such a representation is called a *lexicographic decomposition*, the P_q's *blocks*, Q the *index*, and $P = \sum_{q\in Q} P_q$ is *lexicographically nondecomposable* if either $|Q| = 1$ or each $|P_q| = 1$.

Here are the main results.

Theorem 3.5 [Ga67], [Ke85], [Ha84], [Ri85a]

(i) *With respect to orientations of its comparability graph, a lexicographically nondecomposable ordered set has, up to duality, precisely one orientation.*

(ii) *If $P = \sum_{q\in Q} P_q$, if P'_q is an orientation of the comparability graph of $P_q, q \in Q$, and Q', too, is an orientation of the comparability graph of Q, then $P' = \sum_{q\in Q} P'_q$ is an orientation of the comparability graph of P.*

(iii) *Moreover, all orientations of a comparability graph are constructed in this way.*

Although we are much further away from a completely satisfying structure theory for upward drawings the ingredients of one stem from a classification and structure theory for ordered sets based on the twin constructions of retraction and direct product [DuRi81]. Basically, what we are after is a theory that describes the orientations of a covering graph.

How to describe orientations P, Q with the same covering graph G?

Given orientations P, Q of a covering graph, how to describe Q in terms of P?

One approach, is to consider G as a conglomeration of ingredient graphs, each to be oriented, and pieced together to produce P. To this end, we treat an oriented analogue of the retract. A subset S of an ordered set P is an *isotone retract* of P if there is an isotone retraction g of P onto S, that is, $a \leq b$ implies $g(a) \leq g(b)$ and g is a retraction of the covering graph of P to the covering graph of S. We also write $S \trianglelefteq P$. For a class $\mathbf{K}$ of ordered sets let $\downarrow \mathbf{K}$ stand for all orders, obtained by pushdowns, from $P \in \mathbf{K}$.

Theorem 3.6 [BaRi89] *Let P, Q be ordered sets with the common covering graph G. As graphs let $G \trianglelefteq \prod_i G_i$. Then there are orientations P_i of G_i such that, as ordered sets, $P \trianglelefteq \downarrow \prod_i P_i$.*

4 Planarity

4.1 Planar lattices

The theory of planar lattices is well understood [KeRi75].

Its starting point is the innocent observation that every planar lattice has a *doubly irreducible* element, that is, an element with precisely one upper cover and precisely one lower cover.

Theorem 4.1 [BFR71] *Every planar lattice (with at least three elements) has a doubly irreducible on the left boundary of any planar upward drawing.*

Indeed, choose a maximal join irreducible element a on the left boundary of a planar upward drawing of the lattice L. Suppose that $b \succ a$, where b is on the left boundary, and $c \succ a$, where c is not on the left boundary. There are chains C, maximal from c to the top $\top$, and D, maximal from b to the bottom $\perp$. As these chains can neither cross nor have a common point, we conclude that a must have been meet irreducible, after all, whence doubly irreducible.

Planarity, for lattices, serves even as an algorithm for lattice testing [KeRi75].

Theorem 4.2 [Bi67] *A bounded planar ordered set is a lattice.*

The key point is that a *four-cycle* $\{a_i < c_i | i = 1, 2\}$ with no element between the a_i's and the b_i's, together with top and bottom, has no planar upward drawing (see Figure 31).

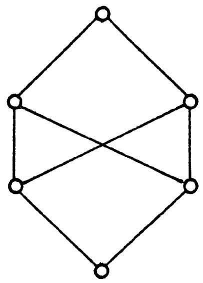


Figure 31: Every finite bounded ordered set, which is not a lattice, contains this ordered set

An important generalization is a *dismantlable lattice*, that is, a lattice L whose elements can be labelled $a_1, a_2, \ldots, a_n$ such that each $a_i, i = 1, 2, \ldots, n-2$ is doubly irreducible in $L \backslash \{a_1, a_2, \ldots, a_{i-1}\}$. For a reason analogous to planar ordered sets, *every bounded dismantlable ordered set is a lattice.* Moreover, a simple induction shows that, every planar lattice is dismantlable (see Figure 32).

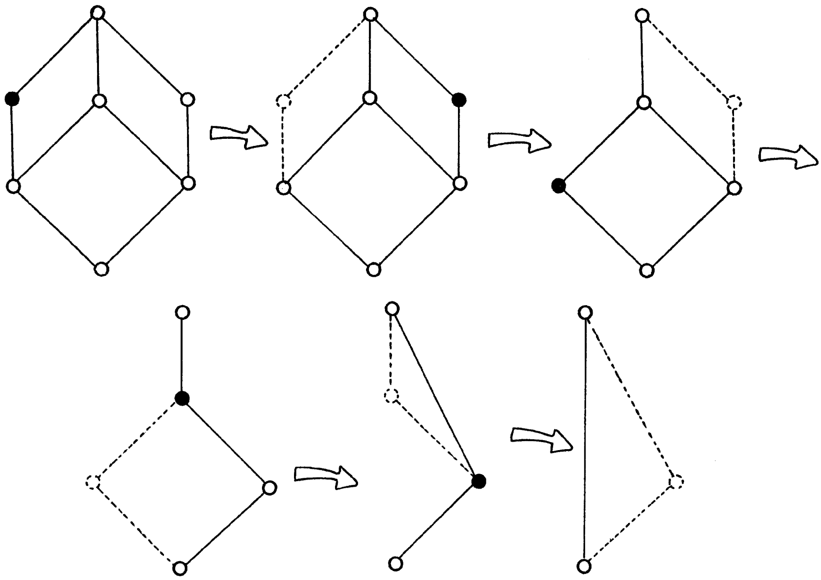


Figure 32: Dismantling a planar lattice

More general than planar lattices, they are, not surprisingly, simpler to characterize.

Theorem 4.3 [KeRi74], [Aj73] *A lattice is dismantlable if and only if it contains no cycle.*

A subset $\{a_1, a_2, \ldots, a_n, c_1, c_2, \ldots, c_n\}, n \geq 3$, is a *cycle* if, $c_i > a_i, c_i > a_{i+1}$, $i = 1, 2, \ldots, n-1$, and $c_n > a_n, c_n > a_1$ are the only comparabilities.

Dismantlable lattices have numerous combinatorial connections [Bj80], [Cr84], [DuRi78], [DuRi79], [Ke81], [KoLo83], [Ri76b]. Unlike planar lattices their dimension is unbounded [Ke81] (cf. Figure 19); there are intriguing links to matroids [Bj80], [Cr84], [KoLo83]. Such lattices and especially, generalizations of dismantlability, even shed light on the fixed point property [DuRi79] [Ri76a], [Ru89].

Several problems about dismantlable lattices have resisted solution during the last two decades [Ri76b], [KeRi75]. *In any dismantlable lattice, is width $\geq$ dimension?* Not apparently of direct combinatorial significance, this next problem, if positively settled, would imply a positive solution to the last one. Its two parts are essentially identical.

(i) *Is every dismantlable lattice order embeddable in a modular dismantlable lattice?*

(ii) *Is every dismantlable lattice order embeddable in the subgroup lattice of the direct product of two cyclic groups, each of prime power order?*

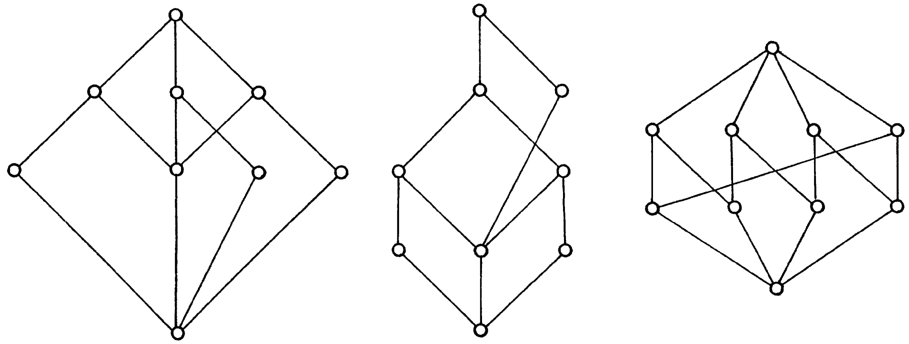


Figure 33: (i) Dismantlable, nonplanar lattices; (ii) nondismantlable lattices

The terms of these questions delineate, too, a dichotomy between *algebra* and *order*. For instance, what is, and what should, a "sublattice"? From an algebraic viewpoint it is customary that a *sublattice* is a subset closed with respect to the algebraic operations: (finitary) *supremum* and (finitary) *infimum*; from an order theoretical viewpoint we distinguish *order sublattice* as a subset which, with the induced ordering is itself a lattice (see Figure 34(i)). Analogously, a *homomorphism* is routinely defined as a $\{\vee, \wedge\}-$ preserving operation, from one lattice to another, while the natural choice of "morphism" from an order theoretical viewpoint is, simply, an *isotone* map, that is, an order-preserving map from the one lattice to the other (see Figure 34(ii)) [Ri78].

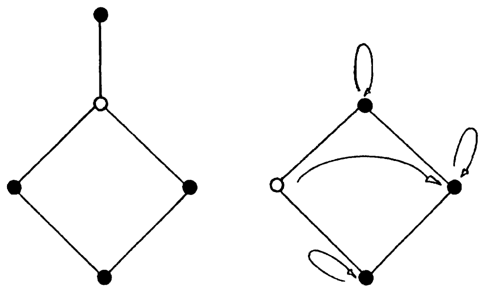


Figure 34: (i) An order sublattice; (ii) an isotone map

Even the well known fact that a planar lattice has *dimension* ≤ 2 has an extra dividend. Fix a planar upward drawing of a planar lattice L. The *left linear extension* L_{left} whose ith element a_i is the (unique) "left-most" element from among the minimals in $P\backslash\{a_1, a_2, \ldots, a_{i-1}\}$. The *right linear extension* L_{right} is constructed analogously and $L = L_{left} \cap L_{right}$ (see Figure 35).

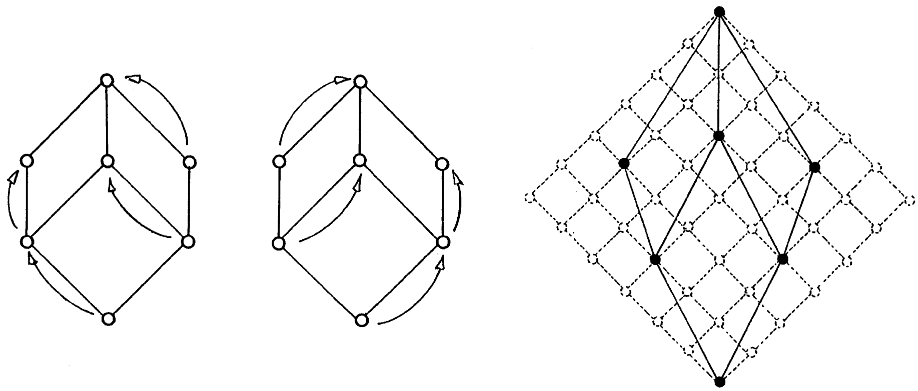


Figure 35: The intersection of the left and right linear extensions of a planar upward drawing of a lattice induces an order embedding of the lattice in the two-dimensional product $L_{left} \times L_{right}$

This linear extension construction scheme is the very first illustration of a *greedy linear extension*, that is a linear extension of an ordered set which proceeds according to the slogan: "climb as high as you can", that is, $a_i > a_{i-1}$ unless every upper cover of a_{i-1} has a predecessor not yet selected. These linear extensions are at the heart of the *jump number problem* which emerged from problems in operations research on scheduling theory and which has attracted numerous researchers who have written many articles about it, during the last decade (cf. [Ri86]).

Here, in outline, are the fundamental theorems about planar lattices.

For comparable elements $a < b$ in an upward drawing, say a is *visible from* b if it is possible to include a monotonic arc from b to a which crosses no other line or point. We denote the closed interval $\{a \leq x \leq b\}$ by $[a, b]$. An $[a, b]$-*component* is a connected component, in the covering graph, of the open interval $\{a < x < b\}$ of L. The closed interval $[a, b]$ has a *left dangle* if there is an element c to the left of a, $c \notin [a, b]$, and either $c < a$ or $c > b$. A *right dangle* is defined analogously.

Theorem 4.4 (Visibility) [KeRi75] *Fix a planar upward drawing of a planar lattice L and let $a < b$ in L. Then b is not visible from a if and only if*

(i) $b \not\succ a$,

(ii) *there is a unique* $[a, b]$*-component, and*

(iii) $[a, b]$ *has a left and a right dangle.*

The *permutation* and *reflection* transformations are illustrated in Figure 36.

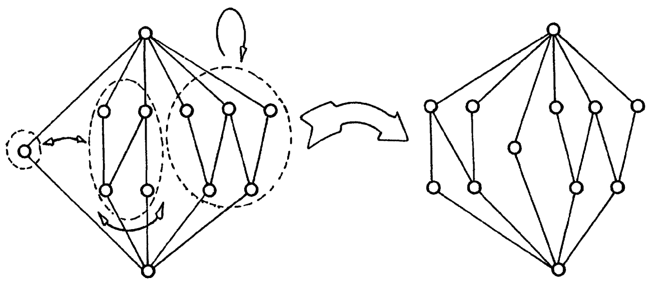


Figure 36: (i) Permutation; (ii) reflection

Theorem 4.5 (Transformation) [KeRi75] *Fix planar drawings Δ, Δ' of a planar lattice. Then Δ' can be obtained from Δ by a sequence of permutations and reflections of the components.*

The most striking of the fundamental results is this "forbidden (order) sublattice" characterization which highlights, too, our *order-theoretical* platitudes contrasting order sublattices and algebraic sublattices.

Theorem 4.6 (Characterization) [KeRi75] *A lattice is planar if and only if it contains no order sublattice isomorphic to a lattice from the minimal list* $\mathcal{L}$(nonplanar).

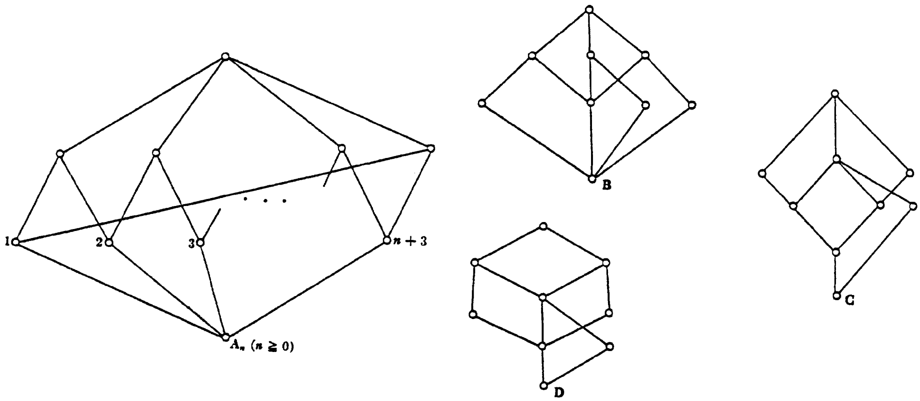


Figure 37: Some of the lattices from the forbidden list $\mathcal{L}$(*nonplanar*)

An early and important application of the theory of planar lattices is the complete description of the minimal, dimension three ordered sets [Ke77]. Although the details are involved the strategy of the proof is this. To begin with, if $dim(P) \geq 3$ and the completion is nondismantlable, then P, itself, must contain a cycle, and any cycle (with at least six elements) is minimal of dimension three. Otherwise, the completion is a dismantlable nonplanar lattice, whence it contains an order sublattice from the list $\mathcal{L}(nonplanar)$. From dismantlability it follows, too, that every element of this lattice is the supremum of at most two supremum irreducible elements and the infimum of at most two infimum irreducibles. Now, to each and every minimal nonplanar lattice, adjoin, for each of its elements, an additional pair of supremum irreducibles and an additional pair of infimum irreducibles. This ordered set still has dimension three and, as every order contains a minimal one, the proof ranges over all the subsets, whose completion still produces a minimal nonplanar lattice.

Theorem 4.7 (Dimension Three) [Ke77] *An ordered set has dimension* ≤ 2 *if and only if it contains no subset isomorphic to a member of the minimal list* $\mathcal{L}$(dimension three).

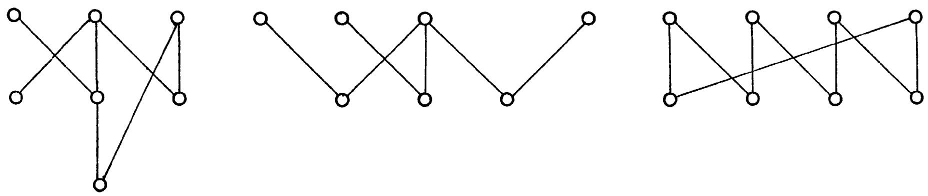


Figure 38: Some of the orders from the forbidden list $\mathcal{L}$(*dimension three*)

4.2 Planar orders

The theory of planar ordered sets is much less advanced than the theory of planar lattices. An indication is the difficulty to find any procedure, at all — that is, a finite one — to test whether an n-element ordered set is planar. Indeed, with some considerable effort we may derive an algorithm from these following observations, illustrated in Figure 39. Start with a planar upward drawing of a planar ordered set P.

(i) *There is a planar upward drawing in which all y-coordinates are distinct (just shake the elements about).*

(ii) *By successively sweeping a vertical from the right boundary of this planar upward drawing, elastically stretching its edges, the drawing is transformed into another planar upward drawing in which all vertices lie along a vertical.*

(iii) *These vertically aligned vertices induce a linear extension of P.*

(iv) *There is a bounded number of planar upward drawings of a planar n-element ordered set with elements aligned vertically, and of equivalence classes of monotonic edges of such "snake-like curves".*

The promised finitary procedure is this: for an arbitrary linear extension (of which there are at most $n!$) test whether there is a planar upward drawing with edges monotonically "snaking" around the vertically aligned elements, of which there is a finite number, assembled according to "combinatorial equivalence". If P is planar, then, for one of the linear extensions, there will be such a planar upward drawing; otherwise, P is nonplanar.

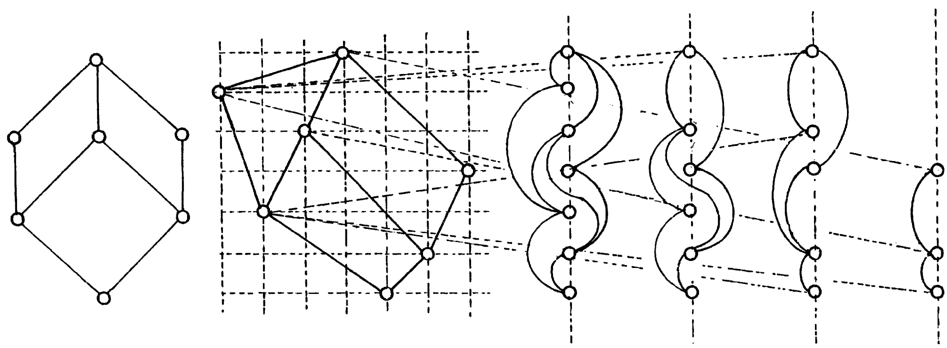


Figure 39: A finite procedure to test whether an ordered set has a planar upward drawing

4.3 Blocking relations

A quite pleasing metaphor arises from recent work in ice flow analysis, a natural and important theme in the hardy Canadian landscape. Massive ice flows in the Arctic and St. Lawrence River are routinely monitored by *Environment Canada* and its arm the *Ice Centre.* Ice flows consist of vast areas (even many thousands of kilometers on a side) of ice, largely of recent vintage (a few years old), interspersed with quite old ice (many years old). The very old ice is dangerous for boats and oil rigs, and indeed, any man-made vessel at all. As a matter of fact, every effort is made to keep away from old ice for its effect, upon collision, may be devastating — remember the Titanic! Ocean currents, wind, and temperature affect the icebergs' direction of flow. In fact, icebergs may change position and velocity substantially — even within hours.

The ice flows are monitored in two ways. First, satellite images (up to one thousand kilometers on a side) show the old ice as disjoint convex figures, often closely packed. These images are updated, as often as every three hours, and transmitted to boats (in real-time) as navigation aids for routing. Second, radar apparatus on the boat generates a map of the icebergs up to within a fixed radius of the boat. These radar images are used, in combination with the satellite images, to determine, by pattern matching, just where the boat actually is, within the larger satellite image. The boat's position with respect to the earth's latitude and longtitude is not nearly as important as its position with respect to the moving icefield!

The vessels relying on this data spend large amounts on fuel and, obviously, too, time is of the essence. Thus, shorter routes save fuel but, beware the "short" route which includes five cold months locked within an ice flow! In the long term, efforts are afoot to map the iceflows, over a period of years, to produce an atlas of iceflows to predict their position in the future.

A convenient model based on order, starts with disjoint convex figures, each representing old icebergs, and each assigned a fixed translation velocity (depending on the current, wind, and temperature). Say that a figure A *blocks* a figure B if there is a line joining a point of B to a point of A along the direction assigned to B. The transitive closure of this blocking relation is an order — provided there is no directed cycle. In this case, we call it an *m-directional blocking order*, where m is the number of different directions used among the convex figures (one for each figure).

The first indication that order may well play a role in such motion planning problems is this.

Theorem 4.8 [RiUr88] *There is a one-to-one correspondence between one-directional blocking orders and planar lattices* (see Figure 40).

This, in effect, provides the computational model for the satellite images. The analysis of the radar images contains some surprises, too, to which we turn below.

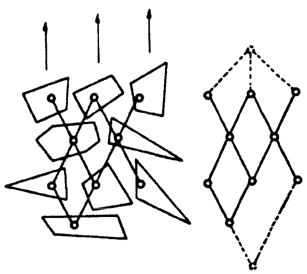


Figure 40: Blocking relations on disjoint convex figures on the plane, each assigned a common direction of motion, correspond to truncated planar lattices

5 Surfaces

Perhaps progress on the planarity problems will come from generalization? The impact of an unusual attack on a difficult problem occasionally leads to unexpected discoveries!

5.1 Dimension

One widely known concept, on which we have already touched, is tempting as generalization: *dimension.* For several reasons I think it is misleading, in this context. Certainly it has served us well in linking lattices to ordered sets, by way of planarity and the completion. Yet, there are scant further links for "higher-dimensional" orders. Indeed, from the drawing and reading perspectives there is actually evidence to reject dimension as a generalization of planarity.

Starting with an upward drawing, a *subdivision* of an edge $a \succ b$ is a three-element chain $a \succ s_{a,b} \succ b$ (replacing the original edge) in which the new interior element $s_{a,b}$ has only the comparabilities induced, above, by a, and below, by b. What affect does subdivision have on planarity? Of course, none at all, in the sense, that a planar ordered set remains planar no matter how many edges are "subdivided". In contrast, the effect on dimension may be devastating.

Theorem 5.1 [Sp88] *For any positive integer n, there exists an ordered set of dimension two, such that subdividing its edges yields an ordered set of dimension at least n.*

Parenthetically, it is worth noting that, for lattices, at least, subdivision does not change the dimension.

Theorem 5.2 [LLNR88] *For any finite ordered set P,*

$$dim(P) = dim(subdivision(completion(P))).$$

5.2 Spherical orders

Ice flows, then, are monitored in two ways: satellite imagery and boat-based radar. The satellite images may be modelled by planar lattices. How can we model the radar imagery? Consider an "illumination" model, viewed from the air, according to which a light source (origin of the boat's radar) is located on a plane disjoint from the collection of disjoint convex figures (icebergs). A figure A *obstructs* a figure B if a ray from the light source passes through B before it passes through A. The transitive closure is a *one-light source order* — provided there are no directed cycles (see Figure 41).

If the light source is sufficiently far away from the collection of disjoint convex figures then the light rays are, in effect, parallel and the one-light source order becomes a one-directional blocking order [FRU92]. More generally, a *spherical ordered set* is a finite ordered

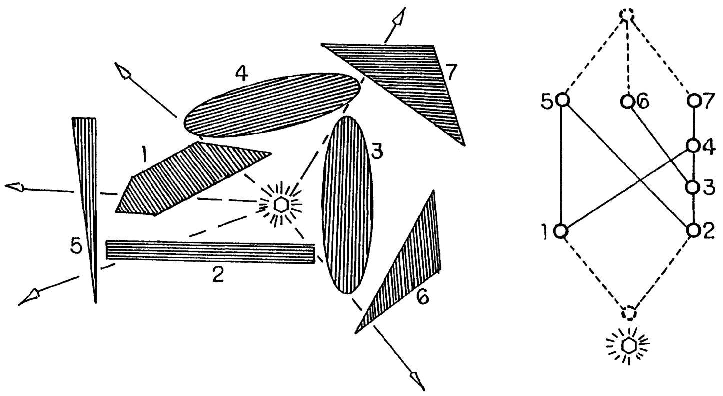


Figure 41: One-light source orders correspond to truncated spherical orders

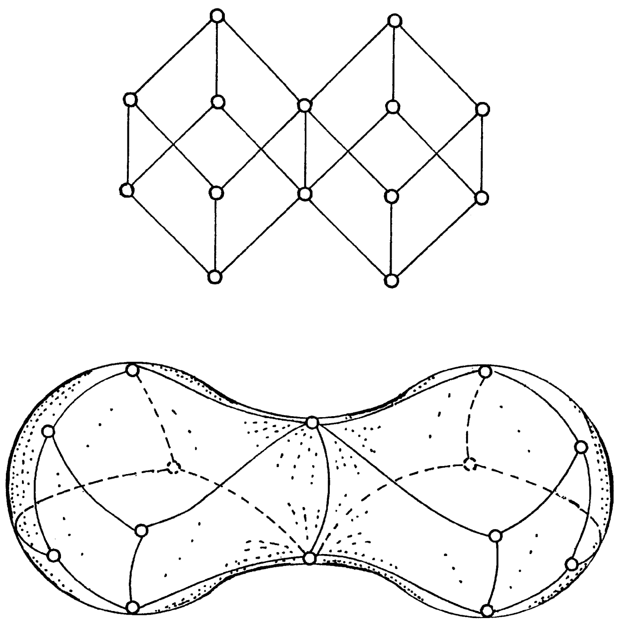


Figure 42: An ordered set, not a one-light source order (that is, not spherical), yet, can be drawn on a "peanut"-like surface

set with bottom and top elements having an upward drawing on the surface of a sphere such that the bottom is located at the south pole, the top at the north pole, all arcs are strictly increasing northward on the sphere, and no pair of arcs cross. The ordered set $\mathbf{Q}_3$ is spherical (see Figure 2(ii)) while the ordered set illustrated in Figure 42 is not.

Here are the basic results.

Theorem 5.3 [FRU92]

(i) *An ordered set is spherical if and only if it has a bottom, a top, and its covering graph is planar.*

(ii) *An ordered set is a one-light source order if and only if it can be obtained from a spherical ordered set by removing its bottom and its top.*

For instance, given a one-light source order, it is easy to see that its covering graph is planar. Place a point $\top$ far away from the convex closure of the disjoint figures and draw a directed arc from each "maximal" to $\top$; place a point $\perp$ at the light source and draw a directed arc from $\perp$ to each "minimal". Locating $\top$ at the north pole, $\perp$ at the south pole and wrapping the rest of the ordered set around the sphere, with edges monotonically upward will produce an upward drawing on the sphere, a spherical ordered set.

Spherical orders are not apparently linked uniformly to any other familiar combinatorial, order-theoretical parameter. As an example, the dimension can be small, or large [ReRi91]. More interesting, however, is this.

Theorem 5.4 [ELR91], [ReRi91] *There are ordered sets with (planar) upward drawings on a surface topologically equivalent to a sphere, and yet which have no (planar) upward drawings, at all, on a simple sphere, that is, on a ball.* (See Figure 42.)

5.3 Handles and genus

Little by little we are led, almost ineluctably, to the apparently distant realm of topological graph theory. Here we will find several beautiful results, easily compensating our modest efforts.

Every surface is topologically equivalent to a sphere with handles; its *genus* is the number of handles that must be added to obtain its homeomorphism type. The *graph genus* of an (undirected) graph is the smallest number g such that the graph can be drawn, without edge crossings, on a surface with genus g, that is, on a sphere with g handles [He79]. The *order genus* of an ordered set P is the smallest integer g such that it can be drawn, without edge crossings, on a surface with genus g, in such a way that, whenever $a < b$ in P, the z-coordinate of a is smaller than the z-coordinate of b, and all edges of P are monotonic with respect to the z-coordinate (see Figure 43). It is worth emphasizing the difference between graph embeddings on such surfaces and order embeddings on them. The monotonicity requirement is so exacting that, unlike graphs, there may, *a priori*, even be an ordered set

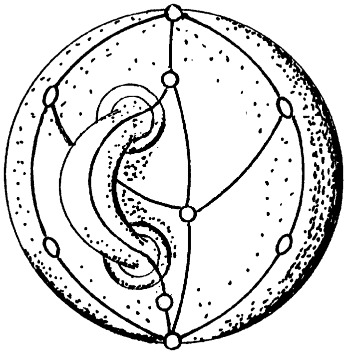


Figure 43: A lattice with order genus one

of genus $g > 0$ with no upward drawing, without edge crossings, on a surface of genus g *realizable* in $\mathbf{R}^3$.

Such drawings of ordered sets on surfaces, without edge crossings, obviously constitute a higher-dimensional analogue of "planarity". This generalization of planarity justifies itself even if only by the next result which furnishes the first nontrivial orientation invariant.

Theorem 5.5 [ELR91] *The order genus of an upward drawing equals the graph genus of its covering graph.*

As the order genus depends only on the covering graph, every orientation of a fixed (covering) graph has the same order genus.

Theorem 5.6 [ELR91] *Order genus is an orientation invariant.*

The proof idea of Theorem 5.5 is basically an easy one. Remember "Flatland"?

Consider the case of genus zero. Suppose the covering graph G of an ordered set P has genus zero. This means that G can be drawn, without edge crossings, on the surface of a sphere and, in particular, that it can be drawn (as an undirected graph), without edge crossings, on the plane. To each vertex a of such a planar drawing, assign a nonnegative integer $h(a)$, its height, that is the size of the largest chain from it to a minimal element. Thus, every minimal element has height one.

Now, supposing that G is drawn in the $z = 0$ plane of $\mathbf{R}^3$. Replace each vertex a of G whose coordinates are $(x, y, 0)$ by the vertex in $\mathbf{R}^3$ with coordinates $(x, y, h(a))$ and draw a straight edge in $\mathbf{R}^3$ for each pair of vertices joined by a covering edge in G. This produces a drawing of P on a surface, much like a topographic drawing of it, topologically equivalent to a hemisphere which, in turn, has genus zero. See Figure 44.

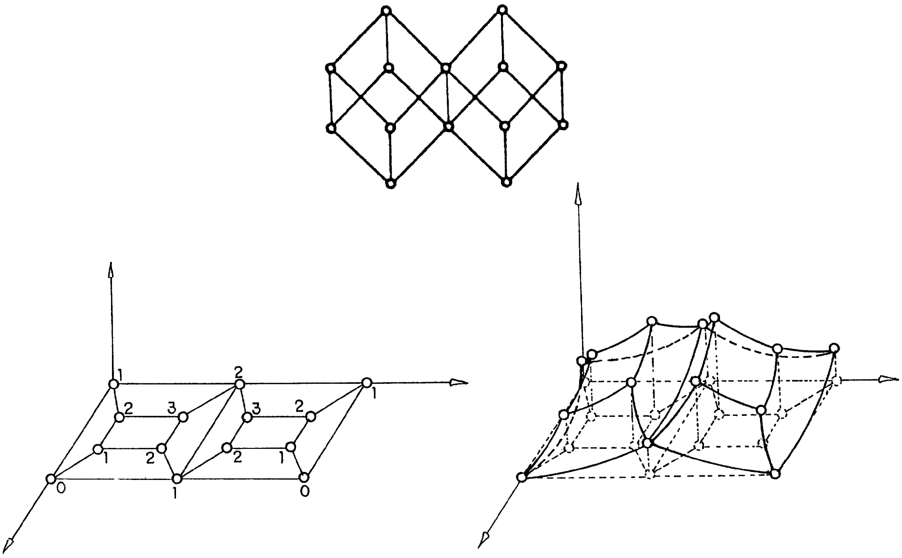


Figure 44: (i) The drawing of an ordered set; (ii) a drawing of its planar covering graph on the horizontal plane; (iii) a "lifting" of its covering graph onto a surface topologically equivalent to a hemisphere

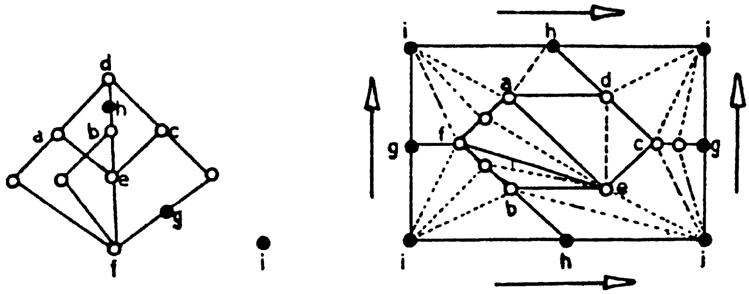


Figure 45: (i) The drawing of an ordered set; (ii) a triangulated polygon model of it, with opposite edges identified to form a torus

This approach may be generalized to a surface of any genus, as any surface has a polygon model according to which certain designated pairs of edges are identified, and this polygon maybe drawn in the $z = 0$ plane of $\mathbf{R}^3$ and, subsequently, lifted, much as in the genus zero case. Parenthetically, it comes as a striking realization that topological graph theorists hardly ever seem to work with the sphere model and its handles, although the end results are formulated in terms of them; instead, the techniques and algorithms are in terms of the plane, polygon model (see Figure 45).

5.4 Lattices

Loosely speaking, order genus is a higher-dimensional analogue of planarity — in the obvious sense that, an embedding on a surface with handles and without crossing lines, is "locally planar". Apart from the fact that, for lattices, planarity is equivalent to dimension at most two, there is no further apparent connection between order genus and order dimension. For instance, the k-spider (see Figure 46) has order dimension at most three (for apart from its top and bottom, its drawing is a tree, whose order dimension is always at most three [TrMo74]). In contrast, the order genus of the k-spider grows with k.

Theorem 5.7 *The order genus of the k-spider is* $\lceil \frac{k-2}{4} \rceil$.

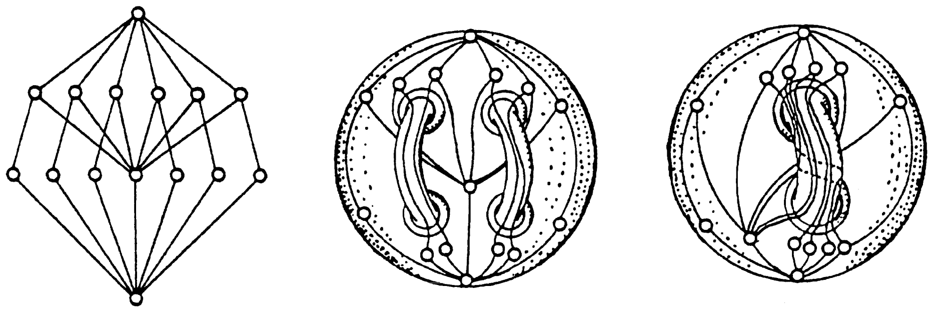


Figure 46: (i) The "k-spider", k=6; (ii) an embedding of the six-spider with two handles; (iii) an embedding of the six-spider with one handle — its order genus is one

There are planar orders with high order dimension [Ke81], although they cannot, of course, be lattices (see Figure 19(iii) and Figure 47). There are, however, orders with small order genus (zero) and large order dimension (see Figure 48).

Evidence from the theory of planar lattices lent credibility to the question whether every lattice of dimension three contains a join irreducible or a meet irreducible of degree at most three [NR89b]?

It is false [ReRi91].

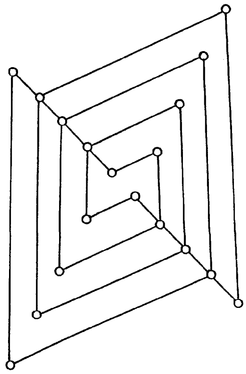


Figure 47: A planar ordered set with dimension six

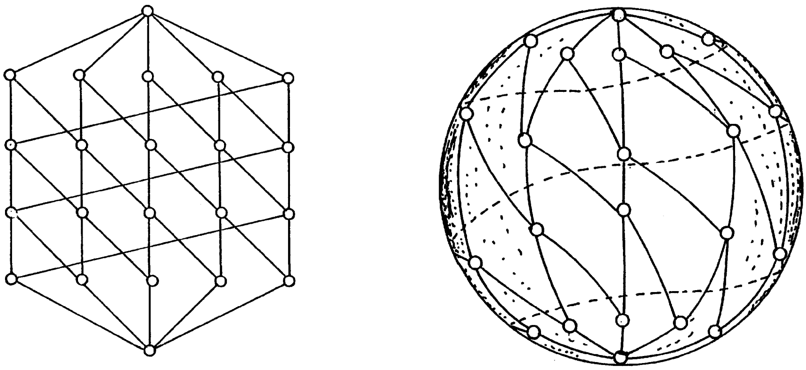


Figure 48: (i) An ordered set of dimension five; (ii) an upward drawing of it on the sphere

However, in the light of surface embeddings of upward drawings, such a result holds in terms of order genus.

Theorem 5.8 [ReRi91] *Every lattice contains a join irreducible element of degree at most* $4genus+3$.

Some estimates about the number of edges of the covering graph are known; for instance, a planar lattice with $n \geq 3$ vertices has at most $2n - 4$ edges. This theorem can, in turn, be used to bound the number of edges in terms of the genus.

Theorem 5.9 [ReRi91] *For any lattice*

$$|edges| \leq (4genus + 3)|vertices|.$$

Order genus is not a hereditary property. Thus, a subset of an ordered set with small order genus may have large order genus (see Figure 49). It is surprising that this hereditary property does hold true for lattices.

Theorem 5.10 [ReRi91] *Let K and L be lattices. If K is an order sublattice of L then the order genus of K is at most the order genus of L.*

This should be compared, too, with the earlier result that, *an order sublattice of a planar lattice is planar* [Ke87b], cf. [NRU92].

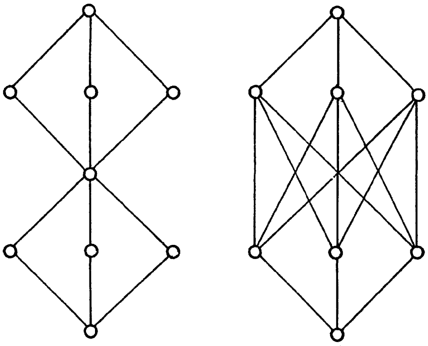


Figure 49: (i) An order with order genus zero; (ii) a subset with genus one

5.5 Cylinders and pants

It is tempting to stroll along the landscape of topological graph theory, now that we have touched down upon its rich surface.

Although every ordered set has an order genus, the "minimal" surface on which the drawing is embedded may not be recognizable as a sphere with handles for, *a priori*, it need not even be representable in $\mathbf{R}^3$. There is, however, an embedding possible on a sphere, in which the handles are all disjoint and located without intersection of its constituent surfaces vertically on the surface of the sphere (see Figure 50). This representation of the ordered set may, however, use more handles than its order genus.

Theorem 5.11 [NR89a] *Every ordered set has an upward drawing on a vertical, many-holed torus.*

The proof idea adapts the Theorem 1.11 about three-channel drawing, with bending and stretching.

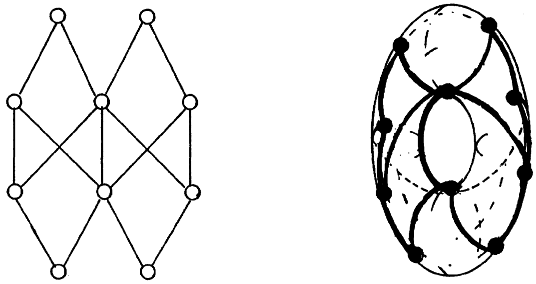


Figure 50: Drawing orders on an upright torus

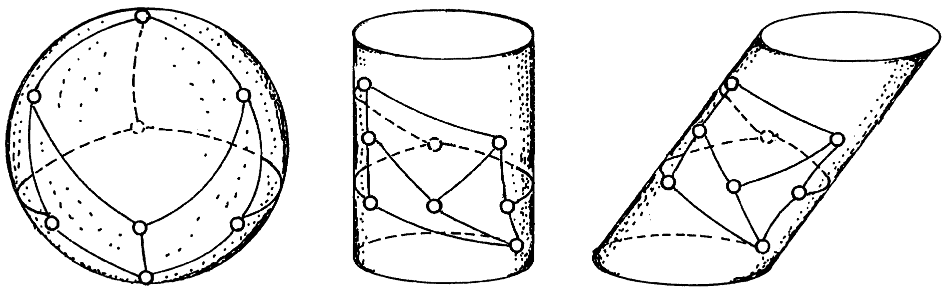


Figure 51: Orders drawable on the surface of a horizontal cylinder are planar although nonplanar orders are drawable on tilted cylinders

As it turns out such surfaces have appeared also in the exotic area of "superstring theory" [GSW87], where they are viewed as aspects of attire, pairs of "pants" or "trousers", sewn together vertically. Such surfaces are viewed in theoretical physics as models of space-time.

Another intriguing direction of research concerns the precise"orientation" of the orientable surface in $\mathbf{R}^3$. Thus, on a horizontal plane ($z = 0$) only antichains can be embedded, because edges in an upward drawing must be monotonic arcs (with respect to the z-coordinate. On the other hand, on any inclined plane we can draw any planar order.

In this same sense consider the orders that can be drawn on the surface of a cylinder. As long as the right cylinder's central axis is tilted away from the horizontal, any ordered set of genus = 0 can be drawn. What is astonishing, however, is that on the surface of the right cylinder, with horizontal axis, only planar orders can be drawn [Ri93c] (see Figure 51)! Thus, an orientable surface may be distinguished according to its position in $\mathbf{R}^3$, by the orders drawable on it. For instance, while $\mathbf{Q}_3$ can be drawn on the surface of any cylinder with non-horizontal axis, it cannot be drawn on the surface of the horizontal cylinder!

This may herald a new approach to the classification of orientable surfaces ...

References

[Aj73] Ajtai, M., On a class of finite lattices, *Period. Math. Hungar.* **4** (1973), 217–220.

[Al65] Alvarez, L., Undirected graphs as graphs of modular lattices, *Canad. J. Math.* **17** (1965), 923–932.

[BFR71] Baker, K.A., Fishburn, P., and Roberts, F.S., Partial orders of dimension 2, *Networks* **2** (1971), 11–28.

[BaRi89] Bandelt, H.J., and Rival, I., Diagrams, orientations, and varieties, *Order* **6** (1989), 119–132.

[BLR90] di Battista, G., Liu, W.-P., and Rival, I., Bipartite graphs, upward drawings, and planarity, *Inform. Process. Lett.* **36** (1990), 317–322.

[BaNa88] di Battista, G., and Nardelli, E., Hierarchies and planarity theory, *IEEE Trans. Systems Man Cybernet.* **18** (1988), 1035–1046.

[Bi67] Birkhoff, G., *Lattice Theory, 3rd ed.*, American Mathematical Society, Providence, RI 1967, Ex. 7(a), p.32.

[Bj80] Björner, A., Shellable and Cohen-Macaulay partially ordered sets, *Trans. Amer. Math. Soc.* **260** (1980), 159–183.

[BrNe91] Brightwell, G., and Nešetřil, J., On the dimension of partially ordered sets, *Discrete Math.* **88** (1991), 135–156.

[Cr84] Crapo, H., Selectors. A theory of formal languages, semimodular lattices, branchings and shelling processes, *Adv. Math.* **54** (1984), 233–277.

[Cz91] Czyzowicz, J., Lattice diagrams with few slopes, *J. Combin. Theory, Ser. A* **56** (1991), 96–108.

[CPR90a] Czyzowicz, J., Pelc, A., and Rival, I., Planar ordered sets of width two, *Math. Slovaca* **40** (4) (1990), 375–388.

[CPR90b] Czyzowicz, J., Pelc, A., and Rival, I., Drawing orders with few slopes, *Discrete Math.* **82** (1990), 233–250.

[CPRU90] Czyzowicz, J., Pelc, A., Rival, I., and Urrutia, J., Crooked diagrams with few slopes, *Order* **7** (1990), 133–143.

[Di50] Dilworth, R.P., A decomposition theorem for partially ordered sets, *Ann. of Math.* **51** (1950), 161-166.

[DuRi78] Duffus, D., and Rival, I., Crowns in dismantlable partially ordered sets, in: *Combinatorics* (*Keszthely 1976*) (A. Hajnal and V.T. Sós, eds.) Colloq. Math. Soc. János Bolyai **18**, North-Holland, Amsterdam 1978, 271–292.

[DuRi79] Duffus, D., and Rival, I., Retracts of partially ordered sets, *J. Austral. Math. Soc. Ser. A* **27** (1979), 495–506.

[DuRi81] Duffus, D., and Rival, I., A structure theory for ordered sets, *Discrete Math.* **35** (1981), 53–118.

[DuRi83] Duffus, D., and Rival, I., Graphs orientable as distributive lattices, *Proc. Amer. Math. Soc.* **88** (1983), 197–200.

[ElRi85] El-Zahar, M.H., and Rival, I., Greedy linear extensions to minimize jumps, *Discrete Appl. Math.* **11** (1985), 143–156.

[ELR91] Ewacha, K., Li, W., and Rival, I., Order, genus and diagram invariance, *Order* **8** (1991), 107–113.

[FaSc87] Faigle, U., and Schrader, R., Set-up optimization problems with matroid structure, *Order* **4** (1987), 43–54.

[FRU92] Foldes, S., Rival, I., and Urrutia, J., Light sources, obstructions, and spherical orders, *Discrete Math.* **102** (1992), 13–23.

[Ga67] Gallai, T., Transitiv orientierbare Graphen, *Acta Math. Acad. Sci. Hungar.* **18** (1967), 25–66.

[GSW87] Green, M.B., Schwarz, J.H., and Witten, E., *Superstring Theory*, Cambridge, 1987.

[Gr58] Grötsch, H., Ein Dreifarbensatz für dreikreisfreie Netze auf der Kugel, *Wiss. Z. Martin-Luther-Univ. Halle-Wittenberg* (*Math. Natur. Reihe*) **8** (1958), 109–119.

[Ha84] Habib, M., Comparability invariants, in: *Ordres: Descriptions et Rôles* (M. Pouzet and D. Richard, eds.), Ann. Discrete Math. **23**, North-Holland, Amsterdam 1984, 371–386.

[He79] Henle, M., *A Combinatorial Introduction to Topology*, Freeman 1979.

[HoTa74] Hopcroft, J., and Tarjan, R.E., Efficient planarity testing, *J. Assoc. Comput. Mach.* **21** (4) (1979) 549–568.

[HuLu90] Hutton, M., and Lubiw, A., *SIAM Symp. Discrete Appl. Math.*, San Francisco 1990.

[JNR87] Jégou, R., Nowakowski, R., and Rival, I., The diagram invariant problem for planar lattices, *Acta. Sci. Math.* (*Szeged*) **51** (1987), 103–121.

[Ke77] Kelly, D., The 3-irreducible partially ordered sets, *Canad. J. Math.* **29** (1977), 367-383.

[Ke81] Kelly, D., On the dimension of partially ordered sets, *Discrete Math.* **35** (1981), 135–156.

[Ke85] Kelly, D., Comparability graphs, in: *Graphs and Order* (I. Rival, ed.), NATO ASI Ser. C 147, Reidel, Dordrecht 1985, 3–40.

[Ke87a] Kelly, D., Fundamentals of planar ordered sets, *Discrete Math.* **63** (1987), 197–216.

[Ke87b] Kelly, D., private communication (cf. [NRU92]).

[KeRi74] Kelly, D., and Rival, I., Crowns, fences and dismantlable lattices, *Canad. J. Math.* **26** (1974), 1257–1271.

[KeRi75] Kelly, D., and Rival, I., Planar lattices, *Canad. J. Math.* **27** (1975), 636–665.

[KeTr82] Kelly, D., and Trotter, W.T., Dimension theory for ordered sets, in: *Ordered Sets* (I. Rival, ed.), NATO ASI Ser. C 83, Reidel, Dordrecht 1982, 171–211.

[KlRo75] Kleitman, D.J., and Rothschild, B.L., Asymptotic enumeration of partial orders on a finite set, *Trans. Amer. Math. Soc.* **205** (1975), 205–220.

[KoLo83] Korte, B., and Lovász, L., Structural properties of greedoids, *Combinatorica* **3** (1983), 359–374.

[LLNR88] Lee, G.J., Liu, W.-P., Nowakowski, R., and Rival, I., Dimension invariance of subdivision, *Technical Report, University of Ottawa* TR-88-30, 1988, 14pp.

[LEC67] Lempel, A., Even, S., and Cederbaum, I., An algorithm for planarity testing of graphs, in: *Theory of Graphs*, (*International Symposium, Rome 1966*) (P. Rosenstiehl, ed.), Gordon and Breach, New York 1967, 215–232.

[LR91a] Liu, W.-P., and Rival, I., Inversions, cuts, and orientations, *Discrete Math.* **87** (1991), 163–174.

[LR91b] Liu, W.-P., and Rival, I., Enumerating orientations of ordered sets, *Discrete Math.* **88** (1991), 239–247.

[Mo72] Mosesjan, K.M., Strongly basable graphs, *Akad. Nauk Armyan. SSR Dokl.* **54** (1972), 134–138; **54**, 241–245; **55**, 83–86. (*Russian*)

[NeRö78] Nešetřil, J., and Rödl, V., On a probabilistic graph-theoretical method, *Proc. Amer. Math. Soc.* **72** (1978), 417–421.

[NeRö87] Nešetřil, J., and Rödl, V., The complexity of diagrams, *Order* **3** (1987), 321–330.

[NR89a] Nowakowski, R., and Rival, I., Bending and stretching orders into three channels, technical report, University of Ottawa, 1989.

[NR89b] Nowakowski, R., and Rival, I., private communication, 1989 (cf. [ReRi91]).

[NRU92] Nowakowski, R., Rival, I., and Urrutia, J., Lattices contained in planar orders are planar, to appear in *Algebra Universalis.*

[Or62] Ore, O., *Theory of Graphs*, American Mathematical Society, Providence, RI 1962.

[PeRi91] Pelc, A., and Rival, I., Orders with level diagrams, *European J. Combin.* **12** (1991), 61–68.

[Pl76] Platt, C.R., Planar lattices and planar graphs, *J. Combin. Theory Ser. B* **21** (1976), 30–39.

[PrTa88] Preparata, F.P., and Tamassia, R., Fully dynamic techniques for point location and transitive closure in planar structures, *Proc. IEEE Symp. Found. Comp. Sci.* **29** (1988), 558–567.

[Pr86] Pretzel, O., Orientations and reorientations of graphs, in: *Combinatorics and Ordered Sets* (I. Rival, ed.), *Contemp. Math.* **57**, American Mathematical Society, Providence, RI 1986, 103–125.

[ReRi91] Reuter, K., and Rival, I., Genus of orders and lattices, in: *Graph-Theoretic Concepts in Computer Science* (R. Möhring, ed.), Lecture Notes in Comput. Sci. **484** Springer-Verlag, Berlin-Heidelberg-New York 1991, 260–275.

[Ri76a] Rival, I., A fixed point theorem for finite partially ordered sets, *J. Combin. Theory, Ser. A* **21** (1976), 309–318.

[Ri76b] Rival, I., Combinatorial inequalities for semimodular lattices of breadth two, *Algebra Universalis* **6** (1976), 303–311.

[Ri78] Rival, I., Lattices and partially ordered sets, in: *Colloq. CNRS Probab. Combin. et Théorie des Graphes*, Orsay **260** (1978), 349–351.

[Ri82] Rival, I., The retract construction, in: *Ordered Sets* (I. Rival, ed.), NATO ASI Ser. C 83, Reidel, Dordrecht 1982, 97–122.

[Ri85a] Rival, I., The diagram, in: *Graphs and Order* (I. Rival, ed.), NATO ASI Ser. C 147, Reidel, Dordrecht 1985, 103–133.

[Ri85b] Rival, I., The diagram, *Order* **2** (1985), 101–104.

[Ri86] Rival, I., Stories about order and the letter N (en), in: *Combinatorics and Ordered Sets* (I. Rival, ed.), Contemp. Math. **57**, American Mathematical Society, Providence, RI 1986, 263–285.

[Ri89] Rival, I., Graphical data structures for ordered sets, in: *Algorithms and Order* (I. Rival, ed.), NATO ASI Ser. C 255, Kluwer, Dordrecht 1989, 3–31.

[Ri93a] Rival, I., *Introducing Ordered Sets*, to appear in 1993.

[Ri93b] Rival, I., Problems about planar orders, in: NATO ASI (N. Sauer, ed.), to appear (1993).

[Ri93c] Rival, I., Order, iceflow and invariance, in: *The Birkhoff Symposium* (R. Wille, ed.), to appear (1993).

[RiUr88] Rival, I., and Urrutia, J., Representing orders by translating convex figures in the plane, *Order* **4** (1988), 319–339.

[RiZa85] Rival, I., and Zaguia, N., Antichain cutsets, *Order* **1** (1985), 235–247.

[Ru89] Rutkowski, A., The fixed point property for small sets, *Order* **6** (1989), 1–14.

[Sa85] Sands, B., Problem 2.7, in: *Graphs and Order* (I. Rival, ed.), Reidel, 1985, p.531.

[Sp88] Spinrad, J., Edge subdivision and dimension, *Order* **5** (1988), 143–147.

[Ste87] Stephan, J., *Liniendiagramme von Verbänden*, Diplomarbeit, TH Darmstadt, 1987.

[Sto80] Storer, J.A., The node-cost measure for embedding graphs in the planar grid, in: *Proc. 12th ACM Symp. Theory Comput.*, 1980, 201–210.

[Ta87] Tamassia, R., On embedding a graph in the grid with the minimum number of bends, *SIAM J. Comput.* **16** (1987), 421–443.

[Th89] Thomassen, C., Planar acyclic oriented graphs, *Order* **5** (1989), 349–361.

[TrMo74] Trotter, W.T., and Moore, J.I., The dimension of planar posets, *J. Combin. Theory Ser. B* **22** (1977), 54–67.

Hyperidentities

Dietmar SCHWEIGERT
Mathematisches Institut
Universität Kaiserslautern
DW-6750 Kaiserslautern
Germany

Dla Magdy, Dagusi i Krzysia

Abstract

If one considers free algebras as semigroups of composition of terms or more specifically as clones of terms, the identities of these semigroups (respectively clones) can be interpreted as hyperidentities. Hyperidentities contain hypervariables which stand for terms, and describe the manipulation of these terms. We present a logic for hyperidentities and more generally for hybrid identities, a completeness theorem and deal with solid varieties. Separating hyperidentities for various semigroups and clones are presented next. The congruences of free algebras and of the clones of terms are described. Furthermore, many open questions and problems are included.

Introduction

The concept of a free algebra plays an essential role in universal algebra and in computer science. Manipulation of terms, calculations and the derivation of identities are performed in free algebras. Word problems, normal forms, systems of reductions, unification and finite bases of identities are topics in algebra and logic as well as in computer science.

A very fruitful point of view is to consider structural properties of free algebras. A.I. Malcev initiated a thorough research of the congruences of free algebras. Henceforth congruence permutable, congruence distributive and congruence modular varieties are intensively studied. Many Malcev type theorems are connected to the congruence lattice of free algebras.

Here we consider free algebras as semigroups of compositions of terms and more specifically as clones of terms. The properties of these semigroups and clones are adequately described by hyperidentities. Naturally, many theorems of "semigroup" or "clone" type can be derived.

I. G. Rosenberg and G. Sabidussi (eds.), Algebras and Orders, 405–506.

This topic of research is still in its beginning and therefore some of the concepts and results cannot be presented in a final and polished form. Furthermore numerous problems and questions which are of importance for the further development of the theory of hyperidentities remain open.

Contents

Preliminaries

An identity is a pair of terms where the variables are bound by the all quantifier. Let us take the following medial identity as an example

$$\forall u \forall x \forall y \forall w \quad (x \cdot x) \cdot (y \cdot w) = (u \cdot y) \cdot (x \cdot w).$$

An identity can by considered as a notion in a first order language with equality.

Let us look at the following hyperidentity

$$\forall F \forall u \forall x \forall y \forall w \quad F(F(u,x), F(y,w)) = F(F(u,y), F(x,w)).$$

A hyperidentity can be considered as a notion in a second order language with equality. A second order language allows the quantification of predicate or operator variables. We consider the operator variables F in a very specific way. Primarily all our operator variables are restricted to functions of a given arity; in our example to binary functions. Secondly we bind the interpretation of F to term functions. Therefore such kind of operator variables are called hypervariables. As is customary, we will not write quantifiers in front of identities and hyperidentities.

Let us consider the variety of distributive lattices. Then the list of all binary terms consists of

$$e_1^2(x,y) = x, \quad e_2^2(x,y) = y, \quad x \wedge y, \quad x \vee y.$$

Let us replace the binary hypervariable F in the above hyperidentity by a binary term leaving the variables unchanged. For $x \wedge y$ we get the identity

$$(u \wedge x) \wedge (y \wedge w) = (u \wedge y) \wedge (x \wedge w)$$

which holds for the variety of distributive lattices. In the other three cases we get the identities

$$u = u, \quad w = w \text{ and } (u \vee x) \vee (y \vee w) = (u \vee y) \vee (x \vee w)$$

which also hold for distributive lattices. We say that the above hyperidentity holds in the variety D of distributive lattices as every binary term yields an identity which holds in D. Let us take another example: the variety of abelian groups (where we write the operation as addition). Every binary term, for instance $((x + y) + (x + x)) + y$, can be written in a normal form $ax + by, a, b \in \mathbb{N}_0$; in our case $3x + 2y$. The above hyperidentity holds for the variety of abelian groups because it is transformed into the following identity interpreting F:

$$a(au + bx) + b(ay + bw) = a(au + by) + b(ax + bw)$$

The reader will recognize that we consider only a small fragment of a second order logic. Of course all these restrictions reduce the expressive power of a second order language. But nevertheless, by hyperidentities one can express more than by identities.

Part 1

Hyperidentities

1.1 Hyperterms

The four sections 1.1–1.4 are based on the paper by Graczyńska and Schweigert [GraSch 90].

Notation Our nomenclature is basically the same as in [Grätzer 79]. We consider varieties of algebras of a given type. A type of algebras τ is a sequence $(n_0, n_1, \ldots, n_\gamma, \ldots)$ of positive integers, $\gamma < O(\tau)$, where $O(\tau)$ is an ordinal, called the order of τ. For every $\gamma < O(\tau)$ we have a symbol f_γ for an n_γ-ary operation. Moreover, for every γ with $n_\gamma > 0$ the symbol F_γ is associated. F_γ is called an n_γ-ary hypervariable.

Definition 1.1.1 Let τ be a given type. The *n-ary hyperterms of type* τ are recursively defined by:

(1) the variables $x_1, \ldots, x_n$ are n-ary hyperterms;

(2) if $T_1, \ldots, T_m$ are n-ary hyperterms and F is an m-ary hypervariable of type τ, then $F_\gamma(T_1, \ldots, T_m)$ is an n-ary hyperterm of type τ.

$H^n(\tau)$ is the smallest set containing (1) which is closed under finite application of (2). $H(\tau) = \bigcup(H^n(\tau) : n \in \mathbb{N})$ is called the set of all *hyperterms of type* τ (where $\mathbb{N}$ is the set of all positive integers).

A *hyperidentity of type* τ is a pair of hyperterms (T_1, T_2), which is also denoted by $T_1 = T_2$.

The *free algebra* in *countably many* variables of a variety V of type τ is denoted by $T(V)$ and its elements t are called *terms*. If V is generated by the algebra $\mathbf{A}$ we write $T(\mathbf{A})$ instead of $T(V)$.

Definition 1.1.2 Let (T_1, T_2) be a hyperidentity of type τ and let V be a variety of type μ. If every n_γ-ary hypervariable occurring in (T_1, T_2) is replaced by an n_γ-ary term $t_\gamma \in T(V)$ leaving the variables $(x_i : i \in \mathbb{N})$ unchanged in (T_1, T_2), then the resulting identity (t_1, t_2) is called a *transformation* of the hyperidentity (T_1, T_2).

For a more formal definition consider 1.7.

Example Let $F(F(u,x), F(y,v)) = F(F(u,y), F(x,v))$ by a hyperidentity of type (2). Let V be the variety of abelian groups $(G : +, -, 0)$ of type $(0, 1, 2)$. Then $(u+x)+(y+v) = (u+y)+(x+v)$ is a transformation of the above hyperidentity. Let $ax + by, a, b \in \mathbb{Z}$ be a binary term of $T(V)$. Then

$$a(au + bx) + b(ay + bv) = a(au + by) + b(ax + bv)$$

is another example of a transformation of the above hyperidentity.

If E is a set of hyperidentities of type τ, then the set of all transformations of E for a variety V of type τ is denoted by $I_V(E)$.

Definition 1.1.3 A variety V of type μ *satisfies the hyperidentity* (T_1, T_2) of type τ if the set $I_V((T_1, T_2))$ of all transformations of (T_1, T_2) is contained in the set of identities which hold for V.

Example The hyperidentity $F(F(u,x), F(y,v)) = F(F(u,y), F(x,v))$ is satisfied by the variety of abelian groups.

