A Logical Approach to Computational Theory Building

with applications to sociology

Jaap Kamps

A Logical Approach to Computational Theory Building

ILLC Dissertation Series 2000-02

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam
phone: +31-20-5256051
fax: +31-20-5255206
e-mail: illc@wins.uva.nl
http://www.illc.uva.nl/

A Logical Approach to Computational Theory Building
with applications to sociology

Academisch Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof.dr. J.J.M. Franse
ten overstaan van een door het college voor promoties ingestelde
commissie, in het openbaar te verdedigen in de
Aula der Universiteit
op vrijdag 10 maart 2000, te 12.00 uur

door

Jacob Kamps

geboren te Kampen.

Promotor:
prof.dr. W.E. Saris, Universiteit van Amsterdam

Co-promotor:
dr. J.M.F. Masuch, Universiteit van Amsterdam

Further members of the reading committee:
prof.dr. J.F.A.K. van Benthem, Universiteit van Amsterdam
prof.dr. M.T. Hannan, Stanford University
prof.dr. T.A.F. Kuipers, Rijksuniversiteit Groningen
prof.dr. R.J. Mokken, Emeritus, Universiteit van Amsterdam
dr. L. Pólos, Erasmus Universiteit Rotterdam
dr. Y.-H. Tan, Erasmus Universiteit Rotterdam

The investigations were supported in part by the Netherlands Organization for Scientific Research (NWO grants # 030-50-334; # SIR-12-2918; and # SIR-12-3591). This book was set by the author using the LaTeX typesetting system and was printed and bound by PrintPartners Ipskamp.

Printed in the Netherlands
03 02 01 00 5 4 3 2 1

ISBN 90–5776–044–4

Contents

Illustrations

Tables

Acknowledgments

We gratefully acknowledge the permission of editors and publishers to reproduce (parts of) the following papers used in this thesis:

Chapter 2 is partly based on: J. Kamps. Formal theory building using automated reasoning tools. In A. G. Cohn, L. K. Schubert, and S. C. Shapiro, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixth International Conference (KR'98)*, pages 478–487. Morgan Kaufmann Publishers, San Francisco CA, 1998.

Chapter 3 is an extended version of: J. Kamps and L. Pólos. Reducing uncertainty: A formal theory of *Organizations in Action*. *American Journal of Sociology*, 104:1776–1812, 1999. ©1999 by The University of Chicago.

Chapter 4 is an extended version of: J. Kamps. On criteria for formal theory building: Applying logic and automated reasoning tools to the social sciences. In J. Hendler and D. Subramanian, editors, *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*, pages 285–290. AAAI Press/The MIT Press, Menlo Park CA, 1999. ©1999 by the American Association for Artificial Intelligence.

Chapter 5 is an extended version of: J. Kamps. On the process of axiomatizing scientific theories: Using justification criteria in the context of discovery. In *Proceedings of the AISB'99 Symposium on Scientific Creativity*, pages 49–58. The Society for the Study of Artificial Intelligence and Simulation of Behaviour, 1999.

Chapter 6 is a shorter version of: J. Kamps and M. Masuch. Partial deductive closure: Logical simulation and management science. *Management Science*, 43 (9):1229–1245, 1997. ©1997 by the Institute for Operations Research and the Management Sciences.

Chapter 7 was published as: J. Kamps and G. Péli. Qualitative reasoning beyond the physics domain: The density dependence theory of organizational ecology. In B. Bredeweg, editor, *Ninth International Workshop on Qualitative Reasoning*, pages 114–122, University of Amsterdam, Amsterdam, The Netherlands, 1995.

Preliminaries

Formal Logic

We will assume that the reader is familiar with the basic notions of classical logic. See introductory texts (Barwise and Etchemendy 1992, 1994) or standard textbooks such as (Tarski 1946; Enderton 1972; Chang and Keisler 1990).

FIRST-ORDER LOGIC (WITH EQUALITY)

Syntax A logic is based on a particular set of symbols. For first-order logic these are a fixed set of the logical symbols (variables; connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$; equality symbol $=$; quantifiers $\forall$ and $\exists$; and parenthesis and commas) and a varying set of non-logical symbols (individual constants; functions; and predicates).

There are strict rules that define which sequences of symbols are well-formed formulas of the language. A formula with no free variables is called a sentence.

Semantics A (formal) model consists of a universe (a non-empty set) and an interpretation function that is a mapping between the non-logical symbols (i.e., constants, functions, and predicates of the language) and elements of this universe. The logical symbols have a fixed interpretation.

This allows for a strict definition of relative truth, i.e., whenever a sentence φ is satisfied (or true) in a model $\mathcal{A}$, in symbols, $\mathcal{A} \models \varphi$. We will use $\Sigma \models \varphi$ to denote that sentence φ is satisfied in every model of the set of sentences Σ, i.e., φ is a logical consequence of Σ.

Proof Theory We use $\Sigma \vdash \varphi$ to denote that sentence φ is deducible from the set of sentences Σ, i.e., there exists a proof of φ from Σ. Such a (formal) proof is a finite sequence of sentences ending with φ and such that each sentence is either a logical axiom (i.e., a tautology, quantifier or identity axiom); belongs to Σ; or can be inferred from earlier sentences using an inference rule.

There exist sets of inference rules that are sound and complete, i.e., such that $\Sigma \vdash \varphi$ if and only if $\Sigma \models \varphi$.

In this thesis we will assume that the underlying logic is standard first-order logic—generally conceived as *the* logic by outsiders.[1] Although a reader lacking some proficiency in formal logic may not be able to follow all detailed arguments or their underlying motivation, we made particular efforts to ensure that the main points can still be understood.

Artificial Intelligence

The automatization of logical inferencing has a long history (Newell and Simon 1956; Beth 1958b). The decision to use standard first-order logic makes immediately available a wide variety of computer applications. We do extensively use generic tools from the field of automated reasoning.[2]

AUTOMATED REASONING TOOLS

Automated Theorem Provers are programs designed for finding proofs of theorems (usually by proving that a set of sentences is unsatisfiable). We mainly used OTTER (McCune 1994b), a resolution-style theorem prover for first-order logic with equality. Other automated theorem provers we used include metafor (Ó Nualláin 1993), SPASS (Weidenbach 1997), and bliksem (De Nivelle 1999).

Automated Model Generators are programs designed for enumerating the finite (small) models of a set of sentences. We mainly used MACE (McCune 1994a), a model generator for first-order logic with equality, based on a Davis-Putnam procedure for propositional satisfiability testing. Other automated model generators we used include Finder (Slaney 1994b) and SEM (Zhang and Zhang 1995).

Problem Library There exists a comprehensive library of automated reasoning problems, called TPTP (Sutcliffe, Suttner, and Yemenis 1994). TPTP uses a uniform format for representing problems, which can be translated into the appropriate format required by specific programs.

A general introduction to automated reasoning is (Wos, Overbeek, Lusk, and Boyle 1992) or the more formal treatment of (Fitting 1996). Automated reasoning is a subfield of artificial intelligence. For a general introduction, see (Genesereth and Nilsson 1987) or the classic collection (Feigenbaum and Feldman 1963), which is especially recommended since it originated from a business school.

The substantive theories we investigated are qualitative theories. Their representation is related to research on formal theories of the everyday physical world (Hayes 1979, 1985b; Davis 1998) and common-sense reasoning (McCarthy 1959; Hobbs and Moore 1985; Hobbs, Blenko, Croft, Hager, Kautz, Kube, and Shoham 1985; Davis 1990). One chapter deals explicitly with qualitative reasoning techniques (Weld and de Kleer 1990; Kuipers 1994).

[1] The notions we use generally do not depend on this choice and can easily be generalized to other logics.

[2] The programs we will refer to as automated model generators are sometimes called a (finite) domain enumerator, model generation program, model searcher, model finder, or model builder.

Since, in our view, the axiomatization of theories is not restricted to justification only, there are some links with machine discovery (Langley, Simon, Bradshaw, and Żytkow 1987; Thagard 1988; Shrager and Langley 1990).

