Grounded Consequence for Defeasible Logic

"Antonelli applies some of the techniques developed in Kripke's approach to the paradoxes to generalize some of the most popular formalisms for non-monotonic reasoning, particularly default logic. The result is a complex and sophisticated theory that is technically solid and attractive from an intuitive standpoint." – John Horty, *Committee on Philosophy and the Sciences, University of Maryland, College Park*

This is a monograph on the foundations of defeasible logic, which explores the formal properties of everyday reasoning patterns whereby people jump to conclusions, reserving the right to retract them in the light of further information. Although technical in nature, the book contains sections that outline basic issues by means of intuitive and simple examples.

G. Aldo Antonelli is Professor of Logic and Philosophy of Science at the University of California, Irvine.

Grounded Consequence for Defeasible Logic

G. ALDO ANTONELLI
University of California, Irvine

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
40 West 20th Street, New York, NY 10011-4211, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521842051

First published 2005

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data

Antonelli, G. Aldo.
Grounded consequence for defeasible logic / G. Aldo Antonelli.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-84205-0 (hardback)
1. Defeasible reasoning. I. Title.
BC199.D38A58 2005
160 – dc22 2005003176

ISBN-13 978-0-521-84205-1 hardback
ISBN-10 0-521-84205-0 hardback

Contents

List of Figures *page* vii
Foreword ix

1 The Logic of Defeasible Inference 1
1.1 First-order logic 1
1.2 Consequence relations 4
1.3 Nonmonotonic logics 9
1.4 Skeptical versus credulous reasoning 18
1.5 Floating conclusions 20
1.6 Conflicts and modularity 24
1.7 Assessment 27

2 Defeasible Inheritance over Cyclic Networks 29
2.1 Background and motivation 29
2.2 Graph-theoretical preliminaries 35
2.3 Constructing extensions 38
2.4 Non-well-founded networks 43
2.5 Extensions and comparisons 49
2.5.1 Decoupling 49
2.5.2 Zombie paths 50
2.5.3 Infinite networks 51
2.6 Proofs of selected theorems 54

3 General Extensions for Default Logic 59
3.1 Introductory remarks 59
3.2 Categorical default theories 62
3.3 Examples 65
3.4 Grounded extensions 69

3.5 Examples, continued 72
3.6 Proofs of selected theorems 75
4 Defeasible Consequence Relations 86
4.1 Defeasible consequence 86
4.2 Alternative developments 90
4.2.1 Seminormal theories 91
4.2.2 Optimal extensions 93
4.2.3 Circumspect extensions 95
4.3 Conclusions and comparisons 96
4.3.1 Existence of extensions 97
4.3.2 Defeasible consequence – again 99
4.3.3 Floating conclusions, conflicts, and modularity 101
4.4 Infinitely many defaults 102
4.5 Proofs of selected theorems 103
Bibliography 113
Index 117

List of Figures

1.1 An inheritance network *page* 13
1.2 The Nixon diamond 20
1.3 Floating conclusions in the Nixon diamond 21
1.4 Horty's "moral" dilemma 23
1.5 Horty's counterexample to floating conclusions 24
2.1 The standard example of preemption 30
2.2 A network with cycles 32
2.3 A network with no creduluous extension 32
2.4 A cycle is spliced into the path *abcde* 37
2.5 Preemption 39
2.6 A net with paths `abcd` and `adeb` each preempting the other 46
2.7 A net illustrating the "decoupling" problem 49
2.8 Zombie paths 51
4.1 Comparison of general extensions and constrained default logic 88
4.2 Nonminimal extensions 94

Foreword

Logic is an ancient discipline that, ever since its inception some 2,500 years ago, has been concerned with the analysis of patterns of valid reasoning. The beginnings of such a study can be traced back to Aristotle, who first developed the theory of the *syllogism* (an argument form involving predicates and quantifiers). The field was further developed by the Stoics, who singled out valid patterns of *propositional* argumentation (involving sentential connectives), and indeed flourished in ancient times and during the Middle Ages, when logic was regarded, together with grammar and rhetoric (the other two disciplines of the *trivium*), as the foundation of humanistic education. However, the modern conception of logic is only approximately 150 years old, having been initiated in England and Germany in the latter part of the nineteenth century with the work of George Boole (*An Investigation of the Laws of Thought*, 1854), Gottlob Frege (*Begriffsschrift*, 1879), and Richard Dedekind (*Was sind und was sollen die Zahlen?*, 1888). Thus modern symbolic logic is a relatively young discipline, at least compared with other formal or natural sciences that have a long tradition.

Throughout its long history, logic has always had a *prescriptive* as well as a *descriptive* component. As a descriptive discipline, logic aims to capture the arguments accepted as valid in everyday linguistic practice. But this aspect, although present throughout the history of the field, has taken up a position more in the background since the inception of the modern conception of logic, to the point that it has been argued that the descriptive component is no longer part of logic proper, but belongs to other disciplines (such as linguistics or psychology). Nowadays logic is, first and foremost, a prescriptive discipline, concerned with the identification, analysis, and justification of valid inference forms.

The articulation of logic as a prescriptive discipline is, ideally, a two-fold task. The articulation first requires the identification of a class of valid arguments. The class thus identified must have certain features: not just any class of arguments will do. For instance, it is reasonable to require that the logical validity of an argument depends on only its *logical form*. This amounts to requiring that the class of valid argument be closed under the relation "having the same logical form as," in that, if an argument is classified as valid, then so is any other argument of the same logical form. If this is the case, then such an identification clearly presupposes, and rests on, a notion of logical form.