Definition 1.1.4 Let (t_1, t_2) be an identity which holds for a variety V. If every n_γ-ary operation symbol f_γ occurring in (t_1, t_2) is replaced by an n_γ-ary hypervariable F_γ leaving the variables unchanged, then the resulting hyperidentity (T_1, T_2) is called the *transformation* of (t_1, t_2).

If Σ is a set of identities of the variety V of type μ, then $H_\mu(\Sigma)$ denotes the set of all transformations of identities in Σ.

Example Let V be a variety of lattices of type (2,2). Let ε be the identity $x = x \vee x$. Then the transformation $H_\mu(\varepsilon)$ which equals $x = F(x,x)$ is the hyperidentity which holds for the variety of lattices. On the other hand the identity ε' which is of the form $x \vee y = y \vee x$ is transformed to the hyperidentity $H_\mu(\varepsilon')$ of the form $F(x,y) = F(y,x)$ which does not hold for a nontrivial variety V of lattices.

1.2 Completeness

Definition 1.2.1 Following G. Birkhoff (comp. [Grätzer 79], [Taylor 79]), we use the following *rules of derivation* for hyperidentities of a given type τ:

(1) $T_1 = T_1$ for every hyperterm $T_1 \in H(\tau)$;

(2) $T_1 = T_2$ implies $T_2 = T_1$, for hyperterms $T_1, T_2 \in H(\tau)$;

(3) $T_1 = T_2, T_2 = T_3$ implies $T_1 = T_3$ for hyperterms $T_1, T_2, T_3 \in H(\tau)$;

(4) $T_i = S_i$ for $i = 1, \ldots, m_\gamma$, implies $F_\gamma(T_1, \ldots, T_{m_\gamma}) = F_\gamma(S_1, \ldots, S_{m_\gamma})$ for hyperterms $T_i, S_i \in H(\tau)$ and m_γ-ary hypervariables F_γ.

(5) $T(x_1, \ldots, x_n) = S(x_1, \ldots, x_n)$ implies $T(R_1, \ldots, R_n) = S(R_1, \ldots, R_n)$ for $T, S, R_1, \ldots, R_n \in H(\tau)$.

Remark 1.2.2 If one considers $H = \bigcup(H(\tau) : \tau \in \mathbf{Q})$, where $\mathbf{Q}$ is the set of all well-ordered sequences, then the above rules hold for hyperidentities in general. In the sequel we shall use also an analogous rule to (5), but for hypervariables. This was the main idea of [Belousov 65] (comp. [Aczél 71], [Taylor 81]).

First, we recursively define the notion of a substitution of a hypervariable by a hyperterm. Let T be a hyperterm of type τ. Consider a hypervariable F_{γ_1}, and a hyperterm R_1 of type τ, both of the arity m. We define the term T^*, called the substitution of the hypervariable F_{γ_1} in the term T by the hyperterm R_1, as follows:

(1^0) If T is a variable, then T^* is equal to T;

(2^0) If T has the form $F_\gamma(T_1, \ldots, T_m)$, then T^* has the form:

$$R_1(T_1^*, \ldots, T_n^*) \text{ if } \gamma = \gamma_1;$$

and

$$F_\gamma(T_1^*, \ldots, T_n^*) \text{ if } \gamma \neq \gamma_1.$$

The rule (6) is called a *hypersubstitution*, and is defined in the following way:

(6) $T_1 = T_2$ implies $T_1^* = T_2^*$ for any $T_1, T_2 \in H(\tau)$ and any simultaneous hypersubstitution of hypervariables in T_1 and T_2 by a hyperterm of the same arity.

Example Consider the hyperidentity $Q(Q(x, y, z), y, z) = Q(x, y, z)$ and the hyperterm $T(x, y, z) = F(G(x, y), z)$. By rule (6) we derive $F(G(F(G(x, y), z), y), z) = F(G(x, y), z)$. The latter hyperidentity of type (2,2) is also called a hyperconsequence of type (2,2) from the former hyperidentity of type (3).

Remark 1.2.3 Note that rule (6) commutes with all rules of derivation (1)-(5) (i.e. if Σ is a set of identities closed under the rule (6), then all consequences of Σ by the rules (1)-(6) are consequences of Σ by the rules (1)-(5)).

Given a variety V of type τ, $\mathrm{Id}(V)$ denotes the set of all identities satisfied in V (see [Grätzer 79], p. 169, 170). $E_\tau(V)$ denotes the set of all hyperidentities of type τ which are satisfied by the variety V. Furthermore $E(V)$ denotes the set of all hyperidentities of any type which hold for V. Obviously, $E_\tau(V) \subseteq E(V)$. Furthermore, if $V_1 \subseteq V_2$ then $E_\tau(V_2) \subseteq E_\tau(V_1)$. $H_\tau(\Sigma) :=$ set of all transformations of identities of Σ to hyperidentities of type τ which may or may not hold in V.

Proposition 1.2.4 *Let V be a nontrivial variety of lattices of type* (2,2). *Then $E_\tau(V)$ is properly contained in $E(V)$.*

Proof In [Penner 81] it is proved that for any positive integer m there exists a hyperidentity (T_1, T_2) which is satisfied in V but does not follow from hyperidentities involving at most m-ary hypervariables. For $m = 2$ we have the statement of Proposition 1.2.4.

Definition 1.2.5 The set V of all varieties V of type μ which satisfy a set E of hyperidentities of type τ is called a *hypervariety C of type* (τ, μ). We say that E *defines* C. If $\tau = \mu$, then C is called a *hypervariety of type* τ.

Completeness Theorem *A set Σ of hyperidentities of type τ can by represented in the form $E_\tau(K)$, for some variety K of type τ, if and only if Σ is closed under rules* (1)-(6).

Proof This theorem is a slight modification of G. Birkhoff's theorem (see [Birkhoff 35]) for sets of identities. The proof is similar to that of [Grätzer 79], p. 171. Obviously the set of hyperidentities of type τ of the variety K must be closed under rules (1)-(6).

Take a set Σ of hyperidentities of type τ, closed under rules (1)-(6). Consider the set $I_V(\Sigma)$ of identities of type τ for a fixed variety V of type τ. Then the set $\Sigma_1 = I_V(\Sigma)$ is closed under the rules of inference (i)-(v) of [Grätzer 79], p. 170. Consider the variety K of type τ, constructed as in [Grätzer 79], p. 171. Then Σ_1 is the set of identities of K. Moreover $\Sigma = H_\tau(\Sigma_1) = E_\tau(K)$, because of the assumption that Σ is closed under the rule (6).

1.3 Solid varieties

We say that a hyperidentity is satisfied by an algebra $\mathbf{A}$, if it is satisfied in the variety generated by $\mathbf{A}$.

An algebra $\mathbf{A}$ is solid if every identity satisfied in $\mathbf{A}$ is transformed into a hyperidentity which is satisfied in $\mathbf{A}$.

Definition 1.3.1 Let Σ be the set of all identities which hold for the variety V of type τ. V is called *solid* if $E_\tau(V) = H_\tau(\Sigma)$.

Theorem 1.3.2 *Let Σ be the set of all identities of the variety V of type τ. V is solid if* *only if* $\Sigma = I_V(E_\tau(V))$.

Proof Note that by definition we have:

(1⁰) $E_\tau(V) \subseteq H_\gamma(\Sigma)$;

(2⁰) $I_V(E_\gamma(V)) \subseteq \Sigma$.

To prove the necessity, assume that V is a solid variety. Let ε be an identity from Σ. By Definition 1.3.1, the transformation $H_\gamma(\varepsilon)$ is a hyperidentity of type τ, satisfied in V. We conclude that $\varepsilon \in I_V(H_\tau(\varepsilon)) \subseteq I_V(E_\tau(V))$ and thus $I_V(E_\tau(V)) = \Sigma$, by (2^0).

For sufficiency, assume that we have $I_V(E_\gamma(V)) = \Sigma$. By ($1^0$) we need only to prove the inclusion $H_\tau(I_V(E_\tau(V))) \subseteq E_\tau(V)$. To show this, take a hyperidentity E from the set $E_\tau(V)$ of all hyperidentities of V. Then $I_V(E)$ is contained in Σ. Now consider $H_\tau(I_V(E))$. Any element of $H_\tau(I_V(E))$ can be obtained as an element of the closure of the set $\{E\}$ by rule (6), which is contained in the set $E_\tau(V)$–closed under (6), by the completeness theorem. Thus we conclude that V is a solid variety.

The above results also hold if we restrict ourselves to bases of hyperidentities and identities.

Remark 1.3.3 The completeness theorem can be reformulated in the following way:

Let Σ be a set of hyperidentities of type τ. The following conditions are equivalent:

(1) Σ is closed under the rules (1)-(6).

(2) $\Sigma = H_\gamma(\mathrm{Id}(K))$ for some solid variety K of type τ.

Theorem 1.3.4 *A variety V of type τ is solid if and only if it is closed under the condition:*

$$\begin{array}{l} \textit{Let } \mathbf{A}; \textit{ be an algebra of } V, \textit{ of type } \tau = (n_1, n_2, \ldots, n_\gamma, \ldots : \gamma < O(\tau)). \\ \textit{If } t_\gamma \textit{ is the realization of an } n_\gamma\textit{-ary term operation of type } \tau \textit{ in } \mathbf{A}, \\ \textit{then } \mathbf{A} = (A; t_1, t_2, \ldots, t_\gamma, \ldots : \gamma < O(\tau)) \textit{ is an algebra of } V. \end{array} \tag{1.3.4}$$

Proof Let V be a solid variety. Consider the algebra $\mathbf{A} = (A; t_1, t_2, \ldots, t_\gamma, \ldots : \gamma < O(\tau))$. The identities of V are transformed into hyperidentities of V and hence hold for the term functions t_γ. Especially they hold for $\mathbf{A}$. Hence $\mathbf{A} \in V$. Let condition (1.3.4) hold for V. Then the identities of V hold for all term functions of the appropriate arity and hence are transformed into hyperidentities, i.e. V is a solid variety.

1.4 Derived algebras

Notation 1.4.1 Let K be a class of algebras of a given type $\tau = (n_0, n_1, \ldots, n_\gamma, \ldots)$. The algebra $\mathbf{B}$ is called a derived algebra of $\mathbf{A} = (A; f_0, f_1, \ldots, f_\gamma, \ldots)$ if there exist term operations $t_0, t_1, \ldots, t_\gamma, \ldots$ of type τ such that $\mathbf{B} = (A; t_0, t_1, \ldots, t_\gamma, \ldots)$. For a class K of algebras of type τ we denote by $\mathbf{D}(K)$ the class of all derived algebras of type τ of K. We use the closure operator $\mathbf{D}$ to reformulate Theorem 1.3.4.

Theorem 1.3.4′ *Let V be a class of algebras of a given type τ. V is a solid variety if and only if V is closed under homomorphic images* $\mathbf{H}$, *subalgebras* $\mathbf{S}$, *direct products* $\mathbf{P}$ *and derived algebras* $\mathbf{D}$, *i.e.,*

$$\mathbf{H}(V) \subseteq V; \quad \mathbf{S}(V) \subseteq V; \quad \mathbf{P}(V) \subseteq V; \quad \mathbf{D}(V) \subseteq V.$$

Problem 1.4.2 Describe the semigroup generated by the operators $\mathbf{H}, \mathbf{S}, \mathbf{P}, \mathbf{D}$. Compare [Pigozzi 72].

Theorem 1.4.3 *Let V be a class of algebras of given type τ. V is a solid variety if and only if $V = \mathbf{HSPD}(V)$.*

Proof

(a) $\mathbf{DP}(V) \subseteq \mathbf{PD}(V)$. For $\mathbf{B} \in \mathbf{DP}(V)$ we have $\mathbf{B} = (A; t_0, t_1, \ldots, t_\gamma, \ldots)$ with $\mathbf{A} = (A; f_0, f_1, \ldots, f_\gamma, \ldots)$ and $\mathbf{A} = \prod \mathbf{A}_i, \mathbf{A}_i = (A_i; f_0, f_1, \ldots, f_\gamma, \ldots)$. Consider $\mathbf{B}_i := (A_i; t_0, t_1, \ldots, t_\gamma, \ldots)$; then we have $\mathbf{B} = \prod \mathbf{B}_i$ and hence $\mathbf{B} \in \mathbf{PD}(V)$.

(b) $\mathbf{DS}(V) \subseteq \mathbf{SD}(V)$. For $\mathbf{B} = (B; t_0, t_1, \ldots, t_\gamma, \ldots) \in \mathbf{DS}(V)$ we have $\mathbf{C} = (B; f_0, f_1, \ldots, f_\gamma, \ldots)$ is a subalgebra of some algebra $\mathbf{A} = (A; f_0, f_1, \ldots, f_\gamma, \ldots)$. As $(B; t_0, t_1, \ldots, t_\gamma, \ldots)$ is a subalgebra of $(A; t_0, t_1, \ldots, t_\gamma, \ldots)$ we have $\mathbf{B} \in \mathbf{SD}(V)$.

(c) $\mathbf{DH}(V) \subseteq \mathbf{HD}(V)$. Let $\mathbf{B} = (B; t_0, t_1, \ldots, t_\gamma, \ldots) \in \mathbf{DH}(V)$. Then there is a homomorphic image $f[\mathbf{A}] = (f[A]; f_0, f_1, \ldots, f_\gamma, \ldots)$ of an algebra $\mathbf{A}$ with $f[A] = B$. But $(B; t_0, t_1, \ldots, t_\gamma, \ldots)$ is also a homomorphic image of $(A; t_0, t_1, \ldots, t_\gamma, \ldots)$ because $f[A] = B$ and $f(t_\gamma(x_1, \ldots, x_{n_\gamma})) = t_\gamma(f(x_1), \ldots, f(x_{n_\gamma}))$. Now we have $\mathbf{DHSP}(V) \subseteq \mathbf{HSPD}(V)$. (Observe that for some V we have $\mathbf{DS}(V) \not\subseteq \mathbf{SD}(V)$.)

Remark In the sense of [Schweigert 87a] the derived algebra $\mathbf{B}$ of an algebra $\mathbf{A}$ has the property that $T(\mathbf{B})$ is a surjective image of a clone homomorphism from the clone $T(\mathbf{A})$ onto the clone $T(\mathbf{B})$. Also weak endomorphisms [Goetz 66, Schweigert 85a] induce such clone homomorphisms.

Example 1.4.4 The variety U of semigroups of type (2) defined by the following identities, is a solid variety:

$$\begin{aligned} x \circ x &= x \\ x \circ (y \circ z) &= (x \circ y) \circ z \\ (u \circ x) \circ (y \circ v) &= (u \circ y) \circ (x \circ v). \end{aligned}$$

Proof One can show that $F_2 = \{x, y, x \circ y, y \circ x, x \circ y \circ x, y \circ x \circ y\}$ is the set of all binary terms of the variety U. Furthermore, for these terms the transformed identities: $F(x, x) = x$, $F(x, F(y, z)) = F(F(x, y), z)$ and $F(F(u, x), F(y, v)) = F(F(u, y), F(x, v))$ hold as hyperidentities for U.

Remark 1.4.5 The transformation of some identities for an algebra $\mathbf{A}$ always leads to hyperidentities which hold for $\mathbf{A}$ (for example $x = x$ or $x \circ x = x$).

1.5 Weak isomorphisms

The notion of weak homomorphism and weak isomorphism has been introduced by Marczewski and Goetz ([GlaMic 77], [Schweigert 84]). For these definitions we have to consider the clone $T(\mathbf{A})$ of all term functions of an algebra $\mathbf{A}$.

Definition 1.5.1 Let $\mathbf{A} = (A, \Omega_1)$ and $\mathbf{B} = (B, \Omega_2)$ be algebras not necessarily of the same type and let $h : A \to B$ be a mapping. Let $\varphi \in T(\mathbf{A})$ and $\psi \in T(\mathbf{B})$ be of the same arity n. Then φ and ψ are in the *relation* R_h, i.e. $(\varphi, \psi) \in R_h$, if $h(\varphi(x_1, \ldots, x_n)) = \psi(h(x_1, \ldots, h(x_n))$.

Definition 1.5.2 Let $\mathbf{A} = (A, \Omega_1)$ and $\mathbf{B} = (B, \Omega_2)$ be algebras not necessarily of the same type. The mapping $h : A \to B$ is called a *weak homomorphism* of $\mathbf{A}$ into $\mathbf{B}$ if

(i) for every $\varphi \in T(\mathbf{A})$ there is a $\psi \in T(\mathbf{B})$ with $(\varphi, \psi) \in R_h$,

(ii) for every $\alpha \in T(\mathbf{B})$ there is a $\beta \in T(\mathbf{A})$ with $(\beta, \alpha) \in R_h$.

Remark 1.5.3 It is easy to show that (i) and (ii) can be replaced by the weaker conditions (a) and (b).

(a) for every $\omega \in \Omega_1$ there is a $\psi \in T(\mathbf{B})$ with $(\omega, \psi) \in R_h$,

(a) for every $\eta \in \Omega_2$ there is a $\varphi \in T(\mathbf{A})$ with $(\varphi, \eta) \in R_h$.

If $h : A \to B$ is a homomorphism of the algebra $\mathbf{A}$ into the algebra $\mathbf{B}$ of the same type, then h is also a weak homomorphism, because we have $h(\omega_A(x_1, \ldots, x_n)) = \omega_B(h(x_1), \ldots, h(x_n))$ for every operation $\omega_A \in \Omega_1$ and the corresponding operation $\omega_B \in \Omega_2$.

A weak homomorphism $h : A \to B$ is called a *weak isomorphism* if h is bijective.

Definition 1.5.4 A weak homomorphism $h : A \to B$ is called a *near isomorphism* if h is the identity map.

Example 1.5.5 Let $\mathbf{B} = [\{1, 0\}; \wedge, \vee, ^{`}, 0, 1]$ be the Boolean algebra on the set $\{0, 1\}$. Let $\mathbf{R} = [\{1, 0\}; +, 0, \cdot, 1]$ be the commutative ring on the set $\{0, 1\}$, where the addition is modulo 2. Then $\mathbf{B}$ and $\mathbf{R}$ are near isomorphic. In particular, we have:

(a)
1) $x \vee y = (x + y) + x \cdot y$ 4) $0 = 0$
2) $x \wedge y = x \cdot y$ 5) $1 = 1$
3) $x^{`} = x + 1$

(b)
1) $x + y = (x^{`} \wedge y) \vee (y \wedge x^{`})$ 3) $0 = 0$
2) $x \cdot y = x \wedge y$ 4) $1 = 1.$

Hence conditions (a) and (b) of Remark 1.5.3 are fulfilled.

Lemma 1.5.6 *Let* $\mathbf{A} = (A, \Omega_1)$ *and* $\mathbf{B} = (B, \Omega_2)$ *be algebras not necessarily of the same type. If* $h : A \to B$ *is a weak isomorphism then there is an isomorphism then there is an isomorphism* $\alpha : A \to B$ *for an algebra* $\mathbf{B}^* = (B, \Omega_1)$ *and a near isomorphism* $g : B \to B$ *from* $\mathbf{B}^*$ *onto* $\mathbf{B}$ *such that* $h = g \circ \alpha$.

Proof We define the operation ω_{B^*} of $\mathbf{B}^*$ by setting

$$\omega_{B^*}(b_1, \ldots, b_n) = h(\omega_A(h^{-1}(b_1), \ldots, h^{-1}(b_n))).$$

Furthermore we define $\alpha(a) := h(a)$ for every $a \in A$. Then α is bijective. Put $b_i = h(a_i)$, $i = 1, \ldots, n$. Then we have

$$\begin{aligned} \alpha(\omega_A(a_1, \ldots, a_n)) &= h(\omega_A(a_1, \ldots, a_n)) = h(\omega_A(h^{-1}(b_1), \ldots, h^{-1}(b_n)) \\ &= \omega_{B^*}(b_1, \ldots, b_n) = \omega_{B^*}(h(a_1), \ldots, h(a_n)) \\ &= \omega_{B^*}(\alpha(a_1), \ldots, \alpha(a_n))). \end{aligned}$$

Hence α is an isomorphism. For the identity map $g : B \to B$ and the corresponding relation R_g the following holds:

(a) For $\omega_{B^*} \in \Omega_1$ we have a term function $\psi \in T(\mathbf{B})$ with $(\omega_A, \psi) \in R_h$ such that $h(\omega_A(a_1, \ldots, a_n)) = \psi$. Therefore

$$\alpha(\omega_A(a_1, \ldots, a_n)) = \psi(h(a_1), \ldots, h(a_n))$$

hence

$$\omega_{B^*}(\alpha(a_1), \ldots, \alpha(a_n)) = \psi(h(a_1), \ldots, h(a_n)),$$

and hence

$$\omega_{B^*} = \psi,$$

i.e. $(\omega_{B^*}, \psi) \in R_g$.

(b) is proved similarly.

Remark 1.5.7 A weak isomorphism $h : A \to B$ for the algebras $\mathbf{A} = (A, \Omega_1)$ and $\mathbf{B} = (B, \Omega_2)$ also defines a map

$$\bar{h} : T(\mathbf{A}) \to T(\mathbf{B}) \text{ by } (\bar{h}(\psi))(b_1, \ldots, b_n) := h(\psi(h^{-1}(b_1), \ldots, h^{-1}(b_n))).$$

This map $\bar{h}$ is a clone isomorphism and $\bar{h}$ is compatible with the composition of term functions, permutation of variables and with fictitious variables. $\bar{h}$ also preserves the arity of term functions ([GlaMic 77], [Schweigert 84]).

Notation 1.5.8 Let $\mathbf{A} = (A, \Omega_1)$ and $\mathbf{B} = (A, \Omega_2)$ be two algebras not necessarily of the same type. Let $\mathbf{A}$ be near isomorphic to $\mathbf{B}$ and $\bar{h} : T(\mathbf{A}) \to T(\mathbf{B})$ the corresponding map for the clones of term functions. If $\epsilon \equiv (\varphi = \psi)$ is an equation which holds for the algebra

A, then $\bar{h}(\epsilon) \equiv (\bar{h}\varphi = \bar{h}\psi)$ is an equation which holds for the algebra **B**. $\bar{h}(\epsilon)$ is called the transformation of the equation $(\varphi = \psi)$ by $\bar{h}$.

Example 1.5.9 Let $\mathbf{B} = [\{0,1\}; \wedge, \vee, ^\backprime, 0, 1]$ be the Boolean algebra and $\mathbf{R} = [\{0,1\}; +, 0, \cdot, 1]$ the commutative ring on the set $\{0,1\}$. By 1.5.5, **B** is near isomorphic to **R**. Some axioms for the Boolean algebra are transformed in the following way.

$$
\begin{array}{lrcl}
B1: & x \wedge (y \wedge z) & = & (x \wedge y) \wedge z \\
\bar{h}(B1): & x \cdot (y \cdot z) & = & (x \cdot y) \cdot z \\
B2: & x \wedge y & = & y \wedge x \\
\bar{h}(B2): & x \cdot y & = & y \cdot x \\
B3: & x \wedge x & = & x \\
\bar{h}(B3): & x \cdot x & = & x \\
B4: & x \wedge (y \vee x) & = & x \\
\bar{h}(B4): & x \cdot ((y + x) + (y \cdot x)) & = & x \\
B5: & x \wedge (y \vee z) & = & (x \wedge y) \vee (x \wedge z) \\
\bar{h}(B5): & x \cdot ((y + z) + (y \cdot z)) & = & ((x \cdot y) + (x \cdot z)) + ((x \cdot y) \cdot (x \cdot z)) \\
B6: & x \wedge x^\backprime & = & 0 \\
\bar{h}(B6): & x \cdot (x + 1) & = & 0.
\end{array}
$$

Remark 1.5.10 If Σ is an equational basis for the equational theory of **A**, then the transformation $\bar{h}(\Sigma)$ of Σ by h is not an equational basis for **B** except in special cases. Therefore it is necessary to add some further equations to $\bar{h}(\Sigma)$ to get an equational basis for **B**.

Notation 1.5.11 Let ω be an n-place operation of the algebra $\mathbf{B} = (B, \Omega_2)$. Then $\bar{h}^{-1}(\omega)$ is a term function of the algebra **A** and hence may be presented by a term $\psi(x_1, \ldots, x_n)$ of the term algebra of **A**. $\bar{h}(\psi)$ is a term function of **B** and can be presented by a term $\varphi(x_1, \ldots, x_n)$. Obviously the equation $\varphi(x_1, \ldots, x_n) = \omega(x_1, \ldots, x_n)$ holds for **B**. We denote this equation by π_ω and consider the set $\{\pi_\omega | \omega \in \Omega_2\}$ of equations.

Example 1.5.12 Consider the operations $+$ and $\cdot$ of the ring R. Then we have

$$\pi_+ : (((x+1) \cdot y) + (x \cdot (y+1))) + (((x+1) \cdot y) \cdot (x \cdot (y+1))) = x + y$$

$$\pi_\bullet : x \cdot y = x \cdot y \quad \text{(which can be dropped because of triviality).}$$

For the following result compare also [Felscher 68] p. 148, Theorem 4.

Theorem 1.5.13 *Let Σ_1 be an equational basis for the equational theory of the algebra $\mathbf{A} = (A, \Omega_1)$. If $\mathbf{B} = (A, \Omega_2)$ is near isomorphic to $\mathbf{A}$ by h, then $\Sigma_2 = \bar{h}(\Sigma_1) \cup \{\pi_\omega | \omega \in \Omega_2\}$ is an equational basis for the equational theory of* **B**.

Proof We show that any equation of $\varphi(x_1, \ldots, x_n) = \psi(x_1, \ldots, x_n)$ holding for **B** can be derived from Σ_2.

$$(h^{-1}(\varphi))(x_1, \ldots, x_n) = (h^{-1}(\psi))(x_1, \ldots, x_n)$$

is an equation holding for **A** and hence can be derived by a sequence $(\sigma_1, \ldots, \sigma_k)$ of equations from Σ_1, the properties of the equality sign and the substitution [Grätzer 79], p. 381. Transforming this sequence by $\bar{h}$ we have $(\bar{h}(\sigma_1), \ldots, \bar{h}(\sigma_k))$ a sequence of equations from $\bar{h}(\Sigma_1)$ which proves the equality

$$(h(h^{-1}(\varphi)))(x_1, \ldots, x_n) = (h(h^{-1}(\psi)))(x_1, \ldots, x_n).$$

By the equations from $\{\pi_\omega | \omega \in \Omega_2\}$ we have that

$$\varphi(x_1, \ldots, x_n) = (h(h^{-1}(\varphi)))(x_1, \ldots, x_n)$$

and

$$\psi(x_1, \ldots, x_n) = (h(h^{-1}(\psi)))(x_1, \ldots, x_n).$$

Example 1.5.14 We wish to show that $x+y = y+x$ holds for **R**. A proof of this equation in **B** is the following ($B2$` is the dual of $B2$):

$$(x` \wedge y) \vee (x \wedge y`) =_{(B2`)} (x \wedge y`) \vee (x` \wedge y) =_{(B2)} (y` \wedge x) \vee (x` \wedge y) =_{(B2)} (y` \wedge x) \vee (y \wedge x`).$$

By transformation we get the following proof.

$$\begin{aligned}
x + y &=_{\pi_+} (((x + 1) \cdot y) + (x \cdot (y + 1))) + (((x + 1) \cdot y) \cdot (x \cdot (y + 1))) \\
&=_{\bar{h}(B2`)} ((x \cdot (y + 1)) + ((x + 1) \cdot y) + (((x \cdot (y + 1)) \cdot ((x + 1) \cdot y)) \\
&=_{\bar{h}(B2)} ((y + 1) \cdot x)) + ((x + 1) \cdot y) + (((y + 1) \cdot x) \cdot (x + 1) \cdot y)) \\
&=_{\bar{h}(B2)} ((y + 1) \cdot x) + ((y \cdot (x + 1)) + (((y + 1) \cdot x) \cdot (y \cdot (x + 1))) \\
&=_{\pi_+} y + x
\end{aligned}$$

Theorem 1.5.15 *Let* $\mathbf{A} = (A, \Omega_1)$ *and* $\mathbf{B} = (B, \Omega_2)$ *be of finite type and weakly isomorphic to one another. The equational theory of* **B** *is finitely based if and only if the equational theory of* **A** *is finitely based.*

Proof By Lemma 1.5.6 we have that Theorem 1.5.13 holds also in the case of a weak isomorphism. Furthermore, if Σ_1 is finite and the type of **B** is finite, then $\bar{h}(\Sigma_1) \cup \{\pi_\omega | \omega \in \Omega_2\}$ is a finite basis for the equational theory of **B**.

Definition 1.5.16 A weak homomorphism $h : A \to B$ is called *type-preserving* if the algebras **A**, **B** are of the same type.

Notation 1.5.17 Let K be a class of algebras of type τ. Let $\mathbf{W}(K)$ be the class of all algebras which are images of algebras in K by type-preserving weak isomorphisms.

Problem Describe solid varieties with the class operator **W**. Can **W** replace the class operator **D**?

Definition 1.5.18 A congruence relation θ of an algebra **A** is *totally invariant* if $(a, b) \in \theta$ implies $(h(a), h(b)) \in \theta$ for every type-preserving weak endomorphism h of **A** and every $a, b \in A$.

One observes that $\Sigma \subseteq \text{Id}\ \tau$ is a solid theory if and only if Σ is a totally invariant congruence of $F_\tau(X)$ [Schweigert 89].

The closures with respect to the deduction rules (1)-(6) correspond to the properties of Σ as follows.

Σ is closed under (1)-(3) $\Leftrightarrow$ Σ is an equivalence relation on $F_\tau(x)$
Σ is closed under (1)-(4) $\Leftrightarrow$ Σ is a congruence relation
Σ is closed under (1)-(5) $\Leftrightarrow$ Σ is a fully invariant congruence
Σ is closed under (1)-(6) $\Leftrightarrow$ Σ is a totally invariant congruence.

1.6 Types

Let us consider two weakly isomorphic algebras $\mathbf{A} = (A, \Omega)$ and $\mathbf{B} = (A, \Omega)$. They have isomorphic clones $T(\mathbf{A}), T(\mathbf{B})$ on the set A. But these clones may be generated by different fundamental operations. The fundamental operations determine the identities. Therefore, the identities of $\mathbf{A}$ may appear very different from the identities of $\mathbf{B}$. In this respect the type of a variety plays an essential role.

Remark 1.6.1 Let V be a variety of type τ. Let the type τ be contained in the type μ. Then the set $H_\tau(V)$ of hyperidentities of type τ is contained in the set $H_\mu(V)$ of hyperidentities of type μ. For the variety of semilattices SL one can present an increasing sequence of types.

$$\mu_1 \subseteq \mu_2 \subseteq \ldots \mu_{i-1} \subseteq \mu_i \subseteq \ldots$$

such that there are hyperidentities in $H_{\mu_i}(SL)$ which cannot be implied by $H_{\mu_{i-1}}(SL)$ [Penner 81].

Notation 1.6.2 Let $Q = T$ be a hyperidentity of type μ. (*) Assume that we have hyperterms $R_i, i \in I$, of type τ available such that any hypervariable of type μ in $Q = T$ can be hypersubstituted by hyperterms R_i of type τ. The result of such a hypersubstitution (rule 6) is called a hyperconsequence of type τ from the hyperidentity $Q = T$ of type μ. If the assumption (*) holds we say that type τ is compatible with type μ. As the reader will observe this assumption it is usually fulfilled.

Examples 1.6.3 $(Q = T) : T(T(x, y, z), y, z) = T(x, y, z)$ is a hyperidentity of type (3). Consider $R = F(G(x, y), z)$ a hyperterm of type (2, 2). Then

$$(M = N) : F(G(F(G(x, y), z), y), z) = F(G(x, y), z)$$

is a hyperconsequence of type τ from the hyperidentity $Q = T$ of type μ.

Consider the variety D of distributive lattices of type (2, 2). Then the hyperidentity $F(F(x, y), y, z) = F(x, y, z)$ is of type (3) and holds for all ternary terms of D. Obviously also every hyperconsequence $M = N$ of type (2) holds in D.

Theorem 1.6.4 *Let V be a variety of type τ, and let τ be compatible with the type μ. A hyperidentity $P = Q$ of type μ holds for V if and only if every hyperconsequence $M = N$ of type τ holds for V.*

Proof Let $P = Q$ of type μ hold for V and let the hyperidentity $M = N$ of type τ be a hyperconsequence of $P = Q$. Then any hypersubstitution of the hypervariables yields an identity of type τ which is also implied by $P = Q$ and hence holds for V. $M = N$ holds for V. Let every hyperconsequence of type τ hold for V and let $p = q$ be the result of hypersubstituting the hypervariables in $P = Q$ by the appropriate terms. Transform $p = q$ into a hyperidentity $M = N$ of type τ. Obviously $M = N$ is a hyperconsequence of $P = Q$ and holds for V by hypothesis. Hence $p = q$ is an identity for V. Hence $P = Q$ is a hyperidentity for V.

Remark The above result shows that for a variety of type τ it is sufficient to consider hyperidentities of type τ. To consider a hyperidentity of other types for V may be useful as this hyperidentity may stand for a huge set of hyperconsequences of type τ. Hence such a hyperidentity of type μ is a short notation for a set of hyperidentities of type τ. Again a hyperidentity of type τ may be considered as a short notation for a possibly infinite set of identities of type τ. This is one of the essential properties of hyperidentities.

Example Consider the solid variety B of regular bands defined by

$$\begin{aligned} x \circ (y \circ z) &= (x \circ y) \circ z \\ x \circ x &= x \\ (u \circ x) \circ (y \circ w) &= (u \circ y) \circ (x \circ w) \end{aligned}$$

B is of type (2). We add as an additional fundamental operation $\square$ with $x \square y = x$ and have $\bar{B}$ of type $(2,2)$. This operation reflects the projection $e_1^2, e_1^2(x,y) = x$, and is contained in $T(B)$ anyhow. Let $\bar{B}$ denote this variety of type $(2,2)$ with $T(B) = T(\bar{B})$. Claim: $\bar{B}$ is not solid.

If $\bar{B}$ were solid, then $F(x,y) = x$ would be a hyperidentity. Now $e_2^2(x,y) = y$ would imply $x = y$. $\bar{B}$ is an example of a variety which is equivalent to solid variety but is not solid itself. On the other hand a reduct of a solid variety is solid.

We call two varieties V, W equivalent if they can be generated by weakly isomorphic algebras.

Problem 1.6.5 Let V be a variety. Under what conditions is V equivalent to a solid variety W?

This problem can be considered from a syntactic point of view as well as from a semantic one. For the semantic point of view we have

Theorem 1.6.6 *If V is equivalent to a solid variety W, then every subdirectly irreducible derived algebra* $\mathbf{A}$ *from V is weakly isomorphic to a subdirectly irreducible algebra* $\mathbf{C}$ *of* V.

Proof Let $\mathbf{A} = (A;\ t_1, \ldots, t_{n_\gamma}, \ldots)$ be a subdirectly irreducible derived algebra from V. $\mathbf{A}$ must be constructed from an algebra $\bar{\mathbf{A}}$ of V. Then there is an algebra $\bar{\mathbf{B}} = (B;\ f_1, \ldots, f_{n_\gamma}, \ldots)$ in W with a weak isomorphism h from $\bar{\mathbf{A}}$ to $\bar{\mathbf{B}}$. We have a derived algebra $\mathbf{B} = (B;\ t_1^h, \ldots, t_{n_\gamma}^h, \ldots)$, where h describes the transformation of the operations into $t_1, \ldots, t_{n_\gamma}, \ldots$. We have that $\mathbf{B}$ is weakly isomorphic to $\mathbf{A}$. As W is solid, $\mathbf{B}$ is a subdirectly irreducible algebra of W. Then there exists a subdirectly irreducible algebra $\mathbf{C}$

in V such that $\mathbf{B}$ is weakly isomorphic to $\mathbf{C}$ because W is equivalent to V. Hence $\mathbf{A}$ is weakly isomorphic to $\mathbf{C}$.

Definition 1.6.7 The *solid envelope* $s(V)$ of a variety V is the smallest solid variety of the same type containing V.

Another definition would be $s(V) = \mathbf{HSPD}(V)$. It follows that $s(V)$ is generated by the subdirectly irreducible derived algebras from V. From Theorem 1.6.4 we get:

Theorem 1.6.8 *V of type τ is equivalent to a solid variety W of some type μ if and only if V is equivalent to its solid envelope $s(V)$.*

1.7 Transformations

Definition 1.7.1 Let τ be a given type and $H(\tau)$ the set of all hyperterms of type τ. Let V be a variety of type μ and $W(\mu)$ the set of all terms of type μ. The mapping $\sigma : H(\tau) \to W(\mu)$ is defined recursively by

$$\sigma(x_i) = x_i \text{ for every variable } x_i,$$
$$\sigma(F_\delta(x_1, \ldots, x_n) = t_\delta(x_1, \ldots, x_n),$$

where to every n_δ-ary hypervariable F_δ of type τ an n_δ-ary term t_δ of type μ is assigned. σ is then extended by the construction of hyperterms to $H(\tau)$.

Example Let $F(F(x, y, z), y, z)$ be a hyperterm of type (3) and let $t(x, y, z) = x \wedge (y \vee z)$ be a term of type (2,2) of the variety of lattices. Then

$$\sigma(F(F(x, y, z), y, z) = (x \wedge (y \vee z)) \wedge (y \vee z).$$

If we consider the hyperidentity $F(F(x, y, z), y, z) = F(x, y, z)$, then by σ we get the identity

$$(x \wedge (y \vee z)) \wedge (y \vee z) = x \wedge (y \vee z).$$

Notation 1.7.2 The mapping σ in Definition 1.7.1 is called a (τ, μ)-transformation.

Definition 1.7.3 A hyperidentity of type τ *holds* in a variety of type μ if every (τ, μ)-transformation of the hyperidentity yields an identity which holds for V. (For $\tau = \mu$ compare Definition 1.1.2)

Remark For a given variety V we wish to consider all hyperidentities of any type which hold for V. How does one have to choose the hypervariables F_δ? A rough estimate would be to have a fundamental hyperterm $F_\delta(x_1, \ldots, x_{n_\delta})$ for every term $t_\delta(x_1, \ldots, x_{n_\delta})$ for the variety. With this crude construction we may get a type which furnishes us with all hyperidentities for V. An alternative is to consider a type which is the set of all ordered sequences.

Definition 1.7.4 A type ν is called a *general type* for a variety V if the set of all hyperidentities for V is equivalent to the set of all hyperidentities of type ν which hold for V.

(Of course, two sets of hyperidentities are equivalent to one another if they can be derived from one another by (1)-(6)).

The general type can be used to present a completeness theorem which holds for hyperidentities of any type. In order to stress the semantic aspect one can use the following.

Definition 1.7.5 $D_\nu(V)$ is the set of algebras $(A;(t_i)i\epsilon I)$ of type ν, where t_i are term operations of V of type ν. $D_\nu(V)$ is called the *derived variety* of V of type ν. One can say that a hyperidentity ϵ of type ν *holds* for V if and only if the set of transformations of ϵ yields identities of $D_\nu(V)$.

It is up to the reader to reformulate and generalize some of the results from the preceding section.

Part 2

Iterative hyperidentities

2.1 Iterative hyperidentities

One can define operations on the set $H(\tau)$ of all hyperterms of type τ (compare Definition 1.1.1) by the hypervariables F in the usual way. We call this algebra $HT(X)$ of type τ a hyperterm algebra generated by the variables $x \in X$.

As an example consider the hyperterm algebra $HT(2)$ in two variables x, y and a binary hypervariable F for the variety of semilattices

$$F(x,y), F(y,x) \qquad \text{(we have } F(x,x)=x, F(y,y)=y\text{)}$$
$$F(F(x,y),x), F(F(y,x),y) \qquad \text{(we have } F(F(x,y),y)=F(x,y),\ldots\text{).}$$

Here we use the hyperidentities as given in Example 1.4.4.

Slightly more formally we use the following

Notation 2.1.1 Let V be a variety of type τ. Let F be an n-ary hypervariable with respect to τ. Then $F(x_1,\ldots,x_n)$ is called a *fundamental hyperterm.* The *hyperterm algebra* $HT(V)$ is the set of all hyperterms for type τ closed under the application of all fundamental hyperterms as operations on the hyperterms.

Let us denote the set of all fundamental hyperterms by $FT(V)$. Obviously we have that for every map

$$\alpha : FT(V) \to T(V)$$

there exists an extension

$$\beta : HT(V) \to T(V)$$

where β is surjective.

Remark 2.1.2 If V is a solid variety, then the free algebra coincides with the hyperterm algebra of V. If V is not solid and if $s(V)$ is the solid envelope of V, then the free algebra of $s(V)$ coincides with the hyperterm algebra of V.

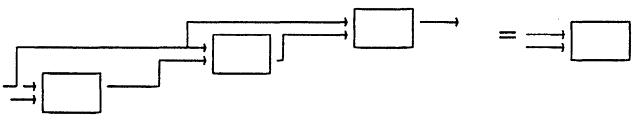
Problem 2.1.3 Determine the hyperterm algebra for the variety of your choice.

Problem 2.1.4 Determine the hyperterm algebra in three generators x, y, z and the binary hypervariables F, G for the variety of modular lattices.

Problem 2.1.5 If one can decide the equality of two hyperterms in a finite number of steps in the variety V, we say that the "hyperword problem" of V is solvable. Let V be a variety with a solvable hyperword problem. Under what conditions is the word problem of V solvable?

Problem 2.1.6 Consider the reverse problem to 2.1.5.

Example 2.1.7 Let  be a symbol for an arbitary switching circuit realizing some Boolean function $f : \{0,1\}^2 \to \{0,1\}$.



We have the hyperidentity $F(F(F(x,y),y),y) = F(x,y)$, which we will write in the short form

$$F^3(x,y) = F(x,y).$$

If we wish to prove this in a syntactic way, we have to list up all 16 Boolean terms

$$x, y, x^{`}, y^{`}, x \wedge y, x \vee y, x^{`} \wedge x, x \vee x^{`}, x^{`} \wedge y, x \wedge y^{`}, x^{`} \vee y, x \vee y^{`}$$
$$(x^{`} \wedge y) \vee (x \wedge y^{`}), (x^{`} \vee y) \wedge (x \vee y^{`}), x^{`} \vee y^{`}, x^{`} \wedge y^{`}$$

and check every term, for instance $x \wedge y$

$$((x \wedge y) \wedge y) \wedge y = x \wedge y.$$

If one wishes to prove this in a semantic way, one can proceed as follows. Consider the semigroup of polynomial functions of the algebra $\mathbf{B} = (\{0,1\}; \wedge, \vee, {}^{`}, 0, 1)$. This is the semigroup T_2 of all transformations on $\{0,1\}$. It fulfills the semigroup identity

$$p \circ p \circ p = p$$

or written in another way

$$p^3 = p.$$

(Transformations: $f_1(v) = x, f_2(x) = 0, f_3(x) = 1, f_4(x) = x'$, where we have

$$f_i^3(x) = f_i(x), \quad i = 1, 2, 3, 4.)$$

In the following we will prove that any identity of the semigroup of polynomial functions of an algebra A yields a set of hyperidentities for A. For instance for the Boolean algebra we would have

$$\begin{aligned} F_1^3(x) &= F_1(x) \\ F_2^3(x,y) &= F_2(x,y) \\ F_3^3(x,y,z) &= F_3(x,y,z) \end{aligned}$$

where F_i is a hypervariable of arity $i, i = 1, 2, 3, \ldots$.

2.2 Iterations of functions

Proposition 2.2.1 *Let $f : A \to A$ be a function, $|A| = n$. There exists a least natural number $\lambda(f)$ (the* index *of f) such that*

$$f^{\lambda(f)+1}[A] = f^{\lambda(f)}[A].$$

Proposition 2.2.2 *Let $f : A \to A$ be a function, $|A| = n$. Then there exists a least natural number $\pi(f)$ (the* period *of f) such that*

$$f^{\lambda(f)+\pi(f)} = f^{\lambda(f)}.$$

Propositions 2.2.1 and 2.2.2 date back to Frobenius [Frobenius 1895].

Proposition 2.2.3 *Let S be a semigroup of functions on A, $|A| = n$. Let $f, g \in S$ such that $f^{\lambda(f)+\pi(f)} = f^{\lambda(f)}$ and $g^{\lambda(g)+\pi(g)} = g^{\lambda(g)}$. Then we have*

$$h^{\max(\lambda(f),\lambda(g))+\mathrm{lcm}(\pi(f),\pi(g))} = h^{\max\{\lambda(f),\lambda(g)\}}$$

for every $h \in \{f, g\}$.

Here $\mathrm{lcm}(\pi(f), \pi(g))$ denotes the least common multiple of the integers $\pi(f), \pi(g)$.

Definition 2.2.4 Let S be a semigroup of functions on A, $|A| = n$.

$$\begin{aligned} \lambda_S &:= \max\{\lambda(f) | f \in S\} \quad \text{is called the } \textit{index} \text{ of } S. \\ \pi_S &:= \mathrm{lcm}\,\{\pi(f) | f \in S\} \quad \text{is called the } \textit{period} \text{ of } S. \end{aligned}$$

We denote by F, G, H variables which stand for the functions $f, g, h, \ldots$ in S. Obviously the functions in S fulfill the equation

$$F^{\lambda_S+\pi_S} = F^{\lambda_S}.$$

Definition 2.2.5 An equation $F^r = F^s, r < s$, for the semigroup S is called *irreducible* if for every equation $F^k = F^l, k < l$, which holds for S we have $r \leq k$ and $s \leq l$.

Obviously $F^{\lambda_S+\pi_S} = F^{\lambda_S}$ is an irreducible equation for S.

Proposition 2.2.6 *Let U be a subsemigroup of a semigroup S of functions on A, $|A| = n$. For the equations $F^{\lambda_S+\pi_S} = F^{\lambda_S}$ and $F^{\lambda_U+\pi_U} = F^{\lambda_U}$ which are irreducible for S and U, respectively, we have that*

$$\lambda_U \leq \lambda_S \text{ and } \pi_U \text{ divides } \pi_S.$$

Proof By definition of the index and the period.

Proposition 2.2.7 *Let S be a semigroup of functions on A, $|A| = n$, let $f \in S$ be such that $\pi(f) = rt$. Then there exists $g \in S$ with $\lambda(g) \leq \lambda(f)$ and $\pi(g) = t$.*

Proof We consider the cyclic group of permutations on $A_f := f^{\lambda(f)}[A]$ which is generated by $f/_{A_f}$ and we put $g := f^r$. Now we have

$$\begin{aligned} g^{\lambda(f)+t} &= f^{r(\lambda(f)+t)} = f^{r\lambda(f)+rt} = f^{r\lambda(f)+\pi(f)} \\ &= f^{\lambda(f)+\pi(f)+(r-1)\lambda(f)} = f^{\lambda(f)+(r-1)\lambda(f)} = f^{r\lambda(f)} = g^{\lambda(f)}. \end{aligned}$$

Proposition 2.2.8 *Let S be a semigroup of functions on A, $|A| = n$, with the equation $F^{\lambda_S+\pi_S} = F^{\lambda_S}$. For every prime power p^m which divides π_s there exists $g \in S$ such that $g^{\lambda(g)+p^m} = g^{\lambda(g)}$.*

Proof There exists a set $\{f_1, \ldots, f_k\}$ of functions such that $\pi_S = \operatorname{lcm}\{\pi(f_i) | i = 1, \ldots, k\}$. Because of the definition of lcm we have that p^m divides $\pi(f_j)$ for some $j \in \{1, \ldots, k\}$. Now we apply Proposition 2.2.7 to the function f_j with $\pi(f_j) = p^m \cdot s$.

Notation $\gamma(n) := \max(\operatorname{lcm}(x_1, \ldots, x_m))$ denotes the maximum of the least common multiple of $x_1, \ldots, x_m$ taken over all partitions of n, $n = x_1 + \ldots + x_m$, $m = 1, \ldots, n$. For $n = 1, \ldots, 301$, the values of $\gamma(n)$ can be found in the table of [Nicolas 69], p. 187.

Theorem 2.2.9 *Let $f : A \to A$, $|A| = n$. Then we have*

$$\pi(f) \leq \gamma(n - \lambda(f)).$$

Proof The order of the permutation $f/_{A_f}$, where $A_f := f^{\lambda(f)}[A]$, is the least common multiple of the length $x_1, \ldots, x_m$ of m disjoint cycles representing the permutation [Hall 59]. On $|A_f|$ numbers every partition $|A_f| = x_1 + \ldots + x_m$ corresponds to a permutation. The maximal order of these permutations is $\gamma(|A_f|)$ and hence we have $\gamma(n - \lambda(f)) \geq \gamma(A_f) \geq \pi(f)$.

Corollary 2.2.10 $\lambda(f) + \pi(f) \leq n - 1 + \gamma(n - \lambda(f))$.

The above formula gives an useful estimate for the size of powers in a finite semigroup of functions. Indeed we have $\lambda(f) + \pi(f) \sim \gamma(n - \lambda(f))$ for large numbers.

Notation 2.2.11 (*The Transformation Semigroup of a Finite Set*) For $n \geq 1$ let $\langle n \rangle$ be the set $\{0, 1, \ldots, n-1\}$. By T_n we shall denote the set of all transformations of $\langle n \rangle$,

$$T_n = \{f | f : \langle n \rangle \to \langle n \rangle\}.$$

T_n is a monoid with the composition of transformations as multiplication; its unit element is the identical transformation of $\langle n \rangle$.

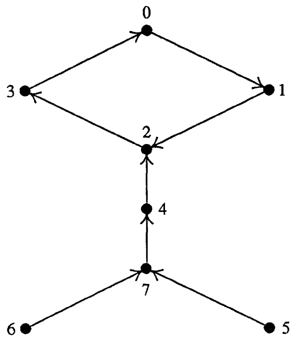
Let $\kappa(n)$ be the least common multiple of $\{1, 2, \ldots, n\}$.

Theorem 2.2.12 *Let k, l be two natural numbers, $k > l > 0$. Then $f^k = f^l$ for all transformations $f \in T_n$ if and only if $l \geq n - 1$ and $k \equiv l (\mathrm{mod}\ \kappa(n))$.*

Proof [Reischer, Simovici] Let f be a fixed element of T_n. We shall consider the directed graph $G_f = (\langle n \rangle, E_f)$ having $\langle n \rangle$ as set of vertices; the set of edges E_f is given by $E_f = \{(x, f(x)) | x \in \langle n \rangle\}$. Since the out-degree of each vertex $x \in \langle n \rangle$ is 1, it is clear that G_f consists of oriented cycles to which trees may be attached by their roots. For instance the graph of the transformation $f \in T_8$ given by the table

x	0	1	2	3	4	5	6	7
$f(x)$	1	2	3	0	2	7	7	4

is presented here. Let b be the length of the longest attached branch. If $l \geq b$, for any $x \in \langle n \rangle b$, $f^l(x)$ will be a vertex on a directed cycle. Therefore, if $\kappa(n)$ is the least common multiple of the cycle lengths we shall have $f^k(x) = f^l(x)$, for all $x \in \langle n \rangle$, if $k \equiv l (\mathrm{mod}\ \kappa)$. Varying the transformations we get the necessity of the statement. The proof of the sufficiency is similar.



2.3 Monoids of polynomial functions

The following results are essentially from [Schweigert 79]. Let $P_1(\mathbf{A})$ denote the semigroup of unary polynomial functions of the algebra $\mathbf{A}$.

Theorem 2.3.1 *Let V be a variety generated by algebras $\{\mathbf{B}_i | i \in I\}$. Let $M(V)$ be the variety of monoids generated by $\{P_1(\mathbf{B}_i) | i \in I\}$. If $\mathbf{B}$ is an algebra of V, then $P_1(\mathbf{B})$ is a monoid of $M(V)$.*

Proof Let $\mathbf{A}$ be an algebra of V and $f : \mathbf{A} \to \mathbf{B}$ a surjective homomorphism. Then f can be extended to surjective homomorphism $g : P_1(\mathbf{A}) \to P_1(\mathbf{B})$. [Lausch, Nöbauer (3.3.1)].

Let $\mathbf{A} = \prod_{i\in I} \mathbf{A}_i$ be a direct product then $P_1(\mathbf{A})$ is isomorphic to a subdirect product of $P_1(\mathbf{A}_i), i \in I$ (cf. Lausch, Nöbauer [LauNöb 73, (3.4.1)]. If $\mathbf{B}$ is a subalgebra of $\mathbf{A}$ we consider the subsemigroup U of $P_1(\mathbf{A})$, $U = \{\psi \in P_1(\mathbf{A}) | \psi(x_1) \in \mathbf{B}[x_1]\}$, where $\mathbf{B}[x_1]$ is the polynomial algebra of $\mathbf{B}$ in the indeterminate x_1. Then $U \to P_1(\mathbf{B})$ is a surjective semigroup homomorphism. Therefore, $P_1(\mathbf{B})$ is isomorphic to a homomorphic image of a subsemigroup of $P_1(\mathbf{A})$.

Under what conditions on V does the converse of Theorem 2.3.1 hold?

Example 2.3.2 A lattice L is distributive if and only if $P_1(L)$ is idempotent [Schweigert 75].

Example 2.3.3 Given the distributive lattice $D = (\{0,1\}; \wedge, \vee)$ we have the following equational base for $P_1(D)$

$$p^2 = p,$$
$$p \circ q \circ p = p \circ q.$$

From these equations we can derive the following hyperidentities for the variety of distributive lattices.

$$F(F(x_1, \ldots x_k), x_2, \ldots, x_k) = F(x_1, \ldots, x_k)$$
$$F(G(F(x_1, \ldots, x_k), x_2, \ldots, x_l), x_2, \ldots, x_k) = F(G(x_1, \ldots, x_l), x_2, \ldots, x_k)$$

for every $k, l \in \mathbb{N}$. If we consider hyperidentities of a fixed type (2,2), then we have $k = l = 2$.

Example 2.3.4 An equational base of the semigroup T_2 of all transformations on the set $\{0, 1\}$ is given by

$$\begin{aligned} x^3 &= x \\ xyx^2 &= xy \\ xy^2 &= yxyx. \end{aligned}$$

Theorem 2.3.5 [Volkov 89] *The semigroup T_n of all transformation on an n-element set has no finite basis of identities for $n \geq 3$.*

Example 2.3.6 The semigroup T_n $(n \geq 3)$ fulfills the identity

$$x^{n-1}yx^{n-2} = x^{n-1}yx^{n-2+\kappa(n)},$$

where $\kappa(n)$ is the least common multiple of $1, \ldots, n$.

Theorem 2.3.7 *The following are equivalent for an algebra $\mathbf{A}$.*

(1) *The monoid equation*

$$p_1^{k_1} \circ \ldots \circ p_n^{k_n} = q_1^{h_1} \circ \ldots q_m^{h_m}$$

holds for $P_1(\mathbf{A})$.

(2) *The hyperidentity*

$$T_1^{k_1} \circ \ldots \circ T_n^{k_n} = S_1^{h_1} \circ \ldots \circ S_m^{h_m}$$

holds for the variety **HSP(A)**.

Proof Consider the monoid equation

$$p_1^{k_1} \circ \ldots \circ p_n^{k_n} = p_{n+1}^{h_1} \circ \ldots \circ p_{n+m}^{h_m}.$$

It follows that for every polynomial function $p_i(x)$ which has some representation as a word $p_i(x) = w(x, a_{i1}, \ldots a_{ik})$, $a_{ij} \in \mathbf{A}$, $j = 1, \ldots, k$, the above monoid equation holds. As the element a_{ij} can be selected arbitrarily from $\mathbf{A}$, we may replace these elements formally by a variable x_{ij} and the above monoid equation holds for every term function t_i of any arity. Hence, we conclude that the hyperidentity

$$T_1^{k_1} \circ \ldots \circ T_n^{k_n} = T_{n+1}^{h_1} \circ \ldots \circ T_{n+m}^{h_m}$$

holds for the variety **HSP(A)**. Obviously also the reverse direction holds.

Notation 2.3.8 A hyperidentity is called *iterative* if it is constructed by the iteration of hyperterms in a fixed variable x_i. Iterative hyperidentities are connected to semigroups. It is not important in which variable x_i the iteration is executed but one cannot change the variable during the steps of the iteration.

Remark 2.3.9 The set of all iterative hyperidentities for a variety V of type τ is closed under the rules (1), (2), (3), (6) of Definition 1.2.1 and Remark 1.2.2, and

(4′) $T = S$ implies $F(x_1, \ldots, x_{i-1}, T, x_{i+1}, \ldots, x_m) = F(x_1, \ldots, x_{i-1}, S, x_{i+1}, \ldots, x_m)$ for hyperterms $T, S \in H(\tau)$ and an n-ary hypervariable $F, i = 1, \ldots, m$;

(5′) $T(x_1, \ldots, x_n) = S(x_1, \ldots, x_n)$ implies $T(x_{\pi(1)}, \ldots, x_{\pi(n)}) = S(x_{\pi(1)}, \ldots, x_{\pi(n)})$ for every permutation π on $\{1, \ldots, n\}$.

A set Σ of iterative hyperidentities of type τ is called a *basis* for the set of all iterative hyperidentities for a variety V if every iterative hyperidentity of type τ of V can be derived by (1), (2), (3), (4′), (5′), (6) from Σ. (Consider Example 2.3.3).

2.4 Lattices and abelian groups

We are considering the symmetric semigroup S of the 1-place functions on an n-element set. The order of an element $f \in S$ is the least number k such that the elements of the cyclic subsemigroup $\{f, f^2, \ldots, f^k\}$ are different. We have $f^{k+1} = f^m$ for some m with $0 < m \leq k$ and we put $p = k + 1 - m$.

The following results are contained in [Schweigert 85]. In this section we consider $T(\mathbf{A})$ as the semigroup of term functions of $\mathbf{A}$ with respect to composition in the first variable

x_1. Denote by M_3 the modular, non-distributive lattice with 5 elements, and by N_5 the non-modular lattice with 5 elements.

Lemma 2.4.1 *The semigroup equation $\varphi^3 = \varphi^5$ holds for the semigroup of all 1-place monotone functions of the lattice N_5.*

Proof We consider the cases for all monotone functions f where the image of f consists of $|\mathrm{Im}\ f| = n$ elements. If $|\mathrm{Im}\ f| = 1$ we have $f^2 = f$, if $|\mathrm{Im}\ f| = 2$ we can have $f^3 = f^2$, if $|\mathrm{Im}\ f| = 3$ we have $f^4 = f^3$, and if $|\mathrm{Im}\ f| = 5$ we have again $f^2 = f$. In the case $|\mathrm{Im}\ f| = 4$ we consider two subcases. If Im f is a chain then we can have $f^4 = f^3$. If Im f is not a chain then there are monotone functions f with $f^3 = f$. By Proposition 2.2.3 we have altogether that $f^3 = f^5$. One notices that this also holds for congruence preserving monotone functions.

Lemma 2.4.2 *The semigroup equation $\varphi^3 = \varphi^9$ holds for the semigroup of all 1-place monotone functions of the lattice M_3.*

Proof If $|\mathrm{Im}\ f| = 1$ we have $f^2 = f$, if $|\mathrm{Im}\ f| = 2$ we can have $f^3 = f^2$, if $|\mathrm{Im}\ f| = 3$ we can have $f^3 = f^2$, and if $|\mathrm{Im}\ f| = 5$ we have the group of lattice automorphisms with $f^7 = f$. In the case $|\mathrm{Im}\ f| = 4$ there are functions with $f^4 = f^3$. Altogether we have $\varphi^3 = \varphi^9$ by Proposition 2.2.3.

Theorem 2.4.3 *The equation $\varphi^3 = \varphi^5$ holds for the semigroup $T(N_5)$ of the term functions of the lattice N_5. The clone equation $\varphi^3 = \varphi^9$ holds for the semigroup $T(M_3)$ of the term functions of the lattice M_3.*

Proof A function $f : N_5 \to N_5$ is a polynomial function of N_5 if and only if f is congruence preserving and monotone [Wille 77]. Hence for the semigroup of 1-place polynomial functions the equation $\varphi^3 = \varphi^5$ holds. By Lemma 4.1 in [Schweigert 83] this holds also for every term function of N_5 and hence for the clone of term functions of N_5. A function $f : M_3 \to M_3$ is a polynomial function of M_3 if and only if f is monotone [Schweigert 74]. Then by the same arguments the equation $\varphi^3 = \varphi^9$ holds for the variety $T(M_3)$.

Remark Let T be a variety of semigroups with an equation $\varphi^m = \varphi^{m+k}$ and S a subvariety of T with an equation $\varphi^n = \varphi^{n+s}$ with $n \leq m$, $s \leq k$. By Proposition 2.2.3 we conclude that s divides k. This consequence can be used to study varieties containing $T(N_5)$ or $T(M_3)$. In particular, $T(N_5)$ does not generate a subvariety of $\mathbf{HSP}(T(M_3))$ nor $T(M_3)$ a subvariety of $\mathbf{HSP}(T(N_5))$.

In [Schweigert 83] we have shown that a lattice L is distributive if and only if the variety $T(L)$ of semigroups is idempotent, i.e. $\varphi^2 = \varphi$.

Theorem 2.4.4 *Let V a non-trivial variety such that in the variety $\mathbf{HSP}(T(V))$ of semigroups the equation $\varphi^2 = \varphi$ holds. Then V is not congruence permutable.*

Proof We assume that V is congruence permutable and consider the term $p(x, y, z)$ with $p(x, z, z,) = p(z, z, x) = x$. For $\varphi(x, y, z) = p(y, x, z)$ we have $p(y, p(y, x, z), z) = p(y, x, z)$ and therefore $y = p(y, z, z)p(y, p(y, z, z), z) = p(y, y, z) = z$, a contradiction.