Sociology

We will extensively discuss the formalization of theories from the field of sociology. For a general introduction to sociology, see (Elster 1989). In some chapters, we will use theories from a remarkable approach to sociological theory development that attracted much attention in the fifties and sixties. This approach, based on the work of Zetterberg (1955, 1965), attempted to put ordinary language theories in an axiomatic form.[3] Although these theories are relatively simple from our current point of view, their perspicuous form makes them excellent illustrations for discussing various issues that arise during the formalization of social science theories.

The other substantive theories we deal with belong to sociology's subfield of organization theory.

(Some) Organization Theory

Rational adaptation theories Mainstream organizational theories regard organizations as agents that adapt rationally to changing environments. Specifically, these theories try to explain the structure of organizations by analyzing their adjustment to external influences. We will extensively discuss the axiomatization of a particular rational adaptation theory (Thompson 1967), which discusses the adjustments due to external influences that introduce uncertainty in the organization's decision making process. These organization theories describe organizations from an *individual* viewpoint.

Environmental selection theories Complementarily, a change in environmental conditions also affects the whole *population* of organizations. For example, if resource conditions deteriorate, the total population of organizations will decline (regardless of efforts of individual organizations to avoid this fate). Two chapters deal with the theory of organizational ecology (Hannan and Freeman 1977, 1989). Organizational ecology describes the processes by which organizational populations grow and decline due to changing environmental conditions. It abstracts from the rational behavior of individual organizations, and assumes that organizational populations are solely dependent on the environmental conditions.

For overviews of organization theory, see (Perrow 1986; Scott 1998) or see (Grandori

[3]This so-called axiomatic theory approach used ordinary language exclusively and based its theoretical inferencing on an intuitive notion of deduction. It is important to note that the virtues of these axiomatic theories have not been disputed. Criticism primarily focused on the validity of its ordinary language deductions (Costner and Leik 1964). There have been attempts to give a causal interpretation of these ordinary language statements, and to investigate under which conditions valid inferences can be made in the resulting causal models (Blalock 1969).

1987) for an overview that attempts to stress commonalities instead of differences.[4]

Philosophy

Any effort to formalize scientific theories is bound to encounter various issues of philosophical interest. For a general introduction to philosophy, see (Quine and Ulian 1970; Glymour 1992b). The most relevant subfield is philosophy of science. See (Bechtel 1988) for an overview, or more substantial texts (Nagel 1961, especially chapters 13 and 14) and (Hempel 1966; Rudner 1966).

There are important relations with various issues treated in the rich logico-philosophical literature. Specifically, there is an obvious link with the philosophical movement known as the logical empiricism or logical positivism (Ayer 1959; Neurath, Carnap, and Morris 1970). Moreover, our approach is closely related to application-oriented, formal approaches in the philosophy of science (for example, Kyburg 1968; Goodman 1977; Balzer, Moulines, and Sneed 1987). And last, but not least, there turn out to be interesting parallels with the philosophical treatment of mathematical discovery, especially with (Lakatos 1976), based on the critical philosophy of (Popper 1959, 1963) and the revival of mathematical heuristics by (Pólya 1945, 1954a,b, 1962, 1965).

[4]Current research in organization theory is extremely fragmented, therefore it is recommendable to start with reading some of the classics (for example, Simon 1947; March and Simon 1958; Cyert and March 1963; Thompson 1967). Alternatively, it is often insightful to read research on the edge between organization theory and the more coherent field of economics (for example, Barney and Ouchi 1986). A great last resort is (Nicholson 1995).

Chapter 1

Introduction & Overview

> If we like, we can look on theory as a game. The winner is the man who can deduce the largest variety of empirical findings from the smallest number of general propositions, with the help of a variety of given conditions. ... A science whose practitioners have been good at playing it has achieved a great economy of thought. No longer does it face just one damn finding after another. It has acquired an organization, a structure. ... But if theory is a game, it must like other games be played according to the rules, and the basic rules are that a player must state real propositions and make real deductions. Otherwise, no theory!
>
> —George Caspar Homans, *The Nature of Social Science*

The common availability of powerful computers can radically change the way science is performed. To realize this potential we need to make scientific knowledge accessible to our computers. Over the last decades, advances in computer science have allowed for the large-scale, computational representation of scientific data. The same progress has not yet been achieved for the most distinguished form of scientific knowledge—our scientific theories. Before we can computationally reason with theories, we have to find a suitable representation for them. This thesis presents *a logical approach to computational theory building*, that is to say, we will investigate the representation of scientific theories in formal logic.

Scientific theories are usually viewed as systems of propositions: the theory makes particular claims that follow from a number of underlying assumptions. The argumentation for such a claim usually involves both assumptions about the empirical reality, and general rules of inferencing (whose validity does not depend on the case at hand). Logic is the scientific discipline that traditionally investigates these general rules of inferencing. Formal logic provides rigorous formalisms that can play at least a normative role for deciding the validity of arguments.[1] The representation of theories as logical arguments is one of the least original thoughts in this thesis. It dates back long before the invention of computers, and can be traced to the writings of Aristotle, who provided the first explicit formal logic. In light of these historical facts, perhaps a more appropriate title for this thesis would have been *a computational approach to logical theory building.*[2]

The merits of a logical representation of theories are well-understood and frequently elaborated, and it has been shown that interesting methodological issues can be raised and discussed under the assumption that theories are axiomatized in formal logic. However, it should be noted that, so far, logical axiomatizations of significant scientific theories from the empirical sciences have rarely been constructed. In other words, the practical use of formal logic to represent actual theories from the empirical sciences is far less understood. These practical aspects will be the main focus throughout this thesis, and one of its main lines will be the logical formalization of actual scientific theories. In particular, we will provide or discuss formal renditions of actual theories or theory fragments in each chapter of this thesis.[3] For these applications of formal logic, we will use social science theories from the field of sociology and, especially, its subfield of organization science. This is, perhaps, a surprising choice because the paradigm examples of scientific theories in the empirical sciences belong to other domains, in particular to the domain of physics. The field of sociology is even renowned for its lack of systematic theory. Nevertheless, in as far as sociology is a science, it should adhere to the same methodological principles as the other sciences (for example, whether a theoretical argument in sociology is valid should be judged by the same criteria as an argument in physics). As has been widely argued before (Nagel 1956; Rudner 1966; Popper 1969; Neurath 1970), the "logic" of the social sciences is not different from the other sciences.

[1] The precise relation between logic and thinking is quite complex and need not concern us here. Even though human reasoning frequently defies logical principles, it can usually be reconstructed as a logical argument, for example by making explicit the underlying tacit assumptions (Henle 1962).

[2] Some have argued against logical representations of theories because they are 'unrealistic,' in the sense that there is little correspondence between logical and informal notions of theories as they occur in the literature. More specifically, logical theories are commonly thought to be only 'syntactic,' having no place for the 'theoretical intuition' or the 'mental models' behind the theory. To what extent this is a valid objection remains to be seen (at least, we will later present clear cases that undermine this view). Here, we just want to make the following observation: even if this notion may seem unnatural for human theorizing (especially outside the fields of logic and mathematics), it does seem a *natural* notion for computational theorizing.

[3] Whether these formal theory fragments will qualify as full axiomatizations of significant scientific theories remains to be decided, but they will suffice for substantiating the points we want to address.

1.1 Overview

The chapters in this thesis are relatively self-contained, allowing for separate chapters to be read independently. This thesis is structured as follows:

1.1.1 Introduction and Overview

This chapter will continue with an overview of the rest of this thesis.

1.1.2 Formal Theory Building Using Automated Reasoning Tools

Chapter 2 prepares the ground for a computational methodology for axiomatizing scientific theories. Its first goal is to explore how logical criteria for evaluating theories can be established, in practice, using generic automated reasoning tools. Its second goal is to investigate which logical criteria correspond to natural questions to be asked about theories from the social sciences.

