

# *Conjectures*

an inquiry concerning the  
**logic of induction**

Proefschrift ter verkrijging van de graad van  
doctor aan de Katholieke Universiteit Brabant,  
op gezag van de rector magnificus, prof. dr. L.F.W. de Klerk,  
in het openbaar te verdedigen ten overstaan van een  
door het college van dekanen aangewezen commissie  
in de aula van de Universiteit op  
donderdag 6 april 1995 om 16.15 uur

door

**Pieter Adriaan Flach**

geboren op 8 april 1961 te Sneek

*Promotoren:* Prof.dr. R.A. Meersman  
Prof.dr. J.-J.Ch. Meyer

*Voor mijn ouders*

© 1995 by Peter A. Flach  
All rights reserved. Published 1995  
Printed in the Netherlands  
99 98 97 96 95 5 4 3 2 1

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK,  
DEN HAAG

Flach, Pieter Adriaan

Conjectures : an inquiry concerning the logic of  
induction / Pieter Adriaan Flach. - [S.l - s.n.]  
Proefschrift Katholieke Universiteit Brabant  
Tilburg. -  
Met lit. opg.  
ISBN 90-9008119-4  
NUGI 855  
Trefw.: kunstmatige intelligentie / logica /  
inductie.

## CONTENTS

*Preface* ix

1. INTRODUCTION AND OVERVIEW	1
<i>— in which the motivation for this study will be given, its scope will be delineated, preliminary concepts and notation will be defined, and an overview will be provided of what is still to come —</i>	
§1. Motivation and scope of the thesis	1
<i>Philosophy 2; Logic 2; Artificial intelligence 3</i>	
§2. Preliminaries	3
<i>Terminology 3; Logical preliminaries 5</i>	
§3. Overview of the thesis	9
<i>Backgrounds 9; Foundations 10; Practice 10</i>	

### *Part I: Backgrounds*

2. THE PHILOSOPHY OF INDUCTION	13
<i>— in which the philosophical backgrounds of this study are discussed, leading to the conclusion that the dichotomy between explanatory and confirmatory induction proposed and defended in this thesis is already present, albeit implicitly, in the work of Peirce and Hempel —</i>	
§4. The ‘Problem of Induction’	13
§5. The logic of abduction	16
§6. The logic of confirmation	18
<i>Adequacy conditions 18; The confirmation paradox 20; The satisfaction criterion of confirmation 21</i>	
§7. Discussion	22
<i>The logical form of abduction 23; Adequacy conditions for abduction 24; The confirmation paradox 26; The satisfaction criterion of confirmation 27; Degree of confirmation and its relation to the qualitative concept 28</i>	
§8. Summary and conclusions	31
3. APPROACHES TO COMPUTATIONAL INDUCTION	35
<i>— in which it is demonstrated how problems from machine learning can be reformulated as problems of explanatory and confirmatory induction —</i>	
§9. Concept learning from examples	35
<i>The Version Space model 35; Attribute-value languages 38; Concept learning in predicate logic 39</i>	

## Contents

§10. Induction of logic programs	41
<i>Top-down induction 43; Bottom-up induction 43</i>	
§11. Induction of integrity constraints	44
<i>Induction as nonmonotonic inference 44; Discovery of clausal theories 45; Induction of attribute dependencies 46</i>	
§12. Discussion	47
<i>Induction of logic programs as explanatory reasoning 47; Induction of integrity constraints as confirmatory reasoning 48; Incremental induction 50</i>	
§13. Summary and conclusions	51
4. THE ANALYSIS OF NON-DEDUCTIVE REASONING	53
— <i>in which the logical tool of a consequence relation is introduced through the work of Kraus, Lehmann &amp; Magidor</i> —	
§14. Deduction: the logic of satisfaction-preservation	54
<i>Non-deductive logics 56</i>	
§15. An analysis of plausible reasoning	56
<i>Plausible consequence relations 57; Rule systems for plausible reasoning 58; Characterisation of plausible consequence relations 60</i>	
§16. Discussion	62
<i>The KLM approach as a methodology of descriptive logic 62; The pragmatics of consequence relations 64; Comparing forms of reasoning 66</i>	
§17. Summary and conclusions	67

## Part II: Foundations

5. OUTLINE OF A DESCRIPTIVE THEORY OF INDUCTION	71
— <i>in which it is argued that the logical foundations of induction be based on a generalised notion of preservation semantics, two qualities that may be preserved in inductive arguments being explanatory power and regularity of interpretations</i> —	
§18. What is a logical system?	72
<i>What is a semantics? 72; What is a proof procedure? 78; The complete picture 79</i>	
§19. Elements of an inductive logic	80
<i>Conjectural consequence relations 81; Preservation semantics 81</i>	
§20. Summary and conclusions	83
6. PROPERTIES OF CONJECTURAL CONSEQUENCE RELATIONS	85
— <i>in which the study of general patterns of inductive reasoning is commenced, leading to a catalogue of rules for explanatory and confirmatory reasoning</i> —	
§21. The concept of an admissible formula	85

## Contents

§22. Adequacy conditions for confirmatory reasoning	86
§23. Adequacy conditions for explanatory reasoning	89
§24. Additional rules for conjectural reasoning	92
<i>Predictions 92; Verification, Right Extension, and Incrementality 93; Falsification, Right Excess, and Consistency 95; Convex consequence relations 96</i>	
§25. Summary and conclusions	100
Appendix: properties of conjectural discovery procedures	102
7. RULE SYSTEMS FOR CONJECTURAL REASONING 105	
— <i>in which various forms of conjectural reasoning will be axiomatised and semantically characterised</i> —	
§26. Explanatory reasoning	105
<i>Reverse deductive consequence relations 106; Explanatory consequence relations 108; Weaker notions of explanation 113</i>	
§27. Regularity-based confirmatory reasoning	113
<i>Preferential confirmatory consequence relations 114; Closed-world reasoning as preferential confirmatory reasoning 119</i>	
§28. Consistency-based confirmatory reasoning	120
<i>Weak confirmatory reasoning 121; Consistency-based confirmatory reasoning with partial models 122</i>	
§29. Summary and conclusions	128