Let $\kappa(k)$ denote the least common multiple of $\{1,\ldots,k\}$.

Theorem 2.4.5 *Let V be an arithmetical variety such that $\varphi^n = \varphi^m$ $(m > n > 0)$ holds for the variety* $\mathbf{HSP}(T(V))$ *of semigroups. If* $\mathbf{A}$ *is a simple algebra of V, then* $\mathbf{A}$ *is finite, $|A| \le n+1$ and $\kappa(|A|)$ is a divisor of $m-n$.*

Proof $\mathbf{A}$ is local polynomially complete [Penner 81] and hence every permutation π of the carrier set A is a local polynomial function. π cannot be of order greater than $m-n$. Therefore $\mathbf{A}$ is finite and the order of π divides $m-n$. For $A = \{a_1,\ldots,a_k\}$ we consider the polynomial function ψ defined by $\psi(a_i) = a_{i-1}$, $i = 2,\ldots,k$ and $\psi(a_1) = a_1$. We have $\psi^{k-1} = \psi^k$. From this identity we have $\varphi^n = \varphi^m$ only in the case $n \ge k-1$.

Theorem 2.4.6 *Let V be a congruence permutable variety such that $\varphi^n = \varphi^m$ $(m > n > 0)$ holds for the semigroup $T(V)$. If* $\mathbf{A}$ *is a finite simple algebra and p is a prime number which divides* $|\mathbf{A}|$, *then p is a divisor of $m-n$.*

Proof By a theorem of R. McKenzie, $\mathbf{A}$ is either polynomially complete or affine [Pixley 77], p. 602. In the first case the theorem follows from 2.4.5, in the second case we know that $\mathbf{A}$ is polynomially equivalent to a module with $p \cdot x = 0$. We consider the polynomial function $\psi(x) = x+1$ and have $\psi^p(x) = x$, hence p divides $m-n$.

Theorem 2.4.7 *If* $\mathbf{G}$ *is a finite subdirectly irreducible abelian group, then the equation $\varphi^n = \varphi^{n+p^n(p-1)}$ holds for $T(\mathbf{G})$ with $|G| = p^n, p$ a prime number.*

Proof Let ψ be a 1-place polynomial function of $\mathbf{G}$. Then ψ is of the form $\psi(x) = a^t x^k$, where $G = \langle a \rangle$ and $0 \le t,k \le p^n - 1$. We have $\psi^m(x) = a^{t(1+k+\cdots+k^{m-1})}x^{k^m}$. Now we consider the following cases. If $k = 0$ then $\psi(x) = a^t$ and we have $\psi^2 = \psi$. If $k = 1$ then $\psi^m(x) = a^{tm}x$. For $m = p^n + 1$ we have $(a^t)^{p^n+1} = a^t$ and hence $\psi^{p^n+1} = \psi$. If $k > 1$ we consider two subcases. If $(k,p) = 1$, then we have by Fermat $k^{\varphi(p^n)} \equiv 1 \pmod{p^n}$ for the Euler function φ. We have also

$$\frac{k^{\varphi(p^n)} - 1}{k-1} = 1 + k + \cdots + k^{\varphi(p^n)-1} \equiv 0 \pmod{p^n}.$$

We conclude that

$$\psi^{\varphi(p^n)}(x) = a^{t(1+k+\cdots+k^{\varphi(p^n)-1})}x^{k^{\varphi(p^n)}} = a^{t\cdot 0}x = x.$$

Hence we have $\psi^{\varphi(p^n)+1} = \psi$. For the subcase $k = p^s \cdot r$ with $1 \le s < n$ we have

$$\psi^m(x) = a^{t(1+p^s r+\cdots+(p^s r)^{m-1})}x^{(p^s r)^m}.$$

Here we have $(p^s r)^m \equiv 0 \pmod{p^n}$ for $m \ge n-s$. If we put $m = n$, then we have $\psi^n = \psi^{n+1}$. Altogether we have $\psi^1 = \psi^{1+1}, \psi^1 = \psi^{1+p^n}, \psi^1 = \psi^{1+\varphi(p^n)}$ and $\psi^n = \psi^{n+1}$. By Proposition 2.2.3 it follows that $\psi^n = \psi^{n+p^n(p-1)}$.

Corollary 2.4.8 *If* $\mathbf{G}$ *is a simple group of order p, p a prime number, then $\varphi = \varphi^{p^2-p+1}$ holds for $T(\mathbf{G})$.*

Theorem 2.4.9 *Let V be a finitely generated variety of groups. V is a variety of abelian groups generated by simple groups if and only if $\varphi^n = \varphi$ holds for $T(V)$ for some $n \in \mathbb{N}$, $n > 1$.*

Proof One direction is implied by Corollary 2.4.8. On the other hand, if $\varphi^n = \varphi$ holds for $T(V)$, then we have the equation $[x, y, \ldots, y] = [x, y]$ because of the term function $\psi(x, y) = x^{-1}y^{-1}xy = [x, y]$. We show that every finite group in V is abelian. Let $\mathbf{G}$ be a minimal counter example. If $\mathbf{G}$ is simple, then $\varphi^n = \varphi$ holds only in the case that $\mathbf{G}$ is abelian, otherwise $\mathbf{G}$ would be polynomially complete [LauNöb 73, p. 41]. If $\mathbf{N}$ is a non-trivial normal subgroup of $\mathbf{G}$, then by hypothesis $\mathbf{G}/\mathbf{N}$ and $\mathbf{N}$ are abelian and hence by [Hall 59], Corollary 9.2.1 (p. 141), $\mathbf{G}$ is solvable. There are elements $b \in N$ and $a \in G$ such that $[a, b] \neq e$. Because $\mathbf{G}$ is solvable we have $[a, b] \in N$ ([Hall 59], Theorem 9.2.1 (p. 138)). As $\mathbf{N}$ is abelian it follows that $e = [[a, b], b] = [a, b, b] = [a, b, ..., b] = [a, b]$ a contradiction. We have still to show that V is generated by simple groups, but this follows from the proof of Theorem 2.4.7.

Theorem 2.4.10 *Let $(G; +)$ be a finite elementary abelian p-group and let* $\mathrm{End}(\mathbf{G})$ *be the endomorphism ring of* $\mathbf{G}$. *The equation $\varphi^n = \varphi$ for some $n \in \mathbb{N}$, $n > 1$, holds for* $\mathrm{End}(\mathbf{G})$ *if and only if $|G| = p$.*

Proof If $|G| = p$, then every endomorphism ψ is of the form $\psi(x) = kx$, $k = 0, \ldots, p-1$, as $\mathbf{G}$ is a cyclic group and $\psi(0) = 0$. If $k = 0$, we have $\psi^2 = \psi$ and if $k = 1$, we have $\psi^2 = \psi$. In all other cases ψ is an automorphism of order p, hence $\psi^{p+1} = \psi$. On the other hand assume that $\psi^n = \psi$ holds and that $|G| = p^n$ with $n > 1$. Consider $G = \langle g_1 \rangle + \langle g_2 \rangle + \ldots + \langle g_k \rangle$ as a direct product of simple p-groups $\langle g_i \rangle$. The map $\bar{f}(g_1) = g_2$ and $\bar{f}(g_i) = e$, $i = 2, \ldots, k$, can be extended to an endomorphism $f : G \to G$. We have $f^3 = f^2$. Therefore an equation $\varphi^n = \varphi$ cannot hold for any $n \in \mathbb{N}$, $n > 1$.

In the following we use the notation of [Schweigert 79].

Theorem 2.4.11 *Let $(G; +)$ be a group of order p, p a prime number, and let R be a subring of* $\mathrm{End}(\mathbf{G})$. *Then the R-module $\mathbf{G}$ is prepolynomially complete.*

Proof The polynomial functions $\psi : G^n \to G$ are of the form $\psi(x_1, \ldots, x_n) = a_1x_1 + \ldots + a_nx_n + a_{n+1}$, where $a_i \in G = \{0, 1, \ldots, p-1\}$. On the other hand every function of this form is a polynomial function. We conclude that the clone $P(\mathbf{G})$ of the polynomial functions of $\mathbf{G}$ is the clone of all quasilinear functions of $G = \{0, \ldots, p-1\}$. Hence $P(\mathbf{G})$ is maximal, and therefore $\mathbf{G}$ is prepolynomially complete both as a group and as an R-module.

2.5 A criterion for primality

Two semigroups $\mathbf{A}, \mathbf{B}$ can be defined by different sets $\mathrm{Id}(\mathbf{A})$, $\mathrm{Id}(\mathbf{B})$ of semigroup identities if and only if $\mathbf{A}$ and $\mathbf{B}$ generate different varieties of semigroups. In this case one can find identities which hold for $\mathbf{HSP}(\mathbf{A})$ but not for $\mathbf{HSP}(\mathbf{B})$ and vice versa. We call these identities separating identities (compare also 3.5). For an example consider [Schweigert 75].

The aim of this section is to find separating identities for the full transformation monoid and their maximal submonoids. These identities yield hyperidentities by which we can characterize primal algebras. Many results of this section are due to Denecke and Pöschel [DenPös 88 a, b].

Theorem 2.5.1 *Let A be a finite set, $|A| = n$, and let H be a proper subsemigroup of the full transformation semigroup H_A on A. Then* $\mathrm{Var}(H)$ *is properly contained in* $\mathrm{Var}(H_A)$.

Proof We give a sketch of the proof which in its main part is due to P.P. Pálfy. We start with some known facts which can be proved by group theoretic methods.

Fact 1: Let H be a semigroup, G a group (both finite). If $G \in \mathrm{Var}(H)$, then G belongs to the variety generated by the subgroups (i.e. subsemigroups which are groups) of H.

Fact 2: If $S_n \in \mathbf{HS}(G_1 \times G_2)$ (G_1, G_2 finite groups), then $S_n \in \mathbf{HS}(G_1)$ or $S_n \in \mathbf{HS}(G_2)$. Consequently, if $S_n \in \mathrm{Var}\{G_i | i \in I\}$ ($= \mathbf{HSP}\{G_i | i \in I\}$), where $G_i (i \in I)$ are finite groups and I is finite, then $S_n \in \mathbf{HS}(G_i)$ for some $i \in I$.

Fact 3: The subgroups G of H_A are of the following form

$$G \leq \{f \in H_A | \mathrm{Im}\, f = \mathrm{Im}\, e,\ \ker f = \ker e\},$$

where $e \in H_A$ is the identity element of the group G, in particular, e is idempotent (here and in the following e does not necessarily denote the identical function on A). The mapping $f \mapsto f|\mathrm{Im} f$ is an embedding of G into $S_{\mathrm{Im} f}$. In particular we have

$$|G| \leq |\mathrm{Imf}|! \leq n!.$$

Now we can prove the theorem. Suppose H is a proper subsemigroup of H_A such that $\mathrm{Var}(H) = \mathrm{Var}(H_A)$. Without loss of generality we can assume that H is maximal in H_A. Let $\{G_i | i \in I\}$ be the set of the subgroups of H. Since $S_A \in \mathrm{Var}(H_A) = \mathrm{Var}(H)$ we get from Fact 1 and Fact 2 that $S_A \in \mathbf{HS}(G_i)$ for some $i \in I$. Thus $n! = |S_A| \leq |G_i|$, and by Fact 3 ($|G_i| \leq n!$) we have $|S_A| = |G_i|$. Consequently, $S_A = G_i$, i.e. H contains S_A. Since every $f \in H_A$ with $|\mathrm{Imf}| = n$ together with S_A generates all elements of H_A, there is a single maximal subsemigroup of H_A containing S_A, namely

$$H = S_A \cup \{f \in H_A |\ |\mathrm{Imf}| \leq n-2\}.$$

Obviously, $\lambda(f) \leq n-2$ for all $f \in H$. Thus, by 2.4, H but not H_A satisfies the hyperidentity $\psi^{n-2}(x) = \psi^{n-2+\kappa(n)}(x)$, i.e. the semigroup identity $x^{n-2} = x^{n-2+\kappa(n)}$, in contradiction to $\mathrm{Var}(H) = \mathrm{Var}(H_A)$.

Before we proceed to the general case we will point out the methods. These methods were developed by Reischer, Schweigert and Simovici in [ReSchSi 87] for the case $A = \{0, 1, 2\}$.

Our starting point is the Slupecki criterion and Iablonskiĭ's list of the 18 maximal clones. Ultimately we wish to describe the maximal submonoids of unary operations by hyperidentities. For this it is enough to consider relations up to isomorphisms.

Let $(A;\rho_1)$ and $(B;\rho_2)$ be relational systems of the same type. A bijective map $\alpha : A \to B$ is called an isomorphism if

$$(a_1,\dots,a_n) \in \rho_1 \quad \text{iff} \quad (\alpha(a_1),\dots,\alpha(a_1)) \in \rho_2.$$

By $\mathrm{End}\rho$ denote the set of unary operations preserving ρ. Let α be an isomorphism.

Proposition 2.5.2 *The monoid* $\langle \mathrm{End}\rho_1, \circ\rangle$ *is isomorphic to the monoid* $\langle \mathrm{End}\rho_2, \circ\rangle$.

Proof For $f : A \to A$ put $f^{(\alpha)}(x) = \alpha(f(\alpha^{-1}(x))$ for all $x \in B$. It is almost immediate that $f \to f^{(\alpha)}$ is the required isomorphism. (This may be extended to O_A and to clones preserving relations but we need only the particular form).

We need the following form of the Slupecki criterion. $O_A^{(n)}$ denotes the set of all n-place functions on A, and $O_A := \bigcup_{n\in N} O_A^{(n)}$. For $X \subseteq O_A$ denote by $[X]$ the clone generated by X, and for $n > 0$ set $[X]^{(n)} := [X] \cap O_A^{(n)}$.

Theorem 2.5.3 *Let A be a finite set and $|A| > 2$. A set X of operations on A is complete if and only if*

(i) *X contains an essentially at least binary surjective operation, and*

(ii) $[X]^{(1)} = O_A^{(1)}$ *(i.e. X generates all unary operations on A).*

Actually, Slupecki proved only sufficiency for f binary. There are several proofs and it has been generalized in several directions (condition (ii) may be weakened to: $[X]^{(1)} = M$, where M may be the alternating group, etc.). For our purposes we first replace $[X]^{(1)} = O_A^{(1)}$ by "$[X]^{(1)}$ is not included in a maximal submonoid of $\langle O^{(1)}; \circ\rangle$", and then describe the maximal submonoids functionally. For the description of the maximal submonoids we use Iablonskiĭ's list of maximal clones. It is clear that the maximal submonoids are among the $C^{(1)}$, where C is one of the 18 maximal clones. We eliminate 5 of them. The list is the following:

Lemma 2.5.4 *Let $A = \{0,1,2\}$. There are 13 maximal submonoids of* $\langle O_A^{(1)}, \circ\rangle$:

(1) *The monoid $\{ax + b | b \in A\}$ of linear functions;*

(2) *The three monoids* $\mathrm{End}\{i,j\}$, *where* $0 \le i < j \le 2$;

(3) *The three monoids* $\mathrm{End}(\le)$, *where $\le$ is a chain (= linear order) on A;*

(4) *The three monoids* $\mathrm{End}\ \theta$, *where θ is a non-trivial equivalence relation on A;*

(5) *The three monoids* $\mathrm{End}\rho_{ij}$, *where* $\rho_{ij} = \mathrm{Pol}(A^2\backslash\{(i,j),(j,i)\})$ $(0 \le i \le j \le 2)$ *(so-called central relation).*

Proof The monoids $C^{(1)}$ with C a maximal clone not on the list are: $\mathrm{End}\{i\}$ $(i \in A)$ and the monoid of self-dual maps (i.e. maps satisfying $f(x+1) = f(x)+1$). The clone C of essentially unary or non-surjective operations has $C^{(1)} = O_A^{(1)}$ and so may be omitted. Let $i \in A$ then $f \in \text{End}\{i\}$ iff $f(i) = i$. It is easy to see that $f \in \text{End}\rho_{jk}$ where $\{i,j,k\} = A$ (indeed, if $(a,b) \in \rho_{jk}$ and $a \neq b$, we have $a = i$ or $b = i$ and so $f(a) = i$ or $f(b) = i$, proving $(f(a), f(b)) \in \rho_{jk}$ hence $\text{End}\{i\} \subsetneq \text{End}\rho_{jk}$). Here $\text{End}\rho_{jk}$ contains the constant j and so the inclusion is proper. Similarly we prove that each self-dual f is linear. The inclusion is again proper. We give a hyperidentity for all the above monoids. The monids of linear functions satisfies $F^2 = F$. This can be verified directly (as for $f(x) = ax + b$ we have $f^{(n)}(x) = a^n x + b(1 + a + \ldots + a^{n-1})$ where $a^7 = a$ and $(1 + a + \ldots + a^6 = 1)$, and follows also from a more general result in [Schweigert 83]. Note that a selfmap f of A does not satisfy $f^4 = f^2$ exactly if $f(x) \in \{x+1, x+2\}$. Indeed if the diagraph of f is not a cyclic permutation, then it has at most a cycle of length 2. Suppose that it has a cycle of length 2, say $\{0,1\}$. On the cycle clearly f^4 and f^2 agree. If $f(2) \neq 2$ then $f^2(2) \in \{0,1\}$ and again $f^4(2) = f^2(2)$. If f has no cycle then $f^2(2)$ is a fixed point of f for each $x \in A$ and so $f^4 = f^2$. Now it suffices to verify that neither of $x+1$ and $x+2$ belong to the monoids listed above in (2) - (5) (which is almost immediate). Thus we are led to the following

Theorem 2.5.5 *Let $|A| = 3$ and let X be a set of operations on A. Then X is complete if and only if*

(i) *X contains an essentially at least binary surjective operation, and*

(ii) *the foundation $[X]^{(1)}$ satisfies neither $F^4 = F$ nor $F^4 = F^2$.*

Now we proceed to the general case. By e we denote the identity map.

Proposition 2.5.6 *Let $H \leq H_A$ be a subsemigroup such that the algebra $(A; H)$ has a proper subalgebra with carrier $B \subset A$. Let $l = \max\{\lambda(f) | f \epsilon H\}$ and $n_B = \max\{|B|, |A \backslash B|\}$. Then*

$$H \models \varphi^l = \varphi^{l+\kappa(n_B)}.$$

Moreover,

$$H \models \varphi^{n-1} = \varphi^{n-1+\kappa(n)}, \quad n = |A|.$$

Proof First consider the permutations $f \in H \cap S_A$. Since every such f preserves the subset B (and consequently also $A \backslash B$), the cycles of f have a length which belongs to $\{1, 2, \ldots, |B|\}$ or $\{1, 2, \ldots, |A \backslash B|\}$. Thus $f^{\kappa}(B) = e$, consequently $f^l = f^{l+\kappa(|B|)}$. Now let $f \in H \backslash S_A$ (i.e. $\lambda(f) \geq 1$). Then the permutation $f = f|\text{Im } f^{\lambda(f)}$ on $I = \text{Im} f^{\lambda(f)}$ preserves the subsets $B \cap I$ and $(A \backslash B) \cap I$. Thus

$$g^{\kappa(n')} = e \quad (n' = \max\{|B \cap I|, |(A \backslash B) \cap I|\}),$$

and we get

$$f^{\lambda(f)} = g^{\kappa(n')}(f^{\lambda(f)}) = f^{\lambda(f)+\kappa(n')}.$$

Because of $n' \leq n_B$ we get $f^l = f^{l+\kappa(n_B)}$ by Lemma 2.2.3. Since $l \leq n-1$, $n_B \leq n-1$, again by 2.2.3. It follows that also $\varphi^{n-1} = \varphi^{n-1+\kappa(n-1)}$ holds in H (this however directly follows from 2.2.11 too).

Proposition 2.5.7 *Let $H \leq H_A$ be a subsemigroup such that the algebra $(A; H)$ has a non-identical automorphism $s \in S_A$ which consists of r cycles of length $p \geq 2$ ($n = |A| = pr$, $1 \leq r \leq n-1$). Then*

$$H \models \varphi^{r-1} = \varphi^{r-1+\kappa(n)}.$$

Moreover,

$$H \models \varphi^{n-2} = \varphi^{n-2+\kappa(n)}.$$

Proof If $f \in H \cap S_A$, then we have $f^{\kappa(n)} = \text{id}$; consequently $f^{r-1} = f^{r-1+\kappa(n)}$. Now let $f \in H \backslash S_A$. We will show that $\lambda(f) \leq r-1$. At first we note that f maps every cycle of the permutation s onto a cycle of s. In fact, $b = s^i(a)$ implies

$$f(b) = f(s^i(a)) = s^i(f(a)) \neq f(a)$$

for $1 \leq i \leq p-1$. Since s has only r cycles (of equal length p), we get Im $f^{r-1} =$ Imf^r. Consequently $\lambda(f) \leq r-1$. By 2.2.11 and 2.2.3 we get $f^{r-1} = f^{r-1+\kappa(n)}$. Finally, since $r-1 \leq n-2$, we get $f^{r-2} = f^{r-2+\kappa(n)}$ for all $f \in H$.

Proposition 2.5.8 *Let H be a subsemigroup of H_A ($n = |A|$) such that the algebra $(A; H)$ has a non-trivial congruence relation θ. For $B \subset A$ with $2 \leq |B| \leq n-1$ let θ_B be the equivalence relation with blocks (congruence classes) $B, \{c\} (c \in A \backslash B)$ (i.e., B is the only non-trivial block). Then we have:*

(i) *if θ is not of the form θ_B ($B \subset A$), then*

$$H \models \varphi^{n-2} = \varphi^{n-2+\kappa(n)};$$

(ii) *if $\theta = \theta_B$ for some $B \subset A$, $2 \leq |B| \leq n-1$, then*

$$H \models \varphi^{n-1} = \varphi^{n-1+\kappa(n-1)}.$$

Proof (ii): For $f \in H \backslash S_A$ the indicated identity is fulfilled trivially. For $f \in H \cap S_A$, however, f preserves B since every permutation in H must map any block onto a block; thus $H \cap S_A$ satisfies the identity in (ii).

(i): For $f \in H$ with $\lambda(f) \leq n-2$ (in particular for $f \in H \cap S_A$) the identity in (i) is satisfied. If there were some function $f \in H$ with $\lambda(f) = n-1$, then f should have the following form. There is an $a \in A$ such that $a, f(a), f^2(a), \ldots, f^{n-1}(a) = f^n(a)$ are all the elements of A. Since $f \in H$ preverves θ, every non-trivial block of θ must be of the form $\{f^i(a), f^{i+1}(a), \ldots, f^{n-1}(a)\}$, $i \in \{1, 2, \ldots, n-2\}$. Consequently, θ is of the form θ_B in contradiction to the assumption in case (i).

We need the following result due to G. Rousseau [Rousseau 67].

Theorem *A function $f \in O_A$ is Sheffer if and only if the algebra $(A; f)$ has no proper subalgebras, no non-identity automorphism and is simple.*

Theorem 2.5.9 (Denecke, Pöschel) *The algebra $(A; f \in O_A)$ of prime power cardinality n is primal if and only if it satisfies none of the following hyperidentities:*

(i) $$\varphi^{n-1} = \varphi^{n-1+\kappa(n-1)},$$

(ii) $$\varphi^{n-2} = \varphi^{n-2+\kappa(n)}$$

(φ a unary operation symbol).

It is easy to see that $n = |A|$ is a prime power if and only if $k(n-1) \neq \kappa(n)$.

Proof If $\mathbf{A} = (A; f)$ is primal then $H_A = T(\mathbf{A})$. However, H_A satisfies neither (i) nor (ii) (since $\kappa(n-1) \neq \kappa(n)$). Conversely, if $\mathbf{A}$ is not primal, then, by 3.1, $\mathbf{A}$ has a proper subalgebra – and therefore satisfies (i) by 2.5.7 – or $\mathbf{A}$ has a non-trivial congruence – and therefore satisfies (i) or (ii) by 2.5.8 – or $\mathbf{A}$ has a non-trivial automorphism, say s. If s consists of cycles of equal length, then, by 2.5.7, $\mathbf{A}$ satisfies (ii). Otherwise some power of s has fixed points. Since the fixed points of an automorphism constitute a subalgebra of $\mathbf{A}$, $\mathbf{A}$ satisfies (i) by 2.5.8. Consequently, if $\mathbf{A}$ is not primal then (i) or (ii) are satisfied.

With arguments of the same kind the following can be shown:

Theorem 2.5.10 *An algebra $\mathbf{A} = (A; f \in O_A), |A| \geq 2$, is primal if and only if it does not satisfy the following unary hyperidentity:*

$$\varphi_2\varphi_1^{n-2}\varphi_2\varphi_1^{n-1}(x) = \varphi_2\varphi_1^{n-2+\kappa(n)}\varphi_2\varphi_1^{n-1}(x).$$

Theorem 2.5.11 (Denecke, Pöschel) *Let $\mathbf{A} = (A; \Omega)$ be a finite algebra ($|A| \geq 2$). Then $\mathbf{A}$ is primal if and only if it does not satisfy the binary hyperidentity*

$$\psi^{n^2-2}\psi^{\top}\psi^{n^2-1}(x_2, x_1) = \psi^{n^2-2+\kappa(n2)}\psi^{\top}\psi^{n^2-1}(x_1, x_2),$$

where

$$\psi(x_1, x_2) = (\varphi(x_1, x_2), \varphi' x_1, x_2)), \quad \psi^{\top}(x_1, x_2) = \varphi'(x_1, x_2)).$$

With the following concept these results can be seen from another point of view [Schweigert 89].

Notation 2.5.12 We call a hyperidentity of the form $F^r = F^s$, $r < s$, for the algebra $\mathbf{A} = (A, \Omega)$ *irreducible* if $F^r = F^s$ is an irreducible equation for the semigroup $T_1(\mathbf{A})$ of one-place term functions (see Definition 2.2.5).

We continue to use the notation $\kappa(n)$ for the least common multiple of $1, \ldots, n$.

Theorem 2.5.13 *Let $\mathbf{A} = (A, \Omega)$ be an algebra with an essentially at least binary surjective operation, $|A| = n = p^m, p^m$ a prime power. Then $\mathbf{A}$ is primal if and only if*

$$(^*) \quad F^{n-1} = F^{n-1+\kappa(n)}$$

is an irreducible hyperidentity for **A**.

Proof Let **A** be primal, $A = \{1, \ldots, n\}$ and $g : A \to A$ defined by $g(x) = x - 1$ for $x \neq 1$ and $g(1) = 1$. Then we have $\lambda(g) = n - 1$ and hence $\lambda(T_1(\mathbf{A})) = n - 1$. The permutation group on A has the exponent $\kappa(n)$ [Hall] p. 54, and hence $\pi(T_1(\mathbf{A})) = \kappa(n)$. We have that $(^*)$ is an irreducible hyperidentity.

Conversely, let $(^*)$ hold as an irreducible hyperidentity. Hence the hyperidentities

$$(^{**}) \quad F^{n-2+\kappa(n)} = F^{n-2} \text{ and } F^{n-1+\kappa(n-1)} = F^{n-1}$$

do not hold as $\kappa(n-1) < \kappa(n)$ for n a prime power. By the above results of Denecke and Pöschel **A** is primal.

Remark The results of Denecke and Pöschel are proved by Rosenberg's completeness theorem. One may ask whether one can find an elementary proof. Indeed this is the case for $n = 2$.

Proposition 2.5.14 *If* $(^*)$ *is irreducible, then* $T_1(\mathbf{A})$ *contains a cyclic permutation of order* p^m.

Proof Because of 2.2.4 we have functions $f_1, \ldots, f_k$ such that $\kappa(n) = \operatorname{lcm}\{\pi(f_i) : i = 1, \ldots, k\}$. For the prime number $n = p^m$ we have a function f with $\pi(f) = p^m \cdot s$ and hence by Proposition 2.2.8 a function g with $\pi(g) = p^m$. Now g is a permutation on $A_g := g^{\lambda(g)}[A]$ and therefore consists of disjoint cycles such that the lcm of the length of these cycles is $\pi(g) = p^m$. We conclude that g is a cyclic permutation on A consisting of a single cycle of length p^m.

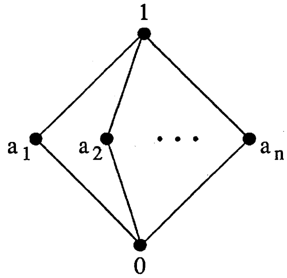
Proof for the case $n = 2$. By the above Proposition 2.5.14 all permutation of A, $|A| = 2$, are in $T_1(\mathbf{A})$. Furthermore because of $\lambda(T_1(A)) = 1$ at least one constant function is in $T_1(A)$, and hence both of them are. $T_1(\mathbf{A})$ contains all one-place functions and **A** is primal. One should mention that this also included a proof for $n = 2$ of the result of Denecke and Pöschel as Theorem 2.5.13 is equivalent to their results.

2.6 Algebraic monoids

Most results in this chapter are due to Reichel and Schweigert [ReiSch 91]. We consider the following

Representation-Problem Let M be a monoid and V a variety. Is there an algebra $\mathbf{A} \in V$ such that M is isomorphic to the monoid $P_1(\mathbf{A})$ of 1-place polynomial functions of **A**?

Proposition 2.6.1 *Every finite monoid is isomorphic to a submonoid of the monoid* $P_1(L)$ *of the modular lattice* M_n *for some* $n, n \geq 3$.



Proof Consider $M_n = \{0, 1, a_1, \ldots, a_n\}$ and the monoid $K = \{f | f : M_n \to M_n, f(0) = 0, f(1) = 1\}$ of self-maps on the set A: Every such self-map is order-preserving and the monoid K is the symmetric monoid S_n.

For $n \geq 3$ this lattice M_n is order-polynomially complete [Schweigert 74], and hence S_n is isomorphic to a submonoid of $P_1(M_n)$. Every finite monoid is isomorphic to some submonoid of the symmetric monoid S_n for some n.

The construction of Proposition 2.6.1 is a bit crude. The monoid M may be a very small submonoid of $P_1(M_n)$. In the following we present a construction which is rather tight. To prepare the proof of this result we need some constructions on graphs.

Notation End G denotes the monoid of endomorphisms of the graph G. End L is the monoid of order-endomorphisms of the lattice L. An order-endomorphism of L is a monotone function $f : L \to L$. An order-endomorphism which preserves 0 and 1 is called $\{0,1\}$-endomorphism of L. $\mathrm{End}_{\{0,1\}}L$ is the monoid of the $\{0,1\}$-endomorphisms of L.

Proposition 2.6.2 *The monoid* End G *of a simple graph* G *is isomorphic to the monoid* $\mathrm{End}_{\{0,1\}}L$ *of* $\{1,0\}$*-endomorphisms of a lattice* L *constructed from* G.

Proof According to the construction in the proof of 2.6.1 we define an embedding $\varphi : G \to L$ by

$$\begin{aligned} \varphi(P_j) &= A_j, & j &= 1, \ldots, p \\ \varphi(a_i) &= B_i, & i &= 1, \ldots, q. \end{aligned}$$

Let $f \in$ End G. Then we define a monotone function $f^\frown$ on L by

$$\begin{aligned} f^\frown(A_j) &:= \varphi(f(P_j)) \quad j = 1, \ldots, p \\ f^\frown(B_i) &:= \varphi(f(a_i)) \quad i = 1, \ldots, q \\ f^\frown(1) &:= 1, \\ f^\frown(0) &:= 0. \end{aligned}$$

Obviously $f^\frown$ is monotone and preserves 0 and 1. On the other hand, for every $\{0,1\}$-endomorphism of L we have an endomorphism of G and hence End $G \simeq \mathrm{End}_{\{0,1\}}L$.

Proposition 2.6.3 *Every finite monoid is isomorphic to the monoid* $\mathrm{End}_{\{0,1\}}L$ *of* $\{0,1\}$-*endomorphisms of a finite atomistic lattice* L *of height* 2.

Proof According to Hedrlín and Pultr [HedPul 64] for every finite monoid M there exists a finite simple, undirected graph G such that the monoid End G is isomorphic to M.

Theorem 2.6.4 *Every finite monoid is isomorphic to a submonoid of the monoid* $P_1(L)$ *of polynomial functions of a finite atomistic lattice* L *of height* 3.

Proof We consider the lattice L constructed above and add two special elements a, b to L such that for a, b the lower neighbor is 0 and the upper neighbor is 1. According to [Schweigert 74] L is an order polynomially complete lattice, and hence every order endomorphism is a unary polynomial function of L. Hence $\mathrm{End}_{\{0,1\}}L$ is a submonoid of $P_1(L)$. We consider the connected simple graph G presented in [Frucht 50] with q vertices $P_1, \ldots, P_q$ and p edges $a_1, \ldots, a_p$. To this graph G corresponds a lattice L with $p+q+2$ elements $1, 0, A_1, \ldots, A_p$, $B_1, \ldots, B_q$. The vertices are considered as the atoms $A_1, \ldots, A_p$ of L and the edges are the upper neighbors $B_1, \ldots, B_q$ of the atoms. An element B_i is the join of at least two atoms, namely the vertices incident with this edge. We add a greatest element 1 and a least element 0 to have an atomistic lattice L of height 3.

Definition 2.6.5 A monoid M is called *algebraic* if there exists an algebra $\mathbf{A}$ such that $P_1(\mathbf{A}) \simeq M$. A variety S of monoids is *algebraic* if there exists a variety V such that $S = \mathbf{HSP}(T_1(V))$.

Example 2.6.6 The variety $W = \{p^2 = p,\ p \circ q = p\}$ of monoids is not algebraic.

Proof Assume that V is a variety with $W = \mathbf{HSP}(T_1(V))$. Let $\mathbf{A}$ be a non-trivial algebra from V. Then $p_1(x) = x$, $p_2(x) = a$ are two polynomial functions of $\mathbf{A}$. Obviously $p \circ q = p$ does not hold for $p = p_1$ and $q = p_2$. But if $p \circ q = p$ holds for $T_1(V)$ it also holds for $P_1(\mathbf{A})$ according to Theorem 2.3.7.

Problem 2.6.7 Characterize the varieties of monoids which are algebraic.

Problem 2.6.8 Given the variety V of your choice. Which are algebraic monoids for V?

2.7 The k-ary monoids of term operations

There are various ways to construct semigroups by the composition of functions on a set A. Let us consider $f : A^k \to A^k$ defined by $f = (f_0, \ldots, f_{k-1})$, where $f_i : A^k \to A$ are k-place functions. Similarly we consider $g = (g_0, \ldots, g_{k-1})$ with $g_i : A^k \to A$ and define the composition $fg := (fg)$

$$(fg)(x_0, \ldots, x_{k-1}) = (f_0(g_0(x_0, \ldots, x_{k-1}), \ldots, f_{k-1}(g_{k-1}(x_0, \ldots, x_{k-1})).$$

This composition is associative and has an identity e defined by $e(x_0, \ldots, x_{k-1}) = (x_0, \ldots, x_{k-1})$.

For the algebra $\mathbf{A}$ the k-ary monoid $C_k(\mathbf{A})$ of term operations is defined in the following way. $C_k(\mathbf{A})$ consists of all maps $F : A^k \to A^k$ such that each component of F is defined by some k-ary term of the algebra $\mathbf{A}$.

In particular, $C_k(\mathbf{A})$ contains as components the projections p_j^k, $p_j^k(x_1, \ldots, x_k) = x_j$, and hence the identity function $e : A^k \to A^k$. Furthermore, $C_k(\mathbf{A})$ is closed under composition.

$C_k(\mathbf{A})$ is a submonoid of End $\mathbf{A}^k$, the monoid of all endomorphisms. This section reports on results of Hyndman, McKenzie and Taylor on this topic.

If $\mathbf{A}$ is finite, with $|A| = n$, then the monoid equation $u^{m-1} = u^{m-1+\kappa}$ holds in $C_k(\mathbf{A})$, where $m = n^k$, and κ is the least common multiple of all positive integers $\leq n^k$.

The following table presents equations of this form (see [HyMcKTa 92]).

	$k=2$		$k=3$		$k=4$	
semilattices	$m=2$	$\kappa=2$	$m=5$	$\kappa=6$	$m=10$	$\kappa=12$
distributive lattices	$m=2$	$\kappa=2$	$m=5$	$\kappa=6$	$m=10$	$\kappa=60$
modular lattices	$m=2$	$\kappa=2$	$m=5$	$\kappa=6$		?
lattices	$m=2$	$\kappa=2$		∞		∞
$\mathbf{Z}_2$-modules	$m=2$	$\kappa=6$	$m=3$	$\kappa=4\cdot3\cdot7$	$m=3$	$\kappa=4\cdot3\cdot5\cdot7$
Boolean algebras	$m=3$	$\kappa=12$	$m=7$	$\kappa=8\cdot3\cdot5\cdot7$	$m=15$	$\kappa=2^4\cdot3^2\cdot5\cdot7\cdot11\cdot13$

The next lemma presents a conceptually alternative description of $C_k(\mathbf{A})$.

Lemma 2.7.1 *For any algebra* $\mathbf{A}$ *and any* k*, the monoid* $C_k(\mathbf{A})$ *is dually isomorphic to the monoid* End $F(k)$ *of all endomorphisms of the free algebra* $F(k)$ *on* k *generators, for the variety* $V = \mathbf{HSP}(\mathbf{A})$.

Proof If an element Q of $C_k(\mathbf{A})$ is defined by the k-tuple of terms $(q_0, \ldots, q_{k-1})$, then there exists a unique endomorphism σ of $F(k)$ that maps each free generator x_i to the term q_i (considered as an element of the free algebra $F(k)$). This is a bijective correspondence that reverses multiplication.

In the following we will fix an algebra $\mathbf{A}$ and look for monoid equations $w = w'$ and positive integers k such that $C_k(\mathbf{A})$ satisfies $w = w'$. We cannot expect that this identity holds for fixed $\mathbf{A}$, w and w' if we consider $k+1$. (End k denotes the monoid of all self-maps of a k-element set.)

Lemma 2.7.2 *If* $\mathbf{A}$ *is any algebra with* $|A| \geq 2$ *and* $k \geq 1$*, then* End k *is dually isomorphic to a submonoid of* $C_k(\mathbf{A})$.

Lemma 2.7.3 *Let the algebra* $\mathbf{A}$ *have more than one element. For any monoid equation* $w = w'$ *and any* $k > 0$*, if* $C_k(\mathbf{A})$ *satisfies* $w = w'$*, then the dual of* End k *satisfies* $w = w'$.

Proposition 2.7.4 *If* $\mathbf{A}$ *is an algebra with* $|A| \geq 2$ *and* $C_k(\mathbf{A})$ *satisfies* $u^r = u^s$, *then* $r - s$ *is a common multiple of all the positive integers* $\leq k$.

The proof follows from Lemma 2.7.3 and the fact that End k contains permutations of every order $\leq k$.

Corollary 2.7.5 *If* $\mathbf{A}$ *is an algebra with* $|A| \geq 2$ *and* $C_k(\mathbf{A})$ *satisfies* $u^r = u^s$, *then* $r = s$, *or* $r \geq k - 1$ *and* $s \geq k - 1$.

The corollary follows Proposition 2.7.3 and the fact that End k contains the function f with $f(0) = 1, f(1) = 2, \ldots, f(k-2) = f(k-1) = k-1$.

Connection to hyperidentities.

The connection to hyperidentities may be pointed out by the following example. We consider the monoid equation $u^r = u^s$ which is satisfied by some $C_k(\mathbf{A})$. Then for every component $i = 0, \ldots, k-1$ we can derive a hyperidentity of the form $T_i^r(x_0, \ldots, x_{k-1}) = T_i^s(x_0, \ldots, x_{k-1})$ where T_i is a hyperterm (which corresponds to the ith component of u).

Let L be a language. The terms of L are defined in the usual recursive manner as follows:

(i) Each v_i (variable of L) is a term.

(ii) If $t_0, \ldots, t_{k-1}$ are terms, then ${}_jF_i^k t_0, \ldots, t_{k-1}$ is a term for every $i < k$ and every $j \geq 0$, where ${}_jF_i^k$ is any k-ary operation symbol.

Whenever we confine ourselves to $j = 0$ we write only F_i^k.

$[w]_i^k$ is now defined by recursion on the length of w as follows:

(iii) $[e]_i^k = v_i$.

(iv) $[u_j w]_i^k = {}_jF_i^k[w]_0^k \ldots [w]_{k-1}^k$.

If we take w to be e, $[u_j]_i^k$ is defined by condition (iv) to be ${}_jF_i^k v_0 \ldots v_{k-1}$. The monoid words w contain only the variable $u = u_0$, then $[w]_i^k$ will contain only the operation symbols ${}_0F_i^k = F_i^k$. In this case, we can define

(v) $[e]_i^k = v_i$.

(vi) $[uw]_i^k = F_i^k[w]_0^k \ldots [w]_{k-1}^k$.

The following Lemma 2.7.6 is a generalization of condition (iv). If we take $v = u_j$ in the lemma, we come back to condition (iv). In general, if s is a term whose variables are among $v_0, \ldots, v_{k-1}$, and $t_0, \ldots, t_{k-1}$ are any terms, then $s(t_0, \ldots, t_{k-1})$ denotes the result of simultaneously substituting t_i for v_i in s $(i = 1, \ldots, k-1)$. The following two properties completely describe the substitution.

(vii) $v_i(t_0, \ldots, t_{k-1}) = t_i \quad (i < k)$.

(viii) For any m-ary operation symbol F we have

$$(Fs_0 \ldots s_{m-1})(t_0, \ldots, t_{k-1}) = Fs_0(t_0, \ldots, t_{k-1}) \ldots s_{m-1}(t_0, \ldots, t_{k-1}).$$

Lemma 2.7.6 *For any monoid-theoretic terms v and w we have*

$$[vw]_i^k = [v]_i^k \left[[w]_0^k, \ldots, [w]_{k-1}^k\right].$$

Proof By induction on the length of v. If v is a single letter, then the statement reduces to condition (iv). Let, $v = u_j v'$. Then we have

$$\begin{aligned}
w[vw]_i^k &= [u_j v' x]_i^k =_j F_i^k [v'w]_0^k \ldots [v'w]_{k-1}^k \\
&= {}_jF_i^k [v']_0^k \left[[w]_0^k, \ldots, [w]_{k-1}^k\right] \ldots [v']_{k-1}^k \left[[w]_0^k, \ldots, [w]_{k-1}^k\right] && (2)\\
&= \left[{}_jF_i^k [v']_0^k \ldots [v']_{k-1}^k\right] \left[[w]_0^k, \ldots, [w]_{k-1}^k\right] && (3)\\
&= [u_j v']_i^k \left[[w]_0^k, \ldots, [w]_{k-1}^i\right] && (4)\\
&= [v]_i^k \left[[w]_0^k, \ldots, [w]_{k-1}^i\right]. && (5)
\end{aligned}$$

Here equation (2) holds by induction, equation (3) holds by condition (viii) in the definition of substitution, equation (4) holds by condition (iv) of the recursive definition of $[w]_i^k$, and equation (5) holds by the given factorization of v.

In the following we consider words w, w' in two letters f and g, and self-maps $\bar{f}$ and $\bar{g}$ of a set A. If $w = h_0 h_1 h_2 \ldots$ (with each h_i either f or g), then $\bar{w}$ will denote the self-map $\overline{h_0} \cdot \overline{h_1} \cdot \overline{h_2} \ldots$ defined by composition of functions. (If w is the empty word, then $\bar{w}$ denotes the identity function).

Lemma 2.7.7 *Suppose that A is an infinite set, and that $\bar{f}, \bar{g} : A \to A$ are injective maps, with $\bar{f}(A) \cap \bar{g}(A)$ a finite set. For any words w and w' with $\bar{w}(A) = \bar{w}'(A)$, we have $w = w'$. Consequently, the maps $\bar{f}$ and $\bar{g}$ are free generators of a free monoid of self-maps of A.*

Let V be a variety. Denote by $F(k)$ the free algebra in V on k generators and by $C_k(V)$ the clone of $F(k)$.

Theorem 2.7.8 *Suppose that V has only finitely many constant operations, and that $F = F(k)$ contains a subalgebra isomorphic to $F = F(2k)$. Then $C_k(V)$ satisfies no nontrivial monoid equation.*

Proof By Lemma 2.7.1 we may consider the theorem for the monoid End F instead of $C_k(V)$. It will suffice to find two functions $\bar{f}$ and $\bar{g}$ in this monoid that satisfy the conditions of Lemma 2.7.7. Let $\mathbf{B}$ be a subalgebra of F which is isomorphic to $F(2k)$. We let $\bar{f}$ map the k free generators of F bijectively to the first k free generators of $\mathbf{B}$, and let $\bar{g}$ map the k free generators of f bijectively to the last k free generators of $\mathbf{B}$. Obviously an element in the ranges of both $\bar{f}$ and $\bar{g}$ must be a constant. As there are only finitely many constants, and as $\bar{f}$ and $\bar{g}$ are clearly injective, Lemma 2.7.7 holds. Therefore End F contains a generic monoid.

Let $\mathbf{A}_k$ denote a primal algebra of k elements. The next theorem shows that $C_k(\mathbf{A}_k)$ and $C_k(\mathbf{A}_{k+1})$ do not satisfy the same equations (and hence that the varieties $\mathbf{HSPA}_k$ and $\mathbf{HSPA}_{k+1}$ do not satisfy the same hyperidentities of the type $[w \approx w']^k$).

Theorem 2.7.9 *Let $n = k^k$ and κ the least common multiple of all positive integers $\leq n$. Then $C_k(\mathbf{A}_k)$ satisfies $u^{n-1} = u^{n-1+\kappa}$, but $C_k(\mathbf{A}_{k+1})$ does not.*

Corollary 2.7.10 *If $\mathbf{A}$ is a finite algebra, then $C_k(\mathbf{A})$ satisfies*

$$u_0^{n-1}u_1u_0^{n-2} = u_0^{n-1}u_1u_0^{n-2+\kappa}$$

for $n = k^k$, and κ the least common multiple of all positive integers $\leq n$.

Theorem 2.7.11 *If $\mathbf{A}$ is a non-trivial vector space over a field K, then*

$$C_k(\mathbf{A}) \simeq M_k(K),$$

where M_k is the monoid of $k \times k$ matrices over K.

The order p of any $k \times k$ Boolean matrix T divides the least common multiple of the integers $2, 3, \ldots, k$, and the index n of T is $\leq (k-1)^2 + 1$. From these results follows

Theorem 2.7.12 *Let p be the least common multiple of all positive integers $\leq k$, and let $n = (k-1)^2 + 1$. For the monoid M_k of Boolean matrices, we have that M_k satisfies*

$$u^n \approx u^{n+p}.$$

Finally we quote results about some lattice varieties.

Theorem 2.7.13 *The variety of modular lattices M satisfies the hyperidentity*

$$[u^5 \approx u^{11}]^3.$$

Theorem 2.7.14 *The variety of distributive lattices D satisfies the hyperidentity*

$$[u^{10} \approx u^{70}]^4.$$

Part 3

Hyperidentities and clone equations

3.1 Clones of functions

Most algebraic structures are connected to certain sets of functions. Bijective functions give rise to the concept of permutation groups on A. If one abstracts from permutation groups one gets the concept of an abstract group. From semigroups of functions one proceeds to the concept of abstract semigroups.

This conceptual development has not fully reached the sets of operations on a set A. We consider the algebraic structure of sets of operations (i.e. functions in several variables on A) concerning composition and manipulation of variables and use the following definition of a clone (closed set) of operations.

Definition 3.1.1 Let H be a set of functions on A. The *clone* $\mathbf{H} = (H, *, \zeta, \tau\Delta, e)$ is an algebra of type (2,1,1,1,0), where the operations are defined in the following way:

(1) $(f * g)(x_1, \ldots, x_m, x_{m+1}, \ldots, x_{m+n-1}) = f(g(x_1, \ldots, x_{m+n}), x_{m+1}, \ldots, x_{m+n-1})$ for an n-ary function f and an m-ary function g;

(2) $(\zeta f)(x_1, \ldots, x_n) = f(x_2, \ldots, x_n, x_1)$ for an n-ary function $f, n > 1$; $(\zeta f)(x_1) = f(x_1)$ for any 1-ary function f;

(3) $(\tau f)(x_1, \ldots, x_n) = f(x_2, x_1, x_3, \ldots, x_n)$ for an n-ary function $f, n > 1$; $(\tau f)(x_1) = f(x_1)$;

(4) $(\Delta f)(x_1, \ldots, x_{n-1}) = f(x_1, x_1, x_2, \ldots, x_{n-1})$ for an n-ary function $f, n > 1$; $(\Delta f)(x_1) = f(x_1)$;

(5) $e(x_1, x_2) = x_1$.

We like to remark that any projection e_i^n, $e_i^n(x_1, \ldots, x_n) = x_i$, is generated and hence contained in any clone. The clone of all functions on the set A is denoted by O_A.

Definition 3.1.2 Let $\mathbf{A} = (A, \Omega)$ be an algebra. The *clone* $T(\mathbf{A})$ of the *term function* of A is the subclone of O_A which is generated by the operations of $\mathbf{A}$. The clone $P(\mathbf{A})$ of *polynomial functions* of $\mathbf{A}$ is the subclone of O_A which is generated by the operations of $\mathbf{A}$ and the constant functions c_a^n, $a \in A$, $n \in \mathbb{N}$, where

$$c_a^n(x_1, \ldots, x_n) = a.$$

There are several approaches to define "abstract" clones without relying on operations and giving a presentation by equations. This includes I.G. Rosenberg, Malcev's preiterative algebra, Preprint Montreal 1976; I.G. Rosenberg, Malcev algebras for universal algebra term, Preprint Montreal 1989; and the work of Trkhimenko 1979. Furthermore, a solution by W. Taylor is fulfilling all requirements within the framework of categories.

In some applications, especially in other branches of mathematics, it is very useful to have the concept of n-clone.

Definition 3.1.3 Let H be a set of n-place functions on a set A. Then an *n-clone* $\mathbf{H} = (H, \circ, \zeta, \tau, \Delta, e)$ is an algebra of type (2,1,1,1,0), where the operations are defined in the following way

(1) $(f \circ g)(x_1, \dots, x_n) = f(g(x_1, \dots, x_n), x_2, \dots, x_n)$;

(2) $(\zeta f)(x_1, \dots, x_n) = f(x_2, \dots, x_n, x_1)$;

(3) $(\tau f)(x_1, \dots, x_n) = f(x_2, x_1, x_3, \dots, x_n)$;

(4) $(\Delta f)(x_1, \dots, x_n) = f(x_1, x_1, x_2, \dots, x_{n-1})$;

(5) $e(x_1, \dots, x_n) = x_1$.

3.2 Clone equations and hyperidentities

In our approach to the notion of an "abstract" clone we introduce the "clone with arity".

Definition 3.2.1 Let H be a set of functions on A. The algebra $\mathbf{H} = (H; *, \xi, \tau, \Delta, e, \Box_n$ $(n \in \mathbb{N}))$ of type $(2, 1, 1, 1, 0, 1, \dots)$ is called a *clone with arity*, where the operations ξ, τ, Δ, e are defined as in 3.1.1 and the operations $\Box_n$ are defined by

(6) $(\Box_n f)(x_1, \dots, x_n) = f(x_1, \dots, x_k)$ if f is a k-ary function with $k \le n$, and
$(\Box_n f)(x_1, \dots, x_n) = f(x_1, \dots, x_1 x_2, \dots, x_n)$ if f is a k-ary function with $k > n$.

Definition 3.2.1 is equivalent to Definition 3.1.1 in the sense that every function $\Box_n f$ can be generated by $*, \xi, \tau, e$. If f is a k-ary function with $k \le n$ we consider $e_1^m * f$ which gives

$$(e_1^m * f)(x_1, \dots, x_{m-k+1} = e_i^m(f((x_1, \dots, x_k), x_{k+1}, \dots, x_{m-k+1})).$$

For $m = n + k - 1$ we have $\Box_n f = e_1^m * f$. If f is a k-ary function with $k > n$, we apply Δ $(k - n)$ times and we get $\Box_n f = \Delta^{k-n} f$.

Equations which hold for a variety of clones are called clone equations. We denote the variables in these equations by $X, Y, Z, X_1, X_2, X_3, \dots$. We are considering the following example $\Delta(\Box_2 X) = e$. Obviously this clone equation holds for every term function of a lattice.

This clone equation $\Delta(\Box_2 X) = e$ yields the hyperidentity $F(x, x) = x$ which holds for any lattice. On the other hand we get from the hyperidentity

$$G(G(x, y, z), y, z) = G(x, y, z)$$

the clone equation

$$\square_3 X * \square_3 X = \square_3 X.$$

3.2.2 Principle of clone equations

1) Every clone equation for a clone $T(V)$ of term functions can be translated into a hyperidentity or a set of hyperidentities for V.

2) Every hyperidentity for V can be translated into a clone equation for the arity operators $\square_n$.

There is an important step between the observation that the unary functions on a set obey the laws of a semigroup and the abstract notion of a semigroup itself given by axioms. This step cannot be done for clones in the same way. The infinite set of axioms which are suggested in the following list are presented only for a further discussion of this problem.

Axioms

A1 $(\square_n X^* \square_m Y)(e_1^{m+n-1}, \ldots, e_{m+n-1}^{m+n-1})$
$= \square_n X(\square_m Y(e_m^{m+n-1}, \ldots, e_m^{m+n-1}), e_{m+1}^{m+n-1}, \ldots, e_{m+1}^{m+n-1})$

A2 $(\xi(\square_n X))(e_1^n, \ldots, e_n^n) = \square_n X(e_2^n, \ldots, e_n^n, e_1^n), \quad n \in \mathbb{N}\backslash\{1\}, \quad n, m \in \mathbb{N}$

A3 $(\tau(\square_n X))(e_1^n, \ldots, e_n^n) = \square_n X(e_2^n, e_1^n, e_3^n, \ldots, e_n^n), \quad n \in \mathbb{N}\backslash\{1\}$

A4 $(\Delta(\square_n X))(e_1^{n-1}, \ldots, e_{n-1}^{n-1}) = \square_n X(e_1^n, e_1^n, e_2^n, \ldots, e_{n-1}^n), \quad n \in \mathbb{N}\backslash\{1\}$

A5 $\xi(\square_1 X) = \tau(\square_1 X) = \Delta(\square_1 X) = \square_1 X$

A6 $\square_m(\square_n X)(e_1^m, \ldots, e_m^m) = \square_n X(e_1^m, \ldots, e_n^m), \quad n \in \mathbb{N} \quad \text{for } n \leq m$

A7 $\square_m(\square_n X)(e_1^m, \ldots e_n^m) = \square_n X(e_1^m, \ldots, e_m^m, e_m^m, \ldots, e_m^m), \quad n, m \in \mathbb{N} \quad \text{for } n > m$

A8 $\square_n X(e_1^n, \ldots, e_n^n) = \square_n X, \quad n \in \mathbb{N}$

A9 $e_1^2(\square_n X), e_2^2) = \square_n X, \quad n \in \mathbb{N}.$

Remark 3.2.3 We have used a lot of abbreviations for these axioms. e_1^n for instance is generated from the nullary operation e contained in every clone. For the construction one uses the operation $\square_n$ and a permutation which effects the exchange of the 1st place with the ith place. Also the various kinds of compositions like $\square_n X(e_1^n, \ldots, e_1^n)$ are thought of as an abbrevation. The description of this by the substitution $*$, the permutations ξ, τ and the other operations is too lengthy. Of course an axiomatization of clones should fulfill similar requirements as abstract semigroups do for semigroups of functions.

Theorem 3.2.4 *Let* $\mathbf{C} = (C; *, \xi, \tau, \Delta, e, \square_n (n \in \mathbb{N}))$ *be an algebra fulfilling the above set of axioms.* $\mathbf{C}$ *is isomorphic to a clone of functions if the following conditions hold:*

(a) *There is a least natural number n such that $\Box_n X = X$ for every $X \in C$;*

(b) $e_i^i \neq e_j^j$ *for all $i, j \in \mathbb{N}$ with $i \neq j$.*

Proof To construct a clone of functions we take as a carrier set $A = \{Z | Z \in C\}$. To every $X \in C$ with a least natural number n for $\Box_n X = X$ we define a function $f_X : A^n \to A$ by $f_X(Z_1, \ldots, Z_n) = X(\Box_n Z_1, \ldots, \Box_n Z_n)$. Furthermore we consider $F_C = \{f_X | X \in C\}$ as a clone of functions on the set A. The map $\alpha : C \to F_C$ is defined by $\alpha(X) = f_X$, and we have to show that α is a clone isomorphism. α is injective because for $f_X = f_Y$ we have

$$X = X(e_1^n, \ldots, e_n^n) = f_X(e_1^n, \ldots, e_n^n) = f_Y(e_1^n, \ldots, e_n^n) = Y(e_1^n, \ldots e_n^n) = Y.$$

α is compatible with $*$ if $f_X * f_Y = f_{X*Y}$ holds. If f_X is m-place we have

$$\begin{aligned}
& f_X * f_Y(e_1^{m+n-1}, \ldots, e_{m+n-1}^{m+n-1}) & \\
= \; & f_X(f_Y(e_1^{m+n-1}, \ldots, e_m^{m+n-1}), e_{m+1}^{m+n-1}, \ldots, e_{m+n-1}^{m+n-1}) & (F_C \text{ is a clone of functions}) \\
= \; & \Box_n X(\Box_m Y(e_1^{m+n-1}, \ldots, e_m^{m+n-1}), e_{m+1}^{m+n-1}, \ldots, e_{m+n-1}^{m+n-1}) & (\text{by definition}) \\
= \; & \Box_n X * \Box_m Y(e_1^{m+n-1}, \ldots, e_{m+n-1}^{m+n-1}) & (\text{by axiom scheme A1}) \\
= \; & X * Y(e_1^{m+n-1}, \ldots, e_{m+n-1}^{m+n-1}) & (\text{by hypothesis}) \\
= \; & f_{X*Y}(e_1^{m+n-1}, \ldots, e_{m+n-1}^{m+n-1}) & (\text{by definition})
\end{aligned}$$

We proved $f_X * f_Y = f_{X*Y}$ only for some special elements of A. But it is easy to see that it holds also for all the other elements if one uses composition and (A1). The other operations like ξ, τ, etc., we get in a similar way.

Comment It is impossible to give an axiomatization of a clone with the property that the models of the axioms are precisely the clones isomorphic to concrete clones. The reason for this is that condition (a) of Theorem 3.2.4 is not preserved under ultraproduct constructions.

Remark 3.2.5 Every subclone D of a clone of functions on a set is again a clone of functions. This holds because the elements of D as a subset of C are functions which are closed under the operations of a clone.

Remark 3.2.6 A countable power of a clone D of finitary functions gives rise to a clone which contains infinitary functions. For this consider $C^\infty = \{(f_1, f_2, f_3, \ldots) | f_i \in C, \; i = 1, 2, 3, \ldots\}$. The sequence $(e_1^1, e_1^2, e_1^3, \ldots,)$ with $e_1^i(x_1, \ldots, x_i) = x_1$, $i = 1, 2, 3, \ldots$, can be considered as a function but of infinite arity.

Remark 3.2.7 Not every homomorphic image of a clone of functions is again isomorphic to a clone of functions. Already A.I. Malcev has found the following clone $M = \{c, a, a^2, a^3, \ldots\}$, where one defines $a^k * a^l = a^{k+l}$, $c =$ unit, $\xi a^l = a^l$, $\tau a^l = a^l$ and $\Delta a^l = a^{l+1}$ for $k, l, n \in \mathbb{N}$. This clone comes up when a clone of functions is factorized by the congruence relation κ, defined by $(f, g) \in \kappa$ if and only if the arity of f is equal to the arity of g. This equivalence relation is compatible with the operations $*, \xi, \tau, \Delta$ and hence a congruence relation (in the sense of universal algebra). Whereas M is isomorphic to the clone of functions of the one-element set, this is not the case for $H = \{c\}$ which is a homomorphic image of M, yet is not isomorphic to the clone of functions of any set.

Finally we would like to pose the following problem and ask for an improvement of the solution presented above.

Problem 3.2.8

1) Describe a variety C which (properly) contains the clones of functions as algebras.
2) The variety C should be axiomatized by a few identities or schemes of identities.
3) The part of identities of C which correspond to hyperidentities should be clearly presented.
4) The subclass of algebras of C which correspond to clones of functions should be clearly presented.

3.3 Varieties generated by clones of polynomial functions

In the following we consider the clone $P(\mathbf{A})$ of polynomial functions of an algebra $\mathbf{A}$ but the results also hold for the clone $T(\mathbf{A})$ of term functions.

Proposition 3.3.1 *Let $f : A \to B$ be a surjective homomorphism from an algebra $\mathbf{A}$ onto an algebra $\mathbf{B}$. Then there is a surjective clone homomorphism from $P(\mathbf{A})$ to $P(\mathbf{B})$ (or respectively from $T(\mathbf{A})$ to $T(\mathbf{B})$).*

Proof We define $\alpha : P(\mathbf{A}) \to P(\mathbf{B})$ recursively by $\alpha(c_a^n) = c_{f(a)}^n$ for every constant function c_a^n, and $\alpha(e_{i(A)}^n) = e_{i(B)}^n$ for every projection e_i^n, $e_i^n(x_1, \ldots, x_n) = x_i$, on A. Every polynomial function $\varphi \in P(\mathbf{A})$ has a representation by a word. Obviously we can extend the definition of α to φ using this word. But α does not depend on the choice of this word and hence is well defined. We also have that α is surjective because to every word of the polynomial algebra $\mathbf{B}[x_1, \ldots, x_n]$ there is a corresponding word of $\mathbf{A}[x_1, \ldots, x_n]$. Clearly α preserves the operations of a clone.