Computational support can be of decisive importance for providing actual formalizations of theories. The automatization of logical inferencing has a long history (Newell and Simon 1956; Beth 1958b).[4] A decision to use standard first-order logic makes immediately available a wide variety of computer applications. More specifically, we can reuse generic tools developed in the field of automated reasoning (Wos et al. 1992). In this chapter, we will investigate whether current automated theorem provers and model generators can be of use to determine some of the familiar criteria for evaluating theories as proposed in logic and philosophy of science, such as the consistency and falsifiability of theories, and the soundness of derivations. This was inspired by earlier papers on the logical formalization of social science theories (Péli, Bruggeman, Masuch, and Ó Nualláin 1994; Péli and Masuch 1997) that use an automated theorem prover to prove the derivability of theorems.[5] By extending the range of criteria we can make a more detailed evaluation of the formal theory.

There are convincing philosophical arguments that social science theories should be evaluated by the same criteria as theories from other disciplines (Rudner 1966; Popper 1969). This does not imply that these criteria are directly applicable to theoretical writings in sociology. More specifically, these criteria

[4]Newell, Shaw, and Simon (1957) report the remarkable achievements of their computer program, the Logic Theory Machine, which discovered proofs of 38 out of the first 52 theorems in the second chapter of the *Principia Mathematica* (Whitehead and Russell 1962, probably theorems *2.02–*2.521 on pp.98–107).

[5]Earlier papers also addressed the consistency of their theories by checking whether the search-space of an automated theorem prover contained a contradiction. Of course, it is likely that the theorem prover would have found a contradiction in case the theory was inconsistent. However, being unable to find a contradiction in this necessarily limited search-space does not in itself prove that the theory under consideration is consistent.

might address issues that do not (yet) appear to be relevant to theorizers in sociology. The main part of this chapter is a case study of the formalization of a sociological theory, Hopkins (1964) "The Exercise of Influence in Small Groups." As it turns out, several of the standard criteria are relevant for this specific theory under consideration. Moreover, we succeeded in computationally evaluating each of these criteria using automated reasoning tools. Specifically, we formally and computationally proved the soundness of theoretical inferences; proved the consistency of the theory; proved the falsifiability and satisfiability (and, therefore, the contingency) of the theorems; proved that additional theorems do belong to the theory (when interpreted as a deductive system); proved that the axioms of the theory are not independent; and, finally, provided minimal axioms sets of the theory (which are proved to be independent). The use of automated reasoning tools is subject to various limitations, think of the undecidability of first-order logic, and practical limitation of CPU-power, memory, and time. However, using only default settings for the programs, none of the actual proof or model searches for determining the criteria required more than five seconds. That is to say, the evaluation of sociological theories seems to be within the computational power of current automated reasoning tools.

Realizing the full potential of these automated reasoning programs requires an intimate knowledge of the internal machinery. First, it requires an understanding of the automated reasoning techniques (see [Fitting 1996] for an overview). For example, interpreting the proofs generated by an automated theorem prover requires, at least, knowledge of the used inference rules (for example, OTTER [McCune 1994b] can use a dozen inference rules). Moreover, since there does not exist a uniform output-format, interpreting generated proofs also requires knowledge of the internal representation format used by the program. If a theorem prover fails to generate a proof, then examining a trace of the failed proof attempt is even far more bewildering. Second, using automated reasoning tools also requires heuristic knowledge on the choice of inference rules and search strategies (there are surprisingly few publications that address this, a notable exception is [Wos 1996]). Most of our problems turned out to be relatively simple, in the sense that they can be solved by the programs using the default settings. At least, this is true for the simplified and cleaned-up final versions of the theories presented in this thesis. Some of the intermediate versions did require serious efforts to be proven automatically, including modification of the used programs themselves.[6] Our main concern is how these types of tools can be used for the formalization of theories. Therefore, we will abstract away from program-specific details as much as possible.

[6]For example, the substantive theories used in this thesis usually have large vocabularies, whereas the distributed version of MACE accepts at most a total of twelve constants, predicates, and functions (including functions resulting from Skolemization). Most programs are distributed with their source-code, allowing modifications to these programs to be implemented by the user. We will gladly spare the reader the details.

1.1.3 A Formal Theory of *Organizations in Action*

Chapter 3 presents an axiomatization of a classic organization theory, Thompson (1967) "Organizations in Action." Thompson's book is one of the classic contributions to organization theory: it provides a framework that unifies the perspective treating organizations as closed systems, with the perspective that focuses on the dependencies between organizations and their environment. This framework has influenced much of the subsequent research in organization theory. "Organizations in Action" is an ordinary language text, in which only the main propositions are clearly outlined. This chapter provides a formal rendition of the first chapters of the book, by reconstructing the argumentation used in the text.

Although the text of "Organizations in Actions" does not contain explicit definitions, the use of terminology in the text strongly suggests strict dependencies between several important concepts. This allowed for the definition of these concepts in terms of a small number of primitive notions of organization theory. In the formal theory, the key propositions of (Thompson 1967) can be derived as theorems. The proofs of these theorems are based on a reconstruction of the argumentation in the text. Additionally, the formal theory explains why the theory is restricted to a particular type of organizations. Moreover, it derives a heretofore unknown implication of the theory that relates Thompson's theory to recent empirical findings and current developments in organization theory.

1.1.4 Criteria for Formal Theory Building

The social sciences are renowned for the richness of their vocabulary (one of the most noticeable differences with theories in other sciences). Social science theories are usually stated using many related concepts that have subtle differences in meaning. As a result, a formal rendition of a social science theory will use a large vocabulary. In the formalization of chapter 3 we started to experiment with the use of definitions as a means to combine a rich vocabulary with a small number of primitive terms.

> Now definitions are unlike theorems and unlike axioms. Unlike theorems, definitions are not things we prove. We just declare them by fiat. But unlike axioms, we do not expect definitions to add substantive information. A definition is expected to add to our convenience, not to our knowledge. (Enderton 1972, p.154)

If dependencies between different concepts are made explicit, we may be able to define some of the concepts in terms of the other concepts, or in terms of a smaller number of primitive concepts. If the theory contains definitions, the defined concepts can be eliminated from the theory by expanding the definitions. Eliminating the defined concepts does not affect the theory, in the sense that the models and theorems of the theory remain the same.

Chapter 4 extends the initial discussion of criteria for evaluating theories in chapter 2. Our earlier discussion did not distinguish between types of premises. Here, we discuss how to augment these criteria for scientific theories containing definitions. We will provide practical operationalizations of criteria for evaluating theories, including the consistency (existence of a model) or inconsistency (derivability of contradictory statements) of the theory; the soundness (existence of a proof) or unsoundness (existence of a counterexample) of conjectures and theorems; the satisfiability and falsifiability of theorems; and the independence of axioms. The tests for the criteria, in practice amounting to particular proof or model searches, can be directly performed by existing automated reasoning tools.

We used the criteria to evaluate a formal rendition of a classic organization theory, the formal theory of "Organizations in Action" discussed in chapter 3. Assessing the criteria allows for an exact evaluation of the merits of a theory. In some cases, this may reveal important facts about the theory, for example, the case study showed that one of the derived statements is unfalsifiable—empirical investigation of it can, at best, confirm its trivial validity.

1.1.5 The Process of Axiomatizing Scientific Theories

Scientific activities are traditionally dichotomized into the context of justification and the context of discovery. Within this distinction, the axiomatization of theories is viewed as the ultimate step in the justification of a theory. Chapter 5 deals with the question to which extent the axiomatization of theories should be considered as strictly justification.