The question of what constitutes a good theory of logical form lies oustide the scope of this book, and hence it is not pursued any further. We shall limit ourselves to the observation that one can achieve the desired closure conditions by requiring that the class of valid arguments be generated in some uniform way from some restricted set of principles. For instance, Aristotle's theory of the syllogism accomplishes this in a characteristically elegant fashion. It classifies: subject–predicate propositions on the basis of their forms into a small number of classes, and one then generates syllogisms by allowing the two premises and the conclusion to take all possible forms.

The second part of the task, however, is much harder. Once a class of arguments is identified, one naturally wants to know what it is that makes these arguments *valid*. In other words, to accomplish this second task, one needs a general theory of *logical consequence*, a theory that was not only unavailable to the ancients, but that would not be available until the appearance of modern symbolic logic – when an effort was undertaken to formalize and represent mathematical reasoning – and that would not be completely developed until the middle of the twentieth century. It is only with the development of the first general accounts of the notion of logical consequence through the work of Alfred Tarski (Der Wahrheitsbegriff in den formalisierten Sprachen, 1935) that modern symbolic logic reaches maturity.

One of the salient features of such an account is a property known as *monotony*, according to which the set of conclusions logically following from a given body of knowledge grows proportionally to the body of knowledge itself. In other words, once a given conclusion has been reached, it cannot be "undone" by the addition of any amount of further information. This is a desirable trait if the relation of logical consequence is to capture the essential features of rigorous mathematical reasoning, in which conclusions follow from premises with a special kind of necessity that cannot be voided by augmenting the facts from which they are derived. Mathematical conclusions follow *deductively* from the premises – they are, in a sense, already *contained in* the premises – and they *must* be true whenever the premises are.

There is, however, another kind of reasoning, more common in everyday life, in which conclusions are reached tentatively, only possibly to be retracted when new facts are learned. This kind of reasoning is *nonmonotonic*, or, as we also say, *defeasible*. In everyday reasoning, people jump to conclusions on the basis of partial information, reserving the right to preempt those conclusions when more complete information becomes available.

It turns out that this kind of reasoning is quite difficult to capture formally in a precise way, and efforts in this direction are relatively new, when compared with the long and successful history of the efforts aimed at formalizing deductive reasoning. The main impetus for the formalization of defeasible reasoning comes from the artificial intelligence community, in which people realized very early on that everyday commonsense inferences cannot quite be represented in the golden standard of modern deductive logic, the first-order predicate calculus. Over the past two or three decades, a number of formalisms have been proposed to capture precisely this kind of reasoning, in an effort that has surpassed the boundaries of artificial intelligence proper, to become a new field of formal inquiry – nonmonotonic logic.

This book aims to contribute to this development by proposing an approach to defeasible reasoning that is in part inspired by parallel developments in philosophical logic, and in particular in the formal theory of truth. The point of view adopted here is the one just mentioned, that the formal study of defeasible reasoning – nonmonotonic logic – has come into its own as a separate field. Accordingly, the emphasis is more on conceptual, foundational issues and less on issues of implementation and computational complexity. (This is not to underestimate the salience of these topics, in fact they are mentioned whenever relevant – they just happen to fall outside the purvue of the book.)

The book is organized as follows. Chapter 1 starts focusing on the development of modern symbolic logic from the point of view of the abstract notion of logical consequence; in particular, we consider those features of logical consequence that aim to capture patterns of defeasible reasoning in which conclusions are drawn tentatively, subject to being retracted in the light of additional evidence. A number of useful nonmonotonic formalisms are briefly presented, with special emphasis on the question of obtaining well-behaved consequence relations for them. Further, a number of issues that arise in defeasible reasoning are treated, including skeptical versus creduluous reasoning, the special nature of so-called *floating conclusions*, and the conceptual distinction between an approach to the *nature* of conflict and the concrete question of how conflicts should be *handled*.

Chapter 2 deals with the problem of developing a direct approach to nonmonotonic inheritance over *cyclic* networks. This affords us the opportunity to develop the main ideas behind the present approach in the somewhat simpler setting of defeasible networks. The main thrust of the chapter is toward developing a notion of *general extension* for defeasible networks that not only applies to cyclic as well as acyclic networks, but also gives a *directly skeptical* approach to nonmonotonic inheritance.

Finally, Chaps. 3 and 4 further develop the approach by extending it to the much richer formalism of default logic. Here, the framework of general extensions is applied to Reiter's *default logic*, resulting in a well-behaved relation of defeasible consequence that vindicates the intuitions of the directly skeptical approach.

ACKNOWLEDGMENTS

I thank Rich Thomason and Jeff Horty for first introducing me to the field of defeasible reasoning, and Mel Fitting and Rohit Parikh for illuminating conversations on this topic. I am indebted to several anonymous referees for the *Artificial Intelligence Journal* and Cambridge University Press for providing useful feedback and criticisms. The critical example 3.5.2 in Chap. 3 was suggested by one of them, and its interpretation in the block world is due to Madison Williams.

I am grateful to Elsevier Publishing for permission to use material in Chaps. 2, 3, and 4 which first appeared in the *Artificial Intelligence Journal* (see Antonelli, 1997, 1999); and to Blackwell for using material in Chap. 1, which first appeared in Floridi (2004) (see Antonelli, 2004).