Proposition 3.3.2 *Let $\mathbf{A} = \prod_{i\in I} \mathbf{A}_i$ be the direct product of a family $\{\mathbf{A}_i | i \in I\}$ of algebras of the variety V. Then $P(\mathbf{A})$ is isomorphic to a subdirect product of $\{P(\mathbf{A}_i) | i \in I\}$.*

Proof We consider the projections $p_i : A \to A_i$ which are surjective homomorphisms from $\mathbf{A}$ onto $\mathbf{A}_i$. By Proposition 3.2.1 we have that $\alpha_i : P(\mathbf{A}) \to P(\mathbf{A}_i)$ are surjective clone homomorphisms. We consider $\prod_{i\in I} P(\mathbf{A}_i)$ and the function $\gamma : P(\mathbf{A}) \to \prod_{i\in I} P(\mathbf{A}_i)$ defined by $p_i(\gamma(\varphi)) = \alpha_i(\varphi)$. We have that γ is a clone homomorphism because of α_i. Let $\gamma(\varphi) = \gamma(\psi)$ for some $\varphi, \psi \in P_n(\mathbf{A})$. If $\varphi(a_1, \ldots, a_n) \neq \psi(a_1, \ldots, a_n)$ for some $a_1, \ldots, a_n \in A$, then $p_j(\varphi)(a_1, \ldots, a_n)) \neq p_j(\psi(a_1, \ldots, a_n))$ for some $j \in I$. We have $\alpha_j(\varphi) \neq a_j(\psi)$ and hence $\gamma(\varphi) \neq \gamma(\psi)$. It is clear that $\gamma(P(\mathbf{A}))$ is a subclone of $\prod_{i\in I}(\mathbf{A}_i)$.

Proposition 3.3.3 *Let $\mathbf{B}$ be a subalgebra of $\mathbf{A}$. Then $P(\mathbf{B})$ is isomorphic to a homomorphic image of a subclone of $P(\mathbf{A})$.*

Proof If $\mathbf{B}$ is a subalgebra of $\mathbf{A}$ we consider the clone of polynomial functions $P = \{\varphi | \varphi \in P(\mathbf{A}), \varphi(x_1, \ldots, x_n) \in \mathbf{B}[x_1, \ldots, x_n]\}$, where $\mathbf{B}[x_1, \ldots, x_n]$ denotes the polynomial algebra of $\mathbf{B}$. It is clear that $\alpha : P \to P(\mathbf{B})$ defined by $\alpha(\varphi) = \varphi|_B$ is a surjective clone homomorphism. On the other hand P is a subclone of $P(\mathbf{A})$.

All the above results also hold for the clones $T(\mathbf{A})$ of term functions.

Theorem 3.3.4 *Let V be a variety and $W = \mathbf{HSP}(T(V))$ the variety of clones generated by $T(V)$. If A is an algebra of V, then $T(\mathbf{A})$ is in the variety W of clones.*

An example of application is

Proposition 3.3.5 *Let V be a variety of lattices. V is the variety of distributive lattices if and only if $\Box_2 x \circ \Box_2 x = \Box_2 x$ $(F^2(x,y) = F(x,y))$ holds for the variety*

$$W = \mathbf{HSP}(T(V))$$

of clones.

3.4 Projection algebras

Let $\mathbf{A} = (A, \Omega)$ be an algebra of type $(n_1, \ldots, n_\delta, \ldots,)$. Then $\mathbf{P} = (A; \{e_1^i\}, i \in I)$ with $e_1^i(x_1, \ldots, x_{ni}) = x_1$, $I = \{x_1, \ldots, n_\delta, \ldots\}$, is a derived algebra from $\mathbf{A}$. If $\mathbf{A}$ has at least an n-ary operation with $n \geq 2$, then every projection is generated.

Definition 3.4.1 $\mathbf{P} = (A; \Omega)$ is called a *projection algebra* if every operation of $\mathbf{P}$ is a projection.

If we have to test whether a given hyperidentity $*$ holds for an algebra $\mathbf{A}$ we also have to test whether $*$ holds for the projection algebras $\mathbf{P}$.

Let V be a variety. Then the trivial subvariety defined by $x = y$ is a solid subvariety. We present here some results by Denecke, Lau, Pöschel and Schweigert on the phenomenon that there are numerous of varieties which have only trivial solid subvarieties [DLPSch 91].

Proposition 3.4.2 *Every congruence modular variety has only a trivial solid subvariety. (In particular, if it is not trivial, then it is not solid).*

Proof Let V be a non-trivial solid variety. Then V contains all projection algebras of the same type. The congruence lattice of a projection algebra $\mathbf{P} = (A; e^i, i \in I)$ is the lattice of all equivalence relations on A. For $|A| \geq 4$ this lattice is not modular.

Proposition 3.4.3 *Let $\mathbf{A}$ be an algebra such that $T(\mathbf{A})$ contains a constant function. Then $\mathbf{HSP}(\mathbf{A})$ has only a trivial solid subvariety.*

Proof For this constant term function $t(x_1, \ldots, x_n) = a$ we would have $d(x) = t(x, \ldots, x) = a$, and hence $d(x) = d(y)$ which is never satisfied in a non-trivial projection algebra.

Instead of testing a hyperidentity by projection algebras one can also take a syntactical point of view. Every valid hyperidentity can be derived from hyperidentities of projection algebras. Hence we like to present a hyperidentity basis for projection algebras. Misusing the language we call the variety of type τ generated by the projection algebra of type τ the variety of projection algebras of type τ (compare 3.2.5 – 3.2.7).

Notation 3.4.4 For $n \geq 1$ let $\Gamma\langle n\rangle$ consist of the n-ary hyperidentities

$$F(x, \ldots, x) = x,$$

$$F(F(x_{11}, \ldots, x_{1n}), \ldots, F(x_{n1}, \ldots, x_{nn})) = F(x_{11}, \ldots, x_{nn}).$$

Theorem 3.4.5 *For $n \geq 1, \Gamma\langle n\rangle$ is a basis for all hyperidentities of type $\langle n\rangle$ for the variety of projection algebras of type $\langle 2\rangle$.*

Notation 3.4.6 By $M(n, m)$ denote the hyperidentity

$$F(G(x_{11}, \ldots, x_{1m}), \ldots, G(x_{n1}, \ldots, x_{nm})) = G(F(x_{11}, \ldots, x_{n1}), \ldots, F(x_{1m}, \ldots, x_{nm})).$$

Theorem 3.4.7 $\Gamma\langle n\rangle \cup \{M(n, n)\}$ *is a basis for all hyperidentities of type $(n, n, n, \ldots)$ of the variety of projection algebras.*

Example $F(x, G(y, z)) = G(F(x, y), F(x, z))$ is a valid hyperidentity (it holds for the variety of lattices). Consider $\Gamma\langle 2\rangle \cup \{M(2, 2)\}$

$$F(x, G(y, z)) =_{\Gamma(2)} F(G(x, x), G(y, z)) =_{M(2,2)} G(F(x, y), F(y, z)).$$

Fact If a hyperidentity holds for some variety, then it can also be derived by the hyperidentities which holds for the variety of projections algebras.

Theorem 3.4.8 *Let $\Gamma = \bigcup_{n \geq 1} \Gamma\langle n\rangle \cup \bigcup_{m,n \geq 1} \{M(n, m)\}$. Then Γ is a countably infinite basis for the hyperidentities of any type for the variety of projection algebras.*

Theorem 3.4.9 *The hyperidentities of any type for the variety of projection algebras are non-finitely based.*

For all these results we have omitted the proofs. Knoebel has shown that the variety RB of rectangular bands is generated by the projection algebras $(A; e_1^2)$ and $(A; e_2^2)$. Therefore the variety RB of rectangular bands satisfies a hyperidentity $S = T$ if and only if $S = T$ can be derived from the hyperidentities of the variety of projection algebras of type $\langle 2\rangle$. Obviously RB is the minimal (non-trivial) solid variety of type $\langle 2\rangle$.

In section 3.5 we are presenting the results of S. Wismath on hyperidentities of the variety RB of regular bands. The above hyperidentity bases are given and the theorem that the hyperidentities of any type are not finitely based is proved.

Remark 3.4.10 Let $\tau = (n_0, \ldots, n_\delta)$ be a type with $n_0 > 1, \ldots, n_\delta > 1, \ldots$. There exists only one minimal solid variety of type τ, namely the variety of projection algebras of this type. These varieties are also described in the work of Plonka [Plonka 66].

3.5 Hyperidentity bases for rectangular bands

The results of this section are due to S. Wismath [Wismath 91] if not quoted otherwise.

We have already recognized that every hyperidentity $S = T$ of a variety V has also to hold for the variety of projection algebras of the same type.

Lemma 3.5.1 [Penner 84] *The variety RB of rectangular bands satisfies the hyperidentity $S = T$ if and only if $S = T$ can be derived from the hyperidentities of the variety of projection algebras of type $\langle 2\rangle$.*

Besides the notations of section 3.4 we use the following:

Notations 3.5.2

$H(V)$	set of all hyperidentities of any type satisfied by the variety V.
$H^m(V)$	set of all hyperidentities in $H(V)$ with at most m hypervariables.
$H_n(V)$	set of all hyperidentities in $H(V)$ with at most n variables.
$H(V)\langle\underline{n}\rangle$	set of all hyperidentities in $H(V)$ with hypervariable of type $\langle n, n, n, \ldots, n\rangle$ (k-factors).
$H(V)\langle\underline{n}\rangle$	set of all hyperidentities in $H(V)$ with hypervariables of type $\langle n, n, n, \ldots\rangle$ (infinitely many factors).

Lemma 3.5.3 *The variety RB of rectangular bands satisfies* $\Gamma(n)$.

We have already seen that both hyperidentities hold for projection algebras. The second hyperidentity essentially says that variables in a type $\langle n\rangle$ hyperidentity for RB which are *nabp* (not accessible by projections) may be eliminated. This is perhaps more clearly seen in the equivalent set of $n+1$ hyperidentities:

$$F(x, \ldots, x) = x,$$
$$F(F(x_{11}, \ldots, x_{1n}), x_2, \ldots, x_n) = F(x_{11}, x_2, \ldots, x_n),$$
$$F(F(x_1, \ldots, x_{n-1}), F(x_{n1}, \ldots, x_{nn})) = F(x_1, \ldots, x_{n-1}, x_{nn}).$$

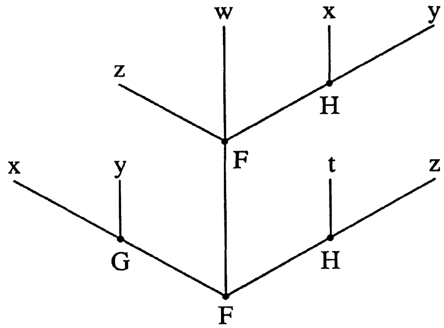
For the case $\langle n\rangle = \langle 2\rangle$ the associative hyperidentity $F(x, F(y, z)) = F(F(x, y), z)$ can be derived from $\Gamma\langle 2\rangle$.

Theorem 3.5.4 *For $n \geq 1, \Gamma\langle n\rangle$ is a basis for all hyperidentities of type $\langle n\rangle$ of the variety RB of rectangular bands.*

It will sometimes be useful to consider the *formation tree* corresponding to hyperterms.

Example $(G(x, y), F(z, u, H(x, y)), H(t, z))$

Formation tree



In particular, if x is a variable of a hyperterm, we record the sequence of hypervariables and turnings in the path in the tree from the root to the variable x, in the string of x. In our example the second occurrence of x is recorded as $(F2, F3, H1, x)$. The height of a variable y is the number of not necessarily distinct hypervariables in its string. The height of a hyperterm is the maximum of the heights of its variables. The above hyperterm is of height 3.

Proof of Theorem 3.5.4 Let $P = Q$ be any hyperidentity satisfied by RB. If P and Q both consist only of a single variable x_i, then $P = Q$ must be trivial. Thus we will assume that at least one of P or Q involves at least one occurrence of the hypervariable F. Now there is a unique hyperterm P^* (Q^*) of height 1, such that $\Gamma\langle n\rangle \vdash P = P^*$ ($Q = Q^*$). For if P involves no occurrences of F, use the idempotent hyperidentity from $\Gamma\langle n\rangle$ to introduce one occurrence of F; if P involves more than one occurrence of F, use the two hyperidentities in $\Gamma\langle n\rangle$ to eliminate all but one occurrence of F.

Since $RB \models \Gamma\langle n\rangle$, we have $RB \models P^* = Q^*$, and $P^* = Q^*$ must model projections. But by definition all variables in $P^* = Q^*$ are abp, so $P^* = Q^*$ must be trivial. Therefore $\Gamma\langle n\rangle \vdash P = Q$, as required.

We will prove that $\Gamma\langle n\rangle$ is a basis for hyperidentities of type $\langle n, n\rangle$ of RB by means of two lemmas. The first one deals with the special case of hyperidentities in which all variables are abp, the second one shows how any hyperidentity may be reduced to one of this special kind using $\Gamma\langle n\rangle$.

Lemma 3.5.5 *If $P = Q$ is a hyperidentity of type $\langle n\rangle$ for the variety of projection algebras and has all variables accessible by projection, then*

$$\{F(x, \ldots, x) = x,\ M(n, n)\} \vdash P = Q.$$

Proof If $P = Q$ has the form $F(-) = x$, then the condition that all variables are abp ensures that all variables in $F(-)$ are x's, so that $P = Q$ is a consequence of idempotence. Hence we may now assume that $P = Q$ has the form $F(\ ) = G(\ )$. Again by the abp condition, after the first occurrences of F at the root of P, there can be no further F's in P. Moreover, if $F \neq G$, we may assume that every branch in P contains exactly one occurrence of G, since more than one G would lead to variables nabp, while if a branch has no G's we can always inflate the final variable on the branch, x say, into $G(x, \ldots, x)$ using the idempotent hyperidentity.

Using these observations, we proceed by induction on the height of $P = Q$. Any hyperidentity of height 1 meeting these conditions must be trivial, so we begin with height 2. Then $P = Q$ would look like

$$F(H_1(\bar{x}_1), \ldots, H_n(\bar{x}_n)) = G(K_1(\bar{y}_1), \ldots, K_n(\bar{y}_n)),$$

where $\bar{x}_1$ and $\bar{y}_i$, $1 \leq i \leq n$, represent n-tuples of variables. If F and G are the same hypervariables, we substitute for F the n-ary projection terms, to obtain n new hyperidentities $H_i(\bar{x}_i) = K_i(\bar{x}_i)$ of height 1, which must then be trivial. Thus $P = Q$ is trivial in this case. If $F \neq G$, the observations above show that we must have $H_i = G$ and $K_i = F$, for all $1 \leq i \leq n$, that is, $P = Q$ is actually $M(n, n)$.

Now consider $P = Q$ of height $k > 2$. If Q also has the form $F(-)$, we use the n projection terms to reduce to n hyperidentities of height $k - 1$ with the same properties. Then, $P = Q$ is a consequence of these, so by induction, $M(n, n)$ and $F(x, \ldots, x) = x$ yield $P = Q$. So we now suppose Q has the form $G(-)$, where $G \neq F$. We give a procedure for forming a new hyperterm P^* from P. As above, every branch of P must contain exactly one occurrence of G. For each such branch, count the number of hypervariables other than G occurring on the path from F to G. Choose any such G for which this number is maximal, say p. Now go back along the branch of this G to the previous hypervariable, say H. Each branch coming out of H must contain an occurrence of G, and by maximality of p these occurrences must also be at height p. So this part of P looks like $H(G(-), \ldots, G(-))$, and we can use the medial identity to change it to $G(H(-), \ldots, H(-))$. In this new identity, we repeat this process, first with any remaining G's at height p, then with G's at lower height. Eventually we reach a new hyperterm P^* of the form $G(-)$, such that $M(n, n) \vdash P = P^*$. Now then hyperidentity $P^* = Q$ still models projections and has all variables abp, and it has the form $G(\ ) = G(\ )$, so by the earlier case it is a consequence of $M(n, n)$ and idempotence. Thus$M(n, n)$ and idempotence yield $P = Q$, as required.

Lemma 3.5.6 *For any hyperterm P there is a hyperterm P^* with no variables which are not accessible by projection* (napb) *such that* $\Gamma\langle n \rangle \vdash P = P^*$.

Proof Obviously if P has no variables nabp, we may take P^* to be P. We show how the hyperidentities in $\Gamma\langle n \rangle$ must be used to eliminate any variable x nabp in P. For any such variable x, there is a hypervariable F and indices $i \neq j$ $(1 \leq i, j \leq n)$ such that the path form the root of P to x involves first F_i, then F_j.

If the two occurrences of F are adjacent, then a part of P looks like

$$-F(-, F(-, R, -), -)-$$

where the second F occurs in the ith place of the first F, R is a hyperterm involving x which occurs in the jth place of the second F, and $-$ indicates other hyperterms in P. We use the idempotent hyperidentity to inflate so that all n entries in the first F have the form $F(-)$, then use the other hyperidentity in $\Gamma\langle n\rangle$ to reduce to

$$-F(-,\ldots,-,-,\ldots-),$$

thereby eliminating the nabp variable x.

If two occurrences of F are separated by one or more other operation symbols, say $G_1,\ldots,G_k$, $k\geq 1$, then P has the form

$$-F(-,\ldots,-,G_1(\ldots G_k(-,\ldots F(-,R-)-)\ldots)\ldots)-.$$

Here we again use idempotence to inflate so that the last hypervariable before the second F has all entries of the form $F(-)$; then use $M(n,n)$ to replace the part $G_k(F(-),\ldots, F(-R-),\ldots F(-))$ by $F(G(-),G(-),\ldots,G(\ ))$. This moves the second occurrence of F one step closer to the first. By repeating this process we eventually reach a stage where the two occurrences of F are adjacent, when the method from above may be used to eliminate x. In this way all nabp variables in P may be eliminated, giving us P^* as required.

Theorem 3.5.7 *$\Gamma\langle n\rangle$ forms a basis for the hyperidentities of type $\langle n,n,n,\ldots\rangle$ for the variety RB of rectangular bands.*

Corollary 3.5.8 *Let $\Gamma=\bigcup_{n\geq 1}\Gamma\langle n\rangle\cup\bigcup_{m,n\geq 1}\{M(n,m)\}$. Then Γ is a countably infinite basis for the hyperidentities of any type for the variety RB.*

Theorem 3.5.9 *The hyperidentities of any type of the variety RB are not finitely based.*

Proof We will prove that for any two positive integers m and n, there is a hyperidentity H such that RB satisfies H, but H is not a consequence of $H^m(RB)\cup H_n(RB)$. Take $k=\max\{m,n\}+1$. Define H to be the following hyperidentity, with one k-ary operation symbol F:

$$(F(x_1,x_2,\ldots,x_k),x_1,\ldots,x_1)=F(F(x_1,x_1,x_3,\ldots,x_k),x_1,\ldots,x_1)\quad (H).$$

Since H models projections, it is clear that $RB\models H$. Now define an algebra $\mathbf{A}=(A;f)$ as follows. Take

$$A=\{a_1,\ldots,a_k,a_1a_2,\ldots,a_{k-1}a_k\},$$

the free rectangular band on the k generators $a_1,\ldots,a_k$. f is k-ary, given by

$$f(x_1,\ldots,x_k)=\begin{cases} x_1x_2, & \text{if } \{x_1,\ldots,x_k\}=\{a_1,\ldots,a_k\}\\ x_1, & \text{otherwise}\end{cases}$$

Using f for the operation symbol in H leads to an identity which does not hold in $\mathbf{A}$, since the evaluation $x_1=a_i, 1\leq i\leq k$, in the identity leads to $a_1a_2=a_1$.

Therefore $\mathbf{A}$ does not satisfy H. However, we claim that $\mathbf{A}$ does not satisfy all the hyperidentities in $H^m(RB)\cup H_n(RB)$.

For if a hyperidentity involves at most n variables, or operation symbols all of arity $\leq m$, since $k > m, n$, it follows that the only **A**-terms used in the hyperidentity amount to projections. So in this case **A** satisfies the hyperidentity if and only if RB does.

Thus we see that H is a hyperidentity satisfied by RB, which is not a consequence of $H^m(RB) \cup H_n(RB)$. Therefore $H(RB)$ is not finitely based.

Theorem 3.5.10 [PadPen 82] *Let SL be the variety of semilattices. Then the following set $H_2^{(2)}$ of hyperidentities is a basis for all hyperidentities of type $\langle 2 \rangle$ for SL.*

$$\begin{aligned} H(2): \quad &(1) \quad F(x, F(y,z)) = F(F(x,y),z), \\ &(2) \quad F(x,x) = x, \\ &(3) \quad F(F(u,x), F(y,w)) = F(F(u,y), F(x,w)). \end{aligned}$$

Proof By the above hyperidentities we can present every hyperterm $T(x_1, \ldots, x_n)$ is a (normal) form $F \ldots F(F(x_a, x_{\pi(1)}), x_{\pi(2)}, \ldots, x_{\pi(t)}, x_b)$. Here we use the associativity (1) for F to put all hypervariables on the left hand side, the commutativity (3) of the variables $x_{\pi(i)}$ inside x_a and x_b, and the idempotency (2) for eliminating $x_{\pi(i)}$ if it appears twice. Inside we can order $\tau(1) < \tau(2) < \ldots < \pi(t)$, $t \leq n-2$. Let $T(x_1, \ldots, x_n) = S(x_1, \ldots, x_m)$ be a hyperidentity of type $\langle 2 \rangle$ which holds for SL. Then both sides can be presented in (normal) forms

$$F \ldots F(F(x_a, x_{\pi(1)}), \ldots, x_{\pi(n-2)}, x_b) = F \ldots F(F(x_c, x_{\mu(1)}), \ldots, x_{\mu(m-2)}, x_d).$$

If we hypersubstitute F by the first projection we have $x_a = x_c$ (and respectively by the second $x_b = x_d$). Now we put $x_a = x_{\pi(1)}$ and $x_c = x_{\mu(1)}$. In this way we show that both sides have to be formally equal. Hence every hyperidentity of type $\langle 2 \rangle$ is implied by $H(2)$.

Problem 3.5.11 Give an example of a variety V of type $\langle 2 \rangle$ such that the hyperidentities of V of type $\langle 2 \rangle$ are not finitely based and V is generated by an algebra $\mathbf{A} = (A; \circ)$ with A as small as possible.

3.6 Normal and regular hyperidentities

Definition 3.6.1 An identity $t_1 = t_2$ is *regular* if the sets of all variables occurring in t_1 and t_2 coincide.

Example $x \cdot y = y \cdot x \cdot y$.

Definition 3.6.2 An identity is said to be *trivializing* if it is of the form $x = y$ (where x, y are different variables) or $x_k = t(x_1, \ldots, x_n)$, where t is a term which is not a variable. An identity $t_1 = t_2$ is *normal* if it is not trivializing.

It is now obvious how we have to define normal and regular hyperidentities. In this chapter we present the beautiful results of Ewa Graczyńska on this topic.

Notations

Σ	= set of identities of type τ
Mod (Σ)	= variety of type τ defined by Σ
$E(K)$:	= set of identities of a variety K
$N(K)$:	= set of normal identities of K
$R(K)$:	= set of regular identities of K
$H(K)$:	= set of hyperidentities of K
$NR(K)$:	$= N(K) \cap R(K)$

One can consider N, R, H as operators on classes of varieties. Let $L(\text{Mod}\, N(V))$ be the lattice of all subvarieties of the variety Mod $N(V)$. Then we have the following results:

Theorem 3.6.3 *If V is not a normal variety ($V \neq$ Mod $N(V)$), then the operator $N : L(V) \to L(\text{Mod}\, N(V))$ is an embedding of the lattice $L(V)$ of subvarieties of V into the lattice $L(\text{Mod}\, N(V))$.*

Proposition 3.6.4 *Let V and W be two varieties of type τ with $N(V) \subseteq E(W)$. If W is not normal, then $W \subseteq V$.*

Proof We present here a proof, proposed by N. Newrly, without using the representation theorem for algebras from Mod $N(V)$. Let f be a not normal identity from $E(W)$. If f is a trivial identity of the form $x = y$, where x and y are different variables, then obviously $W \subseteq V$. Assume that f is of the form $x_k = p(x_1, \ldots, x_n)$, where $1 \leq k \leq n$, and p is a proper term (i.e. not a variable). Consider the identity g of the form $x = p(x, \ldots, x)$, for a variable x. Obviously g is a consequence of f. If $E(V) = N(V)$, then obviously $W \subseteq V$. Otherwise, let e be a not normal identity of V. Assume that e is of the form $x_k = q(x_1, \ldots, x_m)$, with $1 \leq k \leq m$. If q is a variable, then we have $V = W$, because $x = p(x, \ldots, x) = p(y, \ldots, y) = y$ is a proof of $x = y$ from $N(V) \cup \{g\} \subseteq E(W)$. If q is a proper term, take $r = \max(k, m)$. Then

$$\begin{aligned} x_1 &= p(x_1, \ldots, x_{k-1} x_1, x_{k+1}, \ldots, x_r) \\ &= p(x_1, \ldots, x_{k-1}, q(x_1, \ldots, x_r), x_{k+1}, \ldots, x_r) \\ &= q(x_1, \ldots, x_r) \end{aligned}$$

is a proof of $x_1 = q(x_1, \ldots x_r)$ from the set $N(V) \cup \{f\}$. This gives that e is an identity of $E(W)$, and we conclude that $E(V) = E(N(V) \cup \{e\} \subseteq E(W)$, i.e. $W \subseteq V$.

Theorem 3.6.5 *If V is not a normal variety, then the lattice $T(\text{Mod}N(V))$ is isomorphic to the direct product of the lattice $L(V)$ and a two-element lattice.*

Proof Denote by $\mathbf{2} = (\{0, 1\}, \leq)$ the two-element lattice with $0 < 1$. Consider the mapping $h : L(V) \times \mathbf{2} \to L(\text{Mod}\, N(V))$ given by the rule $h(K, 0) = K,\ h(K, 1) =$ Mod $N(K)$, for $K \in L(V)$. Then for $K_1, K_2 \in L(V)$ the following holds:

$$h((K_1, 0) \cap (K_2, 0)) = h(K_1 \cap K_2, 0) = K_1 \cap K_2 = h(K_1, 0) \cap h(K_2, 0),$$

$$h((K_1, 1) \cap (K_2, 1)) = h(K_1 \cap K_2, 1) = \text{Mod}(N(K_1 \cap K_2)) = h((K_1, 1)) \cap h((K_2, 1))$$

by Theorem 3.6.3. For $K_1, K_2 \in L(V)$ we have the following inequalities in $L(\mathrm{Mod}\, N(V))$:

$$K_1 \cap K_2 \leq K_1 \cap \ \mathrm{Mod}\, N(K_2) \leq \ \mathrm{Mod}\, N(K_1) \cap \ \mathrm{Mod}\, N(K_2) = \ \mathrm{Mod}\, N(K_1 \cap K_2).$$

As $K_1 \cap K_2$ is not normal and $\mathrm{Mod}\, N(K_1 \cap K_2)$ covers $K_1 \cap K_2$ in the lattice $L(\mathrm{Mod}\, N(V))$, we get $K_1 \cap K_2 = K_1 \cap \ \mathrm{Mod}\, N(K_2)$. The variety $K_1 \cap \ \mathrm{Mod}\, N(K_2)$ is not normal.

We have:

$$\begin{aligned} h((K_1,0) \cap (K_2,1)) &= h(K_1 \cap K_2, 0)) = K_1 \cap K_2 \\ &= K_1 \cap \ \mathrm{Mod}\, N(K_2) = h((K_1,0)) \cap h((K_2,1)). \end{aligned}$$

Similarly one can show that h is a join-homomorphism.

Finally we have to show that h is surjective. Let $K' \in L(\mathrm{Mod}\, N(V))$. By Proposition 3.6.4 we conclude that there are only two possibilities:

$$\text{(i)} \quad K' \in L(V) \quad \text{or} \quad \text{(ii)} \quad E(K') = N(K').$$

In case (i) we obtain that $K' = H((K',0))$, where $K' \in L(V)$.

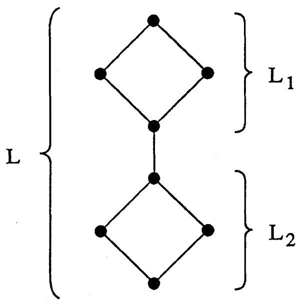
For the second case we consider the variety $K = K' \cap V$. We have $K \in L(V)$. As h is an endomorphism it follows that

$$\begin{aligned} h((K,1)) &= \mathrm{Mod}\, N(K) = \ \mathrm{Mod}\, N(K' \cap V) \\ &= \mathrm{Mod}\, N(K') \cap \ \mathrm{Mod}\, N(V) \\ &= \mathrm{Mod}\, N(K') = \ \mathrm{Mod}\, E(K') = K', \end{aligned}$$

Therefore h is an isomorphism.

A lattice L is called the *double* of its sublattice L_1 if $L_2 = L \backslash L_1$ is also a sublattice of L and there exists and isomorphism $f : L_1 \to L_2$ such that $x \leq f(x)$ in the lattice L.

Remark Theorem 3.6.5 says that the lattice $L(\mathrm{Mod}\, N(V))$ is a double of the lattice $L(V)$, for a given not normal variety V. But the double of a lattice L_1 need not to be isomorphic to the direct product of L_1 and a two-element lattice, as is shown by the following diagram:



The lattice L in the diagram is the double of the lattice L_1 but is not isomorphic to the direct product of L_1 and **2**.

Example Consider varieties of type (1) with one unary operation $f(x)$. Take the variety V defined by the identity $x = f(x)$. This variety is solid and not normal. Moreover, the variety Mod $N(V)$ is also solid. Let us note the following:

Theorem 3.6.6 *Assume that V is a variety of unary type τ and V is solid. Then* Mod $R(V)$ *is solid.*

Proof Assume that V is a solid variety of unary type τ. Let $p(x)$ and $q(y)$ be two polynomial symbols of type τ. If $p(x) = q(x)$ is a hyperidentity of V, then $p(x) = q(x)$ is a hyperidentity of Mod $R(V)$, because any hypersubstitution of a regular identity of unary type is regular. If $p(x) = q(y)$ is a hyperidentity of V, then $p(x) = q(x)$ is a hyperidentity of Mod $R(V)$. Thus $R(V) = H(R(V))$, i.e., Mod $R(V)$ is solid.

Example Consider the trivial variety T of unary type with two unary operations f and g. Then T is solid, but because $f(x) = g(x)$ has a hypersubstitution $x = g(x)$ and is not normal, Mod $N(T)$ is not solid.

Generally, the normal (or regular) part of a solid variety need not be solid. For example, take a trivial variety T of type (2). Then the identity

$$f(f(x,y),z) = f(f(x,z),y)$$

is a normal hyperidentity of T. But it is not a hyperidentity of Mod $N(T)$, therefore Mod $N(T)$ is not solid. The same example shows also that Mod $R(T)$ may fail to be solid.

It is not known whether a theorem similar to Theorem 3.6.6 can be proved for normal varieties. However there are some connections between normal parts of varieties and solid varieties. If the variety Mod $N(V)$ is a cover of a not normal variety V in the lattice $L(\text{Mod}\, N(V))$, then we have the following for solid varieties.

Theorem 3.6.7 *Let V be a solid but not normal variety. Let e be a hyperidentity of V from $H(V)\backslash N(V)$. Then $E(N(H(V)) \cup \{e\}) = H(V)$.*

Proof We present here a syntactic proof which can also be used for the next theorem on the word problem.

The inclusion $\subseteq$ follows from the fact that consequences of hyperidentities of V are hyperidentities of V. For the converse direction we consider the hyperidentity e of the form $x_k = r(x_1, \ldots, x_n)$, $1 \leq k \leq n$. If r is a variable (different from x_k), then the inclusion $\supseteq$ obviously holds. If $r(x_1, \ldots x_n)$ is a proper term, then take a hyperidentity $x_j = q(x_1, \ldots, x_m)$ from the set $H(V)\backslash N(H(V))$. We may assume that $n = m$, otherwise we consider r and q as terms in $l = \max(n,m)$ variables. If $k = j$ and $r(x_1, \ldots, x_n) = q(x_1, \ldots, x_n)$ is a normal hyperidentity of V, then $N(H(V)) \cup \{e\} \vdash x_k = q$. If $k \neq j$ then let $r^*(x_1, \ldots, x_n)$ denote the term $r(x_1, \ldots, x_{k-1}, x_j, x_{k+1}, \ldots, x_n)$, generated from r by replacing x_k by x_j. Then we have $x_k = r \vdash x_j = r^*$, where $x_j = r^*$ is a hyperidentity of V.

If q is not a variable, then $r^* = q$ is a normal hyperidentity of V and one has $e, x_j = r^*$,

$r^* = q$, $x_j = q$ is a proof of $x_j = q$ from the set $N(H(V)) \cup \{e\}$. If q is a variable y different from x_j, then $r(z, \ldots, z) = z$, $r(z, \ldots, z) = r(x_1, \ldots, x_n)$ are hyperidentities of V, for any variable z different from x_k. We have $r(x_1, \ldots, x_n) = r(z, \ldots, z)$, $r(z, \ldots, z) = z$, $x_k = z$, $x_j = y$ for a proof of $x_j = q$.

Remark Theorem 3.6.7 can be reformulated for varieties V in general. Let V be a variety and e a hyperidentity of V. Assume that e is not normal. Then $H(V) = E(N(H(V)) \cup \{e\})$, i.e. each hyperidentity of V can be deduced from all normal hyperidentities of V and any fixed not normal hyperidentity of V.

Let Σ be a set of identities which define a variety $V = \mathrm{Mod}(\Sigma)$. We say that the word problem for the variety $\mathrm{Mod}(\Sigma)$ is solvable if there is an effective procedure to decide whether a word p is equal to a word q. In other words whether $p = q$ is a consequence of Σ.

Consequences of normal identities (hyperidentities) are normal. If the variety $\mathrm{Mod}N(V)$ is not normal, then there exists a not normal identity (hyperidentity) in Σ. From the proof of Theorem 2 in [Graczyńska 84] we get:

Theorem 3.6.8 *The word problem for a variety V is solvable if and only if the word problem for the variety* $\mathrm{Mod}\,N(V)$ *is solvable.*

Proof Necessity is obvious. To prove the sufficiency, let e be a not normal identity from $E(V)$. Then either e is of the form $x = y$ or it is the consequence of an identity of the form $x = r(x, \ldots, x)$ is its consequence, where r is a proper term. Therefore V is trivial in the first case. In the second case $p(x_1, \ldots, x_n) = q(x_1, \ldots, x_n)$ is a consequence of $E(V)$ if and only if $p^*(x_1, \ldots, x_n) = q^*(x_1, \ldots, x_n)$ is an identity of $N(V)$, where $p^*(x_1, \ldots, x_n)$ denotes the term obtained from $p(x_1, \ldots, x_n)$ by substituting x_j by $r(x_j, \ldots, x_j)$ for all $1 \leq j \leq n$, and where q^* is obtained in a similar way.

In the same way we have:

Theorem 3.6.9 *The word problem for a solid variety V defined by a set Σ of hyperidentities is solvable if and only if the word problem for the variety* $\mathrm{Mod}\,N(H(V))$ *is solvable.*

Remark Generally, the operators H and N (H and R) do not commute. For example take the trivial variety T of type (2) and the identity $f(f(x, y), z) = f(f(x, z), y))$. Then this identity is a normal (regular) hyperidentity of T, but it is not a hyperidentity of $\mathrm{Mod}\,N(T)$.

Remark Some properties of the operator H on varieties were considered by Graczyńska and Schweigert [GraSch 90] in connection with Problem 4 of W. Taylor [Taylor 79].

Theorem 3.6.10 *For a variety V, the following conditions are equivalent:*

(i) *The word problem for V is solvable;*

(ii) *The word problem for* $\mathrm{Mod}\,N(V)$ *is solvable;*

(iii) *The word problem for* $\mathrm{Mod}\,R(V)$ *is solvable;*

(iv) *The word problem for* $\mathrm{Mod}\,N(R(V))$ *is solvable.*

3.7 On the unification of hyperterms

This section deals with the unification problems. The results are due to Ewa Graczyńska. Before we proceed with our topic we like to mention the following.

Fact Let V be a variety of type τ. Then the following are equivalent.

(i) $H(N(V)) = N(H(V))$;

(ii) $\mathrm{Mod}\, N(V)$ is solid.

Proof Let $H(N(V)) = N(H(V))$. Take a normal identity $p = q$ from $N(V)$. Consider a substitution $\sigma(p) = \sigma(q)$. Then $\sigma(p) = \sigma(q)$ belongs to $H(N(V)) = N(H(V))$, and hence is a normal hyperidentity of V. Therefore $p = q$ is a hyperidentity of $\mathrm{Mod}\, N(V)$ and we have that $\mathrm{Mod}\, N(V)$ is solid.

Now let $\mathrm{Mod}\, N(V)$ be a solid variety, i.e., $H(N(V)) = N(V)$. As $H(V) \subseteq E(V)$ we have $N(H(V)) \subseteq N(V) \subseteq H(N(V))$. To show the converse, we assume that $p = q$ is a normal identity of V. Any hypersubstitution of $p = q$ is an identity of $\mathrm{Mod}\, N(V)$. Thus $p = q$ is a normal hyperidentity of V. Similarly for H and R.

Notations

$\mathbf{P}(\tau)$ denotes the free algebra of type τ, the algebra of all terms of type τ.
$\mathbf{P}(\tau)/=$ denotes the quotient algebra of $\mathbf{P}(\tau)$ by the equivalence relation ($=$ equation) on terms of type τ.

Definition 3.7.1 Let s and t be two terms of a given type τ. An identity $s = t$ is *unifiable* in a given algebra $\mathbf{A}$ of type τ if and only if there exists a homomorphism $\alpha : \mathbf{P}(\tau) \to \mathbf{A}$ such that $\alpha(s) = \alpha(t)$.

Example Any identity $s = t$ is unifiable in a one-element algebra.

Example Let $\mathbf{A} = (A, (f_t : t \in T))$ be an algebra. The one-element extension $\mathbf{A}^*$ is defined as follows: $A^* = A \cup \{1\}$, where 1 is a new element not belonging to A, and $\mathbf{A}^* = (A^*(f_t^* : t \in T))$ with

$$f_t^*(a_1, \ldots, a_n) = \begin{cases} 1, & \text{if } 1 \in \{a_1, \ldots, a_n\} \\ f_t(a_1, \ldots, a_n), & \text{otherwise} \end{cases}$$

Then any identity $s = t$ is unifiable in $\mathbf{A}^*$. We define α by $\alpha(x) = 1$ for any variable x and have $\alpha(s) = \alpha(t) = 1$ in A^*.

3.7.2 The Unification Problem Let V be a variety of type τ, s and t two terms of type τ. Decide whether $s = t$ is unifiable in the algebra $\mathbf{P}(\tau)/=$. In other words: Is there a substitution α such that $\alpha(s) = \alpha(t)$ is an identity of $E(V)$?

Theorem 3.7.3 *Given two terms* t and t *and a variety* V *of type* τ. *Then*

(1.1) *s and t are unifiable in V if and only if they are unifiable in* Mod $R(V)$;

(1.2) *the unification problem is solvable in V if and only if it is solvable in* Mod $R(V)$.

Proof Let α be a substitution such that $\alpha(s) = \alpha(t)$ is an identity of V. If x is a variable from $\mathrm{Var}(\alpha(s))\backslash\mathrm{Var}(\alpha(t))$ (or $\mathrm{Var}(\alpha(t))\backslash\mathrm{Var}(\alpha(s))$), then we can put $\beta(x) = y$ for any fixed variable y of $\mathrm{Var}(\alpha(t))\backslash\mathrm{Var}(\alpha(s))$. For $\gamma = \beta \circ \alpha$ the identity $\gamma(s) = \gamma(t)$ is of $R(V)$. Hence s and t are unifiable in the variety $\mathrm{Mod}(R(V))$. This procedure shows that from a decision algorithm for the unification problem for $E(V)$ we can get also an algorithm for $R(V)$. The opposite directions are obvious. In the following we call a not normal identity an absorption law [JežMcN].

Theorem 3.7.4 *Let s and t be two terms of type τ. Then*

(2.1) *s and t are unifiable in V if and only if they are unifiable in the variety* Mod $N(V)$;

(2.2) *the unification problem is solvable in V if and only if it is solvable in* Mod $N(V)$.

Proof The sufficiency in (2.1) and (2.2) is obvious. We show the necessity. If $E(V) = N(V)$ then the theorem holds. Let $x = p(x_1, \ldots, x_n)$ be an absorption law satisfied in V. If $\alpha(s) = \alpha(t)$ is an absorption law for the substitution α in V, i.e. $\alpha(s)$ is a variable y and $\alpha(t)$ is not a variable or $\alpha(t)$ is a different variable z, then we define a substitution β with $\beta(w) = p(w)$ for any variable w where $p(w)$ is the term $p(w, \ldots, w)$. For $\gamma = \beta \circ \alpha$ the identity $\gamma(s) = \gamma(t)$ is a normal identity of V. Here again we can get a decision procedure for the unification problem for $N(V)$ by the decision procedure from $E(V)$.

The next theorem shows the role of operators N and R in the problem of description of some special theories which are called permutative by [Siekmann 84] or term finite by Ježek and McNulty [JežMcN].

Definition 3.7.5 Let $E(V)$ be an equational theory of a variety V of type τ. V is called *term finite* if and only if for any term p of type τ the class $[p]/=v$ is finite (i.e. there is only a finite number of terms s such that $s = t$ is an identity of V). If the equational theory $E(V)$ of a variety V is term finite, then we say that V is *term finite.*

From now on we assume that type τ is not empty.

Proposition 3.7.6 [Siekmann 84] *Term finite theories are regular (i.e. if a variety V is term finite, then $E(V) = R(V)$).*

Proposition 3.7.7 (Ježek, McNulty) *Term finite theories are normal (i.e. if V is term finite, then $E(V) = N(V)$).*

Example The variety S of semigroups is term finite.

Example Any variety with an idempotent law is not term finite.

Theorem 3.7.8 *Let V be a term finite variety of type τ. Then*

(3.1) *V is not of the form* Mod $R(W)$ *for any non-regular variety W of type τ,*

(3.2) *V is not of the form* Mod $N(V)$ *for any not normal variety W of type τ.*

Proof Assume that W is a non-regular variety of type τ and $V = \operatorname{Mod} R(V)$. By the lemma of [Plonka 69] one of the following conditions holds (where x and y denote different variables).

(i) $x = y$ is an identity in $E(V)$;

(ii) type τ is unary and $p(x) = p(y)$ is an identity in $E(V)$, where p is a proper term;

(iii) $x = p(x, y)$ is an identity in $E(V)$, where p is a binary term;

(iv) $p(x, x) = p(x, y)$ is an identity in $E(V)$ for some binary term p.

In case (i) take a functional symbol $F(x_1, \ldots, x_n)$ of type τ. Then $x = F(x, \ldots, x) = F(Fx, \ldots, x), x \ldots, x)) = \ldots$ is an infinite sequence of identities in $E(V)$.

In case (ii) we obtain an infinite sequence $p(x) = p(p(x)) = p(p(p(x))) = \ldots$ of identities satisfied in V.

In case (iv) we consider $p(x, x) = p(x, p(x, x)) = p(x, p(x, p(x, x))) = \ldots.$

Case (iii) is similar. This proves (3.1), by contradiction.

For (3.2) we assume that W is a variety of type τ with an absorption law $x = p(x, \ldots, x)$. Such an identity exists, because there are functional symbols of type τ. If $V = \operatorname{Mod} N(W)$, then

$$p(x, \ldots, x) = p(p(x, \ldots, x), x, \ldots, x)) = p(p(p(x, \ldots, x), x \ldots, x), x, \ldots, x) = \ldots$$

is an infinite sequence of $E(V)$. V is not term finite which is a contradiction.

Remark 3.7.9 If type τ is empty, then the trivial variety T of type τ is not term finite. But the varieties $\operatorname{Mod} R(T)$ and $\operatorname{Mod} N(T)$ are term finite.

Corollary 3.7.10 *The variety S of semigroups is not of the form* Mod $R(W)$ *or* Mod $N(W)$ *for any non-regular (or not normal) variety W of type* (2).

Remark 3.7.11 Conditions (3.1) and (3.2) are not sufficient to describe term finite theories. This can be concluded from the properties of the variety of semigroups defined by the identity $x^2 = x^4$.

If we take any equation $p = q$ with the property that p is a proper subterm of q we can construct many such examples.

The unification problem is connected with practical questions appearing in computer programming (for example the procedure presented above is similar to the so-called "occur check" in Prolog programming (cf. Clocksin and Mellish [CloMel 81, p. 224]).

3.8 Boolean clones

Two algebras $\mathbf{A}, \mathbf{B}$ of the same type can be defined by different sets Id $\mathbf{A}$, Id $\mathbf{B}$ of identities if and only if $\mathbf{A}$ and $\mathbf{B}$ generate different varieties. In this case we say that $\mathbf{A}$ and $\mathbf{B}$ can be separated by identities.

In the following we show that non-isomorphic clones on the two-element set can be separated by hyperidentities. Indeed every clone on $\{0,1\}$ is subdirectly irreducible (see Section 4.4), and the non-isomorphic clones on $\{0,1\}$ generate different varieties.

It is obvious that there are non-isomorphic clones on $\langle n \rangle = \{0, 1, \ldots, n\}, n > 2$, which generate the same variety and hence can no longer be separated by hyperidentities (see Section 4.5).

Given a finite nonempty set A let $O_A^{(n)}$ be the set of all n-ary functions $f : A^n \to A$. We put $O_A = \bigcup\limits_{n=1}^{\infty} O_A^{(n)}$ and consider the algebra $\mathbf{O}_A = (O_A; *, \xi, \tau, \Delta, e_1^2)$. The lattice of all subclones of the clone $\mathbf{O}_A$ with $A = \{0, 1\}$ was investigated by E.L. Post (see picture at the end of this section). This lattice is atomic, dually atomic, countably infinite and every clone has a finite basis of generators. Clones which are symmetric in this picture are isomorphic.

There were several attempts to simplify Post's proof from 1920. A very detailed proof of Post's theorem was worked out by Jablonskij, Gawrilov and Kudrjawzev in 1966 [JaGaKu 70].

Other approaches using Malcev type theorems can be found in several papers especially by McKenzie, McNulty and Taylor [McMcTa 87].

An elementary and short approach to the results of Post was presented by Lau in [Lau 91] which uses no theorems of universal algebra.

We present a description in detail for every Boolean clone in the following table. These results are due to Denecke, Malcev and Reschke [DeMaRe 90].

clone	description	generating system (example)
$\mathbf{C}_1$	set of all Boolean functions	$\{\wedge, N\}$
		$\wedge$ conjunction $x \wedge y := xy$
		N negation
$\mathbf{C}_3$	$\{0\}$-preserving functions,	$\{\wedge, +\}$

	i.e., $f(0,\ldots,0)=0$	$+$ addition mod 2
$\mathbf{C}_4$	$\{0\}$- and $\{1\}$-preserving functions, i.e. $f(0,\ldots,0)=0$ and $f(1,\ldots,1)=1$	$\{\vee, g_1\}$ $\vee$ disjunction, $g_1(x,y,z)=x\wedge(y+z+1)$
$\mathbf{M}_1$	monotone functions	$\{\wedge,\vee,c_0^1,c_1^1\}$, c_0^1, c_1^1 unary constant functions
$\mathbf{M}_3$	monotone, $\{0\}$-preserving functions	$\{\wedge,\vee,c_0^1\}$
$\mathbf{M}_4$	monotone, $\{0\}$- and $\{1\}$-preserving functions	$\{\wedge,\vee\}$
$\mathbf{D}_3$	selfdual functions, i.e. $f(x_1,\ldots,x_n)=Nf(Nx_1,\ldots,Nx_n)$	$\{u_2, x+y+z, N\}$ $u_2(x,y,z)=xy\vee xz\vee yz$
$\mathbf{D}_1$	selfdual, $\{0\}$- and $\{1\}$-preserving functions	$\{u_2, x+y+z\}$
$\mathbf{D}_2$	selfdual, monotone functions	$\{u_2\}$
$\mathbf{L}_1$	linear Boolean functions, i.e. $f(x_1,\ldots,x_n)=c_0+a_1x_1+\ldots+a_nx_n$, $a_i\in\{0,1\}$	$\{+,N,c_0^1,c_1^1\}$
$\mathbf{L}_3$	linear, $\{0\}$-preserving Boolean functions	$\{+,c_0^1\}$
$\mathbf{L}_4$	linear, $\{0\}$- and $\{1\}$-preserving Boolean functions	$\{x+y+z\}$
$\mathbf{L}_5$	linear, self-dual functions	$\{x+y+z,N\}$
$\mathbf{P}_6$		$\{\wedge,c_0^1,c_1^1\}$
$\mathbf{P}_3$		$\{\wedge,c_0^1\}$
$\mathbf{P}_1$		$\{\wedge\}$
$\mathbf{P}_5$		$\{\wedge,c_1^1\}$

$\mathbf{O}_9$		$\{N, c_0^1\}$
$\mathbf{O}_8$	clones consisting only of essentially unary functions	$\{\mathrm{id}, c_0^1, c_1^1\}$ id = identity
$\mathbf{O}_4$	functions	$\{N\}$
$\mathbf{O}_6$		$\{\mathrm{id}, c_0^1\}$
$\mathbf{O}_1$		$\{\mathrm{id}\}$
$\mathbf{F}_8^m$	1-separating functions of degree $m \geq 2$, i.e. each m-element subset of $f^{-1}(1)$ has a common i-th component of the value 1	$\{u_m, g_4\}$ $g_4(x,y) = x \wedge Ny$ $u_m(x_1, \ldots, x_{m+1}) = \bigvee_{i=1}^{m+1} x_1 \ldots x_{i-1}x_{i+1} \ldots x_{m+1}$
$\mathbf{F}_5^m$	1-separating of degree $m \geq 2$ $\{0\}$-preserving functions	$\{u_m, g_3\}$ $g_3(x,y,z) = x(y \vee Nz)$
$\mathbf{F}_7^m$	1-separating of degree $m \geq 2$ monotone functions	$\{u_m, c_0^1\}$
$\mathbf{F}_6^m$	1-separating of degree $m \geq 2$, monotone, $\{0\}$-preserving functions	$\{g_2, u_2\}$ $g_2(x,y,z) = x(y \vee z)$
$\mathbf{F}_8^\infty$	1-separating functions, i.e. each subset of $f^{-1}(1)$ has a common component of value 1	$\{g_4\}$
$\mathbf{F}_5^\infty$	1-separating, $\{0\}$-preserving functions	$\{g_3\}$
$\mathbf{F}_7^\infty$	1-separating, monotone functions	$\{g_2, c_0^1\}$
$\mathbf{F}_8^\infty$	1-separating, monotone, $\{0\}$- preserving functions	$\{g_2\}$

Problem 3.8.0 Let $\mathbf{C}_1$ and $\mathbf{C}_2$ be nonisomorphic Boolean clones defined on the same set $A = \{0,1\}$. Are the sets $\mathrm{Id}\mathbf{C}_1$ and $\mathrm{Id}\mathbf{C}_2$ of their identities different?

Case 1 We consider pairs of Boolean clones $\mathbf{C}_1, \mathbf{C}_2$ with $\mathbf{C}_1$ properly contained in $\mathbf{C}_2$. Then $\mathrm{Id}\mathbf{C}_1 \supseteq \mathrm{Id}\mathbf{C}_2$, and in this case Problem 3.8.0 is the question whether this inclusion is proper or not. The answer is positive if there exists a separating hyperidentity – in other words: a hyperidentity of $\mathbf{C}_1$ which is not a hyperidentity of $\mathbf{C}_2$. If $\mathbf{C}_2 = \mathbf{O}_A$ and $\mathbf{C}_1$ is a dual atom in the lattice of all subclones of $\mathbf{O}_A$ then a positive answer for Problem 3.8.0 is given by Denecke and Reichel in [DenRei 88]. The hyperidentity $\varepsilon : F(F(x,y),F(x,y)) = F(F(x,x),F(y,y))$ holds in all dual atoms of the Post lattice but not in $\mathbf{O}_A$, because all binary Boolean functions beside the Sheffer functions satisfy ε. Therefore every subclone of $\mathbf{O}_A$ can be separated from $\mathbf{O}_A$ by a hyperidentity, namely ε. Now we want to find the separating hyperidentities also for the pairs $(\mathbf{C}_1, \mathbf{C}_2)$ with $\mathbf{C}_2 \neq \mathbf{O}_A$. Therefore we have the following definition:

Definition 3.8.1 Let $(\mathbf{C}_1, \mathbf{C}_2)$ and $(\mathbf{C}'_1, \mathbf{C}'_2)$ be two pairs of clones $(\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}'_1, \mathbf{C}'_2 \subseteq \mathbf{O}_A)$ with $\mathbf{C}_1$ properly contained in $\mathbf{C}_2$. Then

$$(\mathbf{C}_1, \mathbf{C}_2) \prec (\mathbf{C}'_1, \mathbf{C}'_2) \;:\Leftrightarrow\; \begin{array}{l} \mathbf{C}'_1 \subseteq \mathbf{C}_1 \text{ or } \mathbf{C}''_1 \subseteq \mathbf{C}_1 \text{ with } \mathbf{C}'_1 \cong \mathbf{C}''_1 \\ \text{and} \\ \mathbf{C}_2 \subseteq \mathbf{C}'_2 \text{ or } \mathbf{C}_2 \subseteq \mathbf{C}''_2 \text{ with } \mathbf{C}'_2 \cong \mathbf{C}''_2. \end{array}$$

Then the following lemma holds:

Lemma 3.8.2 *If $(\mathbf{C}_1, \mathbf{C}_2) \prec (\mathbf{C}'_1, \mathbf{C}'_2)$ with the assumptions in Definition* 3.8.2 *and there exists a hyperidentity ε which holds in $\mathbf{C}_1$ but not in $\mathbf{C}_2$, then it follows that ε holds in $\mathbf{C}'_1$ but not in $\mathbf{C}'_2$.*

This implies that the separation of clones $\mathbf{C}'_1, \mathbf{C}'_2$ with $\mathbf{C}'_1 \not\subseteq \mathbf{C}'_2$ by hyperidentities can be reduced to find separating hyperidentities only for the pairs of clones which are minimal with respect to the relation $\prec$. If $(\mathbf{C}_1, \mathbf{C}_2)$ is such a minimal pair, then $\mathbf{C}_1$ is a maximal subclone of $\mathbf{C}_2$. A positive result was achieved for this case in [Denecke 89]. From that it follows

Theorem 3.8.3 *Any two Boolean clones $\mathbf{C}_1, \mathbf{C}_2$ with $\mathbf{C}_1 \not\subseteq \mathbf{C}_2$ can be separated by hyperidentities, i.e.* $\mathrm{Id}\mathbf{C}_1 \not\supseteq \mathrm{Id}\mathbf{C}_2$.

Case 2 of Problem 3.8.0 We consider pairs of Boolean clones which are incomparable with respect to $\subseteq$.

We define a relation $\prec$ between pairs of incomparable clones in the following manner:

Definition 3.8.4 Let $(\mathbf{C}_1, \mathbf{C}_2)$ and $(\mathbf{C}'_1, \mathbf{C}'_2)$ be two pairs of clones $(\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}'_1, \mathbf{C}'_2 \subseteq \mathbf{O}_A)$ with $\mathbf{C}_1 \not\subseteq \mathbf{C}_2$ and $\mathbf{C}_2 \not\subseteq \mathbf{C}_1$. Then

$$(\mathbf{C}_1, \mathbf{C}_2) \prec (\mathbf{C}'_1, \mathbf{C}'_2) \;:\Leftrightarrow\; \begin{array}{l} \mathbf{C}_1 \subseteq \mathbf{C}'_1 \text{ or } \mathbf{C}_1 \subseteq \mathbf{C}''_1 \text{ with } \mathbf{C}'_1 \cong \mathbf{C}''_1 \\ \text{and} \\ \mathbf{C}'_2 \subseteq \mathbf{C}_2 \text{ or } \mathbf{C}''_2 \subseteq \mathbf{C}_2 \text{ with } \mathbf{C}'_2 \cong \mathbf{C}''_2. \end{array}$$

Then the following lemma holds:

Lemma 3.8.5 *If $(\mathbf{C}_1, \mathbf{C}_2) \prec (\mathbf{C}'_1, \mathbf{C}'_2)$ with the assumptions in Definition* 3.8.4 *and $\mathbf{C}'_1 \not\subseteq$*

$\mathbf{C}_2'$ and $\mathbf{C}_2' \not\subseteq \mathbf{C}_1'$, and there exists a hyperidentity ε which holds in $\mathbf{C}_1'$ but not in $\mathbf{C}_2'$, then it follows that ε holds in $\mathbf{C}_1$ but not in $\mathbf{C}_2$.

This implies that the separation of clones $\mathbf{C}_1, \mathbf{C}_2$ with $\mathbf{C}_1 \not\subseteq \mathbf{C}_2$ and $\mathbf{C}_2 \not\subseteq \mathbf{C}_1$ by hyperidentities can be reduced to finding separating hyperidentities only for the pairs of clones which are maximal with respect to the relation $\prec$

Lemma 3.8.6 *A pair $(\mathbf{C}_1, \mathbf{C}_2)$ with $\mathbf{C}_1 \not\subseteq \mathbf{C}_2, \mathbf{C}_2 \not\subseteq \mathbf{C}_1$, is maximal with respect to $\prec$ if and only if for all clones $\mathbf{K}_1, \mathbf{K}_2$ with $\mathbf{K}_1 \supseteq \mathbf{C}_1$ and $\mathbf{K}_1 \subseteq \mathbf{C}_2$ we have $\mathbf{K}_1 \subseteq \mathbf{K}_2$ (or $\mathbf{K}_1 \subseteq \mathbf{K}_2'$ with $\mathbf{K}_2' \cong \mathbf{K}_2$) or $\mathbf{K}_2 \subseteq \mathbf{K}_1$ (or $\mathbf{K}_2 \subseteq \mathbf{K}_1'$ with $\mathbf{K}_1' \cong \mathbf{K}_1$).*

Denecke, Malcev and Reschke give a method to find all such maximal pairs of incomparable clones [DeMaRe 90, Lemmas 4.5 and 4.6]. Also in this paper one can find a separating hyperidentity for every (with respect to $\prec$) maximal pair of incomparable pairs of Boolean clones. This leads to

Theorem 3.8.7 *Let $(\mathbf{C}_1, \mathbf{C}_2)$ be a pair of Boolean clones with $\mathbf{C}_1 \not\subseteq \mathbf{C}_2'$ and $\mathbf{C}_2 \not\subseteq \mathbf{C}_1'$ for all $\mathbf{C}_2' \cong \mathbf{C}_2$ and all $\mathbf{C}_1' \cong \mathbf{C}_1$. Then $\mathbf{C}_1$ and $\mathbf{C}_2$ can be separated from each other by hyperidentities, i.e. there is a hyperidentity ε which holds in $\mathbf{C}_1$ but not in $\mathbf{C}_2$, and a hyperidentity ε' which holds in $\mathbf{C}_2$ but not in $\mathbf{C}_1$. This means that for the sets of identities of $\mathbf{C}_1$ and $\mathbf{C}_2$ we have* $\mathrm{Id}\mathbf{C}_1 \not\subseteq \mathrm{Id}\mathbf{C}_2$ *and also* $\mathrm{Id}\mathbf{C}_2 \not\subseteq \mathrm{Id}\mathbf{C}_1$.

The following table will give a review of all separating hyperidentities of all Boolean clones. In the first column we list the clones $\mathbf{X}$. The second column gives separating hyperidentities which hold in maximal subclones of $\mathbf{X}$ but not in $\mathbf{X}$ itself. The third column lists hyperidentities which hold in $\mathbf{X}$ but not in clones which are incomparable (w.r. to inclusion) with $\mathbf{X}$.