A first-order logic rendition of a theory gives an explicit, unambiguous exposition of the theory. As discussed in chapter 4, there exists a number of criteria (most importantly, the consistency of theories and the soundness of derivations of theorems) that can be evaluated using generic tools from automated reasoning. We can make a rigorous evaluation of a scientific theory by assessing these criteria. However, we do not view these criteria as rigid, final tests. Quite the opposite, in our experience the criteria are especially useful during the process of formalizing a theory. The tests we suggest to evaluate the criteria do not only prove a criterion, but also present a specific proof or model that is available for further inspection. Specifically, during the construction of a formal theory, the criteria can provide useful feedback on how to revise the formal theory in case of a deficiency. For example, any exposition of a theory in ordinary language makes various theoretical presuppositions. Formalizing a scientific theory requires these presuppositions to be made explicit. To identify these implicit background assumptions is one of the thorniest problems in the formalization of a theory, requiring a deep understanding of the substantive field under consideration. If the appropriate background assumptions are not added to the formal theory, various conjectures may not be derivable. However, if we can find specific counterexamples to such an underivable conjecture, it is immediately clear why our proof attempt has failed. In particular, these counterexamples reveal which implicit (background) assumptions need to be added to the formal theory.

We will give a detailed illustration of this by formalizing a theory fragment from Zetterberg (1965) “On Theory and Verification in Sociology.” As it turns out, we will repeatedly revise the formal rendition of the theory, by adding implicit background knowledge, reformulating the conjectures, and changing the axioms of the theory. This process of formalization is essentially interactive. We attempt to use computational support for those tasks for which computers are better equipped. For example, we use automated reasoning tools for finding proofs or models. Notice that human theorizers have often difficulty in finding counterexamples that were not intended to be models of the theory. Theorizers tend to ignore these models since they conflict with their common-sense or with their understanding of the substantive domain. Fortunately, an automated model generator does not have such a bias. On the other hand, a human theorizer can use this knowledge to distinguish between non-intended models and genuine counterexamples. This decision is crucial because it determines whether we need to revise the premises (in case of a non-intended model) or whether we need to revise the conjecture (in case of a genuine counterexample). The product of our axiomatization attempt, a first-order logic rendition of a theory, is a deductive theory. However, the process of axiomatizing a theory is essentially non-deductive. For example, we may decide to revise the theory to account for a counterexample. Such an attempt to revise the theory is *abductive*. In fact, we use an extended form of abduction, since we may either decide to change the premises to explain the claim (traditional abduction), or decide to change the claim such that it can be explained by the premises as they stand.

Finally, and most importantly, the revision of the formal theory may have an impact on the original theory. This can be of great importance since even a minor modification of the original theory may avoid the costs involved in the empirical testing of incorrect or irrelevant hypotheses. The used criteria facilitate a piecemeal revision of the theory, resulting in a cyclic process of theory development, much like the discussion of the *polyhedra conjecture* in (Lakatos 1976).

1.1.6 Partial Deductive Closure

Chapter 6 presents earlier work on an implemented program to enumerate a certain class of theorems.

In formal logic, the deductive closure of a theory is taken for granted—in fact, the formal definition of a “theory” is a set of sentences closed under the rules of inference (Tarski 1956). In reality, however, it is impossible to generate the complete deductive closure of a premise set, for the resulting set is infinite (since it contains all tautologies, i.e., expressions that are always true and hence follow from any set). In consequence, the complete deductive closure of a premise set is neither realizable nor desirable—only *partial* deductive closures provide useful results. In this chapter, we present an algorithm that performs an efficient partial deductive closure for an important class of formulas, i.e., conditional formulas that relate two relevant domain properties. Statements of this form provide the backbone of any empirical social science; arguably, it is the most important class of empirical statements in the social sciences.

The implemented program is applied to a first-order logic rendition of Hannan and Freeman (1984) "Structural Inertia and Organizational Change" as published by Péli et al. (1994). As it turns out, our algorithm generates more theorems than the original discursive theory of organizational inertia made explicit, some of which are of theoretical interest.

1.1.7 Qualitative Reasoning beyond the Physics Domain

The ability to reason qualitatively about physical systems is important to understanding and interacting with the physical world for both humans and intelligent machines. Accordingly, its study has become an important subject of research in artificial intelligence. Most of the theories in the social sciences are qualitative theories. Chapter 7 does not discuss the formalization of theories in logic, but the modeling of theories using qualitative reasoning techniques (Weld and de Kleer 1990). Qualitative reasoning techniques are almost exclusively applied to the domain of physics (explaining the alternative names 'qualitative physics' and 'naive physics'). This chapter makes a case for the application of qualitative reasoning techniques outside the domain of physics. As a case in point, this chapter provides a qualitative simulation model of the density-dependence theory of organizational ecology (Hannan and Carroll 1992), using the General Architecture for Reasoning about Physics (GARP [Bredeweg 1992]).[7] The simulation model is shown to generate results of theoretical relevance. The chapter discusses differences between the application of qualitative reasoning techniques to physics and to the social sciences.

Reason for inclusion of this chapter is that sociological theories—unlike theories in physics—are usually qualitative descriptions, making them suitable candidates for modeling with these techniques. Moreover, this type of 'cognitive' modeling of qualitative theories has been very influential on the logical formalization view as outlined in earlier chapters. The qualitative reasoning techniques are, for a large part, rooted on artificial intelligence research on common-sense reasoning and, in particular, efforts to formalize knowledge of the everyday physical world (Hayes 1979, 1985b). Hayes' proposal stated explicitly:

> It is *not* proposed to develop a new formalism or language to write

[7]This system largely incorporates the Qualitative Process Theory (QPT [Forbus 1984]). There also exists a mathematically elegant system for reasoning with qualitative differential equations (Kuipers 1994) (to which more expressive systems like QPT can be translated [Farquhar 1993]).

> down all this knowledge in. In fact, I propose (as my friends will have expected) that first-order logic is a suitable basic vehicle for representation. (Hayes 1985b, p.3).

This dictum is, in a sense, precisely what we did in earlier chapters. In this chapter we will use one of the 'special' programming languages that was developed later.[8] To no great surprise, there are direct ways to axiomatize these special qualitative reasoning languages in first-order logic (Davis 1990, 1992).

1.1.8 Discussion and Related Work

Finally, chapter 8 contains some discussion and references to related work.

[8] There were legitimate computational reasons to develop these special languages. These special languages allowed for building complex systems far beyond what computational inferencing in first-order logic allowed. Earlier chapters show that this may be no longer the case.

Chapter 2

Formal Theory Building Using Automated Reasoning Tools

The merits of representing scientific theories in formal logic are well-known. Expressing a scientific theory in formal logic explicates the theory as a whole, and the logic provides formal criteria for evaluating the theory, such as soundness and consistency. On the one hand, these criteria correspond to natural questions to be asked about the theory: is the theory contradiction-free? (is the theory logically consistent?) is the theoretical argumentation valid? (can a theorem be soundly derived from the premises?) and other such questions. On the other hand, testing for these criteria amounts to making many specific proof attempts or model searches: respectively, does the theory have a model? can we find a proof of a particular theorem? As a result, testing for these criteria quickly defies manual processing. Fortunately, automated reasoning provides some valuable tools for this endeavor. This chapter discusses the use of first-order logic and existing automated reasoning tools for formal theory building, and illustrates this with a case study of a social science theory, Hopkins' theory of the exercise of influence in small groups.

2.1 Introduction

The theory building methodology outlined in this chapter is by no means a new one. The use of formal logic to represent scientific theories dates back, at least, to the logical positivists (Ayer 1959). What is novel in our approach is the extensive use of automated reasoning tools. One of the reasons for the demise of positivism was the inability to put philosophy into practice. The formalization of scientific theories requires a huge amount of tedious calculations that exceed manual processing capabilities. The use of computational tools allows us to transcend these limitations, and bring much of the positivist philosophy to life.

This research is part of the concerted effort to revive formal theory building in the social sciences by using formal logics and by taking advantage of the available computational support. Social scientists usually agree that theories should be logical, but they rarely address the issue, eschewing the difficulties of investigating the logical structure of 'discursive theories' (theories expressed in natural language, the standard representation in the social sciences). Rather than engaging ourselves in abstract, philosophical deliberations,[1] we focus on applied case studies. That is, we 'formalize' actual social science theories by rationally reconstructing them and expressing them in logical form. The resulting formalizations are then tested using logical criteria, a task greatly facilitated by the availability of computational tools.