## Part III: Practice

8. KNOWLEDGE DISCOVERY IN DATABASES 131	
— <i>in which an illustration of confirmatory induction is given, and the ideas underlying the system index, a system for inductive data engineering, are discussed</i> —	
§30. Inductive data engineering	131
§31. Attribute dependencies in databases	133
<i>A bit of relational database theory 134; Reformulation in logical terms 136; Using a proof procedure 138</i>	
§32. Induction of attribute dependencies	140
<i>Regularity-based confirmatory induction 142; Consistency-based confirmatory induction 143</i>	
§33. A confirmatory discovery procedure	147
<i>The goal of attribute dependencies: decompositions 148; Evaluating dependencies and decompositions 151; The INDEX system 152</i>	
§34. Summary and conclusions	154

## *Contents*

9. CONCLUSIONS AND FUTURE WORK	155
— <i>in which the main contributions of this thesis pass the review, and in passing the main open problems and opportunities for future work are indicated</i> —	
§35. Philosophy	155
§36. Logic	156
§37. Artificial intelligence	157
<i>Appendix: A session with INDEX</i>	159
<i>Glossary</i>	167
<i>Bibliography</i>	173
<i>Summary</i>	177
<i>Samenvatting</i>	179
<i>Curriculum Vitae</i>	181

## PREFACE

WHEN FOLLOWING THE motivation for this study back to its origin, I should probably start with the Master's thesis I completed 8 years ago, when I was a student in Electrical Engineering at Twente University. Under the stimulating supervision of Leo Veelenturf I was trying to set up a formal framework to compare and analyse various forms of behaviour that was referred to as 'learning from examples'. It was not without reservations that I took up this work, since 'reasoning by induction' seemed such a mythical and unmechanisable process to me; yet, I remember a distinct feeling of disappointment when I became aware that there might be, after all, some rules of the game.

The theme of induction kept intriguing me, and I reworked my Master's thesis into an article which I submitted to the *Artificial Intelligence* journal. When I picked up the subject again in Tilburg I soon discovered that there was quite a bit of similar (and better) work going on, so the rejection of that submission did not come as a great surprise. Even if my early efforts were a scientific failure, they brought me in contact with Ranan Banerji who, since I based my work on his, acted as a referee for the paper. Ranan's warm personality and scholarly attitude has been a continuing stimulus ever since.

Two scientific events that helped to shape my thoughts occurred in early 1991. In January I attended a workshop on non-monotonic reasoning in Amsterdam, organised by Wiebe van der Hoek, John-Jules Meyer and Yao-Hua Tan, where I learned about the work of Kraus, Lehmann and Magidor, that plays such a central role in this thesis. In March I witnessed the genesis of the inductive logic programming community at a workshop organised by Steve Muggleton and Pavel Brazdil in Viana do Castelo (Portugal). Many forms of close scientific co-operation have sprung from that first meeting.

In the meantime, Robert Meersman gave me all the freedom to work on research topics of my choosing. My work on inferring attribute dependencies in relational databases was an attempt to connect to the work on conceptual database design carried out in the Infolab group. The fact that this work is not the main concern of the thesis (although it is treated in some detail in chapter 8) is a result, both of my own esoteric foundational interests and of the liberty provided by my environment, which I hereby gratefully acknowledge.

Another essential ingredient of that environment was embodied in the Institute for Language Technology & Artificial Intelligence (ITK), a collaborative effort of the Computational Linguistics group at the Faculty of Linguistics on the one hand, and the Infolab group at the Faculty of Economics on the other. By way of sad coincidence, not only the start of my thesis work in Tilburg but also its completion more or less coincides with that of ITK as a separate and independent entity. It is in appreciation of ITK's unfortunately short life that this thesis is published in the distinguished ITK dissertation series.

## Preface

I would like to thank my promotors and the other members of the reading committee (Prof.dr. H.C. Bunt (KUB), Dr. L. De Raedt (Leuven), Prof.dr. H. de Swart (KUB), and Prof.dr. K. Schlechta (Marseille)) for their efforts. I would especially like to thank John-Jules Meyer for his unfailing enthusiasm and encouragement, and for his careful proof-reading. I had several stimulating discussions with Luc De Raedt, Harrie de Swart, and Karl Schlechta, and thank them for the many improvements they suggested. Elias Thijsse provided partial help with some of the stuff in chapter 7.

Part of the work was done while I was visiting the Joæef Stefan Institute in Ljubljana and the Katholieke Universiteit Leuven. I thank my hosts, Nada LavraË and Luc De Raedt for making my stays enjoyable and productive. I also gratefully acknowledge the financial support of the European Union, through the Basic Research Action *Inductive Logic Programming* (6020), the PECO network ILPnet (CIPA3510OCT920044), and the Networks of Excellence Compulog and MLnet.

I would like to thank my colleagues for being patient with me while I was writing; Hennie Daniëls, Robert Meersman and Leo Remijn for taking over a few of my lectures; Tineke Kleine and Alice Kloosterhuis for secretarial assistance; and Egon Verharen for providing technical assistance way beyond the call of duty (although MicroSoft products sometimes resist even him...).

Finally, I would like to say ‘thank you’ to all the people who were important for me during the last year (you know who you are), and to my wife Nada who, despite all the nights I was working late, never uttered a single complaint!