$\mathbf{C}_1$	ε_0	$\mathbf{C}_1$ is comparable with all other Boolean clones
$\mathbf{C}_3$	$CHI(\mathbf{C}_4), CHI(\mathbf{M}_3), CHI(\mathbf{L}_3)$	ε_2 for $\mathbf{O}_4, \mathbf{L}_1, \mathbf{L}_5, \mathbf{O}_9, \mathbf{D}_3$
	$CHI(\mathbf{F}_8^2)$	ε_6 for $\mathbf{O}_8, \mathbf{P}_6, \mathbf{M}_1$
$\mathbf{C}_4$	$CHI(\mathbf{M}_4), CHI(\mathbf{D}_1), CHI(\mathbf{F}_5^2)$	ε_2 for $\mathbf{D}_3, \mathbf{L}_1, \mathbf{L}_5, \mathbf{O}_9, \mathbf{D}_4$
		ε_5 for $\mathbf{O}_6, \mathbf{M}_1, \mathbf{M}_3, \mathbf{L}_3, \mathbf{P}_3, \mathbf{P}_5$, $\mathbf{F}_8^m, \mathbf{F}_7^m, m \geq 2, \mathbf{F}_8^\infty, \mathbf{F}_6^\infty$
		ε_6 for $\mathbf{P}_6, \mathbf{O}_8$
$\mathbf{M}_1$	$CHI(\mathbf{M}_3), CHI(\mathbf{P}_6)$	ε_3 for $\mathbf{L}_4, \mathbf{O}_4, \mathbf{C}_3, \mathbf{C}_4, \mathbf{D}_3, \mathbf{D}_1, \mathbf{L}_1, \mathbf{L}_3, \mathbf{L}_5, \mathbf{O}_9$
		ε_{10} for $\mathbf{F}_8^\infty, \mathbf{F}_5^\infty, \mathbf{F}_8^m, \mathbf{F}_5^m, m \geq 2$
$\mathbf{M}_3$	$CHI(\mathbf{M}_4), CHI(\mathbf{P}_5), CHI(\mathbf{F}_7^2)$	ε_3 for $\mathbf{C}_4, \mathbf{D}_3, \mathbf{D}_1, \mathbf{L}_1, \mathbf{L}_3, \mathbf{L}_4, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
		ε_6 for $\mathbf{P}_6, \mathbf{O}_8$

		ε_{10} for $\mathbf{F}_8^m, \mathbf{F}_5^m, m \geq 2, \mathbf{F}_8^\infty, \mathbf{F}_5^\infty$
$\mathbf{M}_4$	$CHI(\mathbf{F}_6^2)$	ε_3 for $\mathbf{D}_3, \mathbf{D}_1, \mathbf{L}_1, \mathbf{L}_3, \mathbf{L}_4, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
		ε_5 for $\mathbf{P}_3, \mathbf{P}_5, \mathbf{O}_6, \mathbf{F}_8^m, \mathbf{F}_7^m, \mathbf{F}_8^\infty, \mathbf{F}_7^\infty$
		ε_6 for $\mathbf{P}_6, \mathbf{O}_8$
		ε_{10} for $\mathbf{F}_5^m, \mathbf{F}_5^\infty$ $(m \geq 2)$
$\mathbf{D}_3$	$CHI(\mathbf{D}_1), CHI(\mathbf{L}_5)$	ε_1 for all other clones besides $\mathbf{D}_2, \mathbf{L}_4, \mathbf{O}_4$
$\mathbf{D}_1$	$CHI(\mathbf{D}_2), CHI(\mathbf{L}_4)$	ε_2 for $\mathbf{L}_5, \mathbf{O}_4$
		ε_1 for all other clones besides $\mathbf{C}_3, \mathbf{C}_4$
$\mathbf{D}_2$	is an atom in the lattice	ε_1 for $\mathbf{L}_1, \mathbf{L}_3, \mathbf{P}_6, \mathbf{P}_3, \mathbf{P}_1, \mathbf{P}_5, \mathbf{O}_8, \mathbf{O}_6,$
	of all Boolean clones	$\mathbf{F}_8^m, \mathbf{F}_5^m, \mathbf{F}_7^m, \mathbf{F}_6^m, \mathbf{F}_8^\infty, \mathbf{F}_5^\infty, \mathbf{F}_7^\infty, \mathbf{F}_6^\infty$ $(m \geq 3)$
		ε_3 for $\mathbf{L}_4, \mathbf{L}_4, \mathbf{O}_9, \mathbf{O}_4$
$\mathbf{L}_2$	$CHI(\mathbf{L}_3), CHI(\mathbf{L}_5), CHI(\mathbf{O}_9)$	ε_4 for $\mathbf{D}_2, \mathbf{P}_1, \mathbf{C}_3, \mathbf{C}_4, \mathbf{M}_1, \mathbf{M}_3, \mathbf{M}_4, \mathbf{D}_3, \mathbf{D}_1,$
		$\mathbf{P}_6, \mathbf{P}_3, \mathbf{P}_5, \mathbf{F}_8^m, \mathbf{F}_5^m, \mathbf{F}_7^m, \mathbf{F}_6^m, m \geq 2$
		$\mathbf{F}_8^\infty, \mathbf{F}_5^\infty, \mathbf{F}_7^\infty, \mathbf{F}_6^\infty$
$\mathbf{L}_3$	$CHI(\mathbf{L}_4), CHI(\mathbf{O}_6)$	ε_2 for $\mathbf{D}_3, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
		ε_4 for $\mathbf{C}_4, \mathbf{M}_3, \mathbf{M}_4, \mathbf{D}_1, \mathbf{D}_2, \mathbf{P}_3, \mathbf{P}_1, \mathbf{P}_5,$
		$\mathbf{F}_8^m, \mathbf{F}_5^m, \mathbf{F}_7^m, \mathbf{F}_6^m, m \geq 2,$
		$\mathbf{F}_8^\infty, \mathbf{F}_5^\infty, \mathbf{F}_7^\infty, \mathbf{F}_6^\infty$
		ε_6 for $\mathbf{M}_1, \mathbf{P}_6, \mathbf{O}_8$
$\mathbf{L}_4$	is an atom in the lattice	ε_1 for $\mathbf{O}_8, \mathbf{O}_6$
	of all Boolean clones	ε_2 for $\mathbf{O}_9, \mathbf{O}_4$
		ε_4 for $\mathbf{M}_1, \mathbf{M}_3, \mathbf{M}_4, \mathbf{D}_2, \mathbf{P}_6, \mathbf{P}_3, \mathbf{P}_1, \mathbf{P}_5,$
		$\mathbf{F}_8^m, \mathbf{F}_5^m, \mathbf{F}_7^m, \mathbf{F}_6^m,$
		$\mathbf{F}_8^\infty, \mathbf{F}_5^\infty, \mathbf{F}_7^\infty, \mathbf{F}_6^\infty, m \geq 2$
$\mathbf{L}_5$	$CHI(\mathbf{L}_4), CHI(\mathbf{O}_4)$	ε_1 for $\mathbf{L}_3, \mathbf{O}_9, \mathbf{O}_8, \mathbf{O}_6$
		ε_4 for all other clones besides $\mathbf{D}_3, \mathbf{L}_1$
$\mathbf{P}_6$	$CHI(\mathbf{P}_3), CHI(\mathbf{P}_5), CHI(\mathbf{O}_8)$	ε_3 for $\mathbf{C}_3, \mathbf{C}_4, \mathbf{D}_3, \mathbf{D}_1, \mathbf{L}_1, \mathbf{L}_3, \mathbf{L}_4, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
		ε_9 for $\mathbf{D}_2, \mathbf{M}_3, \mathbf{M}_4,$
		$\mathbf{F}_8^m, \mathbf{F}_5^m, \mathbf{F}_7^m, \mathbf{F}_6^m, m \geq 2,$
		$\mathbf{F}_8^\infty, \mathbf{F}_5^\infty, \mathbf{F}_7^\infty, \mathbf{F}_6^\infty$

$\mathbf{P}_3$	$CHI(\mathbf{P}_1), CHI(\mathbf{O}_6)$	ε_3 for $\mathbf{C}_4, \mathbf{D}_3, \mathbf{D}_1, \mathbf{L}_1, \mathbf{L}_3, \mathbf{L}_4, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
		ε_6 for $\mathbf{O}_8$
		ε_9 for $\mathbf{M}_4, \mathbf{D}_2, \mathbf{F}_5^m, \mathbf{F}_6^m, \mathbf{F}_5^\infty, \mathbf{F}_6^\infty$ $(m \geq 2)$
		ε_{13} for $\mathbf{P}_5$
$\mathbf{P}_1$	is an atom in the lattice	ε_3 for $\mathbf{D}_3, \mathbf{D}_1, \mathbf{L}_1, \mathbf{L}_4, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
	of all Boolean clones	ε_5 for $\mathbf{L}_3, \mathbf{O}_6$
		ε_6 for $\mathbf{O}_8$
		ε_9 for $\mathbf{D}_2$
$\mathbf{P}_5$	$CHI(\mathbf{P}_1), CHI(\mathbf{O}_6)$	ε_3 for $\mathbf{C}_4, \mathbf{D}_3, \mathbf{D}_1, \mathbf{L}_1, \mathbf{L}_3, \mathbf{L}_4, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
		ε_6 for $\mathbf{O}_8$
		ε_9 for $\mathbf{M}_4, \mathbf{D}_2, \mathbf{F}_8^m, \mathbf{F}_5^m, \mathbf{F}_7^m, \mathbf{F}_6^m, m \geq 2,$ $\mathbf{F}_8^\infty, \mathbf{F}_5^\infty, \mathbf{F}_7^\infty, \mathbf{F}_6^\infty$
		ε_{11} for $\mathbf{P}_3$
$\mathbf{O}_9$	$CHI(\mathbf{O}_8), CHI(\mathbf{O}_4)$	ε_{12} for $\mathbf{L}_4, \mathbf{L}_3, \mathbf{L}_5$
		ε_4 for all other clones besides $\mathbf{L}_1, \mathbf{O}_6$
$\mathbf{O}_8$	$CHI(\mathbf{O}_6)$	ε_3 for $\mathbf{L}_3, \mathbf{L}_4, \mathbf{L}_5, \mathbf{O}_4$
		ε_4 for all clones besides $\mathbf{M}_1, \mathbf{P}_6, \mathbf{O}_9$
$\mathbf{O}_4$	is an atom in the lattice	ε_1 for $\mathbf{L}_3, \mathbf{O}_8, \mathbf{O}_6$
	of all Boolean clones	ε_{12} for $\mathbf{L}_4$
		ε_4 for all other clones besides $\mathbf{D}_3, \mathbf{L}_1, \mathbf{L}_5,$ $\mathbf{O}_9$
$\mathbf{O}_6$	is an atom in the lattice	ε_3 for $\mathbf{L}_4, \mathbf{L}_5, \mathbf{O}_4$
	of all Boolean clones	ε_4 for $\mathbf{C}_4, \mathbf{M}_4, \mathbf{D}_3, \mathbf{D}_1, \mathbf{D}_2, \mathbf{P}_1,$ $\mathbf{F}_5^m, \mathbf{F}_6^m, m \geq 2, \mathbf{F}_5^\infty, \mathbf{F}_6^\infty$
$\mathbf{F}_8^m$	$CHI(\mathbf{F}_8^{m+1})$	ε_2 for $\mathbf{L}_1, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4, \mathbf{D}_3$
$(m \geq 3)$	$CHI(\mathbf{F}_5^m), CHI(\mathbf{F}_7^m)$	ε_6 for $\mathbf{M}_1, \mathbf{P}_6, \mathbf{O}_8$
		ε_7 for $\mathbf{C}_4, \mathbf{M}_3, \mathbf{M}_4, \mathbf{D}_1, \mathbf{L}_3, \mathbf{L}_4$
		ε_8 for $\mathbf{D}_2, \mathbf{F}_6^n, \mathbf{F}_5^n, \mathbf{F}_7^n$ $(n < m)$
		ε_{13} for $\mathbf{P}_5$

$\mathbf{F}_5^m$	$CHI(\mathbf{F}_5^{m+1})$	ε_2 for $\mathbf{D}_3, \mathbf{L}_1, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
$(m \geq 3)$	$CHI(\mathbf{F}_6^m)$	ε_5 for $\mathbf{M}_3, \mathbf{P}_3, \mathbf{O}_6,$
		$\mathbf{F}_8^\infty, \mathbf{F}_7^\infty, \mathbf{F}_8^k, \mathbf{F}_7^m \ (k > m)$
		ε_6 for $\mathbf{M}_1, \mathbf{P}_6, \mathbf{O}_8$
		ε_7 for $\mathbf{M}_4, \mathbf{D}_1, \mathbf{L}_3, \mathbf{L}_4$
		ε_8 for $\mathbf{D}_2, \mathbf{F}_6^n \ (n < m)$
		ε_{13} for $\mathbf{P}_5$
$\mathbf{F}_6^m$	$CHI(\mathbf{F}_7^{m+1})$	ε_2 for $\mathbf{D}_3, \mathbf{L}_1, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
$(m \geq 3)$	$CHI(\mathbf{F}_6^m)$	ε_3 for $\mathbf{C}_4, \mathbf{D}_1, \mathbf{L}_3, \mathbf{L}_4$
		ε_6 for $\mathbf{P}_6, \mathbf{O}_8$
		ε_7 for $\mathbf{M}_4$
		ε_8 for $\mathbf{D}_2, \mathbf{F}_8^k \ (k > m), \mathbf{F}_6^n \ (n < m)$
		ε_{10} for $\mathbf{F}_5^m, \mathbf{F}_8^\infty, \mathbf{F}_5^\infty$
		ε_{13} for $\mathbf{P}_5$
$\mathbf{F}_6^m$	$CHI(\mathbf{F}_6^{m+1})$	ε_2 for $\mathbf{D}_3, \mathbf{L}_1, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
$(m \geq 3)$		ε_3 for $\mathbf{D}_1, \mathbf{L}_3, \mathbf{L}_4$
		ε_5 for $\mathbf{P}_3, \mathbf{O}_6, \mathbf{F}_7^k \ (k > m), \mathbf{F}_7^\infty$
		ε_6 for $\mathbf{P}_6, \mathbf{O}_8$
		ε_8 for $\mathbf{D}_2, \mathbf{F}_8^k \ (k > m)$
		ε_{10} for $\mathbf{F}_5^k \ (k > m), \mathbf{F}_8^\infty, \mathbf{F}_5^\infty$
		ε_3 for $\mathbf{P}_5$
$\mathbf{F}_8^2$	$CHI(\mathbf{F}_8^3)$	ε_2 for $\mathbf{D}_3, \mathbf{L}_1, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
	$CHI(\mathbf{F}_5^2)$	ε_6 for $\mathbf{M}_1, \mathbf{P}_6, \mathbf{O}_8$
	$CHI(\mathbf{F}_7^2)$	ε_7 for $\mathbf{M}_4, \mathbf{L}_4, \mathbf{C}_4, \mathbf{M}_3, \mathbf{D}_1, \mathbf{L}_3$
		ε_{13} for $\mathbf{P}_5$
$\mathbf{F}_5^2$	$CHI(\mathbf{F}_5^3)$	ε_2 for $\mathbf{D}_3, \mathbf{L}_1, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
	$CHI(\mathbf{F}_6^2)$	ε_5 for $\mathbf{M}_3, \mathbf{L}_3, \mathbf{P}_3, \mathbf{P}_5, \mathbf{O}_6, \mathbf{F}_8^m \ (m \geq 3),$
		$\mathbf{F}_7^m \ (m \geq 2), \mathbf{F}_8^\infty, \mathbf{F}_7^\infty$
		ε_6 for $\mathbf{M}_1, \mathbf{P}_6, \mathbf{O}_8$
		ε_7 for $\mathbf{M}_4, \mathbf{D}_1, \mathbf{L}_4$

$\mathbf{F}_7^2$	$CHI(\mathbf{F}_7^3)$	ε_2 for $\mathbf{D}_3, \mathbf{L}_1, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
	$CHI(\mathbf{F}_6^2)$	ε_3 for $\mathbf{C}_4, \mathbf{D}_1, \mathbf{L}_3, \mathbf{L}_4$
		ε_6 for $\mathbf{P}_6, \mathbf{O}_8$
		ε_7 for $\mathbf{M}_4$
		ε_{10} for $\mathbf{F}_8^m$ $(m \geq 3), \mathbf{F}_5^m$ $(m \geq 2), \mathbf{F}_8^\infty, \mathbf{F}_5^\infty$
		ε_{13} for $\mathbf{P}_5$
$\mathbf{F}_6^2$	$CHI(\mathbf{F}_6^3)$	ε_2 for $\mathbf{D}_3, \mathbf{L}_1, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
	$CHI(\mathbf{D}_2)$	ε_3 for $\mathbf{D}_1, \mathbf{L}_4$
		ε_5 for $\mathbf{L}_3, \mathbf{P}_3, \mathbf{P}_5, \mathbf{O}_6, \mathbf{F}_8^m, \mathbf{F}_7^m, m \geq 3,$ $\mathbf{F}_8^\infty, \mathbf{F}_7^\infty$
		ε_6 for $\mathbf{P}_6, \mathbf{O}_8$
		ε_{10} for $\mathbf{F}_5^m$ $(m \geq 2), \mathbf{F}_5^\infty$
$\mathbf{F}_8^\infty$	$CHI(\mathbf{F}_5^\infty)$	ε_2 for $\mathbf{D}_3, \mathbf{L}_1, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
	$CHI(\mathbf{F}_7^\infty)$	ε_6 for $\mathbf{M}_1, \mathbf{P}_6, \mathbf{O}_8$
		ε_7 for $\mathbf{C}_4, \mathbf{M}_3, \mathbf{M}_4, \mathbf{D}_1, \mathbf{L}_3, \mathbf{L}_4$
		ε_8 for $\mathbf{D}_2, \mathbf{F}_5^m, \mathbf{F}_7^m, \mathbf{F}_6^m$ $(m \geq 2)$
		ε_{13} for $\mathbf{P}_5$
$\mathbf{F}_5^\infty$	$CHI(\mathbf{F}_6^\infty)$	ε_2 for $\mathbf{D}_3, \mathbf{L}_1, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
		ε_5 for $\mathbf{P}_3, \mathbf{O}_9, \mathbf{F}_7^\infty$
		ε_6 for $\mathbf{M}_1, \mathbf{P}_6, \mathbf{O}_8$
		ε_7 for $\mathbf{M}_3, \mathbf{M}_4, \mathbf{D}_1, \mathbf{L}_3, \mathbf{L}_4$
		ε_8 for $\mathbf{D}_2, \mathbf{F}_7^m, \mathbf{F}_6^m$ $(m \geq 2)$
		ε_{13} for $\mathbf{P}_5$
$\mathbf{F}_7^\infty$	$CHI(\mathbf{F}_6^\infty)$	ε_2 for $\mathbf{D}_3, \mathbf{L}_1, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$
	$CHI(\mathbf{P}_3)$	ε_3 for $\mathbf{C}_4, \mathbf{D}_1, \mathbf{L}_3, \mathbf{L}_4$
		ε_6 for $\mathbf{P}_6, \mathbf{O}_8$
		ε_7 for $\mathbf{M}_4$
		ε_8 for $\mathbf{D}_2, \mathbf{F}_6^m$ $(m \geq 2)$
		ε_{10} for $\mathbf{F}_5^m$ $(m \geq 2), \mathbf{F}_5^\infty$
		ε_{13} for $\mathbf{P}_5$

$\mathbf{F}_6^\infty$ $\quad CHI(\mathbf{P}_1)$

ε_2 for $\mathbf{D}_3, \mathbf{L}_1, \mathbf{L}_5, \mathbf{O}_9, \mathbf{O}_4$

ε_3 for $\mathbf{D}_1, \mathbf{L}_3, \mathbf{L}_4$

ε_5 for $\mathbf{P}_3, \mathbf{O}_6$

ε_6 for $\mathbf{P}_6, \mathbf{O}_8$

ε_8 for $\mathbf{D}_2$

ε_{13} for $\mathbf{P}_5$

The clone $\mathbf{O}_1$ is separated from all other clones by the hyperidentity $CHI(\mathbf{O}_1)$.

List of hyperidentities

ε_0: $\quad F(F(x,y),F(x,y)) = F(F(x,x),F(y,y))$

ε_1: $\quad F(x,x) = G(F(G(x,y),G(x,y)),F(G(y,x),G(y,x)))$

ε_2: $\quad F(x,x) = F(F(x,x),F(x,x))$

ε_3: $\quad F(x,x,y) = F(F(x,x,y),F(x,x,x),F(y,y,y))$

ε_4: $\quad F(x,y,y) = F(x,y,F(z,z,F(z,z,y)))$

ε_5: $\quad F(x) = x$

ε_6: $\quad F(G(x)) = G(F(x))$

ε_7: $\quad G(F^*,F^*,F^*) = G(F^+,F^+,F^+)$ with

$$F^+ = F(F(x,x,y),F(y,x,x),F(x,y,x))$$
$$F^* = F(F(x,x,G^+),F(G^+,x,x),F(x,G^+,x))$$
$$G^+ = G(G(y,y,x),G(x,y,y),G(y,x,y))$$

ε_8: $\quad G(T_1,T_1) = G(T_2,T_2)$ with

$$T_1 := F(F(x,G^+,y,\ldots,y),F(y,x,G^+,y,\ldots,y),\ldots,F(y,\ldots,y,x,G^+), F(G^+,y,\ldots,y,x)),$$
$$T_2 := F(F(x,y,\ldots,y),F(y,x,y,\ldots,y),\ldots,F(y,\ldots,y,x,y), F(y,y,\ldots,y,x)),$$
$$G^+ := G(G(y,x),G(x,y))$$

where G is binary and T_1, T_2 are 4-ary in the case that ε_8 is a separating hyperidentity for $\mathbf{F}_8^3$ and $\mathbf{D}_2$, i.e. ε_8 holds in $\mathbf{F}_8^3$ and not in $\mathbf{D}_2$, and T_1, T_2 are $(m+1)$ary otherwise.

ε_9: $\quad F(F(x,y,x),F(x,x,y),F(y,x,x)) = F(F(x,F(y,y,y),F(y,y,y)),$

$$F(y,x,F(y,y,y)),F(x,x,x))$$

ε_{10}: $F(F(x,y,x),F(y,y,y),F(y,y,y))=F(F(x,x,x),F(x,y,y),F(y,y,y))$

ε_{11}: $F(x,F(G(x,x),x))=F(x,x)$

ε_{12}: $F(x,y,y)=F(F(F(x,x,x),F(x,x,x),F(x,x,y)),$
$F(F(x,x,x),F(x,y,x),F(x,y,x)),F(F(x,x,x),F(x,y,y),F(x,y,y)))$

ε_{13}: $F(F(F(x,y),x),G(y))=F(F(F(x,x),x),G(y))$

$CHI(\mathbf{C}_3)$: $G(F(x,y),F(x,x))=F(G(G(x,x),G(x,x)),G(G(x,x),G(x,x)))$

$CHI(\mathbf{C}_4)$: $G(x,y,z)=F(G(x,y,z),G(x,y,z),G(x,y,z))$

$CHI(\mathbf{A}_1)$: $F(x,y)=F(F(x,y),F(x,y))$

$CHI(\mathbf{A}_3)$: $F(G(x,y),G(x,y))=G(F(x,F(x,x)),F(F(y,y),y))$

$CHI(\mathbf{A}_4)$: $F(x,y,z)=F(F(x,G(y,y,y),z),y,z)$

$CHI(\mathbf{D}_3)$: $F(F(x,x,x),F(x,x,x),F(x,x,x))$
$=F(F(x,x,y),F(F(x,x,y),x,x),F(x,F(z,z,x),x))$

$CHI(\mathbf{D}_1)$: $F(x,x,x)=F(F(x,x,y),F(F(x,x,y),x,x,),F(x,F(z,z,x),x))$

$CHI(\mathbf{D}_2)$: $F(F(x,z,x),y),F(z,x,F(x,x,y)),F(F(y,x,x),z,x))=F(x,x,x)$

$CHI(\mathbf{L}_1)$: $F(x,x)=F(F(F(x,y),y),F(y,F(y,x)))$

$CHI(\mathbf{L}_3)$: $F(G(x,x),G(x,x))$
$=G(F(G(x,F(y,y)),G(x,F(x,y))),F(G(F(y,x),x),G(F(y,y),x)))$

$CHI(\mathbf{L}_4)$: $F(x,x,G(G(x,y,y),G(y,x,y),G(y,y,x)))$
$=G(x,G(y,G(z,x,y),G(y,z,y),G(z,G(z,z,y),G(y,z,x)))$

$CHI(\mathbf{L}_5)$: $F(x,x,x)=F(x,F(y,F(z,x,y),F(y,z,z)),F(z,F(z,z,y),F(y,z,x)))$

$CHI(\mathbf{P}_1)$: $F(F(x,y,x),F(x,x,y),F(y,x,x))$
$=F(F(x,G(y,y,y),G(y,y,y)),F(x,x,G(y,y,y)),F(x,x,x))$

$CHI(\mathbf{P}_3)$: $G(F(x,y,x),F(x,x,x),F(x,y,x))=F(G(x,x,x),G(G(y,x,x),x,x),G(x,x,x))$

$CHI(\mathbf{P}_5)$: $F(F(x,x),x)=F(x,F(G(x,x),x))$

$CHI(\mathbf{P}_6)$: $F(F(x,G(x,y)),F(G(y,x),x))=F(x,F(G(y,y),x))$

$CHI(\mathbf{O}_9)$: $F(F(x,x),F(y,x))=F(F(x,y),F(y,x))$

$CHI(\mathbf{O}_8)$: $F(x,x) = F(F(x,y), F(y,x))$

$CHI(\mathbf{O}_6)$: $G(F(x,x), F(x,x)) = F(G(x,x), G(G(x,y),x))$

$CHI(\mathbf{O}_4)$: $F(G(x,x,x), G(y,y,y), G(z,z,z))$
$= G(F(x,y,z), F(x,y,z), F(x,y,F(x,y,F(z,z,z))))$

$CHI(\mathbf{O}_1)$: $F(x,x,x) = G(F(x, F(y,x,z), F(y,y,x)))$
$= F(F(F(x,y,z), x, F(y,z,x)), F(F(x,y,z), F(y,x,z), x))$

$CHI(\mathbf{F}_5^2)$: $F(F(x,x,G^+), F(G^+,x,x), F(x,G^+,x)) = F(F(x,x,y), F(y,x,x), F(x,y,x)$
with
$G^+ := G(G(y,y,x), F(x,y,y), G(y,x,y))$

$CHI(\mathbf{F}_6^2)$: $G(F^\oplus, F^\oplus, F^\oplus) = F(F_1^+, F_2^+, F_3^+)$ with
$F^\oplus := F(F(x,x,G^+), F(G^+,x,x), F(x,G^+,x))$
$G^+ := G(G(y,y,x), G(x,y,y), G(y,x,y))$
$F_1^+ := F(F(x,x,x), F(x,y,x), F(x,x,x))$
$F_2^+ := F(F(x,x,x), F(x,x,x), F(x,x,y))$
$F_3^+ := F(F(y,x,x), F(x,x,x), F(x,x,x))$

$CHI(\mathbf{F}_7^2)$: $G(F^\oplus, F^\oplus, F^\oplus) = G(F', F', F')$ with $F' := F(F_1^+, F_2^+, F_3^+)$

$CHI(\mathbf{F}_8^2)$: ε_7

$CHI(\mathbf{F}_6^m)$: $G(h_1) = G'(h_2)$ with

$h_1 :=$ $F(F(x,y,x,y,\ldots,y), F(y,x,x,y,\ldots,y), F(y,y,x,y,\ldots,y), \ldots,$
$F(y,\ldots,y,x)),$

$h_2 :=$ $F(F(x,y,\ldots,y,x), F(y,x,\ldots,y,x), F(y,y,x,y,\ldots,y), \ldots,$
$F(y,\ldots,y,x))$

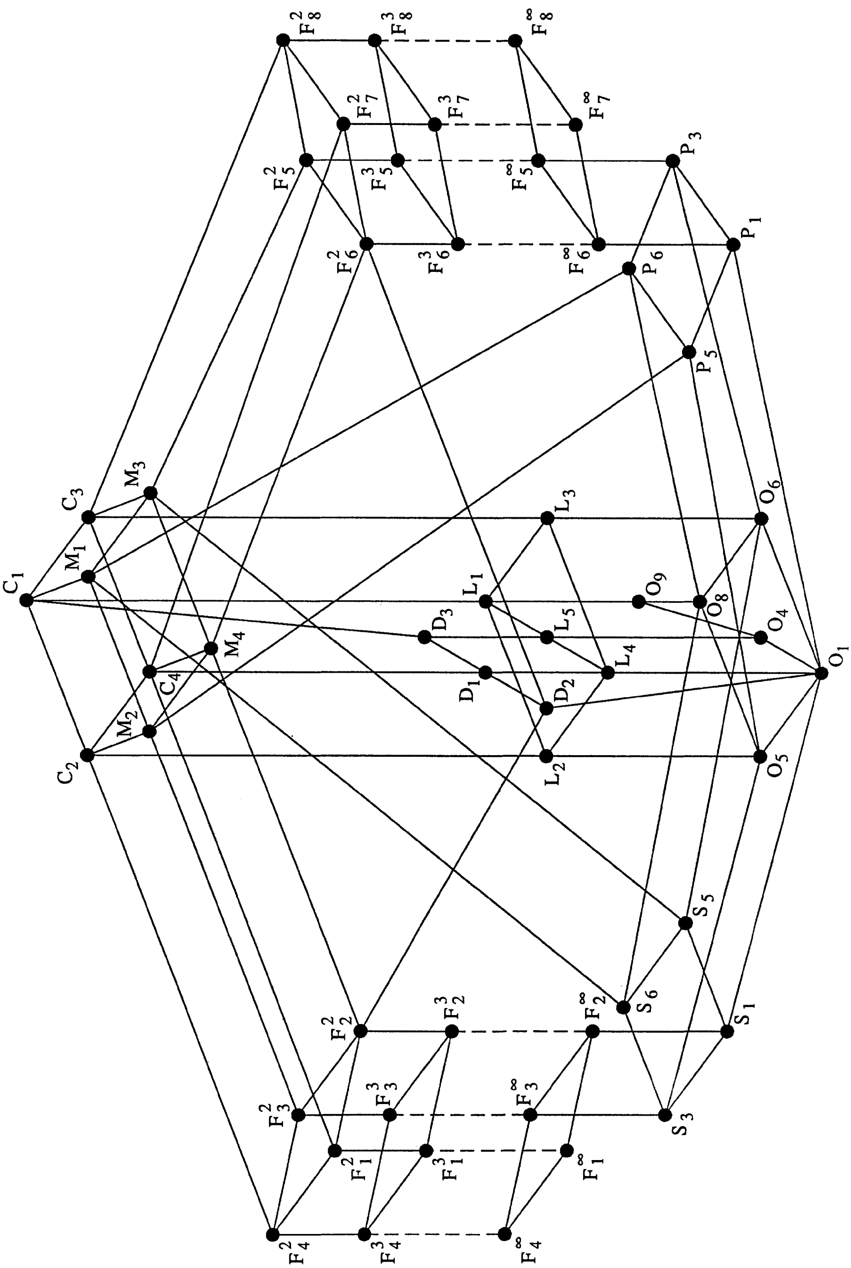
where F is an $(m+1)$-ary operation symbol and G, G' are unary operation symbols

$CHI(\mathbf{F}_5^m)$: $H(T_1) = H'(T_2)$ with T_1, T_2 and G^+ from ε_8,

$CHI(\mathbf{F}_7^m)$: $h_1 = h_2$

$CHI(\mathbf{F}_8^m)$: ε_8 where T_1 and T_2 are $(m+1)$-ary

$CHI(\mathbf{F}_i^\infty)$: $i = 5,6,7,8$ $\{CHI(\mathbf{F}_i^m) : m \geq 3\}$



F_8^2
F_8^3
F_8^∞
F_7^2
F_7^3
F_7^∞
F_5^2
F_5^3
F_5^∞
P_3
F_6^2
F_6^3
F_6^∞
P_1
P_6
P_5
C_3
M_3
L_3
O_6
C_1
M_1
L_1
O_9
O_8
D_3
L_5
O_4
M_4
C_4
L_4
O_1
D_1
D_2
M_2
C_2
L_2
O_5
S_5
S_6
F_2^2
F_2^3
F_2^∞
S_1
F_3^2
F_3^3
F_3^∞
S_3
F_1^2
F_1^3
F_1^∞
F_4^2
F_4^3
F_4^∞

Part 4

Clone congruences

4.1 The lattice of hypervarieties

Proposition 4.1.1 *The set of all hypervarieties of a given type τ froms a lattice $(\mathcal{L}(\tau), \subseteq)$. If V_1, V_2 are two hypervarieties of type τ defined by the closed sets of hyperidentities E_1, E_2, respectively, then $V_1 \wedge V_2$ is the hypervariety defined by the set $[E_1 \cup E_2]$, where $[E]$ denotes the closure of a set E under the rules of inference* (1)-(6) *and $V_1 \vee V_2$ is the hypervariety defined by $E_1 \cap E_2$. $V_1 \wedge V_2 =$ g.l.b. (V_1, V_2) and $V_1 \vee V_2 =$ l.u.b. (V_1, V_2) in $(\mathcal{L}(\tau), \subseteq)$.*

Proof Notice that the set $[E_1 \cup E_2]$ is closed under rule (6). Thus $[E_1 \cup E_2]$ is the smallest set, closed under (1)-(6) and containing the set E_1 and E_2. Thus, the g.l.b. (V_1, V_2) exists in $(\mathcal{L}(\tau), \subseteq)$ and equals $V_1 \wedge V_2$, by the completeness theorem. Obviously the set $E_1 \cap E_2$ is closed under (1)-(6). Thus $E_1 \cap E_2$ is the greatest set, closed under (1)-(6) and contained in E_1 and E_2. Thus, by the completeness theorem, the l.u.b. (V_1, V_2) exists in $(\mathcal{L}(\tau), \subseteq)$ and equals $V_1 \vee V_2$.

Theorem 4.1.2 *The lattice $(\mathcal{L}(\tau), \wedge, \vee)$ of all hypervarieties of type τ is isomorphic to a sublattice of the lattice $(L(\tau), \wedge, \vee)$ of all varieties of type τ.*

Proof We consider the map $k : \mathcal{L}(\tau) \to L(\tau)$ which is defined for a hypervariety C of type τ in the following way. If $C = \{V_i : i \in I\}$, then $k(C) = \bigcup(V_i : i \in I)$, i.e., $k(C)$ is the class of all algebras contained in the varieties of the hypervariety C.

Because C is a hypervariety, $k(C)$ is closed under $\mathbf{H}, \mathbf{S}, \mathbf{P}$ and hence is a variety. From $C_1 \subseteq C_2$ it is easy to see that $k(C_1) \subseteq k(C_2)$, i.e. k is monotone. Now let $k(C_1) = k(C_2)$ for hypervarieties C_1, C_2, and let E_i be the set of hyperidentities of type τ holding in C_i, $i = 1, 2$. Let (T_1, T_2) be a hyperidentity of E_1. As $k(C_1) = k(C_2)$, all algebras of $k(C_2)$ satisfy the hyperidentity (T_1, T_2), i.e. $E_1 \subseteq E_2$. Similarly $E_2 \subseteq E_1$. We conclude that $E_1 = E_2$, and $C_1 = C_2$. Let $\mathbf{A} \in k(C_1) \wedge k(C_2)$ and let E_i be the set of all hyperidentities holding for C_i $i = 1, 2$. Then $H_\tau(\mathbf{A}) \supseteq [E_1 \cup E_2]$. Furthermore, $[E_1 \cup E_2]$ is the set of hyperidentities defining $C_1 \wedge C_2$ by Proposition 4.1.1. Hence $\mathbf{A} \in k(C_1 \wedge C_2)$. Since k is monotone, we conclude that $k(C_1) \wedge k(C_2) = k(C_1 \wedge C_2)$. Now take $\mathbf{A} \in k(C_1 \vee C_2)$. Thus the algebra $\mathbf{A}$ satisfies the hyperidentities of $C_1 \wedge C_2$, i.e. $H_\tau(\mathbf{A}) \supseteq E_1 \cap E_2$, by Proposition 4.1.1. Because $E_1 \cap E_2$ is closed under rules (1)-(6), we conclude also by Proposition 4.1.1 that $\mathbf{A} \in k(C_1) \vee k(C_2)$. Since k is monotone, we have $k(C_1 \vee C_2) = k(C_1) \vee k(C_2)$, i.e. k is a lattice homomorphism.

Remark 4.1.3 The lattice $(\mathcal{L}(\tau), \wedge, \vee)$ is a complete lattice.

Proof Similarly, as in the proof of Proposition 4.1.1 it is easy to see that for a family

$(V_i : i \in L)$ of hypervarieties of type τ, defined by the sets E_i, $i \in I$, of hyperidentities, the hypervarieties $\bigwedge(V_i : i \in I)$ and $\bigvee(V_i : i \in I)$ are defined by the sets of hyperidentities $\bigcup(E_i : i \in I)$ and $\bigcap(E_i : i \in I)$, respectively.

Let V be a variety of type τ. Then $h_\tau(V)$ denotes the hypervariety of type τ and $h(V)$ the hypervariety which is generated by V, i.e. defined by all the hyperidentities of V. Obviously, we have $h(V) \subseteq h_\tau(V)$.

Proposition 4.1.4 *The map $h_\tau : \mathcal{L}(\tau) \to L(\tau)$ defined by $V \to h_\tau(V)$ is a surjective complete join-homomorphism.*

Proof Let C be a hypervariety of type τ, $C = \{V_i : i \in I\}$ where V_i are varieties of type τ for $i \in I$. Take $V = \bigvee(V_i : i \in I)$, the join the family $\{V_i : i \in I\}$ in the lattice $L(\tau)$. Then C is generated by V, i.e. $h_\tau(V) = C$. Hence h_τ is surjective. Obviously h_τ is a monotone map.

To show that $h_\tau(\bigvee(V_i : i \in I)) = \bigvee(h_\tau(V_i) : i \in I)$, notice that the hypervariety C, generated by the join of the family $(V_i : i \in I)$ of varieties of type τ, is defined by the set $\bigcap(E_\tau(V_i) : i \in I)$ of hyperidentities. But this is exactly the join of the hypervarieties $(h_\tau(V_i) : i \in I)$.

Remark 4.1.5 According to the results of [Bergman 81] the map h is not one-to-one in the case of semigroups and groups considered as varieties of the same type.

Proposition 4.1.6 *V is a solid variety if and only if there exists a hypervariety C of the same type, such that $k(C) = V$.*

Proof Let V be a solid variety, i.e. $E_\tau(V) = H_\tau(\mathrm{Id}(V))$. Take the set $\Sigma = E_\tau(V)$ of all hyperidentities of V and the hypervariety $C = \{V_i : i \in I\}$ of the same type as V, defined by Σ, i.e. $E_\tau(V_i) \supseteq \Sigma$, for all $i \in I$. Thus $V \in C$ and $k(C) = \bigcup(V_i : i \in I)$ and $E_\tau(k(C)) = \bigcap(E_\tau(V_i) : i \in I) = E_\tau(V)$, because $V \in C$ and $E_\tau(V_i) \supseteq \Sigma$ for all $i \in I$. V is solid; thus $E_\tau(V) = H_\tau(\mathrm{Id}(V))$. Also, $\mathrm{Id}(V_i) \supseteq I_V(E_\tau(V_i)) \supseteq I_V(E_\tau(V)) = \mathrm{Id}(V)$ for $i \in I$. Thus $V_i \subseteq V$, for all $i \in I$, i.e. $\mathrm{Id}(k(C)) = \bigcap(\mathrm{Id}(V_i) : i \in I) = \mathrm{Id}(V)$, and thus $k(C) = V$.

Now let $V = k(C)$ for some hypervariety $C = \{V_i : i \in I\}$, such that $E_\tau(V_i) \supseteq \Sigma$, for some set Σ of hyperidentities, which is closed under rules (1)-(6) by the completeness theorem. Thus a solid variety W, defined by $I_V(\Sigma)$ belongs to C, i.e. $\mathrm{Id}(W) = I_V(\Sigma)$. Thus $V = k(C) = \bigcup(V_i : i \in I)$ and $\mathrm{Id}(V) = \bigcap(\mathrm{Id}(V_i) : i \in I) = I_V(\Sigma)$, because $\mathrm{Id}(V_i) \supseteq I_V(E_\tau(V_i)) \supseteq I_V(\Sigma)$ for all $i \in I$ and $W \in \{V_i : i \in I\}$ and also $E_\tau(V) = \bigcap(E_\tau(V_i) : i \in I) = \Sigma$, i.e. $\mathrm{Id}(V) = I_V(E_\tau(V)$, and thus V is a solid variety.

Corollary 4.1.7 *Let $S(\tau)$ be the set of all solid varieties of type τ. Then $S(\tau)$ forms a (complete) sublattice of $L(\tau)$.*

This follows from Remark 4.1.3, Theorem 4.1.2 and Proposition 4.1.6.

4.2 Solid kernels and solid envelopes

We have already considered solid envelopes in the first paragraph. Now we wish to study this concept from the point of view of monotone operators.

Notation 4.2.1 Let Σ be the identities of the variety V of type τ and $H_\tau(\Sigma)$ the set of all transformations to hyperidentities. We define the *solid kernel* $k(V)$ of the variety V as the subvariety of V which is given by $\mathrm{Id}(H_\gamma(\Sigma))$.

Example Let D be the variety of distributive lattices. Then the solid kernel $k(V)$ is the trivial variety because we have $x = y$ from $F(x,y) = F(y,x)$ by $x \wedge y = y \wedge x$.

If V is solid we have $k(V) = V$. Let $U \subseteq V$ be varieties for U, V of some type τ and let $\Sigma(U), \Sigma(V)$ denote the identities of U, V respectively. From $U \subseteq V$ we have $\Sigma(V) \subseteq \Sigma(U), H_\gamma(\Sigma(V)) \subseteq H_\gamma(\Sigma(U))$ and $\mathrm{Id}H_\gamma(\Sigma(V)) \subseteq \mathrm{Id}H_\gamma(\Sigma(V))$. Hence $k(U) \subseteq k(V)$.

Theorem 4.2.2 [Graczyńska 89] *Let $L(\tau)$ be the lattice of varieties of type τ and $S(\tau)$ the lattice of solid varieties of type τ. Then $k : L(\tau) \to S(\tau)$ is a meet-homomorphism.*

Proof We have to show that $k(V_1 \cap V_2) = k(V_1) \cap k(V_2)$ for $V_1, V_2 \in L(\tau)$. Obviously k is a monotone map and hence we have $k(V_1 \cap V_2) \subseteq k(V_1) \cap k(V_2)$. For the other direction let ε be a hyperidentity which holds for $k(V_1) \cap k(V_2)$. The hyperidentity ε is a hyperconsequence of $H(V_1)$ as well as $H(V_2)$ and hence also of $H(V_1 \cap V_2)$.

Remark $k(V)$ is the greatest solid variety contained in V. The theorem holds also for complete meets.

We have already seen that for every variety V there exists a least solid variety $s(V)$ which contains V. We call $s(V)$ the *solid envelope* of V.

Remark Let V be a variety which is generated by a set K of subdirectly irreducible algebras A_i, $i \in I$. Then $s(V)$ is generated by $\mathbf{D}(K)$ the set of the derived algebras of V.

We use the fact that $s(V) = \mathbf{HSPD}(V)$.

Example Let D be the variety of distributive lattices. Let $D_2 = (\{0,1\}; \wedge, \vee)$ be the two-element simple lattice. Then we have the following simple derived algebras:

$$E_1 = (\{0,1\}, e_1^2, e_2^2), \qquad S_1 = (\{0,1\}, e_1^2, \wedge), \qquad S_2 = (\{0,1\}, \wedge, e_1^2)$$
$$S_3 = (\{0,1\}; e_1^2, \vee), \qquad S_4 = (\{0,1\}; \vee, e_1^2), \qquad S_5 = (\{0,1\}; \vee, \wedge)$$
$$S_6 = (\{0,1\}; e_1^2, \vee)$$

Nevertheless, it will usually be difficult to find the subdirectly irreducible algebras of $s(V)$.

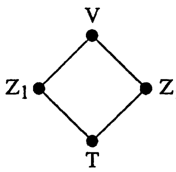
Let $U \subseteq V$ for varieties U, V of some type τ. Then $H_\gamma(V) \subseteq H_\gamma(U)$ for the sets of hyperidentities of type τ which hold for V or U, respectively. Hence $s(U) \subseteq s(V)$.

Theorem 4.2.3 [Graczyńska 89] *Let $L(\tau)$ be the lattice of varieties of type τ* and *$S(\tau)$ the* *lattice of solid varieties of type τ. Then $s : L(\tau) \to S(\tau)$ is a join-homomorphism.*

Proof We have to show that $s(V_1 \vee V_2) = s(V_1) \vee s(V_2)$ for every $V_1, V_2 \in L(\tau)$. Obviously s is monotone and hence $s(V_1 \vee V_2) \leq s(V_1) \vee (V_2)$. Let $\Sigma, \Sigma_1, \Sigma_2$ be the sets of hyperidentities of $V_1 \vee V_2$, V_1 and V_2, respectively. We have $\Sigma \supseteq \Sigma_1 \cap \Sigma_2$ because of the rule (6) and we conclude that $s(V_1) \vee s(V_2) \supseteq s(V_1 \vee V_2)$.

Remark The theorem holds for complete joins.

Example (Graczyńska) Consider the following varieties of semigroups



T the trivial variety of semigroups;

Z_l the variety defined by $xy = x$;

Z_r the variety defined by $xy = y$;

$V = Z_l \vee Z_r$.

V is the solid variety defined by $xyz = xz$ and $x^2 = x$. We have $s(Z_l \cap Z_r) = s(T) = T$ but $s(Z_l) = s(Z_r) = V$ and hence $s(Z_l \cap Z_r) \neq s(Z_l) \cap s(Z_r)$. Furthermore we have $k(Z_l \vee Z_r) = K(V)$ but $k(Z_l) \vee k(Z_r) = T \vee T = T$.

Problem 4.2.4 Let V be a given solid variety of type τ.

a) Describe all varieties W of type τ such that $k(W) = V$.

b) Describe all varieties of W of type τ such that $s(W) = V$.

The extreme cases, where V is trivial or the variety of all algebras of type τ, deserve special interest.

Problem 4.2.5 Let V be a given variety of type τ. The variety W of type τ is called a *flexible complement* of V *if* $k(V) = k(W)$ and $s(V) \cap W = V$. Determine all maximal flexible complements. (As an example consider D, the variety of distributive lattices.)

4.3 Clone congruences

The results of this section are due to Schweigert (compare [Schweigert 87a, 89]).

Definition 4.3.1 Let $\mathbf{H} = (H; *; \xi, \tau, \Delta, e)$ be a clone of functions on a set A. An equivalence relation ζ is called a *clone congruence* of $\mathbf{H}$ if ζ is compatible with the clone operations $*, \xi, \tau, \Delta$.

Example Consider a clone $\mathbf{H}$ of function on a set A. Then $\kappa = \{(f,g) |\ \mathrm{ar}\, f = \mathrm{ar}\, g,\ f, g \in H\}$ is a clone congruence ($\mathrm{ar}\, f = m$ denotes the arity of the function $f : A^m \to A$). Obviously κ is an equivalence relation. Let $(f,g) \in \kappa$, $h \in H$ with $\mathrm{ar}\, f = \mathrm{ar}\, g = m$ and $\mathrm{ar}\, h = n$. Then $(f * h, g * h) \in \kappa$ because $\mathrm{ar}\, f * h = n + m - 1 = \mathrm{ar}\, g * h$. Similarly we have $(h * f, h * g) \in \kappa$. Now let $(d, b) \in \kappa$. For $(f,g) \in \kappa$ we have $(f * d, g * d) \in \kappa$ and we have proved the compatibility of κ with $*$.

Notation 4.3.2 The clone congruence κ of $\mathbf{H}$ is called the *arity congruence* of $\mathbf{H}$. On every clone of functions there are at least three clone congruences κ_0, κ, and κ_1, where κ_0 is the identity and κ_1 the all relation.

Remark 4.3.3 If ζ is a clone congruence of $\mathbf{H}$ with $\zeta \neq \kappa_1$ then $\kappa_0 \subseteq \zeta \subseteq \kappa$.

Proof We assume that $\zeta \not\subseteq \kappa$. Then there are functions $f, g \in H$ ar $f = n > m =$ ar g with $(f, g) \in \zeta$. Let $\bar{f}(x_1, \ldots, x_n) = f(x_1, \ldots, x_1, x_2)$ $\bar{g}(x_1, \ldots, x_m) = g(x_1, \ldots, x_1)$. We have $(\bar{f}, \bar{g}) \in \zeta$ and $(e_n^n * \bar{f}, e_n^n * \bar{g}) \in \zeta$ where $e_n^n(x_1, \ldots, x_n) = x_n$. We conclude that $(e_{n+1}^{n+1}, e_n^n) \in \zeta$. But from this it follows immediately that any two given functions are in the clone congruence ζ and $\zeta = \kappa_1$.

Fact κ is a maximal clone congruence.

Notation Every clone congruence $\zeta \subseteq \kappa$ is called a *proper clone congruence.*

Notations Let $\mathbf{F}(X) = (F(X), \Omega)$ denote the free algebra of the variety V generated by X. Con $\mathbf{F}(X)$ is the lattice of the fully invariant congruences of $\mathbf{F}(X)$. By the terms of $\mathbf{F}(X)$ we define term functions on the set $\mathbf{F}(X)$. $\mathbf{T}(X)$ denotes the clone of all term functions on $\mathbf{F}(X)$. Con $\mathbf{T}(X)$ is the lattice of all proper clone congruences of $\mathbf{T}(X)$.

Theorem 4.3.4 *Every proper clone congruence of the clone* $\mathbf{T}(X)$ *of term functions corresponds to a fully invariant congruence of the free algebra* $\mathbf{F}(X)$. *There is a lattice isomorphism* $h :$ Con $\mathbf{T}(X) \rightarrow$ Con $\mathbf{F}(X)$.

Proof We define a map $h :$ Con $\mathbf{T}(X) \rightarrow$ Con $\mathbf{F}(X)$ in the following way. $(t(x_1, \ldots, x_k), u(x_1, \ldots, x_k)) \in \theta_h$ if and only if $(t, u) \in \theta$ for $t(x_1, \ldots, x_k) \in \mathbf{F}(X)$, $u(x_1, \ldots, x_k) \in \mathbf{F}(X)$, and the corresponding term functions, $t, u \in \mathbf{T}(X)$. The equivalence relation θ_n is compatible with any operation $w \in \Omega$ of the free algebra $\mathbf{F}(X)$ because we have $(w(t_1, \ldots, t_n), w(u_1, \ldots, u_n)) \in \theta$ for $(t_i, u_i) \in \theta$, $i = 1, \ldots, n$. If κ is an endomorphism $\kappa : \mathbf{F}(X) \rightarrow \mathbf{F}(X)$ with $\kappa(x_i) = s_i$, $i = 1, \ldots, n$, then by the substitution property of a clone we have $(t(s_1, \ldots, s_n), u(s_1, \ldots, s_n)) \in \theta$ and hence $(\kappa(t), \kappa(u)) \in \theta_h$. Therefore θ_h is fully invariant. On the other hand if θ_h is a fully invariant congruence of $\mathbf{F}(X)$ and $(t(x_1, x_2, \ldots, x_k), s(x_1, x_2, \ldots, x_m)) \in \theta_h$, then by adding fictitious variables we get pairs (t, s) of term functions on the set A with ar $t =$ ar s. The set θ of these pairs is an equivalence relation on $\mathbf{T}(X)$ which is compatible with the substitution because θ_h is a congruence of $\mathbf{F}(X)$. But θ is also compatible with a permutation π of the variables, as $\kappa(x_i) = x_{\pi(i)}$ extends to an endomorphism and θ_h is fully invariant. The same argument holds for the identification of variables. Hence θ is a proper clone congruence of $\mathbf{T}(X)$. $h :$ Con$T(X) \rightarrow$ Con$F(X)$ with $h(\theta) := \theta_h$ is a lattice isomorphism.

Corollary 4.3.5 *There is a polarity (dual isomorphism) from the lattice* Con $\mathbf{T}(X)$ *of the proper clone congruences to the lattice* $L(V)$ *of all subvarieties of the variety* V.

Notations $\Omega(X)$ denotes the set of all fundamental operations f_δ of $\mathbf{F}(X) = (F(X), \Omega)$. By definition of $\mathbf{T}(X)$ they are contained in $\mathbf{T}(X)$. A *term substitution* β is a map β :

$\Omega(X) \to T(X)$ such that ar $f_\delta =$ ar $\beta(f_\delta)$. We write

$$\beta(f_\delta)(x_1, \ldots, x_n) = \beta(f_\delta(x_1, \ldots, x_n)).$$

As an example consider $(\mathbf{Z}; +)$ with $\beta(x + y) = ax + bx$ for some fixed $a, b \in \mathbf{Z}$.

Proposition 4.3.6 *The variety V is solid if and only if every term substitution $\beta : \Omega(X) \to T(X)$ can be extended to a clone endomorphism $\bar{\beta} : T(X) \to T(X)$.*

Proof If V is solid then we define $\bar{\beta}(t)$ for $t(x_1, \ldots, x_n) \in F(X)$ by the term where every operation symbol is substituted by a term according to the map β. If this map $\bar{\beta}$ is well-defined, then it will obviously be a clone endomorphism. Therefore let us consider the equation $t_1 = t_2$ in $T(X)$, i.e. $t_1(x_1, \ldots, x_n) = t_2(x_1, \ldots, x_n)$. As V is solid, any such term substitution provides a valid equation for V. Therefore we have $\bar{\beta}(t_1) = \bar{\beta}(t_2)$.

We repeat the

Definition A congruence θ of $\mathbf{A} = (A, \Omega)$ is called *totally invariant* if $(a, b) \in \theta$ implies $(h(a), h(b)) \in \theta$ for every type preserving weak endomorphism h of $\mathbf{A}$ and every $a, b \in A$.

Remark A totally invariant congruence is also fully invariant.

Theorem 4.3.7 *Every fully invariant proper clone congruence of the clone $\mathbf{T}(X)$ of term functions corresponds to a totally invariant congruence of the free algebra $\mathbf{F}(X)$. There is a lattice homomorphism $s : \mathrm{Con}_f\mathbf{T}(X) \to \mathrm{Con}_t\mathbf{F}(X)$.*

Proof Let θ be a fully invariant proper clone congruence of $T(X)$. We define $(t(x_1, \ldots, x_k), u(x_1, \ldots, x_k)) \in \theta_s$ if and only if $(t, u) \in \theta$ for the corresponding term functions $t, u \in \mathbf{T}(X)$. We have already shown in the proof of Theorem 4.3.4 that θ_s is a fully invariant congruence of $\mathbf{F}(X)$. Now let $h : F(X) \to F(X)$ be a type preserving weak endomorphism. Let $\bar{h} : T(X) \to T(X)$ be defined by $\bar{h}(t) = s$ if and only if $h(t(x_1, \ldots, x_k)) = s(x_1, \ldots, x_k)$. $\bar{h}$ is compatible with the substitution, i.e., $\bar{h}(u * v) = \bar{h}(u) * \bar{h}(v)$ because

$$h(u(v(x_1, \ldots, x_n), x_2, \ldots, x_{m+n-1})) = h(u(h(v(x_1, \ldots, x_n), x_2, \ldots, x_{m+n-1}))).$$

Obviously $\bar{h}$ is compatible with the other operations ξ, τ, Δ of a clone. h is a clone endomorphism and we have $(\bar{h}(t), \bar{h}(u)) \in \theta$. Hence θ_s is totally invariant.

On the other hand, if θ_s is a totally invariant congruence of $\mathbf{F}(X)$, then from Theorem 4.3.4 it follows that θ is a proper clone congruence. Let $\bar{f} : \mathbf{T}(X) \to \mathbf{T}(X)$ be a clone endomorphism. Let $f : \mathbf{F}(X) \to \mathbf{F}(X)$ be defined by $f(u(x_1, \ldots, x_k)) = v(x_1, \ldots, x_k)$ if and only if $\bar{f}(u) = v$. f is type-preserving. Let $w \in \Omega$ for $\mathbf{F}(X) = (F(X), \Omega)$. Then

$$\begin{aligned} &f(w(u_1(x_1, \ldots, x_{k1}), \ldots, u_n(x_1, \ldots, x_{kn}))) \\ &\quad = f(w)(f(u_1(x_1, \ldots, x_{k1})), \ldots, f(u_n(x_1, \ldots, x_{kn}))) \end{aligned}$$

by the substitution property of f. f is a weak endomorphism. Hence $(\bar{f}(u), \bar{f}(v)) \in \theta$ because $(f(u(x_1, \ldots, x_k), f(v(x_1, \ldots, x_k)) \in \theta_s$. θ is a fully invariant proper clone congruence.

Proposition 4.3.8 *Every totally invariant congruence of the free algebra $\mathbf{F}(X)$ corresponds to a solid subvariety of V.*

Proof Let θ_s be a totally invariant congruence of $\mathbf{F}(X)$. Then θ_s is fully invariant and corresponds to a subvariety U of V. Consider a term substitution $\beta : \Omega(X) \to T_U(X)$ where $T_U(X)$ is the clone of all term functions in the variety U. Consider $\bar{\beta} : T_U(X) \to T_U(X)$ as in Proposition 4.3.6. Let $t_1 \theta t_2$. Hence $(t_1(x_1, \ldots, x_k), t_2(x_1, \ldots, x_k)) \in \theta_s$ and $\bar{\beta}(t_1) = \bar{\beta}(t_2)$. $\bar{\beta}$ is a well-defined clone endomorphism of $T_U(X)$ and U is solid.

Corollary 4.3.9 *There is a polarity (dual isomorphism) from the lattice* $\mathrm{Con}_t\mathbf{F}(X)$ *of all totally invariant congruences of* $\mathbf{F}(X)$ *to the lattice* $\mathbf{L}_s(V)$ *of the solid subvarieties of* V.

Remark The solid kernel $k(V)$ is the greatest solid subvariety of V and the trivial variety is the least solid subvariety of V.

The meet of totally invariant congruences is again a totally invariant congruence. The all congruence is totally invariant. Hence there exists a least totally invariant congruence δ which is the identity relation of Con $\mathbf{F}(X)$ only in case that V is solid.

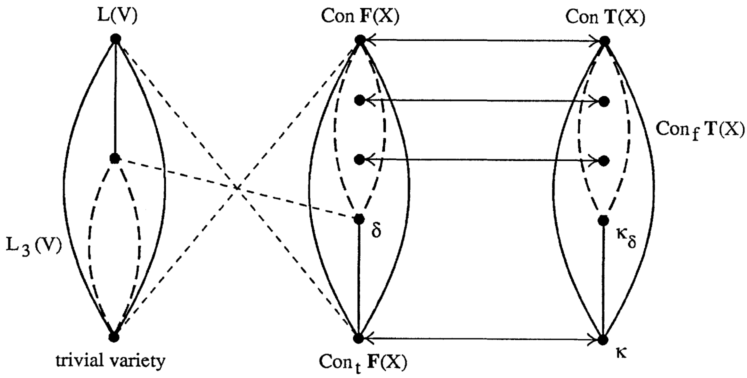


Figure 4.3.10

Con $\mathbf{T}(X)$:	= lattice of the proper clone congruences of the clone $\mathbf{T}(X)$
Con $\mathbf{F}(X)$:	= lattice of the fully invariant congruences of the free algebra $\mathbf{F}(X)$
$L(V)$:	= lattice of all subvarieties of the variety V
$\mathrm{Con}_f\mathbf{T}(X)$:	= lattice of the fully invariant proper clone congruences
$\mathrm{Con}_t\mathbf{F}(X)$:	= lattice of the totally invariant congruences of the free algebra $\mathbf{F}(X)$
κ:	= arity congruence
κ_δ:	= least fully invariant proper clone congruence
δ	= least totally invariant congruence of $\mathbf{F}(X)$

4.4 Subdirectly irreducible clones

Notations Let $O_A = (P_A; *, \xi, \tau, \Delta, e)$ be the clone of all functions on the set A. A function $f \in O_A$ is called a w-function if $x_i = w$ for some i, $1 \leq i \leq n$, implies $f((x_1, \ldots, x_n) = w$. Obviously the set of all w-functions of O_A forms a subclone $\mathbf{H}_w$.

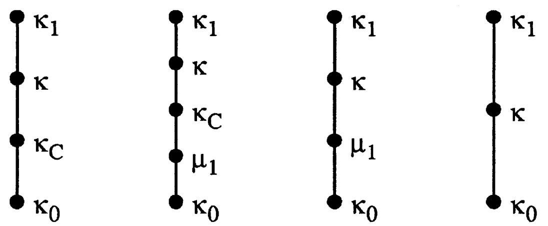
We define the relation κ_w on O_A by taking $(f, g) \in \theta_w$ if and only if either $f = g$ or f, g are functions of different arity which take the constant value w on A. It is easy to check that θ_w is a clone congruence of $\mathbf{H}_w$. With this notation we present

Theorem 4.4.1 *Let* $\mathbf{H}$ *be a subclone of* O_A *such that* $\mathbf{H}$ *properly contains the clone* $\mathbf{H}_w$. *Then the only clone congruences of* $\mathbf{H}$ *are* $\kappa_0, \kappa, \kappa_1$.

Corollary 4.4.2 *Every primal algebra* $\mathbf{A} = (A, \Omega)$ *has a subdirectly irreducible clone* $T(\mathbf{A})$ *of term functions.*

In the following we restrict our consideration to Boolean clones (clones of functions on the set $A = \{0, 1\}$), and define $(f, g) \in \kappa_c$ if and only if $\operatorname{ar} f = \operatorname{ar} g$ and there is an element $c \in \{0, 1\}$ with $f(x_1, \ldots, x_n) = g(x_1, \ldots, x_n) + c$. $(f, g) \in \mu$ if and only if $f = g$ or there is an element $n \in \mathbf{N}$ such that $f, g \in \{c_0^n, c_1^n\}$ with the constant functions $c_0^n(x_1, \ldots, x_n) = 0$ and $c_1^n(x_1, \ldots, x_n) = 1$.

Theorem 4.4.3 [Gorlov] *The congruence lattices of all Boolean clones are of the following form:*



$\mathbf{L}_1, \mathbf{L}_5, \mathbf{O}_4$ $\qquad$ $\mathbf{O}_9$ $\qquad$ $\mathbf{O}_8$ $\qquad$ *all other Boolean clones.*

Corollary 4.4.4 *Every Boolean clone is subdirectly irreducible.*

Now we change the direction of this topic and consider algebras $\mathbf{A}$ with a subdirectly irreducible clone $T(\mathbf{A})$.