In section 2.2 we will explain the value of formalization in the explication of scientific theories and the role tools from automated reasoning can play in this process. In section 2.3 we present a case study that shows how this can be applied in the social sciences, and we end in section 2.4 by summarizing our experiences and drawing conclusions.

2.2 Formal Theory Building

The principal reason for formalizing scientific theories is to clarify and explicate them. Until a scientific theory is expressed in a formal and unambiguous manner, it remains open to many interpretations. Provided that readers can interpret the formalism, a formal exposition of a scientific theory allows them to understand the theory, to distinguish between alternative readings, to gauge its boundaries, and to compare it with alternatives (see also Suppes 1968).

We use the classical, axiomatic-deductive notion of a theory. The premises of a theory consist of universal statements (universal laws or empirical generalizations, possibly supplemented with definitions). The theory itself is the deductive closure of the set of premises (Tarski 1956). Theoretical explanations and predictions correspond to deductions from the set of premises (Popper 1959).

We prefer to use this strict notion of a theory over more liberal ones.[2] The reason for this is simple: our main interest is the justification of theories, and we are therefore interested in strict criteria that can be objectively evaluated. This position has clear limitations; many other aspects of theory building require creativity and insight (activities with which logic is rarely associated). Especially the discovery of theories requires different methods than formal logic and deduction. Most of the research in the social sciences is directed at empirical surveys,

[1]See for example (Adorno, Dahrendorf, Pilot, Albert, Habermas, and Popper 1969) for interesting arguments in favor of, and against the use of formal logic in the social sciences.

[2]Although this conception of theory is mainly inspired by theories in mathematics and physics, it is applicable to all empirical sciences, including the social sciences (Rudner 1966; Popper 1969).

and the main theme of all methodology textbooks is how to perform such empirical research. Our efforts do not replace, but rather complement this empirical research. As soon as a (tentative version of) a theory is formulated, we have some powerful tools for evaluating it. There are several logical criteria available for evaluating a theory formalized in a logical language, such as the consistency of the theory or soundness of a derivation.

Many of these logical criteria correspond to natural questions which we would like to ask about a scientific theory:

Is the argumentation of the theory valid? In logical terms, *can a given conjecture be soundly derived from the premises?* If we can prove a given theorem from the premises, the argumentation is sound. That is, if we consider cases in which the premises hold, then the theorem must also hold (i.e, the theorem is a prediction). Conversely, if we consider cases in which the theorem holds, then the premises give an explanation for the theorem (there may be other explanations, although these must satisfy the same soundness criterion).

For proving the fallaciousness of an argument we use a complementary approach: looking for a counterexample. If we can construct a model in which the premises hold but the conjecture does not hold, we have refuted the conjecture.

Is the theory contradiction-free? In logical terms, *is the theory logically consistent?* An inconsistent theory has an empty domain: empirically testing an inconsistent theory is futile. If we can construct a model of the premise set then the theory is logically consistent. One may want to include the (spelled-out) theorems, but as long as these are soundly derivable from the premises they cannot make the theory inconsistent.

To establish that a theory is inconsistent, we use a complementary approach: showing that it leads to an absurdity. An inconsistent theory is a trivial theory, in the sense that any statement (*and* its negation) is derivable from it. If we can derive a contradiction from some subset of the premises, the theory is inconsistent.

Is the theory falsifiable? *Are the theorems logically contingent?* If no state of affairs can possibly falsify a theory, then it is a waste of time to empirically test the theory. Falsifiability is an essential property of a scientific theory (Popper 1959). If we can construct a model (disregarding all premises) where a theorem of the theory is false, then this theorem is falsifiable. If we can prove a theorem from an empty premise set, the theorem is not falsifiable. A theory that contains at least one falsifiable theorem (and therefore at least one falsifiable premise) is falsifiable.

If we can also construct a model in which the theorem holds (which is always the case for soundly derivable theorems in a consistent theory), then this theorem is satisfiable too. A theorem that is both satisfiable and falsifiable, is contingent:

the validity of the theorem is strictly determined by the premises—it is neither a tautology nor a contradiction.

Does the theory use redundant premises? *Are the premises logically independent?* If a premise can be derived from the other premises, it becomes a theorem and can be removed from the premise set. If we can prove that neither a premise nor its negation is derivable from the other premises, then this premise is independent.

What is the explanatory and predictive power of the theory? *What do the theorems of the theory look like?* Investigating the set of theorems will give insight in a theory's explanatory and predictive power, because the theorems are the predictions of the theory, and their proofs give explanations for them.

What is the domain that the theory describes? *What do the models of the theory look like?* Investigating the models of a theory gives insight into the domain which the theory describes.

In practice, testing for logical criteria requires many derivations involving large sets of formulas. In this endeavor, automated reasoning provides some invaluable tools. We use the following tools:

- OTTER (McCune 1994b), a resolution-style automated theorem prover for first-order logic with equality. This theorem prover can find inconsistencies in the set of input formulas. Its principal use is to construct refutation proofs of conjectures, by feeding it a (consistent) premise set together with the negation of a conjecture. If the program derives an inconsistency, then we have, in fact, proved the conjecture (because the conjecture must hold in all models of the premises).

- MACE (McCune 1994a), a model generator for first-order logic with equality, based on a Davis-Putnam procedure for propositional satisfiability testing. This model generator can find small models of the set of input formulas (for example to prove the consistency of the theory). It can also be used in an attempt to find counterexamples to a conjecture, by feeding it a premise set together with the negation of the conjecture.

There are many other automated theorem provers and model generators available. We choose here to use OTTER and MACE because they are companion programs that can read the same input format. This is a great advantage for our work, since we want to switch between theorem proving and model generation, depending on which type of tool is most suitable for the specific proof/disproof attempt at hand. In the next section, we will explain the use of these tools for formal theory building in detail.

2.3 A Social Science Case Study

Several case studies of formalization using automated reasoning support tools have been performed. These case studies include the following social science theories: Mintzberg's *Contingency Theory* (Glorie, Masuch, and Marx 1990), Thompson's *Organizations in Action* (Pólos 1991), and Hannan and Freeman's *Organizational Ecology* including their theory of organizational inertia (Péli et al. 1994), life history strategies (Péli and Masuch 1997), niche width (Bruggeman 1997; Péli 1997), and age dependence fragment (Hannan 1998).[3] The original theories, as most other social science theories, were stated in natural language. As a result, the main obstacle for formalizing such a discursive theory is their rational reconstruction: interpreting the text, singling out important concepts, distinguishing assumptions and theorems from other parts of the text, and reconstructing the argumentation. This motived our choice for the theory we want to formalize as a case study in this chapter: Hopkins (1964) '*The Exercise of Influence in Small Groups.*' Although this is not a formal theory for it uses natural language exclusively, it is an axiomatic theory in which axioms and theorems are clearly outlined.[4] This greatly facilitates the rational reconstruction of the theory, and allows us to focus on the actual formalization of the theory and the role automated reasoning tools can play in this process.

2.3.1 Hopkins' Theory of Small Groups

This section contains a logical formalization of Hopkins (1964) '*The Exercise of Influence in Small Groups.*' Hopkins investigates the distribution of influence among members of a group, and attempts to explain differences in the amount of influence that members exercise. He follows the standard usage of the concept of a sociological group, which is 'defined' by the interactions of its members (Homans 1950, pp.82–86). If certain persons interact frequently with one another in accord with established patterns, they are viewed as a group (irrespective of the particular activities in which they interact). Two further criteria of a group are that "the interacting persons define themselves as 'members'," and that "the persons in interaction be defined by others as 'belonging to the group'" (Merton 1968, p.340). As is immediately clear, the concept of group is relative and the boundaries of groups are not necessarily fixed.

Hopkins develops his theory for certain kinds of groups, namely small groups. The concept of a 'small' group is characterized by "a rather high degree of visibility," "the absence of stable, well-defined subgroups," and "the interaction takes

[3]Some fragments of the resulting formalizations are included in the TPTP (Thousands of Problems for Theorem Provers) Problem Library (Sutcliffe et al. 1994).