Definition 4.4.5 The algebra $\mathbf{A} = (A, \Omega)$ is called *2-subdirectly irreducible* if $\mathbf{A}$ is a subdirectly irreducible algebra and $T(\mathbf{A})$ is a subdirectly irreducible clone.

Example Every algebra $\mathbf{A} = (\{0,1\}, \Omega)$ is 2-subdirectly irreducible.

Problem 4.4.6 Is every solid variety generated by its 2-subdirectly irreducible algebra?

4.5 Clone-products of algebras

Definition 4.5.1 The algebra $\mathbf{A} = \mathbf{A}_1 * \mathbf{A}_2$ is called a *direct clone-product* of the algebra $\mathbf{A}_1, \mathbf{A}_2$ provided that there exist clone congruences ζ_1, ζ_2 of $T(\mathbf{A})$ such that

(i) for every $f \in T(A)$ there is an $\bar{f} \in T(A)$ with $\bar{f} = \left\{ \begin{array}{l} f/A_i^n \\ e/(A^n \backslash A_i^n) \end{array} \right\}$ such that $(f, \bar{f}) \in \zeta_i, i = 1, 2, e/(A^n \backslash A_i^n)$;

(ii) $\zeta_1 \wedge \zeta_2 = \omega$, where ω is the identity relation;

(iii) $\zeta_1 \vee \zeta_2 = \zeta_2 \circ \zeta_1 = \kappa$ where κ is the arity relation.

e is the first projection $e(x_1, \ldots, x_n) = x_1$.

Definition 4.5.2 If $\mathbf{A}$ is isomorphic to a subalgebra for a direct clone-product of $\mathbf{A}_1, \mathbf{A}_2$, then $\mathbf{A}$ is called a *subdirect clone-product* of $\mathbf{A}_1, \mathbf{A}_2$.

Theorem 4.5.3 *Let* $\mathbf{A}_1 = (A_1, \Omega_1)$ *and* $\mathbf{A}_2 = (A_2, \Omega_2)$ *be algebras of not necessarily the same type. Let* $A_1 \cap A_2 = \emptyset$ *and* $A_1 \cup A_2 = A$. *Let* $T(\mathbf{A})$ *be the clone generated by the functions* f *such that* $f/A_i^n \in T(A_i)$, $i = 1, 2$, *and* $f(a_1, \ldots, a_n) = a_1$ *for* $(a_1, \ldots, a_n) \notin A_1^n$ *and* $(a_1, \ldots, a_n) \notin A_2^n$. *Let* Ω *be a set of generators of* $T(\mathbf{A})$. *Then* $\mathbf{A} = (A, \Omega)$ *is a direct clone-product of* $\mathbf{A}_1, \mathbf{A}_2$.

Proof We define $f\theta_i g$ if and only if f and g have the same arity and $f/A_i^n = g/A_i^n$, $i = 1, 2$, and $f(x_1, \ldots, x_n) = g(x_1, \ldots, x_n)$ for $(x_1, \ldots, x^n) \notin A_2^n$.

By definition, θ_i is an equivalence relation contained in κ. Also by definition the condition (i) is fulfilled. Obviously θ_i is a clone congruence (i.e. compatible with the clone operations).

From $f\theta_1 \wedge \theta_2 g$ it follows that $f(x_1, \ldots, x_n) = g(x_1, \ldots, x_n)$ for every $(x_1, \ldots, x_n) \in A^n$. Hence we have $\theta_1 \wedge \theta_2 = \omega$.

Let $f, g \in T(A)$ and $(f, g) \in \kappa$. Then we consider $h : A_n \to A$ such that $h/A_1^n = f/A_1^n$, $h/A_2^n = g/A_2^n$ and $h(a_1, \ldots, a_n) = a_1$ elsewhere. Obviously we have $f\theta_1 h$ and $h\theta_2 g$. Hence $f\theta_1 \circ \theta_2 h$ and also $\theta_1 \circ \theta_2 = \theta_2 \circ \theta_1 = \kappa$.

Lemma 4.5.4 *Let* $\mathbf{A} = \mathbf{A}_1 * \mathbf{A}_2$ *be a subdirect clone product of* $\mathbf{A}_1$ *and* $\mathbf{A}_2$. *Then* $T(\mathbf{A})$ *is a subdirect product of* $T(\mathbf{A}_1)$ *and* $T(\mathbf{A}_2)$.

This follows from Definition 4.5.1 (ii).

4.6 Clone-unions of algebras

The construction in section 4.5 has a lot of beautiful properties which will fail for the method of clone-union.

Definition 4.6.1 Let $\mathbf{A}_i = (A, \Omega_i)$, $i = 1,2$, be algebras and let $\Omega = \Omega_1 \cup \Omega_2$ and $\tau = \tau_1 \cup \tau_2$. The algebra $\mathbf{A}$ is a *clone-union* of the algebras $\mathbf{A}_1$ and $\mathbf{A}_2$ if the following holds:

(1) $T(A_1) \cap T(A_2) = P$, where P is the clone of projections on A;

(2) $\mathbf{A}$ is weakly isomorphic to (A, Ω) of type τ.

Example 4.6.2 The distributive lattice $D = (\{0,1\}; \wedge, \vee)$ is a clone-union of the semilattices $D_1 = (\{0,1\}; \wedge)$ and $D_2 = (\{0,1\}; \vee)$.

Example 4.6.3 The cyclic group $C_3 = (A; +)$ of order 3, where $C = \{0,1,2\}$ is the clone-union of the groupoids $C_3^{2x+y} = (A; +_1)$ and $C_3^{2x+2y} = (A; +_2)$. The fundamental term function $x +_1 y$ of C_3^{2x+y} is defined in terms of C_3 by $x +_1 y = 2x + 0y$; similarly, $x +_2 y = 2x + 2y$.

This construction may have many disadvantages but can also be considered as a tool to decompose algebras. It is far from being unique in any sense. Nevertheless, let us state the following.

Problem 4.6.4 Can every finite abelian group be presented as a direct product of clone-unions of simple groupoids?

Part 5

Hybrid logic

5.1 Hyperquasi-identities

In the following, the approach to hyperidentities in sections 1.1 and 1.2 is extended to quasi-identities and sentences.

We develop logics containing the hypersubstitution as an additional rule and prove completeness. Compared to section 1.2 we have chosen a different way to these results. Proofs in logics with hypersubstitutions are transformed into proofs in logics without hypersubstitutions and vice versa. This method clearly points out that a logic with hypersubstitution has more expressive power and the proofs are usually shorter. These logics with hypersubstitutions proceed beyond first order (but they are only a fragment of second order logic).

One can strengthen the expressive power of these logics further if operation symbols and hypervariables are admitted simultaneously in the language.

All sentences, quasi-identities and identities are written without quantifiers but are considered as universally closed. We consider quasi-varieties V of algebras of a given type.

Recall that a quasi-identity is an implication of the form

$$(t_0 = s_0) \wedge (t_1 = s_1) \wedge \ldots \wedge (t_{n-1} = s_{n-1})) \rightarrow (t_n = s_n),$$

where $s_0, \ldots, s_n, t_0, \ldots, t_n \in T(V)$. By analogy we have:

Definition 5.1.1 A *hyperquasi-identity* is an implication of the form

$$(T_0 = S_0) \wedge (T_1 = S_1) \wedge \ldots \wedge (T_{n-1} = S_{n-1})) \rightarrow (T_n = S_n),$$

where $T_i = S_i$ are hyperidentities, $i = 1, \ldots, n$.

Definition 5.1.2 A mapping $\sigma : \{F_i | i \epsilon I\} \rightarrow T(V)$ which assigns to every n_i-ary hypervariable an n_i-ary term is called a *hypersubstitution.* Such a mapping σ can be extended to a mapping $\bar{\sigma}$ from the set of hyperterms into $T(V)$ by defining recursively $\bar{\sigma}(x) = x$ for every variable x in $T(V)$, and

$$\bar{\sigma}(F_i(T_1, \ldots, T_n)) = \sigma(F_i)(\bar{\sigma}(T_1), \ldots, \bar{\sigma}(T_n)).$$

In the following, both maps σ, $\bar{\sigma}$ are denoted by σ only, and we call $\sigma(T) = \sigma(S)$ a transformation of the hyperidentity $T = S$ into an identity. Similarly we have transformation σ of hyperquasi-identities into quasi-identities. Z is the set of all these transformations. For a hyperquasi-identity e the set

$$Z(e) = \{\sigma(e) \mid \sigma \in Z\}$$

denotes all transformations of e.

Example Consider the following quasivariety V of type (2):

$(K1)$ $\quad x \circ (y \circ z) = (x \circ y) \circ z,$

$(K2)$ $\quad x \circ x = x,$

$(K3)$ $\quad (u \circ x) \circ (y \circ w) = (u \circ y) \circ (x \circ w),$

$(K4)$ $\quad (x \circ y = y \circ x) \rightarrow x = y.$

This quasivariety is not trivial as it contains for instance the algebra $(\{0,1\}; \circ)$ with $x \circ y = y$. The following is a list of the terms in two variables x, y of $T(V)$:

$$t_1(x,y) = t_2(x,y) = y, \qquad t_3(x,y) = x \circ y, \qquad t_4(x,y) = y \circ x,$$

$$t_5(x,y) = x \circ y \circ x, \qquad t_6(x,y) = y \circ x \circ y.$$

We consider the hyperquasi-identity

$$(F(x,y) = F(y,x)) \to (x = y).$$

If we replace the hypervariable F by the term t_5 this transformation produces

$$(x \circ y \circ x = y \circ x \circ y) \to (x = y).$$

Definition 5.1.3 A quasivariety V of type τ *satisfies a hyperquasi-identity* e if the set $Z(e)$ of quasi-identities holds for V.

Example $(F(x,y) = F(y,x)) \to (x = y)$ holds for the quasivariety V of the preceding example. We would have to consider all terms listed above but confine ourselves to t_5. Now $x \circ y \circ x = y \circ x \circ y$ implies $(x \circ y) \circ (y \circ x) = (x \circ y) \circ (x \circ y)$ by $(K2)$ and $(K1)$, and furthermore $x = y$ by $(K4)$.

Definition 5.1.4 A mapping

$$h : \{f_i \mid i \in I\} \to \{F_i \mid i \in I\}$$

which assigns to every n_i-ary operation symbol f_i an n_i-ary hypervariable F_i is called a *transformation of terms* if the variables $x, y, z, \ldots$ are left unchanged. Of course we extend to the set $T(V)$ of all terms recursively. The set of all these transformations is denoted by Z^{-1}.

Example The quasi-identity $(x \circ y \circ x = y \circ x \circ y) \to (x = y)$ is transformed to the hyperquasi-identity

$$(F(F(x,y),x) = F(F(y,x),y)) \to (x = y).$$

This hyperquasi-identity holds for the above quasivariety V because it can be derived from $h(K1)$, $h(K2)$, $h(K3)$ and $h(K4)$.

Definition 5.1.5 A quasivariety V is called *solid* if every quasi-identity of V can be transformed to a hyperquasi-identity which holds for V.

Notation Let Σ be the set of identities which hold for V. If V is solid, then $Z^{-1}(\Sigma) \subseteq E$, where E is the set of all hyperquasi-identities which hold for V.

Examples (1) The quasivariety V of type 2 with the axioms $K(1) - K(4)$ is solid. (2) Every hyperquasivariety of a given type (i.e. a quasivariety defined by hyperquasi-identities).

5.2 Preservation properties

We quote the following results.

Theorem 5.2.1 [Malcev 71] *A class K of algebras of a type τ is a quasivariety if and only if K is closed under the formation* $\mathbf{S}$ *of subalgebras,* $\mathbf{I}$ *isomorphic images and* $\mathbf{P}_R$ *reduced products.*

$$\mathbf{S}K \subseteq K, \;\; \mathbf{P}_R K \subseteq K.$$

(*Here we put* $\mathbf{P}_R := \mathbf{IP}_R$).

Theorem 5.2.2 *Let K be a class of algebras of type τ. $\mathbf{SP}_R K$ is the class of all models of the set of quasi-identities true in K.*

Theorem 5.2.3 *A class K of algebras of a type τ is a solid quasivariety if and only if we have*

$$\mathbf{S}K \subseteq K, \ \mathbf{P}_R K \subseteq K, \ \mathbf{D}K \subseteq K,$$

where $\mathbf{D}(K)$ is the class of all derived algebras of type τ of K (cf. section 1.4).

Proof by the following lemma.

Lemma 5.2.4 *A quasivariety V of type τ is solid if and only if it is closed under the condition*:

Let $\mathbf{A}$ *be an algebra of* V, *of type* $\tau = (n_1, n_2, \ldots n_\gamma, \ldots : \gamma < O(\tau))$.

$(*)$ *If t_γ is the realization of an n_γ-ary term operation of type τ in $\mathbf{A}$, then $\bar{\mathbf{A}} = (A; t_1, t_2, \ldots, t_\gamma, \ldots : \gamma < O(\tau))$ is an algebra of V.*

Proof Let V be a solid quasivariety. Consider the algebra

$$\bar{\mathbf{A}} = (A; t_1, t_2, \ldots, t_\gamma, \ldots : \gamma < O(\tau)).$$

The quasi-identities of V are transformed into hyperquasi-identities of V and hence hold for the term functions t_τ. In particular, they hold for $\bar{\mathbf{A}}$. Hence $\bar{\mathbf{A}} \in V$. Let the condition $(*)$ hold for V. Then the quasi-identities of V hold for all term functions of the suitable arity and hence are transformed into hyperquasi-identities, i.e., V is a solid variety.

Theorem 5.2.5 *K of type τ is a solid quasivariety if and only if*

$$K = \mathbf{SP}_R \mathbf{D} K.$$

Proof We have to show $\mathbf{DP}_R K \subseteq \mathbf{P}_R \mathbf{D} K$. For $\mathbf{B} \in \mathbf{DP}_R(K)$ we have $\mathbf{B}_0 = (A; t_0, t_1, \ldots, t_\gamma, \ldots)$ with $\mathbf{A} = (A; f_0, f_1, \ldots, f_\gamma, \ldots)$ and $\mathbf{A} = \prod \mathbf{A}_i$, $\mathbf{A}_i = (A_i; f_0, f_1, \ldots, f_\gamma, \ldots)$. Consider $\mathbf{B}_i := (A_i; t_0, t_1, \ldots, t_\gamma, \ldots)$; then we have $\mathbf{B} = \prod \mathbf{B}_i$ and hence $\mathbf{B} \in \mathbf{P}_R \mathbf{D}(K)$.

5.3 Solid classes of models

We are considering a class of relational structures of given type τ. The type of a structure is a sequence $(n_0, n_1, \ldots, n_\gamma, \ldots)$ of positive integers, $\gamma < O(\tau)$, where $O(\tau)$ is an ordinal. For every $\gamma < O(\tau)$ we have a predicate symbol r_γ for an n_γ-ary relation. Moreover, a symbol R_γ is associated to every γ. R_γ is called a hyperpredicate variable.

Definition 5.3.1 An *atomic hyperformula* is an expression of the form $P(T_1, \ldots, T_n)$, where P is an n-place hyperpredicate variable and $T_1, \ldots, T_n$ are hyperterms.

Definition 5.3.2 The *hyperformulas* are built up from the atomic formulas by use of the connective symbols and the quantifier symbol $(R \wedge Q)$, $\forall x_i R$.

Definition 5.3.3 A *hypersentence* is a hyperformula where every variable and every hyperpredicate variable is bound.

Example $\forall P\ \forall x \forall y\ (P(x,y) \to P(y,x))$. We only write: $P(x,y) \to P(x,y)$ dropping all quantifiers.

Notations 5.3.4 Given a class K of models of type τ and a hypersentence $C(R_1, \ldots, R_n)$ of type τ. Let σ be a map of all hyperpredicate variables into the set of quantifier-free formulas. σ transforms $C(R_1, \ldots, R_n)$ into a first-order formula $\sigma(C(R_1, \ldots, R_n))$. Let Z be the set of all these transformations. The hypersentence $C(R_1, \ldots, R_n)$ holds in the class K if for all $\sigma \epsilon Z$, $\sigma(C(R_1, \ldots, R_n))$ is a valid formula of first order for K. We write

$$\models_{\mathbf{hyp}} C(R_1, \ldots, R_n).$$

Similarly we define Σ of hypersentences and a hypersentence U.

Notations 5.3.5 Let $c(r_1, \ldots, r_n)$ be a quantifier-free formula of first order, and let m be the maximum of the arities of the predicate symbols $r_1, \ldots, r_n$. Then we define the *derived relation* r by

$$(x_1, \ldots, x_m) \in r \Rightarrow (x_1, \ldots, x_n) \in c(r_1, \ldots, r_n)$$

or in the usual notation:

$$r(x_1, \ldots, x_m) \Leftrightarrow c(r_1, \ldots, r_n)(x_1, \ldots, x_m).$$

Let $\mathbf{A} = (A, \rho)$ be a relational system of a class K. A derived relational system $\bar{\mathbf{A}} = (A, \bar{\rho})$ is a system where every relation in ρ is substituted by a derived relation of the same arity. $\mathbf{D}(K)$ denotes the class of all derived relational systems of K.

Definition 5.3.6 A class models of type τ is called *solid* if every sentence valid in K holds as hypersentence in K substituting the predicate symbols by hyperpredicate symbols of the same arity.

Notation We denote these transformations by h, the set of these transformations by Z^{-1} and we also write $Z^{-1}[\Sigma]$ of a set Σ of sentences.

Example of a solid model of type (2,2): $(A; p, q)$ with the axioms:

$$(*)\ p(x,y) \to p(y,x), \quad (**)\ q(x,y) \to q(y,x).$$

We show that $(***)\ P(x,y) \to P(y,x)$ is a hypersentence.

Proof If $w(x,y) := p(x,y)$, then $(***)$ holds by $(*)$. All sentences can be built up by the connectives $\neg, \to$.

(a) We assume $w(x,y) \equiv \neg(k(x,y))$ and $(***)$ holds for $k(x,y)$. We have

$$k(y,x) \to k(x,y),$$

and hence

$$\neg k(x,y) \to \neg k(y,x).$$

(b) We assume $w(x,y) \equiv (k(x,y) \to l(x,y))$. By $(***)$ we have

$$\begin{array}{ccc} k(x,y) & \leftrightarrow & k(y,x) \\ \downarrow & & \\ l(x,y) & \leftrightarrow & l(y,x). \end{array}$$

It follows that $k(y,x) \to l(y,x)$ and hence $w(x,y) \to w(y,x)$ q.e.d.

5.4 Completeness for hypersentences

We follow the notation of [Enderton 72] and present the following axiom schemes for a logic of hypersentences.

(1) Tautologies ;

(2) Substitution of variables: $\forall x\ \alpha \to \alpha^x_T$;

(3) $\forall P\ \alpha \to \alpha^P\ C(R_1, \ldots, R_n)$;

(4a) $\forall x\ (\alpha \to \beta) \to (\forall x \alpha \to \forall x \beta)$;

(4b) $\forall P\ (\alpha \to \beta) \to (\forall P \alpha \to \forall P \beta)$;

(5a) $\alpha \to \forall x \alpha$, where x does not occur free in α;

(5b) $\alpha \to \forall P \alpha$, where P does not occur free in α.

Rule of inference: Modus ponens $\dfrac{\alpha, \alpha \to \beta}{\beta}$.

Definition 5.4.1 Let E be a set of hypersentences. The hypersentence e can be *derived from* Σ if there is finite sequence $(\alpha_0, \ldots, \alpha_n)$ of hyperformulas such that $\alpha_n \equiv e$ and for each $0 \leq i \leq n$ either

(a) $\alpha_i \in \Sigma \cup \Lambda$ where Λ denotes the axiom schemes, or

(b) for some $j, k < i$, α_j is obtained by the modus ponens from α_j and α_k.

We write $\Sigma \vdash_{\mathbf{hyp}} e$.

Lemma 5.4.2 *Let E be a set of hypersentences and e a hypersentence. Then*

$$E \vdash_{\mathbf{hyp}} e \textit{ if and only if } \bigcup_{\sigma \in Z} \sigma'(E) \vdash \sigma(e) \textit{ for every } \sigma \in Z.$$

The proofs for Lemma 5.4.1 and Theorem 5.4.2 will be given in section 5.9 in a more general setting.

Remark For quasi-identities we consider the axioms and rules given in [Selman 72]. The above Lemma 5.4.1 and Theorem 4.2 hold also for quasi-identities after changing the notation. This also is the case for the following results of this section 5.4.

Notation Let us denote the substitution of every n_γ-ary predicate symbol by an n_γ-ary hyperpredicate symbol R_γ by the bijective map h. If Σ is a set of sentences then $H(\Sigma)$ denotes the corresponding set of hypersentences. We formalize the

Definition 5.4.3 Let E be the set of hypersentences and Σ the set of sentences which hold for the class K of models. K is a *solid class of models* if $H(\Sigma) = E$. Obviously we have $h(\Sigma) \subseteq E$.

Theorem 5.4.4 *K is solid if and only if* $\bigcup\limits_{\sigma\in Z} \sigma(E) \subseteq \Sigma$.

Proof From $\bigcup\limits_{\sigma\in Z} \sigma(E) \subseteq \Sigma$ we conclude that $h(\Sigma) \subseteq E$ and hence K is solid. On the other hand, assume that we have $h(\Sigma) \subseteq E$. We conclude that $E \vdash_{\mathbf{hyp}} h(\epsilon)$ for every sentence ϵ of Σ. By Lemma 5.4.1 we have $\bigcup\limits_{\sigma'\in Z} \sigma'(E) \vdash \sigma(h(\epsilon))$ for every $\sigma \in Z$. We choose σ such that $\sigma(h(\epsilon)) = \epsilon$ and have $\bigcup\limits_{\sigma'\in Z} \sigma'(E) \vdash \epsilon$ for every sentence ϵ of Σ. As $\bigcup\limits_{\sigma'\in Z} \sigma'(E)$ is closed under the axiom schemes and the modus ponens of the predicate calculus, we have $\bigcup\limits_{\sigma\in Z} \sigma(E) \supseteq \Sigma$.

5.5 Hybrid terms

Terms are built up with variables and operation symbols, hyperterms with variables and hypervariables. If one admits operation symbols and hypervariables simultaneously in a language, then one gets hybrid terms. Therefore the concept of hybrid terms is a generalization of hyperterms. In our approach we restrict the hypervariables (respectively hyperpredicate variables) to a fixed type. By this restriction many problems become solvable, a fact which can also be concluded from Henkin's work on completeness [Henkin 50]. Furthermore we restrict the operator variables which are called hypervariables to terms. These hybrid logics do not have the expressive power of a general second order logic. Nevertheless, proofs may be shorter and axiom systems may become finite in a hybrid logic.

Definition 5.5.1 Let τ be a given type. Then n-ary *hybrid terms* of type τ are recursively defined by:

(1) the variables $x_1, \ldots, x_n$ are n-ary hybrid terms;

(2) if $T_1, \ldots, T_m$ are n-ary hybrid terms and f is an m-ary operation symbol, then $f(T_1, \ldots, T_m)$ is an n-ary hybrid term;

(3) if $T_1, \ldots, T_m$ are n-ary hybrid terms and F is an m-ary hypervariable, then $F(T_1, \ldots, T_m)$ is an n-ary hybrid term.

$B^n(\tau)$ is the smallest set containing (1) which is closed under finite application of (2) and (3). $B(\tau) = \bigcup\{B^n(\tau) | n \in \mathbb{N}\}$ is called the *set of hybrid terms of type* τ. A *hybrid identity* of type τ is a pair of hybrid terms (T_1, T_2); this is also denoted by $T_1 = T_2$.

Definition 5.5.2 Let (T_1, T_2) be a hybrid identity of type τ and V a variety of type τ. If every n_γ-ary hypervariable occurring in (T_1, T_2) is replaced by an n_γ-ary term $t_\gamma \in T(V)$ leaving the variables and operation symbols unchanged in (T_1, T_2), then the resulting identity (t_1, t_2) is called a *transformation of the hyperidentity* (T_1, T_2).

Example Let $F(x \wedge y, z) = F(x, y) \wedge F(y, z)$ be a hybrid identity with a binary hypervariable F and a binary operation symbol. Let V be the variety of distributive lattices of type $(2, 2)$. If we replace $F(x, y)$ by the term $x \vee y$, we get the transformation $(x \wedge y) \vee z = (x \vee z) \wedge (y \vee z)$. To get all four possible transformations, F has to be replaced by the four terms $x, y, x \wedge y, x \vee y$.

Example $F(x, F(y, z)) = F(F(x, y), z)$ is a hybrid identity which does not contain any operation symbol. These hybrid identities are called hyperidentities. If E is a set of hybrid identities of type τ, then the set of all transformations of E for a variety V of type τ is denoted by $I_V(E)$.

Definition 5.5.3 A variety V of type τ *satisfies the hybrid identity* (T_1, T_2) of type τ if the set $I_V((T_1, T_2))$ of all transformations of (T_1, T_2) is contained in the set of identities which hold in V.

Example The hybrid identity $F(x \wedge y, z) = F(x, y) \wedge F(y, z)$ holds for the variety of distributive lattices.

Definition 5.5.4 Let (t_1, t_2) be an identity which holds for a variety V. If one substitutes some n_γ-ary operation symbols f_γ by n_γ-ary hypervariables F_γ leaving the variables unchanged, then the resulting hybrid identity (T_1, T_2) is called a *transformation* of (t_1, t_2).

Example Consider the identity $(x \wedge y) \vee z = (x \vee z) \wedge (y \vee z)$ for the variety of distributive lattices V. If we substitute the operation symbol $\vee$ by the binary hypervariable F, we get the hybrid identity $F(x \wedge y, z) = F(x, z) \wedge F(y, z)$. Of course one can get transformations like $F(x, y) = F(y, x)$ from $x \vee y = y \vee x$ which do not hold as hybrid identities for V.

A transformation of (t_1, t_2) which contains a maximal number of different hypervariables is called general. For instance, $F(G(x, y), z)) = G(F(x, z), F(y, z))$ is a general transformation from the law of distributivity.

We use a slight generalization of the concept of hypersubstitution [GraSch 90, p. 308].

Rule of hybrid substitution:

(6) The hybrid identity (T_1, T_2) implies the hybrid identity (T_1^*, T_2^*) if (T_1^*, T_2^*) is the

result of any simultaneous substitution of hypervariables in T_1 and T_2 by a hybrid term of the same arity.

Together with the rules (1)-(5) which are reformulations from the classical equational logic, the derivation of hybrid identities is defined.

(1) $T_1 = T_1$ for every hybrid term $T_1 \in B(\tau)$;

(2) $T_1 = T_2$ implies $T_2 = T_1$ for hybrid terms $T_1, T_2 \in B(\tau)$;

(3) $T_1 = T_2, T_2 = T_3$ implies $T_1 = T_3$ for hybrid terms $T_1, T_2, T_3 \in B(\tau)$;

(4) $T_i = S_i$ for $i = 1, \ldots, m_\gamma$, implies $F_\gamma(T_1, \ldots, T_{m_\gamma}) = F_\gamma(S_1, \ldots, S_{m_\gamma})$ for hybrid terms $T_i, S_i \in B(\tau)$ and m_γ-ary hypervariables F_γ;

(5) $T(x_1, \ldots, x_n) = S(x_1, \ldots, x_n)$ implies $T(R_1, \ldots, R_n) = S(R_1, \ldots, R_n)$ for T, S, R_1, $\ldots, R_n \in B(\tau)$.

Given a variety V of type τ, $E_\tau(V)$ denotes the set of hybrid identities of type τ which are satisfied in V.

The following is a slight modification of G. Birkhoff's theorem [Grätzer 79].

Completeness theorem *A set Σ of hybrid identities can be presented in the form $E_\tau(K)$ for some variety K of type τ if and only if Σ is closed under the rules* (1)-(6).

5.6 Bases of hybrid identities

The hybrid equational logic has more expressive power than an equational logic. Hence one can expect that some varieties can be described by a shorter system of axioms. Let D be a set of hybrid identities of some variety V. We call D a hybrid basis of identities of V if every identity of V is implied by D. D is a basis of hybrid identities of V if every hybrid identity V is implied. (D is called a basis of hyperidentities if every hyperidentity is implied.)

Proposition 5.6.1 *Let*

$$D = \{x(yz) = (xy)z, \quad xyzw = xzyw, \quad yx^2y = xy^2x, \quad y \cdot G(x) \cdot x^2y = xy \cdot G(x) \cdot yx\}$$

be a set of hybrid identities involving an associative binary operation symbol and a unary hypervariable G. Then D cannot be presented by a finite basis of identities but by a finite hybrid basis of identities.

Proof We replace the hyperterms $G(x)$ by x^k, $k \in \mathbb{N}$, and get an infinite set E of identities. By [Perkins 68] this infinite set E of identities has no finite basis of identities.

Problem 5.6.2 Determine an algebra of minimal cardinality without a finite hybrid basis of identities.

Remark 5.6.3 Let V be a variety which has no finite bases of hyperidentities. Then there exists no finite basis of hybrid identities. Because of the rules (1)-(6) of hybrid logic it is impossible to derive hyperidentities from hybrid identities other than hyperidentities.

Solid varieties are varieties where every identity can be transformed into a hyperidentity. The following results give a new description.

Lemma 5.6.4 *A variety V of type τ is solid if and only if every transformation of any of the identities of V holds as a hybrid identity of V.*

Proof Let (t_1, t_2) be an identity of V and (T_1, T_2) a transformation into a hybrid identity. Let (T_1^*, T_2^*) be a transformation into a hyperidentity with a maximal number of different hypervariables. We replace the appropriate hyperterms in (T_1^*, T_2^*) to get (T_1, T_2). As V is solid, (T_1^*, T_2^*) holds for V and hence also (T_1, T_2).

Corollary 5.6.5 *Let Σ be a basis for the identities of V of type τ. V is solid if and only if every transformation of Σ holds as a hybrid identity of V.*

5.7 Hybrid terms of distributive lattices

Notation 5.7.1 We consider the following set B of hybrid identities of type (2,2) using the binary operation symbols $\wedge, \vee$ and the binary hypervariables F, G.

$(H1)$ $F(x, F(y, z)) = F(F(x, y), z)$

$(H2)$ $F(x, x) = x$

$(H3)$ $F(F(u, x), F(y, w)) = F(F(u, y)F(x, w))$

$(H4)$ $F(G(x, y), z) = G(F(x, z), F(y, z))$

$(H5)$ $F(x, G(y, z)) = G(F(x, y), F(x, z))$

$(E1)$ $x \wedge y = y \wedge x, \quad x \vee y = y \vee x$

$(E2)$ $x \wedge (y \vee x) = x, \ x \vee (y \wedge x) = x$

Remark 5.7.2 An algebra L of type (2,2) is a distributive lattice if the hybrid identities $H1, H2, H4, H5, E1, E2$ hold.

Proof From $H1$ follows the associativity, from $H2$ the idempotency and from $H4$ and $H5$ the distributivity of the lattice operations $\wedge$ and $\vee$ (putting $x := u, y := z,\ z := y = w$ and hypersubstituting F and G by $\wedge$ and $\vee$ respectively).

Remark 5.7.3 The hyperidentity $(H4)$ respectively $(H5)$ implies

$(M1)$ $F(x \wedge y, z) = F(x, z) \wedge F(y, z)$ (if we hypersubstitute G by $\wedge$)

(M2) $F(x \vee y, z) = F(x, z) \vee F(y, z)$

(M3) $F(x, y \wedge z) = F(x, y) \wedge F(x, z)$

(M4) $F(x, y \vee z) = F(x, y) \vee F(x, z)$

Proposition 5.7.4 *Every hybrid term T can be presented as a disjunction of conjunctions of hyperterms, i.e. in dch-form.*

Proof If T is a hyperterm then 5.7.4 holds. If $T = T_1 \vee T_2$ and T_1, T_2 are in dch-form, then T is in dch-form. If $T = T_1 \wedge T_2$, then by the distributive law T can be presented in dch-form. If $T = F(T_1, T_2)$ we apply $(M1)$-$(M4)$ to get a dch-form.

Example Consider $T = G(F(x \wedge y, z), x)$. T can be transformed into dch-form in the following way

$$G(F(x \wedge y, z), x) \longrightarrow_{(M1)} G(F(x, z) \wedge F(y, z), x) \longrightarrow_{(M1)} G(F(x, z), x) \wedge G(F(y, z), x).$$

Notation A hyperterm T is called an F-hyperterm (respectively G-hyperterm), if T contains only hypervariables F (respectively G).

Proposition 5.7.5 *Every hyperterm T can be represented as an F-hyperterm substituted by G-hyperterms.*

Proof One applies $(H4)$ and $(H5)$.

Example

$$\begin{aligned} G(F(x, y), F(u, v)) &\longrightarrow_{(H4)} F(G(x, F(u, u)), G(y, F(u, u))) \\ &\longrightarrow_{(H5)} F(F(G(x, u), G(x, u)), F(G(y, u), G(y, u))) \\ &\longrightarrow_{(H1)} F(F(F(G(x, u), G(x, u)), G(y, u)), G(y, u)). \end{aligned}$$

Remark 5.7.6 As F, G are associative, one may write by abusing the notation

$$F(x_1, \ldots, x_n) := F(F \ldots (F(x_1, x_2), x_3, \ldots, x_n) \ldots).$$

5.8 Unification of hybrid terms of 2-groups

In automatic theorem proving the unification of formulas plays an important role. A unifier of two formulas is a substitution such that the two formulas under this substitution become equal.

The problem of unification has already been studied for second and higher order logics. By a result of [Goldfarb 81] it is shown that unification is undecidable for the second order logic. Hybrid logic is a fragment of second order logic but it is an open question whether unification is decidable. It is obvious that specific examples in hybrid logic can be handled by a transformation to first order logic. If V is a variety where every n-generated free algebra is finite, then the unification of hybrid identities is decidable if and only if the unification of identities is decidable.

There are only a few varieties where the unification problem is explicitly solved. We use an idea of Löwenheim to study the unification in the variety of 2-groups (groups of exponent 2).

$(H1)$ $F(x, F(y, z)) = F(F(x, y), z)$

$(H2)$ $F^3(x, y) = F(x, y)$

$(H3)$ $F(F(u, x), F(y, v)) = F(F(u, y), F(x, v))$

$(M1)$ $F(x + y, u + v) = F(x, u) + F(y, v)$

$(M2)$ $F(0, 0) = 0$

$(A1)$ $x + x = 0$

$(A2)$ $x + y = y + x$

Here we define
$$F^3(x, y) := F(F(F(x, y), y), y).$$
In a more general form we consider a hybrid term $T(x_1, \ldots, x_n)$ and use
$$T(x_1 + y_1, \ldots, x_n + y_n) = T(x_1, \ldots, x_n) + T(y_1, \ldots, y_n). \qquad (M_1)$$
A term for 2-groups can be written in the general form $a_1x_1 + \ldots + a_nx_n$, $a_i \in \{0, 1\}$, $i = 1, \ldots n$. Obviously $a_1(x_1 + y_1) + \ldots + a_n(x_n + y_n) = a_1x_1 + \ldots + a_nx_n + a_1y_1 + \ldots + a_ny_n$.

Similarly we have
$$T(0, \ldots, 0) = 0. \qquad (M_2)$$
We denote
$$T_1^2(x_1, \ldots, x_n) = T(x_1, \ldots, x_{i-1}, T(x_1, \ldots, x_n), x_{i+1}, \ldots, x_n)$$
and recursively
$$T_i^3(x_1, \ldots, x_n) = T(x_1, \ldots, x_{i-1}, T_i^2(x_1, \ldots, x_n), x_{i+1}, \ldots, x_n).$$
We have for $i \in \{1, \ldots, n\}$
$$T_i^3(x_1, \ldots, x_n) = T = (x_1, \ldots, x_n) \qquad (H_2)$$
or explicitly
$$\begin{aligned} &a_1x_1 + \ldots + a_i(a_1x_1 + \ldots + a_i(a_1x_1 + \ldots + a_nx_n) + \ldots + a_nx_n) + \ldots a_nx_n \\ &= a_1x_1 + \ldots + (a_ia_1x_1 + \ldots + a_ia_1x_1 + a_ix_i + a_ia_nx_n + \ldots + a_ia_nx_n) + \ldots + a_nx_n \\ &= a_1x_1 + \ldots + a_ix_i + \ldots + a_nx_n \end{aligned}$$

In the following a hybrid substitution is a finite set $\{v_1|T_2, \ldots, v_n|T_n\}$ of pairs, where $v_1, \ldots, v_n$ are variables and hypervariables, and $T_1, \ldots, T_n$ are hybrid terms such that if v_i is an n_γ-ary hypervariable, then T_i is an n_γ-ary hybrid term.

The result θT of applying a hybrid substitution $\theta = \{v_1|T_1, \ldots, v_n|T_n\}$ to a hybrid term T can be defined recursively in an obvious way. We hypersubstitute the hypervariables and then substitute the variables.

Example Consider the hybrid equation $F(x,y) + z = F(y,x)$ which does not hold as a hybrid identity for 2-groups. Consider $\theta = \{F|+, z|0\}$. The result of $\theta(F(x,y)+z)$ is $x+y$ and of $\theta F(y,x)$ is $y+x$.

Definition 5.8.1 A hybrid substitution θ is a *unifier* to a pair (T_1, T_2) of hybrid terms if $\theta T_1 = \theta T_2$.

Definition 5.8.2 A unifier θ for a pair of hybrid terms is a *most general unifier* if and only if for each unifier σ for the pair there is a hybrid substitution λ such that $\sigma = \lambda \circ \theta$.

To find a non-trivial unifier we use an approach similar to Löwenheim (compare [MarNip 89]). Instead of considering the unification problem for the hybrid equation $T = S$ we study the hybrid equation $T + S = 0$. These equations are equivalent because $T = T + (S + S)$ and $(T + S) + S = 0 + S = S$. Hence we search for a unifier of the hybrid equation $T(x_1, \ldots, x_n) = 0$.

Lemma 5.8.3 *Let $T(x_1, \ldots, x_n) = 0$ be a hybrid equation. Then there exists a non-trivial unifier*

$$\theta = \{x_1|x_1 + T_1^2(x_1, \ldots, x_n), \ldots, x_n|x_n + T_n^2(x_1, \ldots, x_n)\}.$$

Proof We show that θ is a unifier applying (M_1) and (H_2). Let n be odd.

$$\begin{aligned}
& T(x_1 + T_1^2(x_1, \ldots, x_n), \ldots, x_n + T_n^2(x_1, \ldots, x_n)) \\
&= T(x_1 + \underbrace{(x_1 + \ldots + x_1)}_{(n-1)\text{times}} + T_1^2(x_1, \ldots, x_n), \ldots, x_n + \underbrace{(x_n + \ldots + x_n)}_{(n-1)\text{times}} + T_n^2(x_1, \ldots, x_n)) \\
&= T(x_1, \ldots, x_n) + T(T_1^2(x_1, \ldots, x_n), x_2, \ldots, x_n) + \ldots + T(x_1, \ldots, x_{n-1}, T_n^2(x_1, \ldots, x_n)) \\
&= T(x_1, \ldots, x_n) + T(x_1, \ldots, x_n) + \ldots + T(x_1, \ldots, x_n) \quad (n \text{ times}) \\
&= 0.
\end{aligned}$$

Let n be even.

$$\begin{aligned}
& T(x_1 + T_1^2(x_1, \ldots, x_n), \ldots, x_n + T_n^2(x_1, \ldots, x_n)) \\
&= T(x_1 + \underbrace{(x_1 + \ldots + x_1)}_{(n-2)\text{times}} + T_1^2(x_1, \ldots, x_n), \ldots, x_n + \underbrace{(x_n + \ldots + x_n)}_{(n-2)\text{times}} + T_n^2(x_1, \ldots, x_n)) \\
&= T(T_1^2(x_1, \ldots, x_n), x_2), \ldots, x_n) + \ldots + T(x_1, \ldots, x_{n-1}, T_n^2(x_1, \ldots, x_n)) \\
&= T(x_1, \ldots, x_n) + \ldots + T(x_1, \ldots, x_n) \quad (n \text{ times}) \\
&= 0.
\end{aligned}$$

Let $\sigma = \{F|\sigma F, x_1|y_1, \ldots, x_n|y_n\}$ be a unifier and let $\sigma T(x_1, \ldots, x_n)$ be the result by hypersubstituting F without changing the variables.

Lemma 5.8.4 $\theta(\sigma) = \{x_i|x_i + \sigma T_1^2(x_1, \ldots, x_n) + \sigma T(0, \ldots, 0, y_i, 0 \ldots, 0);\ i = 1, \ldots, n\}$ *is a unifier for* $\sigma T(x_1, \ldots, x_n)$.

Proof We have only to consider

$$\sigma T(\sigma T(y_1, 0, \ldots, 0), \ldots, \sigma T(0, \ldots, 0, y_n))$$

because of Lemma 5.8.3. Interpretating $\sigma T(x_1, \ldots, x_n)$ by terms $a_1 x_1 + \ldots + a_n x_n$ we have the result $a_1 a_1 y_1 + \ldots + a_n a_n y_n = a_1 y_1 + \ldots + a_n y_n = 0$.

Theorem 5.8.5 *Every unifier* σ *can be presented by* $\sigma = \sigma \circ \theta(\sigma)$.

Proof We have to consider

$$\begin{aligned} & y_i + \sigma T^2(y_1, \ldots, y_n) + \sigma T(0, \ldots, 0, y_i, 0 \ldots 0) \\ = {} & y_i + \sigma T_i(y_1, \ldots, y_{i-1}, 0, y_{i+1}, \ldots, y_n) + \sigma T(0, \ldots, 0, y_i, 0 \ldots 0) = y_i. \end{aligned}$$

5.9 Hybrid sentences

We use hypervariables and hyperpredicate variables of a fixed type to define hybrid sentences in the usual recursive way.

Definition 5.9.1 An *atomic hybrid formula* is an expression of the form

$$P(T_1, \ldots, T_n),$$

where P is an n-place hyperpredicate variable and $T_1, \ldots, T_n$ are hybrid terms. The *hybrid formulas* are built up from the atomic formulas by the use fo connective symbols and the quantifier symbol: $(\neg R)$, $(R \to Q), \forall x_i R$.

A *hybrid sentence* is a hybrid formula where every variable, every hypervariable and every hyperpredicate variable are bound.

We follow the notation of [Enderton 72] and present the following axiom scheme for a hybrid logic.

(1) Tautologies,

(2) substitution of variables $\forall x\ \alpha \to \alpha^x_T$,

(3) $\forall P\ \alpha \to \alpha^P_c(R_1, \ldots, R_n)$,

(4a) $\forall x\ (\alpha \to \beta) \to (\forall x \alpha \to \forall x \beta)$,

(4b) $\forall P\ (\alpha \to \beta) \to (\forall P \alpha \to \forall P \beta)$, (5a) $\alpha \to \forall x\alpha$, where x does not occur free in α,

(5b) $\alpha \to \forall P\alpha$, where P does not occur free in α.

Rule of inference: Modus ponens

$$\frac{\alpha, \alpha \to \beta}{\beta}$$

for a hybrid term T, a hyperpredicate variable P or respectively a hypervariable, for a hybrid formula $C(R_1, \ldots, R_n)$ and hybrid formulas α, β.

Definition 5.9.2 Let E be a set of hybrid sentences. The hybrid sentence e can be *derived* form Σ if there is a finite sequence $(\alpha_0, \ldots, \alpha_n)$ of hybrid formulas such that $a_n \equiv e$ and for each $0 \leq i \leq n$ either

(a) $a_i \in \Sigma \cup \Lambda$, where Λ denotes the axiom schemes, or

(b) for some $j, k < i$, α_i is obtained by the modus ponens from α_j and α_k. We write $\Sigma \vdash_h e$.

Notations Let $\bar{\sigma} : \{F_i | i \in I\} \to T(L)$ assign to every n_i-ary hypervariable F_i and n_i-ary term t of the language L. Such a map $\bar{\sigma}$ can be extended to a map $\bar{\bar{\sigma}}$ from the set of hybrid terms into $T(V)$. We define furthermore

$$\bar{\bar{\bar{\sigma}}} : \{P_j | j \in J\} \to \{p_j(r_{j_1}, \ldots, r_{j_n}) | j \in J\}$$

which assigns to every hyperpredicate variable an atomic formula of the same arity. Altogether we get a transformation σ which assigns to every hybrid formula a formula of first order. Z denotes the set of all these transformations σ.

Lemma 5.9.3 *Let E be a set of hybrid sentences and e a hybrid sentence. Then*

$$E \vdash_h e \quad \textit{if and only if} \bigcup_{\sigma' \in Z} \sigma'(E) \vdash \sigma(e) \textit{ for every } \sigma \in Z.$$

Proof Let $E \vdash_h e$ and let $(e_1, \ldots, e_n)$, $e_n \equiv e$, be a sequence of hybrid sentences, where e_i either follows from e_j, e_i, $j, k \leq i$, by modus ponens or is from the axiom scheme or from E. We choose a $\sigma \in Z$ and transform every hybrid sentence e_i to a sentence $\sigma(e_i)$. The sequence $(\sigma(e_1), \ldots, \sigma(e_n))$ need not be a derivation in the predicate calculus because axiom schemes (3), (4b), (5b) become meaningless after applying σ. Let us consider a step according to axiom scheme (3) from the hybrid sentence $e_h(T_1, \ldots, T_n)$ to $e_k(T_1, \ldots, T_n)$. Then e_k arises from e_h by replacing hyperpredicate variables P_γ by atomic hybrid formulas T_γ. For every $e_i, 1 \leq i \leq k$, we consider a transformation σ' such that we have a sequence $(\sigma'(e_1), \ldots, \sigma'(e_i))$ with $\sigma'(e_i) = \sigma(e_i)$ We include this sequence before $\sigma(e_i)$ and get i additional members. Proceeding in such a way we finally end with a possibly much longer sentence within the predicate calculus and have $\bigcup_{\sigma' \in Z} \sigma'(E) \vdash \sigma(e)$ for $\sigma \in Z$. On the other hand, if we have $\bigcup_{\sigma' \in Z} \sigma'(E) \vdash \sigma(e)$ for every $\sigma \in Z$ we consider a transformation $h \in Z^{-1}$ such that $\sigma(e)$ is transformed to the hybrid sentence e and $\bigcup_{\sigma' \in Z} \sigma'(E)$ is transformed to a set $\bar{E}$. By axiom scheme (3) it is obvious that $E \vdash_h \bar{E}$ and hence we have $E \vdash_h e$.

Theorem 5.9.4 *$E \models_h e$ if and only if $E \vdash_h e$.*

Proof Let $E \models_h e$ and let K be the class of models which fulfill every hybrid sentence of E. Then K fulfills e by hypothesis and furthermore we conclude that $\bigcup_{\sigma' \in Z} \sigma'(E) \models \sigma(e)$ for every transformation σ. By the completeness of the predicate calculus we have $\bigcup_{\sigma' \in Z} \sigma'(E) \vdash \sigma(e)$ for every transformation σ, and by Lemma 5.9.3, $E \vdash_h e$.

For the reverse direction we use again Lemma 5.9.3 and the correctness of the predicate calculus to get $\bigcup_{\sigma' \in Z} \sigma'(E) \models \sigma(e)$. There is a transformation h such that $h(\sigma(e)) = e$. With this transformation we get a set $\bar{E}$ form $\bigcup_{\sigma' \in Z} \sigma'(E)$. It is obvious that $E \models_h \bar{E}$ and hence $E \models_h e$.

Remark 5.9.5 One may use Lemma 5.9.3 to show that the models of a set E of hybrid sentences are closed under ultraproducts. It is clear that Craig's interpolation theorem holds for hypersentences.

Additional remark One should feel free to interpret the hypervariables by special sets of term functions. For instance in the case of Boolean algebras a binary hypervariable may stand only for monotone term functions generated by the operations join and meet. This will yield a different hybrid logic which may have its own merits. Therefore a manifold of hybrid logics concerning types and restrictions for interpretation are possible and may be of good use in applications (for instance in knowledge representation).

References

[Aczél 71] Aczél, J., Proof of a theorem of distributive type hyperidentities, *Algebra Universalis* **1** (1971), 1–6.

[Asser 81] Asser, G., *Einführung in die mathematische Logik* I–III, Teubner, Leipzig (1981).

[BakPix 75] Baker, K.A., Pixley, A.F., Polynomial interpolation and the chinese remainder theorem for algebraic systems, *Math. Z.* **143** (1975), 165–174.

[BelSlo 71] Bell, J.L., Slomson, A.B., *Models and Ultraproducts*, North-Holland, Amsterdam (1971).

[Belousov 65] Belousov, V.D., Systems of quasigroups with generalized identities, *Uspekhi Mat. Nauk.* **20** (1965), 75–146 (Translation: *Russian Math. Surveys* **20** 75–143).

[Bergman 81] Bergman, G.M., Hyperidentities of groups and semigroups, *Aequationes Math.* **23** (1981), 50–65.

[Berman] Berman, J., Free spectra of 3-element algebras, Preprint, University of Illinois at Chicago, p. 42.

[Birkhoff 35] Birkhoff, G., On the structure of abstract algebras, *Proc. Cambridge Philos. Soc.* **31** (1935), 433–454.

[ChaKei 73] Chang, C.C., Keisler, H.J., *Model Theory*, North-Holland, Amsterdam (1973).

[ClaKra 80] Clark, D.M., Krauss, P.H., Plain para primal algebras, *Algebra Universalis* **11** (1980), 365–388.

[CloMel 81] Clocksin, W.F., Mellish, C.S., *Programming in Prolog*, Springer-Verlag, Berlin, Heidelberg, New York (1981).

[Csákány 83] Csákány, B., All minimal clones on the three element set, *Acta Cybernet.* **6** (1983), 227–238.

[DavPri 90] Davey, B.A., Priestley, H.A., *Introduction to Lattices and Order*, Cambridge University Press, Cambridge (1990).

[Denecke 89] Denecke, K., Boolean clones and hyperidentities in universal algebras, in: *Universal and Applied Algebra* (K. Halkowska, B. Stawski, eds.), World Scientific Publishing Co., Singapore (1989), 23–45.

[DLPSch 91] Denecke, K., Lau, D., Pöschel, R., Schweigert, D., Hyperidentities, hyperequational classes and clone congruences, in: *Contributions to General Algebra*, Vol. 7, Teubner, Wien (1991), 97–118.

[DeMaRe 90] Denecke, K., Malcev, I.A., Reschke, M., Separation of Boolean clones by hyperidentities, Preprint 1990.

[DenPös 88a] Denecke, K., Pöschel, R., A characterization of Sheffer functions by hyperidentities, *Semigroup Forum* **37** (1988), 351–362.

[DenPös 88b] Denecke, K., Pöschel, R., The characterization of primal algebras by hyperidentities, in: *Contributions to General Algebra*, Vol. 6, Teubner, Stuttgart (1988).

[DenRei 88] Denecke, K., Reichel, M., Funktionalgleichungen in zweiwertigen Logiken, *Wiss. Z. Pädag. Hochsch. Potsdam* **32** (1988), 603–606.

[DudGra 81] Dudek, J., Graczyńska E., The lattice of varieties of algebras, *Bull. Acad. Polon. Sci. Sér. Sci. Math.* **29** (1981), 337–349.

[Enderton 72] Enderton, H.B., *Mathematical Introduction to Logic*, Academic Press, New York (1972).

[Evans 79] Evans, T., Some remarks on the general theory of clones, in: *Finite Algebra and Multiple-Valued Logic* (B. Csákány, I.G. Rosenberg, eds.), Colloq. Math. Soc. János Bolyai **28** (1979), 203–244.

[Felscher 68] Felscher, W., Equational maps, in: *Contributions to Mathematical Logic* (H.A. Schmidt et al., eds.), North-Holland, Amsterdam (1968), 121–161.

[FreMcK 87] Freese, R., McKenzie, R., *Commutator Theory for Congruence Modular Varieties*, Cambridge University Press (1987).

[FrKaMá 82] Fried, E., Kaiser, H.K., Márki, L., An elementary approach to polynomial interpolation in universal algebra, *Algebra Universalis* **15** (1982), 40–57.

[Frobenius 1895] Frobenius, G., Über endliche Gruppen, *Sitzungsber. Preuss. Akad. Wiss. Berlin* 1895, 163–194.

[Frucht 50] Frucht, R., Lattices with a given abstract group of automorphisms, *Canad. J. Math.* **2** (1950), 417–419.

[GlaMic 77] Glazek, K., Michalski, J., Weak homomorphisms of general algebras, *Comment. Math. Prace Mat.* **19** (1976/77), 211-228.

[Goldfarb 81] Goldfarb, W.D., The undecidability of the second-order unification problem, *J. Theoret. Comput. Sci.* **13** (1981), 225–230.

[Gorlov 73] Gorlov, W.W., On congruences of Post classes, *Mat. Zametki* **13** (1973), 725–734.

[Graczyńska 84] Graczyńska, E., On bases for normal identities, *Studia Sci. Math. Hungar.* **19** (1984), 317–320.

[Graczyńska 89] Graczyńska, E., On normal and regular identities and hyperidentities, in: *Universal and Applied Algebra* (K. Halkowska, B. Stawski, eds.), World Scientific Publ. Co., Singapore (1989), 107–135.

[Graczyńska 90] Graczyńska, E., On normal and regular identities, *Algebra Universalis* **27** (1990), 387–397.

[GraSch 90] Graczyńska, E., Schweigert, D., Hypervarieties of a given type, *Algebra Universalis* **27** (1990), 305–318.

[Grätzer 79] Grätzer, G., *Universal Algebra*, 2nd ed., Springer-Verlag, Berlin, Heidelberg, New York (1979).

[Hall 59] Hall, M., Jr., *The Theory of Groups*, MacMillan, New York (1959).

[HedPul 64] Hedrlín, Z., Pultr, A., Relations (graphs) with given finitely generated semigroups, *Monatsh. Math.* **68** (1964), 213–217.

[Henkin 50] Henkin, L., Completeness in the theory of types, *J. Symbolic Logic* **15** (1950), 81–91.

[HyMcKTa 92] Hyndman, J., McKenzie, R., Taylor, W., k-ary monoids of term operations, *Semigroup Forum* **44** (1992), 21–52.

[JaGaKu 70] Jablonski, S.W., Gawrilow, G.P., Kudrjawzew, W.B., *Boolesche Funktionen und Postsche Klassen*, Vieweg, Braunschweig (1970).

[JežMcN] Ježek, J., McNulty, G.F., Bounded and well placed theories in the lattice of equational theories. Manuscript.

[Kelly 84] Kelly, D., Complete rules of inference for universal sentences, *Studia Sci. Math. Hungar.* **19** (1984), 347–361.

[Knoebel 73] Knoebel, R.A., Products of independent algebras with finitely generated identities, *Algebra Universalis* **3** (1973), 147–151.

[Kolibiar 84] Kolibiar, M., Weak homomorphism in some classes of algebras, *Studia Sci. Math. Hungar.* **19** (1984), 413–420.

[Lau 91] Lau, D., On closed subsets of Boolean functions I, *J. Inform. Process. Cybernet.* **27** (1991), 167–178.

[LauNöb 73] Lausch, H., Nöbauer, W., *Algebra of Polynomials*, North-Holland, Amsterdam (1973).

[Malcev 71] Malcev, A.I., *Metamathematics of Algebraic Systems*, North-Holland, Amsterdam (1971).

[Malcev 73] Malcev, A.I., *Algebraic Systems*, Springer-Verlag, Berlin (1973).

[MalSch 89] Malcev, I.A., Schweigert, D., Hyperidentities of QZ-algebras, *Siberian Math. J.* **30** (6) (1989), 132–137 (Russian).

[MarNip 89] Martin, U., Nipkow, T., Unification in Boolean rings, *J. Automat. Reason.* **4** (1989), 381–396.

[McMcTa 87] McKenzie, R., McNulty, G., Taylor, W., *Algebras, Lattices, Varieties*, Vol. I, Wadsworth Books (1987), Monterey, CA.

[McKHob 88] McKenzie, R., Hobby, D., The Structure of Finite Algebras, *Contemp. Math.* **76** (1988), Amer. Math. Soc., Providence, RI.

[McKVal 89] McKenzie R., Valeriote, M., *The Structure of Decidable Locally Finite Varieties*, Birkhäuser, Basel (1989).

[McNulty 77] McNulty, G., Fragments of first order logic, I: Universal Horn logic, *J. Symbolic Logic* **42** (1977), 221–237.

[Movsisyan 86] Movsisyan, Yu. M., *Introduction to the Theory of Algebras with Hyperidentities*, University of Erevan (1986) (Russian).

[Neumann 77] Neumann, W.D., On Malcev conditions, *J. Austral. Math. Soc.* **17** (1977), 376–384.

[Newrly 89] Newrly, N., Lattices of equational theories are congruence lattices of monoids with one additional operation, Preprint 1235, TH Darmstadt (1989).

[Nicolas 69] Nicolas, J.L., Ordre maximal d'un lment du groupe S_n des permutations et "highly composite numbers", *Bull. Soc. Math. France* **97** (1969), 129–191.

[PadPen 82] Padmanabhan, R., Penner, P., Bases of hyperidentities of lattices and semilattices, *C.R. Math. Rep. Acad. Sci. Canada* **4** (1982), 9–14.

[Pálfy 84] Pálfy, P.P., Unary polynomials in algebras I, *Algebra Universalis* **18** (1984) 262–273.

[Penner 81] Penner, P., Hyperidentities of lattices and semi-lattices, *Algebra Universalis* **13** (1981), 307–314.

[Penner 84] Penner, P., Hyperidentities of semilattices, *Houston J. Math.* **10** (1984), 81–108.

[Perkins 68] Perkins, P., Bases for equational theories of semigroups, *J. Algebra* **11** (1968), 298–314.

[Pigozzi 72] Pigozzi, D., On some operations on classes of algebras, *Algebra Universalis* **2** (1972), 346–353.

[Pixley 79] Pixley, A.F., A survey of interpolation in universal algebra, in: *Finite Algebra and Multiple-Valued Logic* (B. Csákány, I.G. Rosenberg, eds.), Colloq. Math. Soc. János Bolyai **28** (1979), 203–244.

[Płonka 66] Płonka, J., Diagonal algebras, *Fund. Math.* **58** (1966), 309–321.

[Płonka 69] Płonka, J., On equational classes of abstract algebras defined by regular equations, *Fund. Math.* **64** (1969), 241–247.

[Quackenbush] Quackenbush, R.W., Primality: the influence of Boolean algebras in universal algebra in [Grätzer], 401–416.

[ReiSch 91] Reichel, Schweigert, D., Unpublished manuscript (1991).

[ReSchSi 87] Reischer, C., Schweigert, D., Simovici, D., A completeness criterion by functional equations, *Proc. 17th Internat. Symposium on Multivalued Logic, Boston 1987*, IEEE Computer Soc. Publ. Office.

[ReiSim 82] Reischer, C., Simovici, D., Several remarks on the iteration properties of switching functions, *Proc. 12th Internat. Symposium on Multiple-Valued Logic, Paris 1982*, IEEE Computer Soc. Publ. Office, 244–247.

[RomSmi 85] Romanowska, A., Smith, J., *Modal Theory – an Algebraic Approach to Order*, Heldermann, Berlin (1985).

[Rosenberg 70] Rosenberg, I.G., Über die funktionale Vollständigkeit in den mehrwertigen Logiken, *Rozpravy Československé Akad. Věd. Řada Mat. Přírod. Věd.* **80** (1970), 3–93.

[Rosenberg 76] Rosenberg, I.G., Characterization of Malcev's preiterative algebra, Preprint CRM-**594**, Centre de recherches mathmatiques, Université de Montréal (1976).

[Rosenberg 78] Rosenberg, I.G., Subalgebra systems of direct powers, *Algebra Universalis* **8** (1978), 221–227.

[Rosenberg 86] Rosenberg, I.G., Minimal clones I: the five types, in: *Lectures in Universal Algebra* (L. Szabó, A. Szendrei, eds.), Colloq. Soc. Math. János Bolyai **43** (1986), 405–428.

[RosSch 82] Rosenberg, I.G., Schweigert, D., Locally maximal clones. *Elektron. Informationsverab. Kybernetik* **19** (1982), 7–8 and 389–401.

[Rousseau 67] Rousseau, G., Completeness in finite algebras with a single operation, *Proc. Amer. Math. Soc.* **18** (1967), 1009–1013.

[SauSto 90] Sauer, N., Stone, M.G., Composing functions to reduce image size. *Ars. Combin.* **31** (1991), 171–176.

[Schweigert 74] Schweigert, D., Über endliche, ordnungspolynomvollständige Verbände, *Monatsh. Math.* **78** (1974), 68–76.

[Schweigert 75] Schweigert, D., Über idempotente Polynomfunktionen auf Verbänden, *Elem. Math.* **30** (1975), 30–32.

[Schweigert 79a] Schweigert, D., On prepolynomially complete algebras, *J. London Math. Soc. (2)* **20** (1979), 179–185.

[Schweigert 79b] Schweigert, D., On semigroups of polynomial functions, *Semigroup Forum* **18** (1979), 5–8.

[Schweigert 84a] Schweigert, D., Clone equations and hyperidentities, Preprint no. 86, Universität Kaiserslautern (1984).

[Schweigert 84b] Schweigert, D., On weak isomorphisms and equational theories, in: *Contributions to General Algebra*, Vol. 3, (G. Eigenthaler et al., eds.), Hölder-Pichler-Tempsky, Wien; Teubner, Stuttgart (1985), 335–340.

[Schweigert 85a] Schweigert, D., Clones of term functions of lattices and abelian groups, *Algebra Universalis* **20** (1985), 27–33.

[Schweigert 85b] Schweigert, D., Congruence relations of multialgebras, *Discrete Math.* **53** (1985), 249–253.

[Schweigert 87a] Schweigert, D., On algebras and clones, *Semigroup Forum* **35** (1987), 85–99.

[Schweigert 87b] Schweigert, D., On equations of clones, in: *Mathematical Problems in Computation Theory*, Banach Center Publications **21**, Warsaw (1987), 367–373.

[Schweigert 89] Schweigert, D., Hyperidentities and clone congruences, *Contemp. Math.* **131** (1992), 284–293.

[Schweigert 90] Schweigert, D., Clone homomorphisms and reductions, *J. Inform. Process. Cybernet.* **26** (1990), 237–245.

[Schweigert 91a] Schweigert, D., Hyperidentities and solid models, *Z. Math. Logik Grundlag. Math.*, to appear.

[Schweigert 91b] Schweigert, D., Hybrid terms and sentences, to appear in *Studia Logica.*

[SchSzy 83] Schweigert, D., Szymańska, M., Polynomial functions of correlation lattices, *Algebra Universalis* **16** (1983), 355–359.

[Selman 72] Selman, A., Completeness of calculi for axiomatically defined classes of algebras, *Algebra Universalis* **2** (1972), 20–32.

[Siekmann 84] Siekmann, J.H., Universal unification, in: *7th Internat. Conference on Automated Deduction* (R.E. Shostak, ed.), Lecture Notes in Comput. Sci. **170**, Springer-Verlag, Berlin, New York (1984), 1–42.

[Simelgor 74] Simelgor, E.P., Identities in a finite symmetric semigroup, in: *Modern Algebra, No. 1*, Leningrad. Gos. Ped. Inst., Leningrad (1974), 174–188 (Russian).

[SheVol 85] Shevrin, L.N., Volkov, M.V., Identities of semigroups, *Izv. Vyssh. Uchebn. Zaved. Mat.* **11** (1985), 3–47 (Russian).

[Smith 76] Smith, J.D.H., *Malcev Varieties*, Lecture Notes in Math. **554**, Springer-Verlag, Berlin, Heidelberg, New York (1976).

[Szendrei 86] Szendrei, A., *Clones in Universal Algebra*, Séminaire Math. Supérieures **99**, Les Presses de l'Université de Montréal (1986).

[Taylor 73] Taylor, W., Characterizing Malcev conditions, *Algebra Universalis* **3** (1973), 351–397.

[Taylor 75] Taylor, W., The fine spectrum of a variety, *Algebra Universalis* **5** (1975), 263–303.

[Taylor 79] Taylor, W., Equational logic, *Houston J. Math.* **5** (1979), 1–83.

[Taylor 81] Taylor, W., Hyperidentities and hypervarieties, *Aequationes Math.* **23** (1981), 30–49.

[Volkov 89] Volkov, M.V., On finitely based semigroup varieties, *Mat. Zametki* **45** (1989), 13–23 (Russian), translation *Math. Notes* 45, 187–194.

[Wille 77] Wille, R., Über endliche, ordnungsvollständige Verbände, *Math. Z.* **155** (1977), 103–107.

[Wismath 90] Wismath, S.L., Hyperidentities for some varieties of semigroups, *Algebra Universalis* **27** (1990), 111–127.

[Wismath 91a] Wismath, S.L., Hyperidentities for some varieties of commutative semigroups, *Algebra Universalis* **28** (1991), 245–273.

[Wismath 91b] Wismath, S.L., Hyperidentity bases for rectangular bands and nilpotent semigroups, to appear in *J. Austral. Math. Soc.*.

Abstract Clone Theory

Walter TAYLOR*
Mathematics Department
University of Colorado
Boulder, CO 80309–0395
U.S.A.

Abstract

A *concrete clone* is a family of operations on a set, containing all projection operations and closed under all meaningful compositions. The *abstract clones* discussed here form an axiomatic version of the concrete notion. They stand in the same relation to concrete clones as that of groups to permutation groups. Roughly speaking, each representation of an abstract clone $\boldsymbol{C}$ is an algebra $\boldsymbol{A}$, and the family of all these algebras $\boldsymbol{A}$ is a *variety* $\mathcal{V}$. This correspondence is bijective between isomorphism types of clones $\boldsymbol{C}$ and equivalence classes of varieties $\mathcal{V}$. In this context, surjective clone homomorphisms correspond to the embedding of one variety in another, and injective homomorphisms correspond to the formation of reduct varieties.

The variety $\mathcal{V}$ mentioned here has a completely impartial similarity type (no operations are singled out as fundamental). If an abstract clone $\boldsymbol{C}$ is *presented* by *generators* F_i and *relators* R_j, then the corresponding variety $\mathcal{V}$ can be thought of as the variety with fundamental operations F_i that is defined by certain identities derived from the R_j; this is the connection with ordinary varieties of algebras.

An *interpretation* of one variety $\mathcal{W}$ in another variety $\mathcal{V}$ corresponds to a homomorphism from the clone $\boldsymbol{D}$ of $\mathcal{W}$ to the clone $\boldsymbol{C}$ of $\mathcal{V}$. If $\boldsymbol{D}$ has a finite presentation $\langle F_i, R_j \rangle$, then the existence of such a homomorphism reduces to the satisfaction in $\boldsymbol{C}$ of a certain existential closure of a conjunction of equations derived from F_i and R_j. Equivalently, one says that $\mathcal{V}$ satisfies the *strong Mal'tsev condition* associated to F_i and R_j. Other interesting properties of a variety $\mathcal{V}$ reduce to first-order properties of the associated clone $\boldsymbol{C}$; for example *hypervarieties* are defined by the universal satisfaction of sets of equations in $\boldsymbol{C}$.

Interpretation leads naturally to a quasi-order on the class of all varieties. (Equivalently, clones are quasi-ordered by the existence of homomorphisms.) The associated ordered class has been of some interest in general algebra. Only recently has it been shown (by R. N. McKenzie) to contain a cover pair.

The present paper presents these and other results in detail, and is offered as instruction to those who read it. It closely reflects the author's talks at the meeting on Ordered Sets and Algebras held at the University of Montreal during the summer of 1991. It is a survey of the work of many contributors, known and unknown. Three of the pioneers were Mal'tsev, Lawvere and Tarski. The only new result is a minor improvement of a result of J. Isbell: if $\mathcal{W}$ and $\mathcal{V} \otimes \mathcal{D}_{01}$ are both non-trivial, then so is $\mathcal{V} \otimes \mathcal{W}$. (Here $\mathcal{D}_{01}$ denotes the variety of all distributive lattices with 0 and 1.)

*With the assistance of Mr. John Coleman

I. G. Rosenberg and G. Sabidussi (eds.), Algebras and Orders, 507–530.

Lecture I

Before studying abstract clones we should first review the corresponding concrete notion.

Definition 1 A *clone of operations* (a *concrete clone*) is a family C of finitary operations defined on a set A that is closed under composition and contains all projection functions. More formally, $f \in C$ implies $F : A^k \to A$ for some integer $k \geq 1$. If $\beta \in C$ is k-ary and $\alpha_0, \ldots, \alpha_{k-1} \in C$ are n-ary, then $\beta(\alpha_0, \ldots \alpha_{k-1}) \in C$, where $\beta(\alpha_0, \ldots, \alpha_{k-1})$ is the n-ary operation given by

$$[\beta(\alpha_0, \ldots, \alpha_{k-1})](a_0, \ldots, a_{n-1}) = \beta(\alpha_0(a_0, \ldots, a_{n-1}), \ldots, \alpha_{k-1}(a_0, \ldots, a_{n-1})).$$

Furthermore, $p_i^n \in C$, $0 \leq i < n < w$, where $p_i^n(a_0, \ldots, a_{n-1}) = a_i$.

Note that in this definition we exclude nullary operations. This causes no real loss of generality, since any nullary operation can be represented by the corresponding constant unary function.

The word "clone" was first used in this sense by Philip Hall, in the case of the clone of term operations on a group. He saw a weak analogy between cell growth and the way the term operations of a group were iteratively built up from multiplication and inversion.

Just as semigroups of selfmaps give rise to abstract semigroups, so clones of operations give rise to abstract clones. An important difference between semigroups and clones is that the operation of composition is everywhere defined in semigroups of selfmaps, but not in clones of operations. Two approaches have been developed to meet this difficulty. One is to regard a clone as a special type of category and the other is to regard a clone as a partial algebra. Since both are convenient, we will adopt a synthesis of these two approaches.

To motivate the categorical approach, suppose that C is a clone of operations on a set A. Define a concrete category $\widehat{C}$ with $\mathrm{Obj}(\widehat{C}) = \{A^n : n \geq 1\}$ and $\mathrm{Hom}(A^m, A^n) = \{f : A^m \to A^n \,|\, p_i^n \circ f \in C \text{ for } 0 \leq i < n\}$.

Then $\widehat{C}$ is a subcategory of sets that represents C in an obvious way, i.e., C completely determines $\widehat{C}$ and $C = \bigcup_{m\geq 1} \mathrm{Hom}\,(A^m, A^1)$. This leads to the following definition.

Definition 2 A *clone* $\mathbf{D} = \langle \mathbf{C}, \mathbf{U} \rangle$ is an ordered pair of structures where

I. $\mathbf{C}$ is a category with designated objects $O_1, O_2, \ldots, O_i, \ldots$, $1 \leq i < \omega$, and designated morphisms p_i^j for $0 \leq i < j < n$ such that

1. Each O_j is the j-fold product of O_1 with respect to the morphisms $p_0^j, \ldots, p_{j-1}^j$.
2. The correspondence $j \leftrightarrow O_j$ is a bijection between the positive integers and the objects of $\mathbf{C}$.

II. $\mathbf{U}$ is a partial algebra with (disjoint) universes U_i, $i = 1, 2, \ldots$ with operations p_i^j, F_n^k, $0 \leq i < j < \omega$, $1 \leq m, n < \omega$, such that

1. $p_i^j \in U_j$ is nullary.
2. $F_n^k : U_k \times (U_n)^k \to U_n$.
3. F_n^k, p_i^j satisfy the following axioms:

$$F_n^s(x, F_n^j(y_0, z_0, \ldots, z_{j-1}), \ldots, F_n^j(y_{s-1}, z_0, \ldots, z_{j-1})) =$$
$$F_n^j(F_j^s(x, y_0, \ldots, y_{s-1}), z_0, \ldots, z_{j-1}) \tag{1}$$
$$F_k^j(p_i^j x_0, \ldots, x_{j-1}) = x_i \tag{2}$$
$$F_j^j(y, p_0^j, \ldots, p_{j-1}^j) = y. \tag{3}$$

(There is one axiom (1) for each triple $\langle j, n, s\rangle$ with $j, n, s \in \{1, 2, 3, \ldots\}$, and each equation is required to hold for variables restricted to the appropriate domains, e.g., $x \in U_s$, etc. Similar remarks apply to (2) and (3) as well.)

III. Furthermore $\mathbf{C}$ and $\mathbf{U}$ are related by

1. $U_i = \mathrm{Hom}_{\mathbf{C}}(O_i, O_1)$ for $i = 1, 2, \ldots$.
2. The designated projection morphisms of $\mathbf{C}$ are precisely the projection constants of $\mathbf{U}$.
3. Whenever $F_n^k(g, f_0, \ldots, f_{k-1})$ is defined
$$F_n^k(g, f_0, \ldots, f_{k-1}) = gf \quad \text{in} \quad \mathbf{C}, \tag{4}$$
where f is the unique member of Hom$\mathbf{C}$ such that
$$p_j^k f = f_j \quad \text{for} \quad 0 \leq j < k.$$

Remark A clone that satisfies part I of the preceding definition is essentially the same thing as one of Lawvere's *algebraic theories* [6]. An important aspect of this definition is that a clone is largely determined by part I alone and largely determined by part II alone. In fact, part III describes a duality between categories that satisfy part I of the definition and partial algebras that satisfy part II of the definition. This can be made precise by the following theorem.

Theorem 1 *Suppose that* $\mathbf{C}$ *is a category with distinguished objects* O_i *and morphisms* p_i^j *that satisfies part* I *of the definition. Then there exists a unique partial algebra* $\mathbf{U}$ *with universes* U_i*, operations* F_n^k *and the* p_i^j *as nullary operations that satisfies part* II *of the definition and is related to* $\mathbf{C}$ *via part* III.

Conversely given a partial algebra $\mathbf{U}$ *satisfying part* II *of the definition with universe* U_i *and operations* F_n^k *and* p_i^j*, then there exists a category* $\mathbf{C}$ *with objects* $\mathbf{O}_i$ *satisfying part* I *and related to* $\mathbf{U}$ *via part* III. *Moreover* $\mathbf{C}$ *is unique up to an isomorphism that is the identity map on the hom sets* $\mathrm{Hom}(O_i, O_1) = U_i$.

Proof First suppose that **C** is a category with designated objects and morphisms satisfying (I). (III) suffices to define the universes U_i and the operations F_n^k, so uniqueness is clear. It only remains to verify equations (1), (2) and (3).

(1): Suppose $z_0, \ldots, z_{j-1} \in U_n = \mathrm{Hom}(O_n, O_1)$, $y_0, \ldots, y_{s-1} \in U_j \mathrm{Hom}(O_j, O_1)$ and $x \in U_s = \mathrm{Hom}(O_s, O_1)$. Since the O_k are the products of O_1 in **C** with respect to the P_i^k, there exists unique $z \in \mathrm{Hom}(O_n, O_j)$ and $y \in \mathrm{Hom}(O_j, O_s)$ s.e. $p_m^j \cdot z = z_m$ $(0 \leq m < j)$ and $p_i^s \cdot y = y_i$ $(0 \leq i < s)$. Thus $O_n \xrightarrow{z} O_j \xrightarrow{y} O_s \xrightarrow{x} O_1$ is a sequence of morphisms in $\mathcal{C}$. Finally, by equation (4) we have

$$F_n^s(x, F_n^j(y_0, z_0, \ldots, z_{j-1}), \ldots, F_n^j(y_{s-1}, z_0, \ldots, z_{j-1})) =$$
$$x \cdot (y \cdot z) = (x \cdot y) \cdot z = F_n^j(F_j^s(x, y_0, \ldots, y_{s-1}), z_0, \ldots, z_{j-1}).$$

(2) and (3) are proved similarly, using equation (4) and the properties of products and projections in a category.

Conversely, suppose that **U** is a partial algebra with universes U_i, operations F_n^k and p_i^j satisfying part II of the definition. Fix some distinct objects O_i, $i = 1, 2, \ldots$, and define

$$\mathrm{Hom}(O_s, O_n) = (U_s)^m. \tag{5}$$

(Without loss of generality, suppose $\langle s, m\rangle \neq \langle s', m'\rangle \Rightarrow (U_s)^m \cap (U_{s'})^{m'} = \emptyset$. Otherwise we could simply replace U by an isomorphic partial algebra satisfying this condition.) Thus $U_j - \mathrm{Hom}(O_j, O_1)$ for all j. If

$$x = \langle x_1, \ldots, x_m\rangle \in \mathrm{Hom}(O_s, O_n),$$

and

$$y = \langle y_1, \ldots, y_s\rangle \in \mathrm{Hom}(O_j, O_s),$$

define $x \cdot y = w$ where

$$w \in (U_j)^m = \mathrm{Hom}(O_j, O_n) \tag{6}$$

is given by $w_i = F_j^s(x_i, y_0, \ldots, y_{s-1})$.

Let $\mathbf{C} = \langle \mathrm{Obj}\,\mathbf{C}, \mathrm{Mor}\,\mathbf{C}\rangle$ be given by $\mathrm{Obj}\,\mathbf{C} = \{O_j : 1 \leq j < w\}$ and $\mathrm{Mor}\,\mathbf{C} = \bigcup_{s=1}^{\infty} \bigcup_{n=1}^{\infty} \mathrm{Hom}(O_s, O_n)$. For $1 \leq j < w$ let $1_{O_j} = \langle p_0^j, \cdots p_{g-1}^j\rangle \in \mathrm{Hom}(O_j, O_j)$ we have to show that under these definitions $\mathcal{C}$ is the desired category.

Suppose $x \in \mathrm{Hom}(O_s, O_n)$, $y \in \mathrm{Hom}(O_j, O_s)$ and $z \in \mathrm{Hom}(O_t, O_j)$. Then $(x \cdot y) \cdot z$ and $x \cdot (y \cdot z)$ are both defined. To show that $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ it suffices to show that

$$x_i(y \cdot z) = (x_i \cdots y) \cdot z \qquad (0 \leq i < m),$$

where $x_i \in \mathrm{Hom}(O_s, O_1) = (U_s)'$. It is easy to see that expanding these m equations according to the definition of (equation (6)) yields equation (1), which holds in **U** by assumption. Thus $(x \cdot y) \cdot z = x \cdot (y \cdot z)$. Similarly it is straightforward to check that equations (1), (2) and (3) yield that the 1_{O_j} are the appropriate identities and that O_j is the j-fold product of O_i with respect to $p_0^j, \ldots, p_j^{j-1}$.

Thus $\mathbf{C}$ is a category satisfying part (I), which is by construction related to $\mathbf{U}$ via part (III). Suppose $\mathbf{C}'$ is another category with designated objects O'_j $(1 \le j < w)$ satisfying (I) and related to $\mathbf{U}$ via (III). Thus $\mathrm{Hom}_{\mathbf{C}'}(O'_j, O'_1) = \mathrm{Hom}_{\mathbf{C}}(O_j, O_1) = U_j$ and O'_j is the j-fold product of O'_1 in $\mathbf{C}'$ via the projections $p^j_0, \ldots, p^j_{j-1}$.

Thus, if $x \in \mathrm{Hom}_{\mathbf{C}}(O_j, O_n) = (U_j)^n = (\mathrm{Hom}_{\mathbf{C}'}(O'_j, O'_1))^m$, there exists a unique $\phi(x) = x' \in \mathrm{Hom}_{\mathbf{C}'}(O'_j, O'_n)$ such that $p^j_i \cdot x' = p^j_i \cdot x$ for $0 \le i < j$. Letting $\phi(O_j) = O'_j$ this defines a mapping between $\mathbf{C}$ and $\mathbf{C}'$ which is bijective on objects and on morphisms. Equation (4) completely determines composition in both categories, and so it is straightforward to check that ϕ is a functor and indeed the desired isomorphism. □

In view of this theorem, we adopt the convention of denoting the clone with the same symbol $\mathbf{C}$ as the underlying category. Furthermore, in view of the duality between the category part and the partial-algebra part of the clone, we will denote the clone either as $\mathbf{C} = \langle O_j, p^j_i \rangle$ or as $\mathbf{C} = \langle U_i, F^k_n, p^j_i \rangle$, whichever seems more convenient. When we denote the clone via $\mathbf{C} = \langle U_i, F^k_n, p^j_i \rangle$, then we take equations (5) and (6) as defining the corresponding category.

In order to study clones in any detail, we need a notion of clone homomorphisms. In later lectures we will see that these morphisms provide a unifying framework for the theory of Mal'tsev conditions.

Definition 3 Let $\mathbf{C} = \langle \mathbf{C}, \cdot, O_j, p^j_i, U = \cup U_i, F^k_n \rangle$ and $\mathbf{C}' = \langle \mathbf{C}', \cdot', O'_j, p'^j_i, U' = \cup U'_i, F'^k_n \rangle$ be clones. A *clone morphism* $\phi : \mathbf{C} \to \mathbf{C}'$ is an ordered pair $\phi = \langle \phi_C, \phi_U \rangle$ where $\phi_{\mathbf{C}} : C : \mathbf{C} \to \mathbf{C}'$ is a functor such that

$$\phi_{\mathbf{C}}(O_j) = O'_j, \qquad 1 \le j < w,$$

$$\phi_{\mathbf{C}}(p^j_i) = p'^j_i, \qquad 0 \le i < j < w,$$

and $\phi_{\mathbf{U}} : \mathbf{U} \to \mathbf{U}'$ is a partial algebra homomorphism such that $\phi_{\mathbf{U}}(\mathrm{U}_i) \subseteq U'_i$ for all i.

In analogy with the previous theorem it is straightforward to show that $\phi_{\mathbf{C}}$ is determined by $\phi_{\mathbf{U}}$ and conversely.

Examples

1. Any concrete clone.

2. If $\mathbf{A}$ is an algebra, then clo $\mathbf{A}$ is the clone of all term operations on $\mathbf{A}$.

3. If $\mathcal{V}$ is a variety of type ρ, it has a clone associated with it that can be described in two different ways.

- $\mathrm{Clone}(\mathcal{V}) = \mathrm{Clo}\, \mathbf{F}_{\mathcal{V}}(\omega)$.
- First define $\mathrm{Alg}\,\rho$ to be the category of all algebras of similarity type ρ. Let $\mathrm{Clone}'(\mathcal{V}) =$ [full subcategory of $\mathrm{Alg}\,\rho$ whose objects are $\mathbf{F}_{\mathcal{V}}(v_0, v_1, \ldots, v_{n-1})$, $1 \le n < \omega]^{op}$, where p^j_i is the map $p^j_i : \mathbf{F}_{\mathcal{V}}(v_0) \to \mathbf{F}_{\mathcal{V}}(v_0, \ldots, v_{j-1})$ is given by $p^j_i(v_0) = v_i$.

Note that if $\mathbf{F}_{\mathcal{V}}(n) \xrightarrow{\sigma} \mathbf{F}_{\mathcal{V}}(m) \xrightarrow{\tau} \mathbf{F}_{\mathcal{V}}(k)$ in $\mathrm{Alg}\,\rho$ is given by $\sigma(v_i) = \sigma_i$ and $\tau(v_j) = \tau_j$ then $\tau \cdot \sigma(v_i) = \sigma_i(\tau_0, \ldots, \tau_{m-1})$, since τ is a homomorphism. This is why we need to take the *dual* of the full subcategory of $\mathrm{Alg}\,\rho$ determined by the $\mathbf{F}_{\mathcal{V}}(n)$. Using this observation it is easy to check that $\mathrm{Clone}(\mathcal{V}) \cong \mathrm{Clone}'(\mathcal{V})$.

Lecture II

One purpose of clone theory is to give an overview of certain topics in universal algebra, e.g., the theory of hyperidentities. In the last lecture we introduced the clone, $\mathbf{C}(\mathcal{V})$, of a variety $\mathcal{V}$ and also the clone, $\mathrm{Clone}(\mathbf{A})$, of term functions of an algebra $\mathbf{A}$. These two examples are equivalent, in that $\mathrm{Clone}(\mathbf{A}) \cong \mathbf{C}(\mathbb{HSP}(\mathbf{A}))$. We will see this later when we develop the theory of clone representations.

Further examples of clones

1) Let $\mathbf{R}$ be a ring with unit. Then $\mathbf{M(R)}$, the clone of all matrices over $\mathbf{R}$, is defined as follows. Its underlying category is given by $\mathrm{Hom}(O_j, O_i)$ = the collection of all $(i \times j)$ matrices, with matrix multiplication for composition. Moreover, p_i^j is the $1 \times j$ row matrix $[0\ 0\ \cdots\ 0\ 1\ 0\ \cdots\ 0]$ with all entries zero except the ith one which has a one.

The jth universe, U_j, of $\mathbf{M(R)}$ consists of all $1 \times j$ matrices $[\alpha_0 \cdots \alpha_{j-1}]$ with $\alpha_i \in \mathbf{R}$. We can write such matrices as formal sums

$$\sum_{i=0}^{j-1} \alpha_i x_i.$$

Using this notation F_n^k can be expressed as

$$F_n^k\left(\sum_{i=0}^{k-1} \alpha_i x_i,\ \sum_{j=0}^{n-1} \beta_j^0 x_j, \ldots, \sum_{j=0}^{n-1} \beta_j^{k-1} x_j\right) = \sum_{j=0}^{n-1}\left(\sum_{i=0}^{k-1} \alpha_i \beta_j^i\right) x_j.$$

Let ${}_{\mathbf{R}}\mathcal{M}$ be the variety of left $\mathbf{R}$-modules (with one unary operation f_r for each $r \in \mathbf{R}$ given by $f_r(x) = rx$). Then it is easy to see that

$$\mathbf{C}({}_{\mathbf{R}}\mathcal{M}) \cong \mathbf{M(R)}.$$

2) Given a clone $\mathbf{C} = \langle \mathbf{C}, p_i^j, O_j \rangle$ we can construct a new clone

$$\mathbf{C}^{[k]} = \langle \mathbf{C}^{[k]}, p_i^{j[k]}, O_u^{[k]} \rangle,$$

where $O_j^{[k]} = O_{jk}$, and where $\mathbf{C}^{[k]}$ is the full subcategory of $\mathbf{C}$ determined by the $O_j^{[j]}$, and where $p_i^{j[k]}$ is the unique element of $\mathrm{Hom}_{\mathbf{C}}(O_{jk}, O_k)$ such that $p_s^k p_i^{j[k]} = p_{s+ik}^{jk}$, $0 \le s < k$, $0 \le i < j$. Using the elementary fact that $O_{jk} \cong (O_k)^j$ in $\mathbf{C}$, it is easy to check that $\mathbf{C}^{[k]}$ is a clone. $\mathbf{C}^{[k]}$ *is not a subclone of* $\mathbf{C}$, since the projections of $\mathbf{C}^{[k]}$ are not the projections of $\mathbf{C}$. We leave it as an exercise that $[\mathbf{M(R)}]^{[k]} \cong \mathbf{M}(\mathbf{R}^{[k]})$, where $\mathbf{R}^{[k]}$ is the ring of $k \times k$ matrices over $\mathbf{R}$.

Alternative approaches to clone theory

Various attempts have been made to encode clone theory into varieties of (total) algebras.

1) **Menger algebras** Let $\mathbf{C} = \langle U_1, U_2, \ldots, F_n^k, \ldots, p_i^j \ldots \rangle$ be a clone. Fix $k \geq 1$.

Let $\mathbf{C}_k = \langle U_k, F_k^k, p_i^k \rangle_{i<k}$. $\mathbf{C}_k$ is called the kth *Menger Algebra* associated with $\mathbf{C}$. These are total algebras that contain much of the information of $\mathbf{C}$, but there is some information loss.

Let $\mathcal{V}$ be any variety of algebras. Let $\mathcal{V}^{(k)} = \mathrm{Mod}\,((\mathrm{Eq}\mathcal{V}) \cap \text{equations in } k \text{ variables})$, then $\mathcal{V}^{(0)} \supseteq \mathcal{V}^{(1)} \supseteq \mathcal{V}^{(2)} \supseteq \cdots$ is an infinite descending chain of varieties that intersects to $\mathcal{V}$. Furthermore $[\mathbf{C}(\mathcal{V})]_k \cong \mathbf{C}(\mathcal{V}^{(k)})$. For more on this topic, see Trevor Evans [1].

2) **Neumann type clones** The idea here is to regard an n-ary operation α as an $\aleph_0$-ary operation that only depends on the first n variables. Then the corresponding clone-like structure has a single universe U, projections $p_1, p_2, \ldots$ and a single $\aleph_0$-ary composition operation F. The corresponding abstract definition is as follows:

An *infinitary algebra of clone type* is an algebra $\langle U, p_i (1 \leq i < w), F \rangle$, where the p_i are constants and F is an $\aleph_0$-ary operation on U. A $Clone_{WDN}$ is an infinitary algebra of this type that satisfies the following axioms:

$$\begin{aligned} F(p_i, \alpha_1, \alpha_2, \ldots) &\approx \alpha_i \qquad (i = 1, 2, \ldots) \\ F(\alpha, p_1, p_2, \ldots) &\approx \alpha \\ &\approx F(F(\alpha, \beta_1, \beta_2, \ldots),\ \gamma_1, \gamma_2, \ldots). \end{aligned}$$

There is a faithful functor from the category of clones to the category of Neumann type clones, but this functor is not onto. The problem is that a structure $Clone_{WDN}$ may contain elements that correspond to operations of essentially infinite rank. Technical difficulties have caused this approach to be largely abandoned. See W. D. Neumann [9].

3) **Mal'tsev iterative algebras** This approach has been extensively developed by D. Schweigert. For any clone $\mathbf{C}$, define an algebra $\mathbf{S}(\mathbf{C}) = \langle Z, *, \zeta, \tau, \Delta, e_0^2 \rangle$ of type $\langle 2, 1, 1, 0 \rangle$ as follows. Its universe is $Z = \bigcup U_n$ where the U_n are the universes of C. For $\alpha \in U_n$ and $\beta \in U_n$, define

$$\begin{aligned} \alpha * \beta &= F_{m+n-1}^m(\alpha, F_{m+n-1}^n(\beta, \rho_0^{m+n-1}, \ldots, \rho_{n-1}^{m_n-1}), \rho_n^{m+n-1}, \ldots, \rho_{m+n-2}^{m+n-1}) \\ \zeta(\alpha) &= F_n^m(\alpha, \rho_O^m, \ldots, \rho_{m-1}^m, \rho_0^m) \\ \tau(\alpha) &= F_m^n(\alpha, \rho_1^n, \rho_0^n, \rho_2^n, \rho_3^n, \ldots, \rho_{m-1}^n) \\ \Delta(\alpha) &= F_{n-1}^n(\alpha, \rho_0^{n-1}, \rho_0^{n-1}, \ldots, \rho_{n-2}^{n-1}) \end{aligned}$$

and

$$e_0^2 = p_0^2.$$

Define a $\mathbf{clone}_M$ to be a member of $\mathbb{HSP}(\{\mathbf{S}(\mathbf{C}) : \mathbf{C} \text{ is a clone}\})$. It is not hard to show that if $\mathbf{C}$ and $\mathbf{C}'$ are clones, then $\mathbf{S}(\mathbf{C}) \cong \mathbf{S}(\mathbf{C}')$.

Problem Find a nice axiomatization of the resulting variety.

Representations of clones

Just as an abstract semigroup can be represented as a semigroup of selfmaps, so an abstract clone can be represented as a clone of operations.

Definition 4 A *representation* of a *clone* $\mathbf{C}$ is a clone homomorphism $\varphi : \mathbf{C} \to \mathrm{Clo}(A)$ where $\mathrm{Clo}(A)$ is the clone of all operations on a set A.

A representation is called *faithful* if it is one-to-one. (One-oneness of one map is equivalent to one-oneness for the other map.)

Example Let $\mathcal{V}$ be a variety and let $\mathbf{A}$ be an algebra in $\mathcal{V}$. Define $\varphi_{\mathbf{A}} : \mathbf{C}(\mathcal{V}) \to \mathrm{Clo}(A)$ by

$$\varphi_{\mathbf{A}}(p^{\dagger}) = p^{\mathbf{A}},$$

where $p^{\dagger}$ is the term function corresponding to the term p. Then $\varphi_{\mathbf{A}}$ is a representation of $\mathbf{C}(\mathcal{V})$.

Conversely suppose $\varphi : \mathbf{C}(\mathcal{V}) \to \mathrm{Clo}(A)$ is representation and let

$$\varphi^{\circ} = \langle A, \varphi(Fx_0 \cdots x_{n-1}^{\dagger})\rangle_{F\in \text{ type of } \mathcal{V}}.$$

Then $\varphi^{\circ} \in \mathcal{V}$ and $\varphi \mapsto \varphi^{\circ}$ is inverse to $\mathbf{A} \mapsto \varphi_{\mathbf{A}}$. (The fact that $\varphi^{\circ} \in \mathcal{V}$ will be proved in the next lecture.)

For the purpose of building representations of arbitrary clones, we proceed as follows. For an arbitrary clone $\mathbf{C} = \langle U_1, U_2, \ldots, F_k^n, \ldots, p_i^j, \ldots\rangle$, define $L_n : \mathbf{C} \to \mathrm{Clo}(U_n)$ as follows. For $\alpha \in U_k$, define $[L_n(\alpha)](\beta_0, \ldots, \beta_{k-1}) = F_n^k(\alpha, \beta_0, \ldots, \beta_{n-1})$. It is straightforward to check that L_n is a clone homomorphism. Note that if $\alpha \in U_n$, then

$$[L_n(\alpha)](p_0^n, \ldots, p_{n-1}^n) = F_n^n(\alpha, p_0^n, \ldots, p_{n-1}^n) = \alpha.$$

So if $\alpha, \alpha' \in U_n$ with $\alpha \neq \alpha'$, then $L_n(\alpha) \neq L_n(\alpha')$. Thus L_n is 1–1 on U_n. A slight modification of this proof yields that L_n is 1–1 on U_k for $k \leq n$. L_n is called the *n-ary left regular representation* of $\mathbf{C}$.

Lecture III

Note that if V is the variety of modular lattices with $C(V) = \langle U_1, U_2, \ldots\rangle$, then $|U_1| = 1$, $|U_2| = 4$, $|U_3| = 28$, $|U_4| = \aleph_0, \ldots$. This suggests that U_k is in some sense equivalent to $F_V(k)$. Modulo a technical point about similarity type, this turns out to be true. In order to see this we need to develop a notion of an algebra attached to a clone representation φ that is not so dependent upon similarity type as φ°.

Let $\mathbf{C}$ be a clone. Define a similarity type τ consisting of one k-ary operation symbol F_α for each $\alpha \in U_k$ and for each $k \geq 1$. For $\varphi : \mathbf{C} \to \mathrm{Clo}A$ define an algebra φ^* of type τ by

$$\varphi^* = \langle A, F_\alpha^A := \varphi(\alpha)\rangle_{\alpha\in\bigcup_{i=1}^{\infty} U_i}.$$

Theorem 2 $\mathcal{K} = \{\varphi^* : \varphi$ *is a representation of* $\mathbf{C}\}$ *is a variety of type* τ.

Proof We need to show that $\mathcal{K} = \mathbb{HSP}(\mathcal{K})$. Suppose $\mathbf{A} \in \mathcal{K}$ and $h : \mathbf{A} \twoheadrightarrow \mathbf{B}$ is an onto homomorphism. Then there is a clone representation $\varphi : \mathbf{C} \to \mathrm{Clo}(A)$ such that $\mathbf{A} = \varphi^*$. Define $\psi : \mathbf{C} \to \mathrm{Clo}(B)$ by $[\psi(\alpha)](b_1, \ldots, b_n) = F^{\mathbf{B}}_\alpha(b_1, \ldots, b_n)$. Since $F^{\mathbf{B}}_\alpha(h(a_1), \ldots, h(a_n)) = h(F^{\mathbf{A}}_\alpha(a_1, \ldots, a_n)) = h([\phi(\alpha)](a_1, \ldots, a_n))$ and h is a homomorphism, it is easy to see that ψ is a clone representation and that $\psi^* = \mathbf{B}$. Thus $\mathcal{K} = \mathbb{H}(\mathcal{K})$.

If $\mathbf{A} = \varphi^* \in \mathcal{K}$ and $\mathbf{B} \leq \mathbf{A}$, then clearly $\psi : \mathbf{C} \to \mathrm{Clo}(B)$ given by $\psi(\alpha) = \varphi(\alpha) \mid B^n$ is a clone representation with $\mathbf{B} = \psi^*$. Thus $\mathcal{K} = \mathbb{S}(\mathcal{K})$.

Finally suppose $\mathbf{A}_i = \varphi_i^* \in \mathcal{K}$ for $i \in I$. Define $\varphi : \mathbf{C} \to \mathrm{Clo}(\prod_{i\in I} A_i)$ by

$$[\varphi(\alpha)](\langle a_1^i : \tau \in I\rangle, \langle a_2^i : i \in I\rangle, \ldots, \langle a_n^i : i \in I\rangle)$$
$$= \langle [\varphi_i(\alpha)](a_1^i, \ldots, a_n^i) : i \in I\rangle,$$

then it is easy to see that $\varphi^* = \prod_{i\in I} \mathbf{A}_i$, and hence that $\mathcal{K} = \mathbb{HSP}(\mathcal{K})$. □

We will call this variety $\mathrm{Var}(\mathbf{C})$. Recall that L_n is a representation that is one-to-one on U_i $(i \leq n)$.

Corollary 1 $\mathrm{Var}(\mathbf{C}) \models F_\alpha(v_0, v_1, \ldots, v_{n-1}) \approx F_\beta(v_0, v_1, \ldots, v_{n-1})$ *iff* $\alpha = \beta$, *where* $\alpha, \beta \in U_n$.

Proof Suppose $\alpha, \beta \in U_n$ such that $\alpha \neq \beta$. Then $L_n(\alpha) \neq L_n(\beta)$. Thus $(L_n)^* \not\models F_\alpha(v_0, \ldots, v_{n-1}) \approx F_\beta(v_0, \ldots, v_{n-1})$. Thus $\mathrm{Var}(\mathbf{C}) \not\models F_\alpha(v_0, \ldots, v_{n-1}) \approx F_\beta(v_0, \ldots, v_{n-1})$. The converse is trivial. □

In order to sharpen this corollary we must first relate the terms in the language of $\mathrm{Var}(\mathbf{C})$ with $\mathbf{C}$.

Lemma 1 *Let* p *be a term in the language of* $\mathrm{Var}(\mathbf{C})$. *Then there exists* $\alpha \in \mathbf{C}$ *such that* $\mathrm{Var}(\mathbf{C}) \models p \approx F_\alpha(v_0, \ldots, v_{n-1})$ *where* $\alpha \in U_n$.

Proof If p is a variable, i.e., $p = v_i$, then $\alpha = p_i^n$ works, for any $n > i$. Suppose $p = F_\beta(p_0, \ldots, p_{n-1})$ where $\beta \in U_n$ and the lemma is true for each p_i. For each $i < n$, fix $\alpha_i \in U_{n_i}$ with $\mathrm{Var}(\mathbf{C}) \models p_i \approx F_{\alpha_i}(v_0, \ldots, v_{n_i-1})$. Let $m = \max\{n_i \mid 0 \leq i < n\}$.

For $i < n$ let $\hat{\alpha}_i = F_m^{n_i}(\alpha_i, p_0^m, \ldots, p_{n_i-1}^m)$. Then $\mathrm{Var}(\mathbf{C}) \models p_i \approx F_{\hat{\alpha}_i}(v_0, \ldots, v_{n-1})$ for $i = 1, \ldots, n$. Let $\gamma = F_m^n(\beta, \hat{\alpha}_0, \ldots, \hat{\alpha}_{n-1})$. Then $\mathrm{Var}(\mathbf{C}) \models p \approx F_\gamma(v_0, \ldots, v_{n-1})$. □

Corollary 2 $\varphi : \mathbf{C} \to \mathrm{Clo}A$ *is faithful (one-to-one on every* U_n*) iff* φ^* *is generic in* $\mathrm{Var}(\mathbf{C})$, *i.e.,* $\mathrm{Var}(\mathbf{C}) = \mathbb{HSP}(\varphi^*)$.

Proof φ^* is not generic if and only if $\mathrm{Var}(\mathbf{C}) \not\models p \approx q$ but $\varphi^* \models p \approx q$ for some terms p, q. By the previous lemma, this is true if and only if there are $\alpha, \beta \in U_n$ such that $\mathrm{Var}(\mathbf{C}) \not\models$

$F_\alpha(v_0,\ldots,v_{n-1}) \approx F_\beta(v_0,\ldots,v_{n-1})$ but $\varphi^* \models F_\alpha(v_0,\ldots,v_{n-1}) \approx F_\beta(v_0,\ldots,v_{n-1})$. By Corollary 1 of Theorem 1, Theorem 3 is true if and only if $\alpha \neq \beta$ but $\varphi^* \models F_\alpha(v_0,\ldots,v_{n-1}) \approx F_\beta(v_0,\ldots,v_{n-1})$, if and only if $\alpha \neq \beta$ but $\varphi(\alpha) = \varphi(\beta)$, if and only if φ is not faithful. □

Corollary 3 *Every clone has a faithful representation.* □

Clone(**A**) is not the only clone naturally defined on an algebra **A**. There is also, for example, Pol(**A**), the clone of polynomial functions of **A**. A less obvious example can be defined as follows. Suppose that **A** is in a variety $\mathcal{V}$. Define $\mathrm{Clone}_{\mathcal{V}}(\mathbf{A})$ by

$$\mathrm{Clone}_{\mathcal{V}}(\mathbf{A}) := \{f : \mathbf{A}^n \to \mathbf{A} \mid f \text{ is a } \mathcal{V}\text{-homomorphism}, \ n \geq 1\}.$$

Such concrete clones also yield representations of arbitrary clones. In fact, by the next theorem, we can take **A** to be a unary algebra. This theorem appears (in a slightly different form) in Schein and Trohimenko [13]. See also Sangalli [12].

Theorem 3 *Suppose* **C** *is a clone. Then there exists a unary algebra* **A** *with* $\mathbf{C} \cong \mathrm{Clone}_U \mathbf{A}$, *where* U *is the variety of all unary algebras of the same type as* **A**. *Moreover, we can let* $\mathbf{A} = \langle A, \mathrm{End}\, F_{\mathrm{Var}(\mathbf{C})}(\omega)\rangle$ *where* A *is the universe of the free algebra* $F_{\mathrm{Var}(\mathbf{C})}(\omega)$.

Proof Fix a representation $\phi : \mathbf{C} \to \mathrm{Clo}A$ such that $\phi^* = F_{\mathrm{Var}(\mathbf{C})}(\omega)$. Such a ϕ exists by Theorem 1. By Corollary 2, ϕ is a faithful representation. It is clear by the relevant definitions that the range of ϕ is $\mathrm{Clone}(F_{\mathrm{Var}(\mathbf{C})}(\omega))$, the clone of term functions on $F_{\mathrm{Var}(\mathbf{C})}(\omega)$. Thus it suffices to show that $\mathrm{Clone}(F_{\mathrm{Var}(\mathbf{C})}(\omega)) = \mathrm{Clone}_U(\mathbf{A})$ where $\mathbf{A} = \langle A, \mathrm{End}\, F_{\mathrm{Var}(\mathbf{C})}(\omega)\rangle$. Suppose that $p = p(x_0,\ldots,x_{n-1}) \in \mathrm{Clone}\, F_{\mathrm{Var}(\mathbf{C})}(\omega)$ and that $f \in \mathrm{End}\, F_{\mathrm{Var}(\mathbf{C})}(\omega)$. Then clearly $p(f(a_0),\ldots,f(a_{n-1})) = f(p(a_0,\ldots,a_{n-1}))$, whence $p \in \mathrm{Clone}_U(\mathbf{A})$.

Conversely, suppose that $F \in \mathrm{Clone}_U(\mathbf{A})$. Then $F : \mathbf{A}^n \to \mathbf{A}$ is a homomorphism for some fixed $n \geq 1$. Let $q(x_0,\ldots,x_{n-1}) = F(x_0,\ldots,x_{n-1})$, where the x_i are the free generators of $F_{\mathrm{Var}(\mathbf{C})}(\omega)$ and without loss of generality $m \geq n$. Let p be the n-ary term function given by $p(\alpha_0,\ldots,\alpha_{n-1}) = q(\alpha_0,\ldots,\alpha_{n-1},\ldots,\alpha_{n-1})$. More formally, $p(\alpha_0,\ldots,\alpha_{n-1}) = q(\alpha_0,\ldots,\alpha_{n-1},\ldots,\alpha_{n-1})$. More formally, $p(\alpha_0,\ldots,\alpha_{n-1})$ is the image of $q; (x_0, x_1,\ldots,x_{m-1})$ under the endomorphism of $F_{\mathrm{Var}(\mathbf{C})}(\omega)$ generated by mapping x_i to α_i $(i < n)$ and x_j to α_{n-1} $(j \geq n)$. By freeness this definition of p makes sense. Since F is assumed to commute with all endomorphisms, it is clear that $p = F$, completing the proof. □

Remark Since a clone representation $\phi : \mathbf{C} \to \mathrm{Clo}A$ gives rise to the algebra $\phi^* \in \mathrm{Var}(\mathbf{C})$, we will be somewhat imprecise and simply call the situation $\phi : \mathbf{C} \to \mathrm{Clo}A$ an algebra.

Equivalence of varieties

In order to discuss the relation between $\mathcal{V}$ and $\mathrm{Var}(\mathbf{C}(\mathcal{V}))$, we need a notion of when two varieties are equivalent.

Definition 5 Let $\mathcal{V}$ be a variety of type σ and let $\mathcal{W}$ be a variety of type τ. An *interpretation* of $\mathcal{V}$ in $\mathcal{W}$ is an assignment $D : F \mapsto F_D$ of $\mathcal{W}$-terms F_D to operation symbols F of $\mathcal{V}$, such that

1. $\tau(F_C) = \max\{1, \sigma(F)\}$.
2. If $\sigma(F) = 0$, then $\mathcal{V} \models F_D(x) \approx F_D(y)$.
3. If $\mathbf{A} \in \mathcal{W}$, then $\mathbf{A}^D := \langle A, \dot{F}_D^{\mathbf{A}}(F \in \mathrm{Dom}(\sigma))\rangle \in \mathcal{V}$, where $\dot{F}_D^{\mathbf{A}} = F_D^{\mathbf{A}}$ if $\sigma(F) > 0$, otherwise $\dot{F}_D^{\mathbf{A}}$ = the unique constant $C \in A$ such that $F_D^A(x) \approx C$.

Definition 6 Let $\mathcal{V}$ and $\mathcal{W}$ be varieties. Then $\mathcal{V}$ is *equivalent* to $\mathcal{W}$, written $\mathcal{V} \equiv \mathcal{W}$, iff there is an interpretation, D, of $\mathcal{V}$ in $\mathcal{W}$ and an interpretation, E, of $\mathcal{W}$ in $\mathcal{V}$ such that for all $\mathbf{A} \in \mathcal{V}$ and all $\mathbf{B} \in \mathcal{W}$ we have $\mathbf{A} = \mathbf{A}^{ED}$ and $\mathbf{B} = \mathbf{B}^{DE}$.

By way of illustration, note that this definition covers the well-known equivalence between Boolean algebras and Boolean rings.

Theorem 4 *Let $\mathcal{V}$ and $\mathcal{W}$ be varieties. Then $\mathcal{V} \equiv \mathcal{W}$ iff there is a concrete isomorphism between them, i.e., a category isomorphism $\varphi : \mathcal{V} \to \mathcal{W}$ that commutes with the forgetful functor.*

Proof

($\Rightarrow$) If $D : \mathcal{V} \to \mathcal{W}$ and $E : \mathcal{W} \to \mathcal{V}$ are mutually inverse interpretations, then $\mathbf{A} \mapsto \mathbf{A}^E$ yields the desired concrete isomorphism of $\mathcal{V}$ with $\mathcal{W}$.

($\Leftarrow$) Suppose that $\varphi : \mathcal{V} \to \mathcal{W}$ is a concrete category isomorphism. Let $\psi = \varphi^{-1}$. Suppose that X is a nonempty set and that $h : X \to \mathbf{A} \in \mathcal{W}$. Then $h : X \to \psi(\mathbf{A}) \in \mathcal{V}$. Thus there is a unique $\mathcal{V}$-homomorphism $\hat{h} : F_{\mathcal{V}}(X) \to \psi(\mathbf{A})$ extending h. Thus $\varphi(\hat{h}) = h : \varphi(F_{\mathcal{V}}(X)) \to \varphi\psi(\mathbf{A}) = \mathbf{A}$ is a $\mathcal{W}$ homomorphism extending h. Since φ is an isomorphism, it is clearly unique. Thus $\varphi(F_{\mathcal{V}}(X)) \cong F_{\mathcal{W}}(X)$. Similarly $\psi(F_{\mathcal{W}}(X)) \cong F_{\mathcal{V}}(X)$.

For F a basic $\mathcal{V}$-operation symbol, we define F_D as follows:

Case 1 $\sigma(F) = n > 0$, where σ is the similarity type of $\mathcal{V}$. Then $F^\dagger(v_0, \ldots, v_{n-1}) \in F_{\mathcal{V}}(n) = F_{\mathcal{W}}(n)$ (since φ commutes with the forgetful functor). Thus there is an n-ary $\mathcal{W}$-term F_D such that $F_D^\dagger(v_0, \ldots, v_{n-1}) = F^\dagger(v_0, \ldots, v_{n-1})$ in $F_{\mathcal{W}}(n)$.

Case 2 $\sigma(F) = 0$. Then there is a unary $\mathcal{W}$-term F_D such that $F_D^\dagger(v_0) = F^\dagger(v_0)$ in $F_{\mathcal{W}}(1)$. If $\mathcal{W} \not\models F_D(x) \approx F_D(y)$, then in $F_{\mathcal{W}}(2)$ $F_D^\dagger(v_0) \neq F_D^\dagger(v_1)$. Considering $F_{\mathcal{W}}(1)$ to be a subalgebra of $F_{\mathcal{W}}(2)$, we have that $F_D^\dagger(v_0) = F^\dagger(v_0)$. Let $h : \{v_0\} \to \{v_0, v_1\}$ be given by $h(v_0) = v_1$. In $\mathcal{V}$, h extends to a homomorphism $\hat{h} : F_{\mathcal{V}}(1) \to F_{\mathcal{V}}(2)$. Clearly $\hat{h}(F^\dagger(v_0)) = F^+(v_0)$ in $F_{\mathcal{V}}(1)$, since F is a constant in $\mathcal{V}$. But $\hat{h}$ is also a homomorphism between $F_{\mathcal{W}}(1)$ and $F_{\mathcal{W}}(2)$ with $\hat{h}(F^\dagger(v_0)) = \hat{h}(F_D^\dagger(v_0)) = F_D^\dagger(v_1) \neq F^\dagger(v_0)$, a contradiction. Thus $\mathcal{W} \models F_D(x) \approx F_D(y)$.

Arguing similarly, we can easily check that D is in fact an interpretation of $\mathcal{V}$ in $\mathcal{W}$ such that $\psi(\mathbf{B}) = \mathbf{B}^D$ for any $\mathbf{B} \in \mathcal{W}$. Similarly we can obtain from ψ an interpretation E of $\mathcal{W}$ in $\mathcal{V}$ such that $\varphi(\mathbf{A}) = \mathbf{A}^E$ for all $\mathbf{A} \in \mathcal{V}$, showing that $\mathcal{V} \equiv \mathcal{W}$. □

This characterization of equivalence leads to the following theorem.

Theorem 5 $\mathcal{V} \equiv \mathrm{Var}(\mathbf{C}(\mathcal{V}))$.

Proof We have already seen that $\varphi \mapsto \varphi^{\circ} \in \mathcal{V}$ is a bijective correspondence between representations of $\mathbf{C}(\mathcal{V})$ and algebras in $\mathcal{V}$. Similarly $\varphi \mapsto \varphi^{*}$ is a bijective correspondence between $\mathrm{Var}(\mathbf{C}(\mathcal{V}))$ algebras and clone representations. Furthermore if $\varphi : \mathbf{C}(\mathcal{V}) \to \mathrm{Clo}(A)$ is a representation of $\varphi(\mathcal{V})$, then both φ° and φ^{*} have underlying universe A. Thus it is easy to see that composing these two correspondences, i.e., $\varphi^{\circ} \mapsto \varphi^{*}$, yields the desired concrete isomorphism between $\mathcal{V}$ and $\mathrm{Var}(\mathbf{C}(\mathcal{V}))$, whence $\mathcal{V} \equiv \mathrm{Var}(\mathbf{C}(\mathcal{V}))$. □

Theorem 6 $\mathbf{C} \cong \mathbf{C}(\mathrm{Var}(\mathbf{C}))$.

Proof $\alpha \mapsto F_{\alpha}(v_0, \ldots, v_n)^{\dagger}$ is the desired isomorphism. □

Theorem 7 1. $\mathbf{C} \cong \mathbf{D}$ *iff* $\mathrm{Var}(\mathbf{C}) \equiv \mathrm{Var}(\mathbf{D})$.

2. $\mathcal{V} \equiv \mathcal{W}$ *iff* $\mathbf{C}(\mathcal{V}) \cong \mathbf{C}(\mathcal{W})$. □

Special types of representations

If $\mathbf{A}$ is a concrete category with concrete products and $\mathbf{C}$ is a clone, then an $\mathbf{A}$-*representation* of $\mathbf{C}$ is a clone homomorphism $\varphi : \mathbf{C} \to \mathrm{Clo}(A)$ where $A \in \mathbf{A}$ and $\varphi(U_n) \subset \mathrm{Mor}_{\mathbf{A}}(A^n, A)$ for every $n \geq 1$. For example if $\mathbf{A} = \mathrm{Top}$, then a Top-representation of $\mathbf{C}$ is essentially a *topological* $\mathrm{Var}(\mathbf{C})$*-algebra*.

Suppose that $\mathcal{V}$ and $\mathcal{W}$ are varieties. Then a $\mathcal{W}$-representation of $\mathbf{C}(\mathcal{V})$, $\varphi : \mathbf{C}(\mathcal{V}) \to \mathrm{Clo}(A)$ yields an algebra $\varphi^{\circ} \in \mathcal{V}$ whose universe is A for some $\mathbf{A} \in \mathcal{W}$, and whose operations are $\mathcal{W}$-homomorphism of $\mathbf{A}^n$ into $\mathbf{A}$.

Suppose that F is an n-ary $\mathcal{V}$-operation and G is an m-ary $\mathcal{W}$-operation. The condition that $F^{\mathbf{A}}$ is an n-ary $\mathcal{W}$-homomorphism means that given an $m \times n$ matrix $[a_{ij}]$ of elements from A,

$$(*) \qquad \begin{aligned} & F^{\mathbf{A}}(G^{\mathbf{A}}(a_{00}, \ldots, a_{m-1\ 0}), \ldots, G^{\mathbf{A}}(a_{00}, \ldots, a_{m-1\ n-1})) \\ &= G^{\mathbf{A}}(F^{\mathbf{A}}(a_{00}, \ldots, a_{0\ n-1}), \ldots, F^{\mathbf{A}}(a_{m-1\ 0}, \ldots, a_{m-1\ n-1})), \end{aligned}$$

i.e., the result of first evaluating down the columns with G and then evaluating the resulting vector with F is the same as first evaluating along the rows with F and then evaluating the resulting vector with G. There is an obvious symmetry here (revealed by considering $[a_{ij}]^T$) so that any $\mathcal{W}$-representation of $\mathbf{C}(\mathcal{V})$ is equally a $\mathcal{V}$-representation of $\mathbf{C}(\mathcal{V})$.

In this situation we say that $\mathbf{A}$, equipped with both its $\mathcal{V}$- and its $\mathcal{W}$-operations, is in the variety $\mathcal{V} \otimes \mathcal{W} := \{\varphi^{\circ} \mid \varphi : \mathbf{C}(\mathcal{V}) \to \mathrm{Clo}(\mathbf{A}) \text{ is a } \mathcal{W}\text{-representation}\}$. It is routine to check that $\mathcal{V} \otimes \mathcal{W}$ is in fact a variety. More precisely, the similarity type of $\mathcal{V} \otimes \mathcal{W}$ is the disjoint union of the similarity type of $\mathcal{V}$ and the similarity type of $\mathcal{W}$. Its axioms are the axioms of $\mathcal{V}$ together with the axioms of $\mathcal{W}$, together with an instance of $(*)$ for every n-ary $\mathcal{V}$-operation F and every m-ary $\mathcal{W}$-operation G.

An interesting feature of this definition is that $\mathcal{V} \otimes \mathcal{W}$ is often trivial. For instance let $\mathcal{SL}$ be the variety of semilattices with operation denoted $\vee$, and let $\mathcal{L}$ be the variety of lattices with operations denoted $+, \cdot$. Then in $\mathcal{SL} \otimes \mathcal{L}$,

$$x + y \approx (x + y) \vee (y + x) \approx (x \vee y) + (y \vee x) \approx x \vee y.$$

Similarly $x \cdot y \approx x \vee y$. Whence $x \cdot y \approx x + y$, which only holds in trivial lattices.

Lecture IV

We begin with some interesting facts about the tensor product, $\mathcal{V} \otimes \mathcal{W}$, of two varieties. As we saw last time, it frequently happens that $\mathcal{V} \otimes \mathcal{W} \models x \approx y$. In 1976 J. Isbell proved that if $\mathcal{V}$ satisfies $\mathcal{V} \otimes \mathcal{W} \models x \approx y$ for *some* nontrivial variety $\mathcal{W}$, then $\mathcal{V} \otimes \mathcal{BA} \models x \approx y$ for $\mathcal{BA}$ the variety of Boolean algebras. Here we will prove this result for distributive lattices with 0 and 1 ($\mathcal{D}_{01}$), using Priestley duality. (The Isbell result is obviously a corollary; moreover its original proof used Stone duality in a similar manner.) For the relevant notions of duality, the reader is referred to Brian Davey's article in this volume.

Theorem 8 *Let $\mathcal{V}$ be any variety. The following are equivalent:*

1. *There exists a nontrivial variety $\mathcal{W}$ such that $\mathcal{V} \otimes \mathcal{W} \models x \approx y$.*

2. $\mathcal{V} \otimes \mathcal{D}_{01} \models x \approx y$.

Proof Of course (2) $\Rightarrow$ (1) is trivial. We prove (1) $\Rightarrow$ (2) by contraposition. Supposing that $\mathbf{B}$ is a non-trivial model of $\mathcal{V} \otimes \mathcal{D}_{01}$ and that $\mathbf{A}$ is a non-trivial model of $\mathcal{W}$, we will construct a non-trivial algebra $\mathbf{C}$ in $\mathcal{V} \otimes \mathcal{W}$.

By Priestley duality, we may suppose that the algebra $\mathbf{B}$ has the following form, for some Priestley ordered space P:

(a) $B = \{\alpha \mid \alpha : P \to \{0, 1\},$ continuous & order-preserving$\}$.

(b) the $\mathcal{D}_{01}$-operations of $\mathbf{B}$ are those acquired pointwise from the usual operations on $\{0, 1\}$.

(c) for each $\mathcal{V}$-operation F of $\mathbf{B}$, there is an order-preserving continuous $G : P \to \coprod_n P$ such that, $\forall \lambda \in P, \forall \alpha_1, \ldots, \alpha_n \in B$, $[F(\alpha_1, \ldots, \alpha_n)](\lambda) = \alpha_j(\mu)$, where $G(\lambda) = (j, \mu)$.

To construct non-trivial $\mathbf{C} \in \mathcal{V} \otimes \mathcal{W}$, we define its universe

$$C = \{\gamma \mid \gamma : P \to A, \text{ continuous, order-preserving}\},$$

with continuity and order-preservation understood relative to the discrete topology on A, and w.r.t. the total disorder on A. In other words, $\gamma \in C$ iff γ is continuous and γ is constant on each connected component of P.

$C \neq \emptyset$, because it contains all constant functions. In fact $|C| \geq 2$ for the same reason. Let us first check that C is a subuniverse of $\mathbf{A}^P$. We give the proof that C is closed under the pointwise application of a binary $\mathbf{A}$-operation $+$. The continuity of $\gamma_1 + \gamma_2$ follows from the equation

$$(\gamma_1 + \gamma_2)^{-1}(a) = \bigcup_{b+c=a} [\gamma_1^{-1}(b) \cap \gamma_2^{-1}(c)].$$

For order-preservation, it is obvious that if γ_1 and γ_2 are both constant on a certain connected component of P, then the same is true of $\gamma_1 + \gamma_2$. Thus, we may take $\mathcal{W}$-operations to be defined pointwise on C, and so $\mathbf{C} \in \mathbb{SPA} \subseteq \mathcal{W}$.

Now the $\mathcal{V}$-operations are defined on C by analogy with (c) above, namely:

$$[F(\gamma_1, \ldots, \gamma_n)](\lambda) = \gamma_j(\mu), \quad \text{where} \quad G(\lambda) = \langle j, \mu \rangle.$$

The continuity of $F(\gamma_1, \ldots, \gamma_n)$ is assured by the set-theoretic equation

$$[F(\gamma_1, \ldots, \gamma_n)]^{-1}(a) = \bigcup_{j=1}^{n} G^{-1}[P_j \cap \gamma_j^{-1}(a)],$$

where P_j denotes the jth copy of P in $\coprod_n P$. We also need to show that $F(\gamma_1, \ldots, \gamma_n)$ is constant over any connected component K of P. Observe first that for $\lambda \in K$, $G(\lambda) = \langle j, \mu \rangle$ has constant j, and has μ ranging over a connected component of P. Therefore $\gamma_j(\mu)$ is constant for $\lambda \in K$. In other words $[F(\gamma_1, \ldots, \gamma_n)](\lambda)$ is constant for $\lambda \in K$. Thus F is well defined on C.

We omit the easy proof that the $\mathcal{W}$-operations $+$ commute with the $\mathcal{V}$-operations F. It now remains only to prove that our $\mathcal{V}$-operations actually define an algebra in $\mathcal{V}$. To do this, we will show that $\langle C, \cdots F \cdots \rangle$ lies in $\mathbb{ISP}\{\mathbf{B}\}$.