[4]Note that this is an exceptional case; only part of social science texts state carefully formulated propositions, very few of those attempt to make their underlying assumptions explicit, and even fewer indicate inferences between them.

place when all or most members are present and participating in a common activity" (Hopkins 1964, pp.14–17). Here, the degree of visibility refers to "the extent to which the norms and the role-performances with a group are readily open to observation by others" (Merton 1968, p.373). Hopkins continues by discussing the interaction system of a group, thereby identifying five relevant properties of groups: its rank structure, the frequency of interaction, the level of visibility, the degree of normative consensus, and the amount of influence exercised (pages 21–25). Each of these five group properties can also be interpreted for individual members of a group, and in this way we derive five corresponding properties of the status of members in groups: rank, centrality, observability, conformity, and influence (see Table 2.1). These five properties of the status of members form the

Group Property	Status Property
Rank structure	Rank
Frequency of interaction	Centrality
Level of visibility	Observability
Degree of normative consensus	(Subjective) Conformity
Amount of influence exercised	Influence

Note: This table is reproduced from (Hopkins 1964, p.26).

Table 2.1: Properties of groups and corresponding status properties.

core of Hopkins' theory (and are discussed in detail on pages 26–40 of his book):

Rank is referring to the generally agreed upon worth or standing of a member relative to the other members. A member's relative rank may derive from institutionalized evaluations of well-defined statuses (such as professor and student in a seminar), or from differential evaluations of qualities or performances (such as high- and low-ranking participants in a study group).

Centrality designates how close a member is to the 'center' of the group's interaction network. This refers simultaneously to the frequency with which a member participates in interaction with other members and the number or range of other members with whom he or she interacts.

Observability is referring to a member's actual knowledge of group norms. Visibility refers to a structural property of groups in virtue of which participants are more or less likely to observe the actual norms of the group. Hopkins prefers to disregard a member's structurally given opportunity to know the group norms, but, instead, construes observability as the actual knowledge of group norms. In this way, a member of a group has a greater observability if his or her estimates of group norms and opinions are more accurate.

Conformity is the condition (or degree) of congruence between a member's profile on the relevant norms and the profile of the group-held norm. Conformity refers to a member's innate feelings on group norms, and a member need not necessarily act in accord with his own profile. Conformity should be distinguished from compliance, which does refer to the congruence between a member's actual behavior and group norms.

Influence is the effects of action on the group's normative consensus, relative to the impact of the other members' actions. That is to say, a member has a higher influence, over a given period of time, if the impact of his or her actions on group consensus during that period is greater relative to the other members.

The main theoretical part of Hopkins' book is in the fourth chapter (pages 50–98). Here, Hopkins lists 15 propositions (9 axioms and 6 derived statements). These propositions are about five status properties in groups: rank, centrality, observability, conformity, and influence. As Hopkins states: "The five status-properties—rank, centrality, observability, conformity, and influence—taken two at a time, can be combined to form twenty statements of simple implication" (page 50). Since Hopkins only regards fifteen of these as theoretically or empirically interesting, only these fifteen are discussed. Table 2.2 lists these propositions. Figure 2.1 depicts the nine axioms (Propositions 1, 2, 4, 6–9, 12, and 14) graphically.

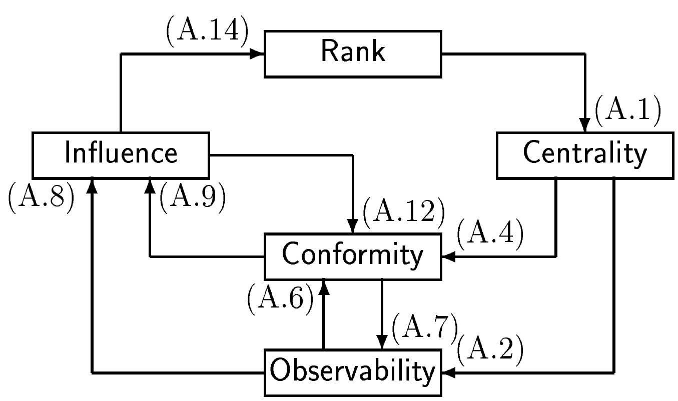

Figure 2.1: Hopkins' axioms (based on page 52).

2.3.2 Formalization

We use first-order logic to formalize Hopkins' propositions. In the formalization we use five unary functions to represent the status properties and the 'greater than' relation (see Table 2.3).

For any member of a small group:	
1.	The higher his rank, the greater his centrality.
2.	The greater his centrality, the greater his observability.
3.	The higher his rank, the greater his observability (derivation).
4.	The greater his centrality, the greater his conformity.
5.	The higher his rank, the greater his conformity (derivation).
6.	The greater his observability, the greater his conformity.
7.	The greater his conformity, the greater his observability.
8.	The greater his observability, the greater his influence.
9.	The greater his conformity, the greater his influence.
10.	The greater his centrality, the greater his influence (derivation).
11.	The greater his influence, the greater his observability (derivation).
12.	The greater his influence, the greater his conformity.
13.	The higher his rank, the greater his influence (derivation).
14.	The greater his influence, the higher his rank.
15.	The greater his centrality, the higher his rank (derivation).

Note: All propositions are from (Hopkins 1964, p.51).

Table 2.2: The propositions of Hopkins' Theory.

$\mathsf{rank}(x)$	the rank of x.
$\mathsf{cent}(x)$	the centrality of x.
$\mathsf{obse}(x)$	the observability of x.
$\mathsf{conf}(x)$	the conformity of x.
$\mathsf{infl}(x)$	the influence of x.
$(x > y)$	x is greater/higher than y.

Table 2.3: Functions and relations used in the formalization.

Using these functions and relations, a proposition like **1**:

The higher his rank, the greater his centrality.

can be formalized in first-order logic as

$$\forall x, y\ [\mathsf{rank}(x) > \mathsf{rank}(y)\ \rightarrow\ \mathsf{cent}(x) > \mathsf{cent}(y)]$$

This translation is, arguably, a natural translation of such a ordinary language proposition into first-order logic.

Hopkins regards Propositions 1, 2, 4, 6–9, 12, and 14 as axioms of his theory. Table 2.4 gives a formalization of these propositions as suggested above.

A.1	$\forall x, y\ [\mathsf{rank}(x) > \mathsf{rank}(y)\ \rightarrow\ \mathsf{cent}(x) > \mathsf{cent}(y)]$
A.2	$\forall x, y\ [\mathsf{cent}(x) > \mathsf{cent}(y)\ \rightarrow\ \mathsf{obse}(x) > \mathsf{obse}(y)]$
A.4	$\forall x, y\ [\mathsf{cent}(x) > \mathsf{cent}(y)\ \rightarrow\ \mathsf{conf}(x) > \mathsf{conf}(y)]$
A.6	$\forall x, y\ [\mathsf{obse}(x) > \mathsf{obse}(y)\ \rightarrow\ \mathsf{conf}(x) > \mathsf{conf}(y)]$
A.7	$\forall x, y\ [\mathsf{conf}(x) > \mathsf{conf}(y)\ \rightarrow\ \mathsf{obse}(x) > \mathsf{obse}(y)]$
A.8	$\forall x, y\ [\mathsf{obse}(x) > \mathsf{obse}(y)\ \rightarrow\ \mathsf{infl}(x) > \mathsf{infl}(y)]$
A.9	$\forall x, y\ [\mathsf{conf}(x) > \mathsf{conf}(y)\ \rightarrow\ \mathsf{infl}(x) > \mathsf{infl}(y)]$
A.12	$\forall x, y\ [\mathsf{infl}(x) > \mathsf{infl}(y)\ \rightarrow\ \mathsf{conf}(x) > \mathsf{conf}(y)]$
A.14	$\forall x, y\ [\mathsf{infl}(x) > \mathsf{infl}(y)\ \rightarrow\ \mathsf{rank}(x) > \mathsf{rank}(y)]$

Table 2.4: A formal version of Hopkins' axioms.

2.3.3 Soundness of Argumentation

Hopkins claims that Propositions 3, 5, 10, 11, 13, and 15 are derivable from the axioms of his theory. Table 2.5 gives a formalization of these propositions. Hopkins (1964, pp.97–98) also indicates which propositions are used in his derivations: Proposition 3 is derived from 1 and 2; Proposition 5 from 1 and 4; Proposition 10 from 2 and 8, or from 4 and 9; Proposition 11 from 12 and 7; Proposition 13 from 1, 4 and 9; and Proposition 15 from 2, 9 and 14, or from 2, 8 and 14.

T.3	$\forall x, y\ [\mathsf{rank}(x) > \mathsf{rank}(y)\ \rightarrow\ \mathsf{obse}(x) > \mathsf{obse}(y)]$
T.5	$\forall x, y\ [\mathsf{rank}(x) > \mathsf{rank}(y)\ \rightarrow\ \mathsf{conf}(x) > \mathsf{conf}(y)]$
T.10	$\forall x, y\ [\mathsf{cent}(x) > \mathsf{cent}(y)\ \rightarrow\ \mathsf{infl}(x) > \mathsf{infl}(y)]$
T.11	$\forall x, y\ [\mathsf{infl}(x) > \mathsf{infl}(y)\ \rightarrow\ \mathsf{obse}(x) > \mathsf{obse}(y)]$
T.13	$\forall x, y\ [\mathsf{rank}(x) > \mathsf{rank}(y)\ \rightarrow\ \mathsf{infl}(x) > \mathsf{infl}(y)]$
T.15	$\forall x, y\ [\mathsf{cent}(x) > \mathsf{cent}(y)\ \rightarrow\ \mathsf{rank}(x) > \mathsf{rank}(y)]$

Table 2.5: A formal version of Hopkins' theorems.

The first derived proposition is number 3. Hopkins claims that "This third proposition derives from the preceding two and its rationale is essentially that of its premises" (page 59). Unfortunately, Hopkins does not further specify his intuitive notion of 'derivability.' Of course, if this proposition is indeed derivable from the two other propositions or axioms, it is no great surprise that the empirical rationale for it can be traced back to the two axioms. Formal logic comes with a strict notion of inference that allows us to verify the derivations in our formal version of the theory.

We can use OTTER to test whether the suggested theorems can be soundly derived from the axioms. We can test whether **T.3** is derivable from the axioms using the input-file (without loss of generality, we use only **A.1** and **A.2**):

```
set(auto).
formula_list(usable).
  % A.1
  all x y (rank(x)>rank(y) -> cent(x)>cent(y)).
  % A.2
  all x y (cent(x)>cent(y) -> obse(x)>obse(y)).
  % negation of T.3
  -( all x y (rank(x)>rank(y) -> obse(x)>obse(y)) ).
end_of_list.
```

OTTER's resolution-style proofs require the conjecture or theorem to be negated in the input-file. If now the theorem (in its original form) is indeed derivable, we will be able to derive a contradiction (proving the theorem by *reductio ad absurdum*). As it turns out, OTTER can derive **T.3** using **A.1** and **A.2**.

Theorem 3 The higher his rank, the greater his observability.

$$\forall x, y\ [\mathsf{rank}(x) > \mathsf{rank}(y)\ \rightarrow\ \mathsf{obse}(x) > \mathsf{obse}(y)]$$

Proof: OTTER can derive **T.3** from **A.1** and **A.2**.[5]

In a similar way, we can try to derive the other theorems.

[5] OTTER produces the following proof:

```
1 [] -(rank(x)>rank(y))|cent(x)>cent(y).
2 [] -(cent(x)>cent(y))|obse(x)>obse(y).
3 [] -(obse($c2)>obse($c1)).
4 [] rank($c2)>rank($c1).
5 [hyper,4,1] cent($c2)>cent($c1).
6 [hyper,5,2] obse($c2)>obse($c1).
7 [binary,6.1,3.1] $F.
```

By two applications of hyper-resolution, and one final application of binary resolution, see (Wos et al. 1992) for details on these inference rules.

Theorem 5 The higher his rank, the greater his conformity.

$$\forall x, y\ [\mathsf{rank}(x) > \mathsf{rank}(y) \ \rightarrow\ \mathsf{conf}(x) > \mathsf{conf}(y)]$$

Proof: OTTER can derive **T.5** from **A.1** and **A.4**.

Theorem 10 The greater his centrality, the greater his influence.

$$\forall x, y\ [\mathsf{cent}(x) > \mathsf{cent}(y) \ \rightarrow\ \mathsf{infl}(x) > \mathsf{infl}(y)]$$

Proof: OTTER can derive **T.10** from **A.2** and **A.8**.

Theorem 11 The greater his influence, the greater his observability.

$$\forall x, y\ [\mathsf{infl}(x) > \mathsf{infl}(y) \ \rightarrow\ \mathsf{obse}(x) > \mathsf{obse}(y)]$$

Proof: OTTER can derive **T.11** from **A.12** and **A.7**.

Theorem 13 The higher his rank, the greater his influence.

$$\forall x, y\ [\mathsf{rank}(x) > \mathsf{rank}(y) \ \rightarrow\ \mathsf{infl}(x) > \mathsf{infl}(y)]$$

Proof: OTTER can derive **T.13** from **A.1**, **A.4**, and **A.9**.

Theorem 15 The greater his centrality, the greater his rank.

$$\forall x, y\ [\mathsf{cent}(x) > \mathsf{cent}(y) \ \rightarrow\ \mathsf{rank}(x) > \mathsf{rank}(y)]$$

Proof: OTTER can derive **T.15** from **A.2**, **A.8**, and **A.14**.

The theoretical argumentation of Hopkins theory is strictly intuitive. He just postulates that propositions are derivable, without specifying the used rules of inference.[6] We have presented a first-order logic rendition of his propositions and tested whether the intuitive arguments would correspond to rigorous proofs in formal logic. As it turns out, we can formally prove all the propositions of which the derivability was claimed. Moreover, the proofs we find correspond exactly with the suggested inferences—confirming his intuitive arguments.

[6] Although, on one occasion, he states that "Proposition 10 derives *logically* from Proposition 2 and 8, or from 4 and 9" (page 74, emphasis added).

2.3.4 Consistency of the Theory

The fact that we can prove all the suggested theorems provides us with some confidence in our particular formalization of the axioms and theorems. We can further investigate our theory by testing whether it is non-contradictory or consistent. Consistency is a crucial property of formal theories, because any statement is derivable from an inconsistent axiom set.[7] We can prove the consistency of the (formal) theory by generating a model of it.

We can use MACE in an attempt to find models of the set of axioms. We can test whether there are such models using the input-file:

```
set(auto).
formula_list(usable).
  all x y (rank(x)>rank(y) -> cent(x)>cent(y)).   % A.1
  all x y (cent(x)>cent(y) -> obse(x)>obse(y)).   % A.2
  all x y (cent(x)>cent(y) -> conf(x)>conf(y)).   % A.4
  all x y (obse(x)>obse(y) -> conf(x)>conf(y)).   % A.6
  all x y (conf(x)>conf(y) -> obse(x)>obse(y)).   % A.7
  all x y (obse(x)>obse(y) -> infl(x)>infl(y)).   % A.8
  all x y (conf(x)>conf(y) -> infl(x)>infl(y)).   % A.9
  all x y (infl(x)>infl(y) -> conf(x)>conf(y)).   % A.12
  all x y (infl(x)>infl(y) -> rank(x)>rank(y)).   % A.14
end_of_list.
```

There exist many models of this axiom set, even models of small cardinalities. If we use $\{0,1\}$ as the universe, then MACE generates over a thousand models in just 2 seconds.[8] Table 2.6 shows one of these models. It is easy to verify

	rank	cent	obse	conf	infl
0	0	0	0	0	0
1	1	1	1	1	1

>	0	1
0	F	F
1	T	F

Table 2.6: A model of Hopkins' axioms.

that the axioms hold in this model, which represents a group with only two members. Proposition 1, for example, is true because the member with the highest rank, which is member 1, also has the highest centrality. In this model, we

[7] Classical logic has the principle *ex falso sequitur quodlibet* that says that everything can be concluded from a contradiction.