For any function $\varphi : A \to \{0,1\}$, define $\bar{\varphi} : \mathbf{C} \to \mathbf{B}$ via $\bar{\varphi}(\gamma) = \varphi \circ \gamma$. Since φ is both continuous and order-preserving (w.r.t. discrete topology and total disorder), it is obvious that $\bar{\varphi}[C] \subseteq B$. The following calculations prove that $\bar{\varphi}$ is a homomorphism:

$$\begin{aligned}
[F(\bar{\varphi}\gamma_1, \ldots, \bar{\varphi}\gamma_n)](\lambda) &= [\bar{\varphi}\gamma_j](\mu) \\
&= \varphi[\gamma_j(\mu)] \quad \text{where} \quad G(\lambda) = \langle j, \mu \rangle,
\end{aligned}$$

$$\begin{aligned}
[\bar{\varphi}F(\gamma_1, \ldots, \gamma_n)](\lambda) &= \varphi([F(\gamma_1, \ldots, \gamma_n)](\lambda)) \\
&= \varphi(\gamma_j(\mu)) \quad \text{where} \quad G(\lambda) = \langle j, \mu \rangle.
\end{aligned}$$

Thus $\bar{\varphi}$ is a homomorphism.

To complete the proof, we will show that if $\Phi = \{\varphi : A \to \{0,1\}\}$ is a family of maps separating points of A, then $\bar{\Phi} = \{\bar{\varphi} \mid \varphi \in \Phi\}$ is a family of homomorphisms that separates points of $\mathbf{C}$. So, suppose that $\gamma_1, \gamma_2 \in C$, $\gamma_1 \neq \gamma_2$. Then $\gamma_1(\lambda) \neq \gamma_2(\lambda)$ for some $\lambda \in P$.

Therefore $\varphi(\gamma_1(\lambda)) \neq \varphi(\gamma_2(\lambda))$ for some $\lambda \in P$ and some $\varphi \in \Phi$. By definition of $\bar{\varphi}$, we have $\bar{\varphi}(\gamma_1)$ $/\bar{\varphi}(\gamma_2)$. Hence $\bar{\Phi}$ separates points of $\mathbf{C}$, and so $\mathbf{C} \in \mathbb{ISP}(\mathbf{B})$. □

There do exist interesting non-trivial varieties $\mathcal{V}$ to which Theorem 8 does not apply, i.e., for which $\mathcal{V} \otimes \mathcal{D}_{01}$ does not satisfy $x \approx y$. A prime example is $\mathcal{S}^{(2)}$ just below.

$[k]$-Powers again

Recall that $\mathrm{Hom}_{\mathbf{C}^{[2]}}(O_n^{[2]}, O_1^{[2]}) = \mathrm{Hom}_{\mathbf{C}}(O_{2n}, O_2) = U_n^{[2]}$ in the clone $\mathbf{C}^{[2]}$, where $\mathbf{C}$ is an arbitrary clone. This means that an n-ary element of $\mathbf{C}^{[2]}$ is defined by a pair of $2n$-ary elements of $\mathbf{C}$. Given a clone representation $\varphi : \mathbf{C} \to \mathrm{Clo}A$ we can define a representation $\varphi^{[2]} : \mathbf{C}^{[2]} \to \mathrm{Clo}A^2$ by

$$\begin{aligned} & [\varphi^{[2]}\langle F_0, F_1\rangle]\langle (a_0^0, a_1^0), \ldots, (a_0^{n-1}, a_1^{n-1})\rangle \\ = \;& \langle [\varphi(F_0)](a_0^0, a_1^0, a_0^1, a_1^1, \ldots, a_0^{n-1}, a_1^{n-1}), [\varphi(F_1)](a_0^0, a_1^0, \ldots, a_0^{n-1}, a_1^{n-1})\rangle. \end{aligned}$$

We leave it to the reader to verify that this is a representation and that all representations of $\mathbf{C}^{[2]}$ are of this form. Thus the variety $\mathcal{V}^{[2]}$ associated to $\mathbf{C}^{[2]}$ contains algebras that are essentially *squares* of algebras in the variety $\mathcal{V}$ associated to $\mathbf{C}$, and only those algebras. From these facts one can prove without too much trouble that, for any variety $\mathcal{V}$, $\mathcal{V}^{[2]}$ is equivalent to $\mathcal{S}^{[2]} \otimes \mathcal{V}$, where $\mathcal{S}$ is the variety with no operations. ($\mathcal{S}$ stands for *sets*.) Thus, in particular, $\mathcal{S}^{[2]}$ provides an example of a variety such that $\mathcal{S}^{[2]} \otimes \mathcal{D}_{01}$ is non-trivial. (Of course 2 can be replaced by any positive integer k.)

Recall that if $\mathcal{V}$ is a variety, then $\mathrm{Spec}(\mathcal{V})$, the spectrum of $\mathcal{V}$, is defined to be $\{n \in \omega \mid \exists\, \mathbf{A} \in \mathcal{V}(|\mathbf{A}| = n)\}$. The above characterization of representations of $\mathbf{C}^{[2]}$ thus yields as an immediate corollary that

$$\mathrm{Spec}(\mathcal{V}^{[2]}) = \{n^2 : n \in \mathrm{Spec}(\mathcal{V})\}.$$

Clone terms

Suppose that $\mathcal{V}$ and $\mathcal{W}$ are varieties such that $\mathcal{V} \neq \mathcal{W}$ but $\mathbf{C}(\mathcal{V}) \cong \mathbf{C}(\mathcal{W})$. Is it possible to say anything meaningful within clone theory about the difference between $\mathcal{V}$ and $\mathcal{W}$, or does the clone perspective entail some inevitable loss of information? We shall see shortly that $\mathcal{V}$ and $\mathcal{W}$ can indeed be distinguished as yielding different *clone presentations* of $\mathbf{C}(\mathcal{V})$ ($\cong \mathbf{C}(\mathcal{W})$). In order to make this precise we need to address the appropriate logical notions.

Definition 7 A *graded set* X is a countable disjoint union of sets $X = X_1 \dot{\cup} X_2 \dot{\cup} \cdots$ indexed by the positive integers.

Definition 8 Let X be a graded set. The set of *clone terms* over X, $T_{\mathrm{clones}}(X)$ is the smallest graded set $T = T_1 \dot{\cup} T_2 \dot{\cup} \cdots$ such that

1. $X_n \subseteq T_n$, $n = 1, 2, 3, \ldots$;

2. $p_i^j \in T_j$, $0 \leq i < j$, $j = 1, 2, 3, \ldots$;

3. if $\beta \in T_k$ and $\alpha_0, \ldots, \alpha_{k-1} \in T_n$, then $F_n^k(\beta, \alpha_0, \ldots, \alpha_{k-1}) \in T_n$, $n, k = 1, 2, 3, \ldots$.

The elements of X are of course called variables. Clone elements are assigned to them.

Definition 9 A *graded assignment* θ of elements from a clone $\mathbf{C}$ to a graded set X is a map $\theta : X \to \mathbf{C}$ that respects rank, i.e., $\theta(X_n) \subset U_n$, $n = 1, 2, 3, \ldots$.

Clearly θ extends recursively to a map $\bar{\theta} : T_{\text{clones}}(X) \to \mathbf{C}$ in the natural way, and we write $\mathbf{C} \models p \approx q[\theta]$ precisely when $\bar{\theta}(p) = \bar{\theta}(q)$. Now for $\theta : X \to \mathbf{C}(\mathcal{V})$ a graded assignment and $p \in T_{\text{clones}}(X)$ we assign an (ordinary) $\mathcal{V}$-term $\bar{p}^\theta$ to p as follows.

1. $\bar{p}_i^{j\theta} = v_i$.

2. if $Q \in X_n$, $\bar{Q}^\theta = F_Q(v_0, \ldots, v_{n-1})$ where F_Q is a $\mathcal{V}$*-term* such that $F_Q^\dagger = \theta(Q)$.

3. $F_n^k(\beta, \alpha_0, \ldots, \alpha_{k-1}) = \bar{\beta}^\theta(\bar{\alpha}_0^\theta, \ldots, \bar{\alpha}_{k-1}^\theta)$ where the right hand side is ordinary substitution in $T_{\mathcal{V}}(\omega)$.

The relationship between these two notions of assignment is given by the following theorem, which is proved by induction on complexity of (clone) terms [cf. Corollary 1 to the first theorem of Lecture III].

Theorem 9 *Let p, q be terms of the same rank. Then*

$$\mathcal{V} \models \bar{p}^\theta \approx \bar{q}^\theta \quad \textit{iff} \quad \mathbf{C}(\mathcal{V}) \models p \approx q[\theta].$$

□

Example Let

$$\begin{array}{rcl} Q & \in & X_2 \\ p & = & F_3^2(Q, F_3^2(Q, p_0^3, p_1^3), p_2^3) \\ q & = & F_3^2(Q, p_0^3, F_3^2(Q, p_1^3, p_2^3)). \end{array}$$

If $\theta(Q) = F$, then

$$\begin{array}{rcl} \bar{p}^\theta & = & F(F(v_0, v_1), v_2) \\ \bar{q}^\theta & = & F(v_0, F(v_1, v_2)). \end{array}$$

Thus F is an associative operation in $\mathcal{V}$ iff $\mathbf{C}(\mathcal{V}) \models p \approx q[\theta]$.

It is also clear that any $\mathcal{V}$-equation arises from a clone equation in this manner. Clone equations form a natural language to discuss hyperidentities.

Homomorphic images

Suppose that $F : \mathbf{C}(\mathcal{V}) \twoheadrightarrow \mathbf{C}(\mathcal{W})$ is an onto clone homomorphism. Note that any representation $\varphi : \mathbf{C}(\mathcal{W}) \to \mathrm{Clo}(A)$ yields a representation of $\mathbf{C}(\mathcal{V})$ by composition:

$$\mathbf{C}(\mathcal{V}) \overset{F}{\twoheadrightarrow} \mathbf{C}(\mathcal{W}) \xrightarrow{\varphi} \mathrm{Clo}\,(A).$$

The assignment $\varphi \mapsto \varphi \circ F$ induces an assignment $\varphi^\circ \mapsto (\varphi \circ F)^\circ$ from algebras in $\mathcal{W}$ to algebras in $\mathcal{V}$. It is easy to check that $\{(\varphi \circ F)^\circ \mid \varphi : \mathbf{C}(\mathcal{W}) \to \mathrm{Clo}(A)\}$ is a subvariety of $\mathcal{V}$. Thus, up to equivalence $\mathcal{W}$ is a subvariety of $\mathcal{V}$.

Clone presentations

Let $\mathcal{V}_0$ be the class of all algebras of a given similarity type. Then $\mathbf{C}(\mathcal{V}_0)$ is a free clone in the sense that any mapping $F : \mathrm{Op}(\mathcal{V}_0) \to \mathbf{C}$ where $\mathrm{Op}(\mathcal{V}_0)$ is the set of operation symbols of $\mathcal{V}_0$, $\mathbf{C}$ is a clone and F respects rank, extends to a unique clone homomorphism $\widehat{F} : \mathbf{C}(\mathcal{V}_0) \to \mathbf{C}$.

Thus in the case that $\mathcal{V}$ is a subvariety of $\mathcal{V}_0$, say $\mathcal{V}_0 = \mathrm{Mod}(\Sigma)$ and $F : \mathrm{Op}(\mathcal{V}_0) \to \mathbf{C}(\mathcal{V})$ is the identity, then $\widehat{F} : \mathbf{C}(\mathcal{V}_0) \twoheadrightarrow \mathbf{C}(\mathcal{V})$ is easily seen to be onto. The kernel of this map is $\ker\widehat{F} := \{\langle p, q\rangle \in [T_{\mathrm{clones}}(\mathrm{Op}(\mathcal{V}_0))]^2 \mid \widehat{F}(O(p)) = \widehat{F}(O(q))$ where O is the identity assignment$\}$. In this case $\Sigma = \{\langle \bar{p}^\theta, \bar{q}^\theta\rangle \mid \langle p, q\rangle \in \ker\widehat{F}\}$. We call this situation a *presentation* of $\mathbf{C}(\mathcal{V})$. Note that if $\mathcal{W}$ is a different subvariety of $\mathcal{V}_0$ with $\mathcal{V} \equiv \mathcal{W}$, then $\mathcal{W}$ yields a different presentation of $\mathbf{C}(\mathcal{V})$, one with kernel (essentially) $\mathrm{Eq}(\mathcal{W})$. It is in this sense that presentations are able to distinguish between equivalent subvarieties of $\mathcal{V}_0$. It is straightforward to extend these notions to cover the case that $\mathcal{V}$ and $\mathcal{W}$ have different similarity type. Elements of $\ker\widehat{F}$ are called *relators* of the presentation $\widehat{F}$.

Example Let $\mathcal{V}_0$ be the variety of all binary algebras with operation F. Let $\mathcal{V}$ be the variety of semigroups. Then $\mathbf{C}(\mathcal{V})$ has generator F and relator

$$F_3^2(F, F_3^2(F, p_0^3, p_1^3), p_2^3) \approx F_3^2(F, p_0^3, F_3^2(F, p_1^3, p_2^3)).$$

Clone morphisms and interpretations

Suppose that $\mathbf{C}(\mathcal{V}) \xrightarrow{F} \mathbf{C}(\mathcal{W})$ is a clone homomorphism. Let Q be any operation symbol of $\mathcal{V}$. For such a Q let α_Q denote a $\mathcal{W}$-term such that

$$F(Qv_0v_1\cdots v_n^\dagger) = \alpha_Q^\dagger. \tag{7}$$

Note that α_Q is unique in the sense that if β_q is another $\mathcal{W}$-term such that (7) holds, then $\mathcal{W} \models \alpha_Q \approx \beta_Q$. It is clear that $D : Q \mapsto \alpha_Q$ yields an interpretation of $\mathcal{V}$ in $\mathcal{W}$. Furthermore, all interpretations of $\mathcal{V}$ in $\mathcal{W}$ are of this form. Note that in this context if $\mathbf{A} \in \mathcal{W}$, then $\mathbf{A} = \varphi^\circ$ for some $\varphi : \mathbf{C}(\mathcal{W}) \to \mathrm{Clo}(A)$, and $\mathbf{A}^D = (\varphi \cdot F)^\circ \in \mathcal{V}$.

Example $\mathcal{V}$ = groups, $\mathcal{BA}$ = Boolean algebras. v_0v_1 can be interpreted as $(v_0 \wedge \neg v_1) \vee (v_1 \wedge \neg v_0)$ and v_0^{-1} can be interpreted as v_0.

Note that if $F : \mathbf{C}(\mathcal{V}) \to \mathbf{C}(\mathcal{W})$ is a clone homomorphism, then $\varphi^\circ \mapsto (\varphi \circ F)^\circ$ is a functor that commutes with the underlying functor. In fact, the converse is true as well.

Theorem 10 *There is a homomorphism $F : \mathbf{C}(\mathcal{V}) \to \mathbf{C}(\mathcal{W})$ iff there is a concrete functor from $\mathcal{W}$ to $\mathcal{V}$.* □

Definition 10 $\mathcal{V} \leq \mathcal{W}$ iff there exists a homomorphism $F : \mathbf{C}(\mathcal{V}) \to \mathbf{C}(\mathcal{W})$.

Example 1 Groups $\leq$ Boolean algebras.

Example 2 For any variety $\mathcal{V}$, $\mathcal{V} \leq \mathcal{BA}$ iff $2 \in \mathrm{Spec}(\mathcal{V})$.

Proof of Theorem 10

($\Rightarrow$) Any concrete functor from $\mathcal{BA}$ into $\mathcal{V}$ clearly sends the 2 element Boolean algebra to a 2 element algebra in $\mathcal{V}$.

($\Leftarrow$) Suppose that $\mathbf{A} \in \mathcal{V}$ with $|A| = 2$. By the primality of $\mathbf{2} \in \mathcal{BA}$ we can represent each $\mathcal{V}$-term Q on A by a Boolean term. But $\mathbf{2}$ is generic in $\mathcal{BA}$ so these same Boolean terms convert any Boolean algebra into a $\mathcal{V}$-algebra, yielding a concrete functor from $\mathcal{BA}$ into $\mathcal{V}$. □

Lecture V

Applications to Mal'tsev conditions

Recall from last lecture that

$$(*) \qquad \mathcal{V} \models \bar{p}^\theta \approx \bar{q}^\theta \quad \text{iff} \quad \mathbf{C}(\mathcal{V}) \models p \approx q[\theta],$$

where p, q are clone terms and θ is a graded assignment of $\mathcal{V}$-operations to variables. In 1954 Mal'tsev proved his famous theorem that a variety $\mathcal{V}$ is congruence-permutable iff there is a $\mathcal{V}$-term $p(x, y, z)$ such that

$$\begin{aligned} \mathcal{V} &\models p(v_0, v_0, v_2) \approx v_2 \\ \mathcal{V} &\models p(v_0, v_2, v_2) \approx v_0. \end{aligned}$$

By $(*)$, this is true iff

$$\mathbf{C}(\mathcal{V}) \models \exists p(F_3^3(p, p_0^3, p_2^3, p_2^3) = p_0^3 \wedge F_3^3(p, p_0^3, p_0^3, p_2^3) = p_2^3).$$

Mal'tsev's result has inspired much research in universal algebra. Many other classes of algebras have been shown to be *Mal'tsev-definable*, i.e., characterized by the existence of a finite set of terms satisfying certain equations. Fundamental examples are Pixley's characterization of arithmetical varieties [10] and Jónsson's characterization of congruence distributivity [5].

Recall that Jónsson proved that a variety $\mathcal{V}$ is congruence-distributive iff for some $n \geq 0$ there exist $\mathcal{V}$-terms $d_0(x, y, z), \ldots, d_n(x, y, z)$ such that the following identities hold in $\mathcal{V}$:

1. $d_0(x,y,z) \approx x$, $d_n(x,y,z) \approx z$,
2. $d_i(x,y,x) \approx x$, $i \leq n$,
3. $d_i(x,x,y) \approx d_{i+1}(x,x,y)$, $i \in n$, i even,
4. $d_i(x,y,y) \approx d_{i+1}(x,y,y)$, $i < n$, i odd.

Note that for n fixed, these conditions are equivalent (by $(*)$) for a variety $\mathcal{V}$ to

$$(**) \qquad \mathbf{C}(\mathcal{V}) \models \exists\, d_0 \exists\, d_1 \cdots \exists\, d_n \Phi(d_0, \ldots, d_n),$$

where Φ is a conjunction of clone equations. But it is well known that there are congruence-distributive varieties requiring n to be arbitrarily large. Thus congruence distributivity for $\mathcal{V}$ is equivalent to $\mathbf{C}(\mathcal{V})$ satisfying an infinite disjunction, $\mathbf{C}(\mathcal{V}) \models \psi_1 \vee \psi_2 \vee \psi_3 \vee \cdots$ where each ψ_n is as in $(**)$.

In order to understand Mal'tsev conditions better, it is helpful to adopt the order-theoretic viewpoint. Recall from the last lecture that $\mathcal{V} \leq \mathcal{W}$ means that there exists a clone homomorphism $F : \mathbf{C}(\mathcal{V}) \to \mathbf{C}(\mathcal{W})$. $\leq$ turns out to be a pre-order which becomes a lattice ordering after identifying $\mathcal{V}$ and $\mathcal{W}$ whenever $\mathcal{V} \leq \mathcal{W} \leq \mathcal{V}$ [2]. This ordering provides a very natural setting to investigate Mal'tsev conditions, which turn out to be certain types of filters. For example, the following theorem is almost immediate.

Theorem 11 *$\mathcal{V}$ is congruence-permutable iff $\mathcal{M} \leq \mathcal{V}$, where $\mathcal{M}$ is the Mal'tsev variety with one ternary operation p satisfying the axioms $\{p(v_0, v_2, v_2) \approx v_0,\ p(v_0, v_0, v_2) \approx v_2\}$.* □

Note that $\mathbf{C}(\mathcal{M})$ has the finite presentation

$$\{p \mid F_3^3(p, p_0^3, p_2^3, p_2^3) = p_0^3,\ F_3^3(p, p_0^3, p_0^3, p_2^3) = p_2^3\}.$$

Thus the class of congruence-permutable varieties forms a principal filter generated by a finitely-presented variety.

For congruence distributivity, let $\mathcal{D}_n$ be the variety with $n+1$ basic ternary operation symbols $d_0(x,y,z), \ldots, d_n(x,y,z)$, defined by Jónsson's equations for congruence distributivity. It is easy to see that $\mathcal{D}_1 \geq \mathcal{D}_2 \geq \mathcal{D}_3 \geq \cdots$ and that each $\mathcal{D}_n$ corresponds to a finitely presented clone $C(\mathcal{D}_n)$. Furthermore $\mathcal{V}$ is congruence-distributive iff $\mathcal{V} \geq \mathcal{D}_n$ for some n. Thus the class of congruence-distributive varieties forms a filter generated by a countable descending chain of finitely-presented varieties.

The next few theorems are due to Taylor [14]. The first reformulates the definition of Mal'tsev conditions in terms of $\leq$.

Theorem 12 (Taylor) *A class $\mathcal{K}$ of varieties is Mal'tsev-definable if and only if there exist finitely presented clones $C(\mathcal{W}_1), C(\mathcal{W}_2), \ldots$ with $\mathcal{W}_1 \geq \mathcal{W}_2 \geq \cdots$ such that $\mathcal{V} \in \mathcal{K}$ iff $\mathcal{V} \geq \mathcal{W}_i$ for some i.* □

Note that $\mathcal{K}$ satisfies

1. $\mathcal{K}$ is a filter with respect to the *lattice* ordering $\leq$.
2. for all $\mathcal{V} \in \mathcal{K}$ there is a finitely-presented $\mathcal{V}_0 \in \mathcal{K}$ with $\mathcal{V} \geq \mathcal{V}_0$.

The following shows that, surprisingly 1. and 2. are also sufficient for a class $\mathcal{K}$ to be Mal'tsev-definable.

Theorem 13 (Taylor) *A class $\mathcal{K}$ of varieties is Mal'tsev-definable if and only if $\mathcal{K}$ satisfies* 1. *and* 2.

Proof Let $S_1, S_2, S_3, \ldots$ be a listing of all finite presentations of varieties in $\mathcal{K}$. Let $T_1 = S_1$, $T_2 = S_1 \cap S_2, \ldots, T_n = S_1 \cap S_2 \cap \cdots \cap S_n$. Let $\mathcal{W}_i$ be the variety presented by T_i. By 1., $\mathcal{W}_i \in \mathcal{K}$. Furthermore if $\mathcal{V} \in \mathcal{K}$, then by assumption there exists a finitely-presented $\mathcal{W} \in \mathcal{K}$ such that $\mathcal{W} \leq \mathcal{K}$. Thus $\mathcal{W}$ is the variety presented by S_i for some i, whence $\mathcal{W}_i \leq \mathcal{W} \leq \mathcal{V}$. Thus $\mathcal{K}$ is Mal'tsev-definable (since clearly $\mathcal{W}_1 \geq \mathcal{W}_2 \geq \cdots$). $\square$

This theorem can be rephrased so as to not explicitly mention clones. Note that $\mathcal{V} \leq \mathcal{W}$ can be factored as $\mathbf{C}(\mathcal{V}) \twoheadrightarrow \mathbf{C}(\mathcal{U}) \hookrightarrow \mathbf{C}(\mathcal{W})$. In lecture IV we saw that $\mathbf{C}(\mathcal{V}) \twoheadrightarrow \mathbf{C}(\mathcal{U})$ implies that $\mathcal{U}$ is, up to equivalence, a subvariety of $\mathcal{V}$. It is easy to see that an embedding $\mathbf{C}(\mathcal{U}) \hookrightarrow \mathbf{C}(\mathcal{W})$ implies that, up to equivalence, $\mathcal{U}$ is a reduct of $\mathcal{W}$. This factorization leads to the following characterization of Mal'tsev definability.

Theorem 14 *A class $\mathcal{K}$ of varieties is Mal'tsev-definable if and only if*

1. *$\mathcal{K}$ is closed under the equivalence of varieties.*
2. *If $\mathcal{U}$ is a subvariety of $\mathcal{V} \in \mathcal{K}$, then $\mathcal{U} \in \mathcal{K}$.*
3. *If $\mathcal{U}$ is a reduct of $\mathcal{W}$ and $\mathcal{U} \in \mathcal{K}$, then $\mathcal{W} \in \mathcal{K}$.*
4. *If $\mathcal{U}, \mathcal{V} \in \mathcal{K}$, then $\mathcal{U} \wedge \mathcal{V} \in \mathcal{K}$.*
5. *If $\mathcal{V} \in \mathcal{K}$, then $\mathcal{V}_0 \in \mathcal{K}$ for some subvariety $\mathcal{V}_0 \subset \mathcal{V}$ with $\mathcal{V}_0 = \mathrm{Mod}(\Sigma_0)$ for some finite* $\Sigma_0 \subset \Sigma = \mathrm{Eq}(\mathcal{V})$. $\square$

In order to make the previous theorems more precise, we must give a brief description of $\mathcal{V} \wedge \mathcal{W}$. So let $\mathcal{V}, \mathcal{W}$ be varieties, where without loss of generality the similarity type of $\mathcal{V}$ is disjoint from the similarity type of $\mathcal{W}$. The similarity type of $\mathcal{V} \wedge \mathcal{W}$ is $\mathrm{type}(\mathcal{V}) \,\dot\cup\, \mathrm{type}(\mathcal{W}) \,\dot\cup\, \{p\}$, where p is a new binary operation symbol. For $\mathbf{A} \in \mathcal{V}$ define $\mathbf{A}'$ of $\mathrm{type}(\mathcal{V} \wedge \mathcal{W})$ to be the algebra with the same universe with operations defined by

$$F^{\mathbf{A}'} = \begin{cases} F^{\mathbf{A}} & \text{if} \quad f \in \mathrm{type}(\mathcal{V}) \\ p_0^2 & \text{if} \quad F = p \\ p_0^n & \text{if} \quad F \in \mathrm{type}(\mathcal{W}) \quad \text{is } n\text{-ary.} \end{cases}$$

Similarly for $\mathbf{B} \in \mathcal{W}$ let $\mathbf{B}'$ be the algebra of $\mathrm{type}(\mathcal{V} \wedge \mathcal{W})$ with the same universe and operations defined by

$$F^{\mathbf{B}'} = \begin{cases} p_0^n & \text{if} \quad F \in \mathrm{type}(\mathcal{V}),\ F\ n\text{-ary} \\ p_1^2 & \text{if} \quad F = p \\ F^{\mathbf{B}} & \text{if} \quad F \in \mathrm{type}(\mathcal{W}). \end{cases}$$

Then $\mathcal{V} \wedge \mathcal{W}$ is defined to be the variety $\mathbb{HSP}\{\mathbf{A}' \times \mathbf{B}' \mid \mathbf{A} \in \mathcal{V},\ \mathbf{B} \in \mathcal{W}\}$. $\mathcal{V} \wedge \mathcal{W}$ is also called $\mathcal{V} \times \mathcal{W}$.

Lemma 2 $\mathbb{HSP}\{\mathbf{A}' \times \mathbf{B}' \mid \mathbf{A} \in \mathcal{V}, \mathbf{B} \in \mathcal{W}\} = \mathbb{I}\{A' \times B' \mid A \in \mathcal{V}, B \in \mathcal{W}\}$.

Proof Let $\mathcal{K} = \{\mathbf{A}' \times \mathbf{B}' \mid \mathbf{A} \in \mathcal{V}, \mathbf{B} \in \mathcal{W}\}$. Suppose that $\prod_{i \in I}(\mathbf{A}'_i \times \mathbf{B}'_i) \in \mathbb{P}(\mathcal{K})$, then

$$\prod_{i \in I}(\mathbf{A}'_i \times \mathbf{B}'_i) \cong \prod_{i \in I}\mathbf{A}'_i \times \prod_{i \in I}\mathbf{B}'_i \cong \left(\prod_{i \in I}\mathbf{A}_i\right)' \times \left(\prod_{i \in I}\mathbf{B}_i\right)' \in \mathcal{K}.$$

Thus $\mathbb{P}(\mathcal{K}) \subset \mathbb{I}(\mathcal{K})$.

Suppose that $\mathbf{C} \leq \mathbf{A}' \times \mathbf{B}' \in \mathcal{K}$. Let $\mathbf{C}_1 = \pi_1(\mathbf{C})$, $\mathbf{C}_2 = \pi_2(\mathbf{C})$. Then $\mathbf{C} \leq \mathbf{C}_1 \times \mathbf{C}_2 \in \mathcal{K}$. Suppose that $\langle c_1, c_2 \rangle \in \mathbf{C}_1 \times \mathbf{C}_2$. Then there exist $a \in A$, $b \in B$ such that $\langle c_1, b \rangle, \langle a, c_2 \rangle \in C$. Thus $\langle c_1, c_2 \rangle = \langle p^{\mathbf{A}'}(c_1, b), p^{\mathbf{B}'}(a, c_2) \rangle = p^{\mathbf{A}' \times \mathbf{B}'}(\langle c_1, b \rangle, \langle a, b_2 \rangle) \in C$ therefore $\mathbb{S}(\mathcal{K}) \subseteq \mathbb{I}(\mathcal{K})$.

Finally suppose that $h : \mathbf{A}' \times \mathbf{B}' \twoheadrightarrow \mathbf{C} \in \mathbb{H}(\mathcal{K})$. Fix $\langle a, b \rangle \in A \times B$. Define $h_1, h_2 : A \times B \to C$ by

$$\begin{aligned} h_1(\langle x, y \rangle) &= h(p(\langle x, y \rangle, \langle a, b \rangle)) = h(\langle x, b \rangle) \\ h_2(\langle x, y \rangle) &= h(p(\langle a, b \rangle, \langle x, y \rangle)) = h(\langle a, y \rangle). \end{aligned}$$

Let $C_i = h_i(A \times B)$, $i = 1, 2$.

Define $(\mathcal{V} \wedge \mathcal{W})$-operations on C_1, C_2 by

$$F^{\mathbf{C}_1}(h_1(\langle x_0, y_0 \rangle), \ldots, h_1(\langle x_{n-1}, y_{n-1} \rangle)) = \begin{cases} h(F^{\mathbf{C}}(h_1\langle x_0, y_0 \rangle, \ldots, h_1\langle x_{n-1}, y_{n-1} \rangle) \\ \text{if } F \in \mathrm{Type}(\mathcal{V}) \\ h_1(\langle x_0, y_0 \rangle) \quad \text{otherwise,} \end{cases}$$

$$F^{\mathbf{C}_2}(h_2(\langle x_0, y_0 \rangle), \ldots, h_2(\langle x_{n-1}, y_{n-1} \rangle)) = \begin{cases} h(F^{\mathbf{C}}(h_1\langle x_0, y_0 \rangle, \ldots, h_1\langle_{n-1}, y_{n-1} \rangle)) \\ \text{if } F \in \mathrm{Type}(\mathcal{W}) \\ h_2(\langle x_0, y_0 \rangle) \quad \text{otherwise.} \end{cases}$$

Suppose that $F \in \mathrm{Type}(\mathcal{V})$ and $\langle x_0, y_0 \rangle, \ldots, \langle x_{n-1}, y_{n-1} \rangle \in A \times B$. Then

$$\begin{aligned} & h_1(F^{\mathbf{A} \times \mathbf{B}}(\langle x_0, y_0 \rangle, \ldots, \langle x_{n-1}, y_{n-1} \rangle)) \\ = \; & h(p^{\mathbf{A} \times \mathbf{B}}(F^{\mathbf{A} \times \mathbf{B}}(\langle x_0, y_0 \rangle, \ldots, \langle x_{n-1}, y_{n-1} \rangle), \langle a, b \rangle)) \\ = \; & h(\langle p^{\mathbf{A}}(F^{\mathbf{A}}(x_0, \ldots, x_{n-1}), a), p^{\mathbf{B}}(F^{\mathbf{B}}(y_0, \ldots, y_{n-1}), b) \rangle) \\ = \; & h(\langle F^{\mathbf{A}}(x_0, \ldots, x_{n-1}), b \rangle) \\ = \; & h(\langle F^{\mathbf{A}}(x_0, \ldots, x_{n-1}), F^{\mathbf{B}}(b, \ldots, b) \rangle) \\ = \; & h(F^{\mathbf{A} \times \mathbf{B}}(\langle x_0, b \rangle, \ldots, \langle x_{n-1} \rangle)) \\ = \; & F^{\mathbf{C}}(h(\langle x_0, b \rangle, \ldots, \langle x_{n-1}, b \rangle)) \\ = \; & F^{\mathbf{C}}(h_1(x_0, y_0), \ldots, h_1(x_{n-1}, y_{n-1})) \\ = \; & F^{\mathbf{C}_1}(h_1(x_0, y_0), \ldots, h_1(x_{n-1}, y_{n-1})). \end{aligned}$$

If $F \in \text{Type}(\mathcal{V} \times \mathcal{W}) - \text{Type}(\mathcal{V})$, then

$$\begin{aligned} h_1(F^{\mathbf{A}\times\mathbf{B}}(\langle x_0, y_0\rangle, \ldots, \langle x_{n-1}, y_{n-1}\rangle)) &= h(F^{\mathbf{A}}(x_0, \ldots, x_{n-1}), b) \quad \text{(as before)} \\ &= h(x_0, b) = h(p(\langle x_0, y_0\rangle, \langle a, b\rangle)) \\ &= h_1(\langle x_0, y_0\rangle) \\ &= F^{\mathbf{C}_1}(\langle x_0, y_0\rangle, \ldots, \langle x_{n-1}, y_{n-1}\rangle). \end{aligned}$$

Therefore h_1 is a homomorphism, whence $\mathbf{C}_1$ is of the form $\mathbf{D}'$ for some $\mathbf{D} \in \mathcal{V}$. Similalry h_2 is a homomorphism and $\mathbf{C}_2$ is of the form $\mathbf{D}'$ for some $\mathbf{D} \in \mathcal{W}$.

Therefore $\mathbf{C}_1 \times \mathbf{C}_2 \in \mathbb{I}(\mathcal{K})$. To complete the proof we must show that $\mathbf{C} \cong \mathbf{C}_1 \times \mathbf{C}_2$. But this is easily checked by the mapping $h(\langle x, y\rangle) \mapsto \langle h, (\langle x, y\rangle), h_2(\langle x, y\rangle)\rangle \in \mathbf{C}_1 \times \mathbf{C}_2$. $\square$

It is also possible to give a syntactical description of $\mathcal{V} \wedge \mathcal{W}$. Let

$$\begin{aligned} \text{Eq}(\mathcal{V}) &= \{\sigma_i \approx \tau_i \mid i \in I\} \\ \text{Eq}(\mathcal{W}) &= \{\mu_j \approx \nu_j \mid j \in J\}. \end{aligned}$$

Then $\mathcal{V} \wedge \mathcal{W}$ is defined by the following equations:

$$\begin{aligned} p(x, x) &\approx x \\ p(p(x, y), p(u, v)) &\approx p(x, v) \\ F(p(x_1, y_1), \ldots, p(x_n, y_n)) &\approx p(F(x_1, \ldots, x_n), F(y_1, \ldots, y_n)) \\ p(F(x_1, x_2, \ldots, x_n), y) &\approx p(x_1, y) \quad F \in \text{Type}(\mathcal{W}) \\ p(x, F(y_1, \ldots, y_n)) &\approx p(x, y_1) \quad F \in \text{Type}(\mathcal{V}) \\ p(\sigma_i, y) &\approx p(\tau_i, y) \quad (i \in I) \\ p(x, \mu_j) &\approx p(x, \nu_j) \quad (j \in J). \end{aligned}$$

It is also possible to define the product of infinitely many varieties via

$$\bigwedge_{i\in I} \mathcal{V}_i = \text{Var}(\prod_{i\in I} \mathbf{C}(\mathcal{V}_i)).$$

We leave it as an exercise for the reader to use the previous characterization of Mal'tsev definability to show that $\mathcal{K} = \{\mathcal{V} \mid 2 \notin \text{Spec}\mathcal{V}\}$ is Mal'tsev-definable. So there exists an infinite descending chain of varieties such that every variety that does not have a 2-element algebra is above one of these varieties. In fact there exists a descending chain $\mathcal{W}_1 \geq \mathcal{W}_2 \geq \cdots$ such that each $\mathcal{W}_i \not\leq \mathcal{BA}$ but $\bigwedge \mathcal{W}_i \leq \mathcal{DL} \leq \mathcal{BA}$ where $\mathcal{DL}$ is the variety of distributive lattices. Thus $\mathcal{BA}$ not completely meet prime. (See W. Taylor [15].)

An old question was whether or not there exists a cover in this lattice of varieties. This was recently settled in the affirmative by McKenzie (1990), who showed that $\mathcal{BA}$ has an upper cover. One reason that this problem remained open for so long is that there is no natural factorization of algebras in $\bigwedge_{i\in I} \mathcal{V}_i$ for I infinite, so it is difficult to directly use $\bigwedge\{\mathcal{V}_i \mid \mathcal{V} < \mathcal{V}_i\}$ to show that a given $\mathcal{V}$ has an upper cover.

McKenzie defined a variety $\mathcal{B}^+$ such that $\mathcal{B}^+ > \mathcal{BA}$ and for any variety $\mathcal{V}$, $\mathcal{V} > \mathcal{BA}$ iff $\mathcal{V} \geq \mathcal{B}^+$.

$\mathcal{B}^+$ has similarity type $\langle \wedge, \vee, \neg, 0, 1, f_0, f_1 \rangle$ with axioms the $\mathcal{BA}$ axioms for $\langle \wedge, \vee, \neg, 0, 1 \rangle$, together with

$$f_0(f_0(1)) = 0, \qquad f_0(1) \vee f_1(1) = 1,$$
$$f_1(f_1(1)) = 0.$$

It turns out that these axioms exclude the 2-element Boolean algebra but not much else. See McKenzie [7]. For related questions see McKenzie and Świerczkowski [8]. Jennifer Hyndman [3] recently found many analogs of $\mathcal{B}^+$. In fact, she found a proper class of covering relations between varieties.

References

[1] Evans, T., Some remarks on the theory of clones, in: *Finite Algebra and Multiple-Valued Logic* (*Szeged 1979*) (B. Csákány and I.G. Rosenberg, eds.), Colloq. Math. Soc. János Bolyai **28**, North-Holland, Amsterdam 1981, 203–244.

[2] García, O.C., and Taylor, W., The lattice of interpretability types of varieties, *Mem. Amer. Math. Soc.* **50** (1984), no.305, 125pp.

[3] Hyndman, J., Cover pairs in the lattice of interpretability types of varieties, Ph.D. Thesis, University of Colorado, Boulder, 1991.

[4] Isbell, J.R., Compatibility and extensions of algebraic theories, *Algebra Universalis* **6** (1976), 37–51.

[5] Jónsson, B., Algebras whose congruence lattices are distributive, *Math. Scand.* **21** (1967), 110–121.

[6] Lawvere, F.W., Functorial semantics of algebraic theories, *Proc. Nat. Acad. Sci.* (*USA*), **50** (1963), 869–872.

[7] McKenzie, R., On the covering relation in the interpretability lattice of equational theories, to appear in *Algebra Universalis.*

[8] McKenzie, R., and Świerczkowski, S., Non-covering in the interpretability lattice of equational theories, to appear in *Algebra Universalis.*

[9] Neumann, W.D., Representing varieties of algebras by algebras, *J. Austral. Math. Soc.* **11** (1970), 1–8.

[10] Pixley, A.F., Distributivity and permutability of congruence relations in equational classes of algebras, *Proc. Amer. Math. Soc.* **14** (1963), 105–109.

[11] Sangalli, A.A.L., A representation of varieties and their morphisms, *Algebra Universalis* **17** (1983), 120–128.

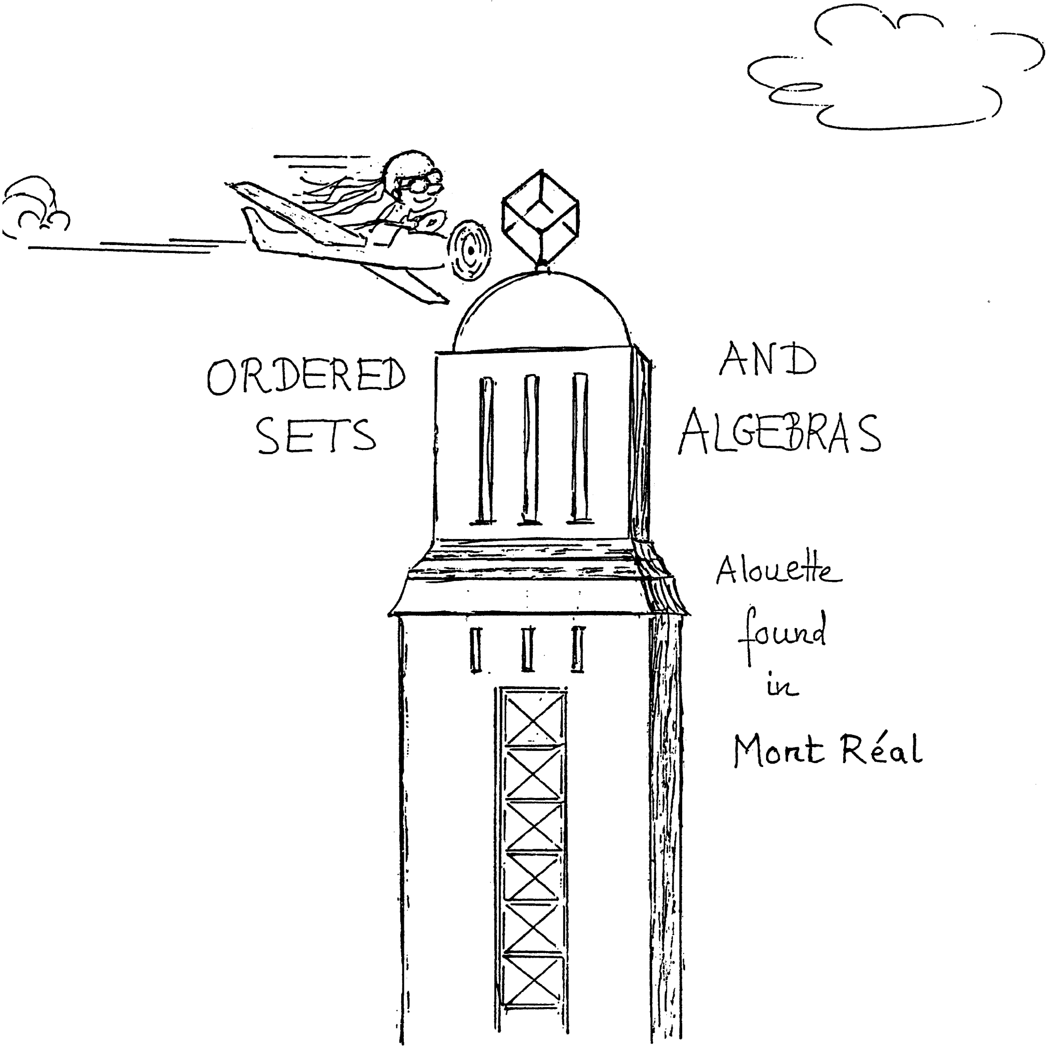
[12] Sangalli, A.A.L., On the structure and representation of clones, *Algebra Universalis* **25** (1988), 101–106.

[13] Schein, B.M., and Trohimenko, V.S., Algebras of multiplace functions, *Semigroup Forum* **17** (1979), 1–64.

[14] Taylor, W., Characterizing Mal'tsev conditions, *Algebra Universalis* **3** (1973), 351–397.

[15] Taylor, W., Mal'tsev conditions and spectra, *J. Austral. Math. Soc. (A)* **29** (1980), 143–152.

[16] Whitlock, H.J., A composition algebra for multiplace functions, *Math. Ann.* **157** (1964), 167–178.



Now Alouette Knows It All

Drawings by Marcel Erné

I. G. Rosenberg and G. Sabidussi (eds.), Algebras and Orders, 531–545.

Chorus

Ordered sets and algebras
Alouette found in Montréal.
Ordered sets and algebras
Now Alouette knows it all.
(Davey)

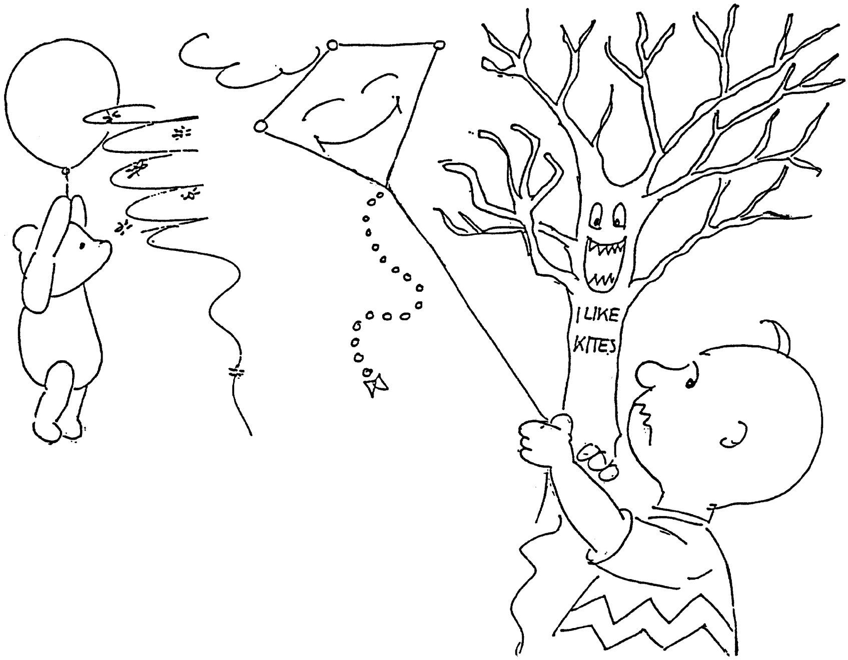
1 From far we came into Québec.
(Now Alouette knows it all.)
Instead of beer they served us sekt.
(Now Alouette knows it all.)
Not champagne for one evening's fling
But one week's beer would be the thing.
(Davey)

2 Bjarni gave us notes for free.
So he could ramble as he pleased.
But I forgot my notes today,
I'll skip this lecture if I may.

(Davey)

3 Be sure that you're not running late,
When Bjarni starts to operate.
The list of errors was quite long.
We can't include them in this song.

(Hartung et al.)



4 Marcel proposed, and that was right,
A tree should not contain a kite,
Told us about a sober Pooh,
And theorems proved by ME and YOU.

(Agliano)

5 In Winnipeg he "dooalized".
But Brian should have realized,
Despite his willingness to croon,
It's time he sang a different "toon".

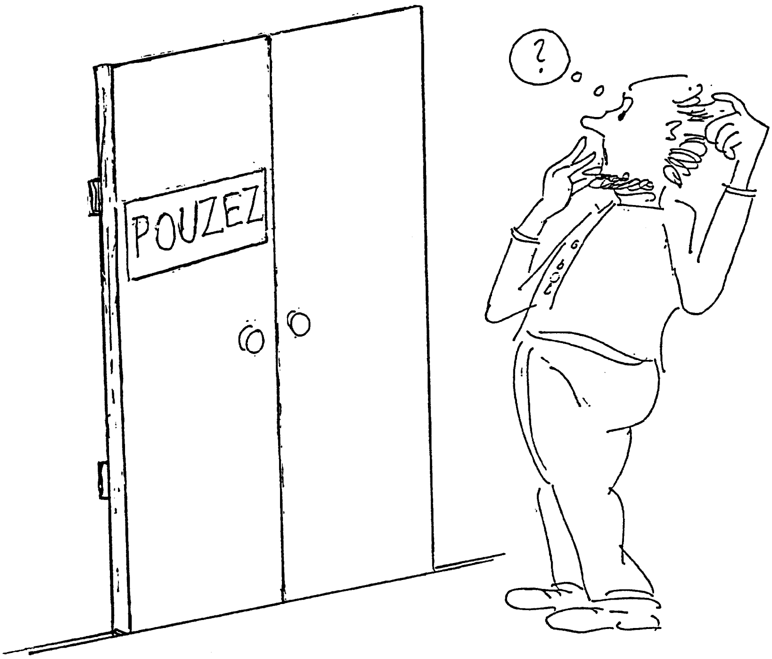
(Quackenbush)

6 He doesn't tell us how it goes,
And simply says follow your nose.
I finally thought, "This proof is sane!",
Then schizophrenia struck again.

(Tischendorf et al.)

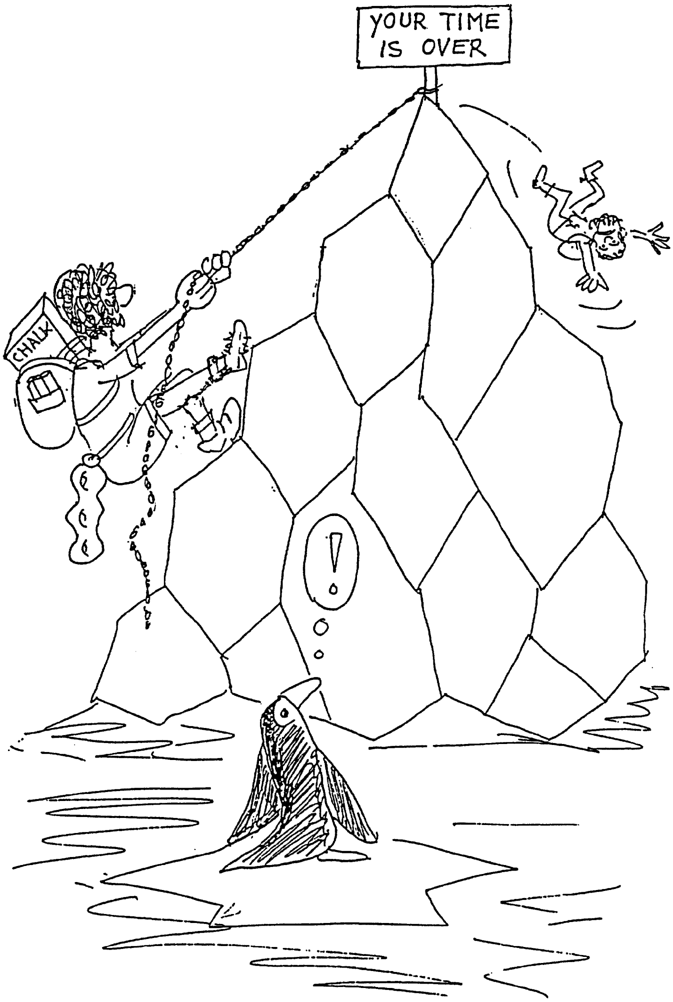
7 Maurice began with metric tricks,
To prove to us his "idée fixe ".
This verse is very hard to rhyme,
I haven't been there since first time.

(Anon.)



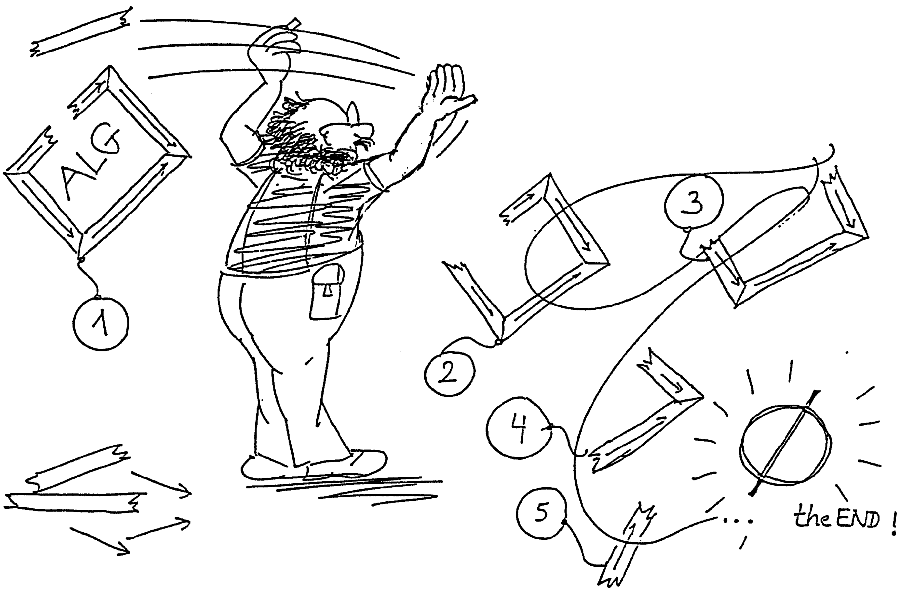
8 But even so I learned so much,
He must possess a magic touch.
You really could not ask for more.
His name appears on every door.

(Agliano, Gould)



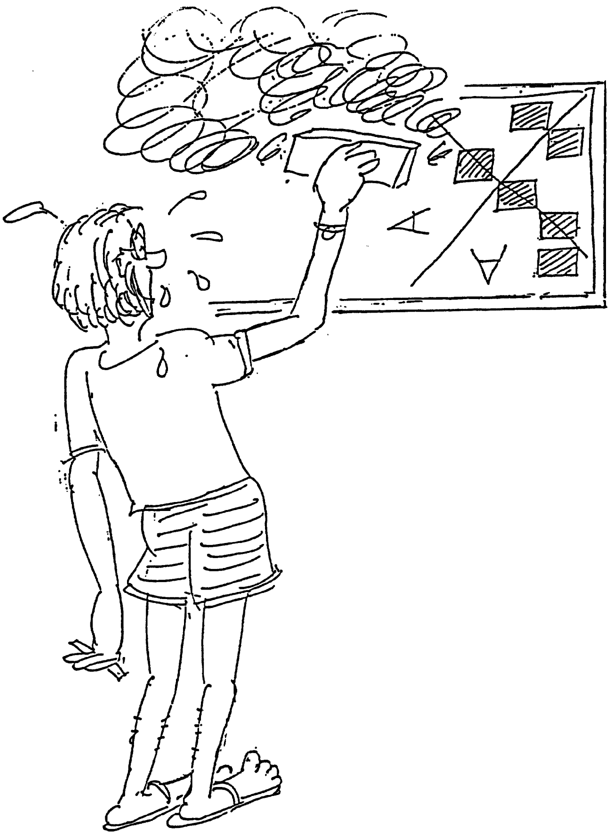
9 For Ivan chalk is but a tool.
His iceberg pictures are so cool.
His diagrams so steep to climb,
He's dragged us into overtime.

(Davey, Quackenbush,
Tischendorf)



10 I found a partial algebra.
With such a thing I won't go far.
Then Peter said unto me, "Friend!
They get much better in the end".

(Davey)



11 Some Pixley theorems were so tough,
That Ralph could not erase them off.
But Alden showed how nice they are,
Once Pixels are rectangular.

(Erné)

12 Out on an island in the sea,
We've all been taught upon the knee.
It floats in on the morning breeze,
That lattices will never Freese.

(Pálfy)

13 Ralph's thongs he's worn some 20 years.
When they are gone he'll be in tears.
A lattice you can get for free.
Could new shoes so expensive be?

(Agliano, Davey)

14 Those abstract clones are Taylor-made,
But intuition starts to fade.
How come he's such a Boulder hunk?
He learned it from a Buddhist monk.
(Davey, Gould, Quackenbush)

15 The atmosphere becomes too tense,
Unless you catch the hyper-sense.
If one more speaker throws a clone,
"Oh! Not again!" you'll hear us moan.
(Burmeister, Davey,
Gould, Volkov)

16 When H, S, P is not enough,
We need a D to make things tough.
And what it means, he'll have to say,
But let's not tell the SPD [auf Deutsch].
(Taylor)

17 When Dietmar tried to show a slide,
Of H, S, P all organized,
They turned the lights off in the room.
He finished off his talk in gloom.

(Coleman)

18 Zis Schwyperterm ist all you need.
Into zis field I vill you lead.
Und if red devils zere you find,
You zimply need ein hyper-mind.

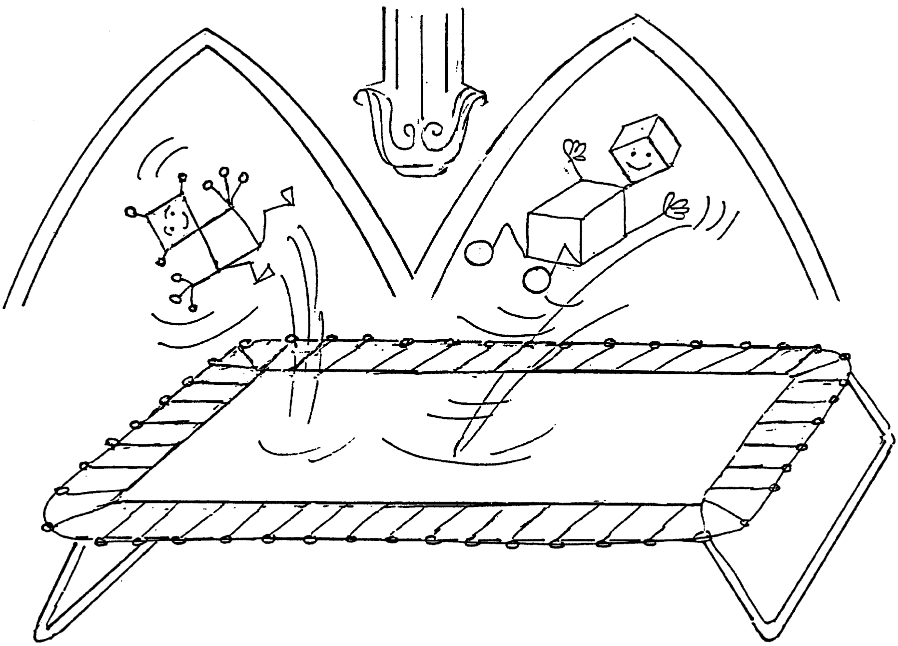
(Davey, Gould, Hartung,
Zickwolff)

19 When Ivo uses only chalk,
He gives a most impressive talk.
But when he reaches for that switch,
The audience begins to twitch.

(Gould)

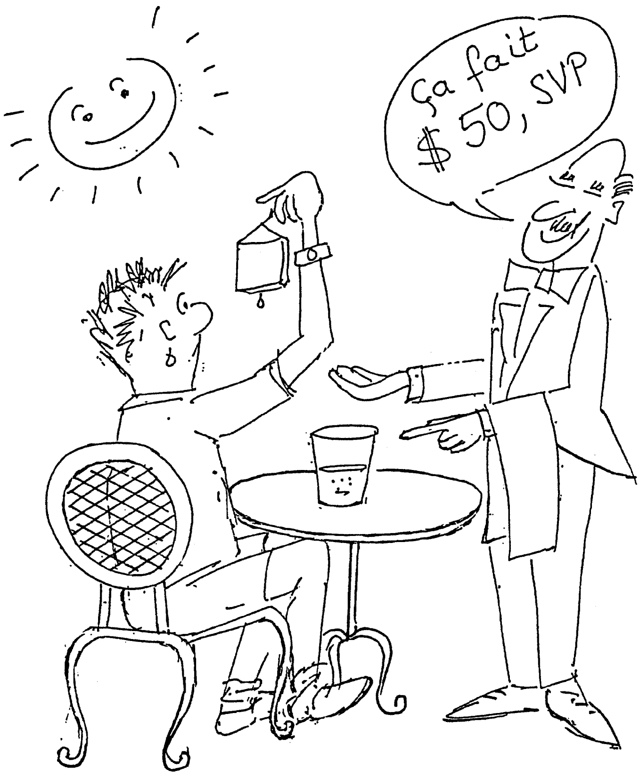
20 We meet for lectures down below,
To join in song far up we go.
We have a conference for two weeks,
And fool around with lattice freaks.

(Fearnley, Gould)



21 The lecture hall you must confess,
Its colour scheme is such a mess.
And let's tear down that stupid screen.
Best use it for a trampoline.

(Davey, Gould, Hartung)

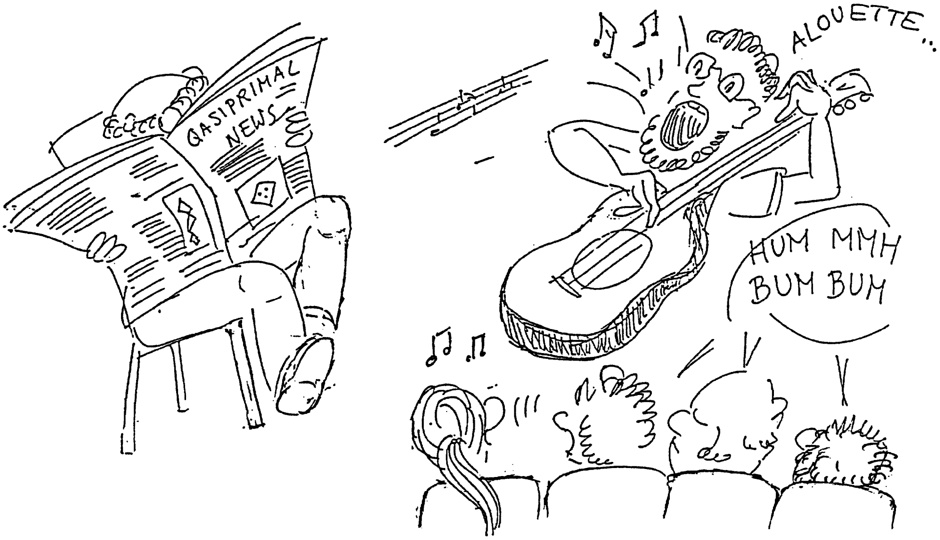


22 For 50 bucks what do we get?
I hardly get my teabag wet.
There's lemonade for you to drink,
But it's all gone within a wink.

(Davey, Gould, Hartung)

23 In this nice town they speak Français.
It's very hard to find your way.
And if we want to get some booze,
We just don't know which words to use.

(Tischendorf)



24 Bob Quackenbush just reads the news,
While Brian sings the lattice blues.
The rest of us just hum along,
And hope to end this silly song.

(Gould)

25 In every verse the truth we bend.
It makes no difference, foe or friend.
So if you find you're in a verse,
Accept your fate, it could be worse.

(Davey)

Index