[8] For example, by invoking MACE with the options '`-n2 -p -m5000`', see (McCune 1994a) for details. MACE generates 1024 models on $\{0,1\}$, we present here one that gives a natural interpretation of the '>' relation. A formal model consists of a universe (here $\{0,1\}$) and a mapping between the non-logical symbols (here a binary relation $>$ and unary functions rank, cent, obse, conf, and infl) and elements of the universe. For example, consider the model in Table 2.6: here, the relation symbol $>$ is interpreted as $(1 > 0)$ is true, and $(0 > 0)$, $(0 > 1)$, and $(1 > 1)$ are false and the function rank is interpreted as $\mathsf{rank}(0) = 0$ and $\mathsf{rank}(1) = 1$.

have one member, 1, with higher rank, centrality, observability, conformity, and influence than another member, 0,—a ‘prototypical’ model of this simple theory. Statements that are derivable from the axioms are necessarily true in all their models. In particular, the theorems we derived earlier are also true in the model of Table 2.6. For example, **T.3** is true because the member with the higher rank (i.e., member 1) has also the higher observability.

Finding an arbitrary model of the axiom set proves that it is consistent. That is, even the single, simple model in Table 2.6 proves that the set of axioms, and therefore the theory, is consistent.

2.3.5 Satisfiability, Falsifiability, and Contingence

We can also try to investigate whether the theorems are falsifiable, i.e., whether it is (logically) possible that the theorem is false. For example, if a statement is tautologically true, its truth is determined by logic and the statement does not make an empirical claim. As a result, empirical testing can only reaffirm its trivial validity. We can also check whether theorems are falsifiable by constructing a model in which the theorem does not hold. Notice that such a model cannot be model of the theory, since a theorem is necessarily true in all models of the premises—otherwise it would not be a theorem. Therefore, we have to ignore the premises of the theory.

Consider theorem 3 *the higher his rank, the greater his observability.*

$$\forall x, y\ [\mathsf{rank}(x) > \mathsf{rank}(y)\ \rightarrow\ \mathsf{obse}(x) > \mathsf{obse}(y)]$$

As it turns out, we can find a model in which the theorem, **T.3**, is false. Even on a small universe like $\{0, 1\}$, MACE generates several of such models (one of which is shown in Table 2.7). Finding this model proves that **T.3** is falsifiable. Notice that this is a model does not respect, for example, the premise **A.1**. It is a model of the language but not a model of the theory (otherwise it would be a counterexample).

	rank	cent	obse	conf	infl
0	0	0	1	0	0
1	1	0	0	0	0

>	0	1
0	F	F
1	T	F

Table 2.7: A model in which **T.3** is false.

In a similar vein, we can also investigate whether the theorems are satisfiable, i.e., whether it is (logically) possible that the theorem is true. For example, if a statement is a contradiction, its truth (or rather its falsity) is also determined by logic. As a result, such a statement does also not make an empirical claim,

	rank	cent	obse	conf	infl
0	0	0	0	1	0
1	1	0	0	0	0

$>$	0	1
0	F	F
1	T	F

Table 2.8: A model in which **T.5** is false.

	rank	cent	obse	conf	infl
0	0	0	0	0	1
1	0	1	0	0	0

$>$	0	1
0	F	F
1	T	F

Table 2.9: A model in which **T.10** is false.

	rank	cent	obse	conf	infl
0	0	0	1	0	0
1	0	0	0	0	1

$>$	0	1
0	F	F
1	T	F

Table 2.10: A model in which **T.11** is false.

	rank	cent	obse	conf	infl
0	0	0	0	0	1
1	1	0	0	0	0

$>$	0	1
0	F	F
1	T	F

Table 2.11: A model in which **T.13** is false.

	rank	cent	obse	conf	infl
0	1	0	0	0	0
1	0	1	0	0	0

$>$	0	1
0	F	F
1	T	F

Table 2.12: A model in which **T.15** is false.

and empirical testing can only reaffirm its trivial falsity. We can check whether a statement is satisfiable by constructing a model in which it does hold. As noted above, theorems are necessarily true in all models of the axiom set. That is, theorem **T.3** is true in the model of Table 2.6, proving that the theorem is satisfiable.

The two models of Tables 2.6 and 2.7 prove that **T.3** is both satisfiable and falsifiable. Theorem **T.3** is a contingent statement—its truth or falsity is not predetermined by the logic, but depends on the validity of the axioms.

We can repeat this for all other theorems (and for the axioms themselves). Of course, all theorems hold in the model of Table 2.6, i.e., this model proves that all theorems are satisfiable. We also succeed in proving the falsifiability of all theorems: the models in Tables 2.8–2.12 prove the falsifiability of **T.5**, **T.10**, **T.11**, **T.13** and **T.15** respectively. As a result, all theorems are satisfiable and falsifiable, and therefore contingent.

2.3.6 Examining the Theorems

Hopkins (1964, p.50) states: "The five status-properties—rank, centrality, observability, conformity, and influence—taken two at a time, can be combined to form twenty statements of simple implication." However, the text only explicitly lists the fifteen statements reprinted in Table 2.2. These five missing pairs of status properties are: conformity-rank, observability-rank, influence-centrality, conformity-centrality, and observability-centrality. The corresponding 'statements of simple implication' are, presumably, those listed in Table 2.13.

For any member of a small group:	
16.	The greater his conformity, the higher his rank.
17.	The greater his observability, the higher his rank.
18.	The greater his influence, the greater his centrality.
19.	The greater his conformity, the greater his centrality.
20.	The greater his observability, the greater his centrality.

Table 2.13: The 'missing' propositions of Hopkins.

We want to investigate the relation between the theory and the missing propositions. Hopkins does not further discuss these propositions, although he states: "Six of the possible derivations from these basic propositions are discussed because of their theoretical or empirical interest, and the other five are not discussed because they are not particularly interesting" (page 52). This suggests that the missing propositions are derivable from the set of axioms. As it turns out, they are all derivable from the axioms.

Theorem 16 The greater his conformity, the higher his rank.

$$\forall x, y\ [\mathsf{conf}(x) > \mathsf{conf}(y) \ \rightarrow\ \mathsf{rank}(x) > \mathsf{rank}(y)]$$

Proof: OTTER can derive **T.16** from **A.9** and **A.14**.

Theorem 17 The greater his observability, the higher his rank.

$$\forall x, y\ [\mathsf{obse}(x) > \mathsf{obse}(y) \ \rightarrow\ \mathsf{rank}(x) > \mathsf{rank}(y)]$$

Proof: OTTER can derive **T.17** from **A.8** and **A.14**.

Theorem 18 The greater his influence, the greater his centrality.

$$\forall x, y\ [\mathsf{infl}(x) > \mathsf{infl}(y) \ \rightarrow\ \mathsf{cent}(x) > \mathsf{cent}(y)]$$

Proof: OTTER can derive **T.18** from **A.14** and **A.1**.

Theorem 19 The greater his conformity, the greater his centrality.

$$\forall x, y\ [\mathsf{conf}(x) > \mathsf{conf}(y) \ \rightarrow\ \mathsf{cent}(x) > \mathsf{cent}(y)]$$

Proof: OTTER can derive **T.19** from **A.9**, **A.14**, and **A.1**.

Theorem 20 The greater his observability, the greater his centrality.

$$\forall x, y\ [\mathsf{obse}(x) > \mathsf{obse}(y) \ \rightarrow\ \mathsf{cent}(x) > \mathsf{cent}(y)]$$

Proof: OTTER can derive **T.20** from **A.8**, **A.14**, and **A.1**.

The five missing propositions are all derivable from the set of axioms. As a result, the formulas in Table 2.14 are further theorems of the theory. A theory

T.16	$\forall x, y\ [\mathsf{conf}(x) > \mathsf{conf}(y) \ \rightarrow\ \mathsf{rank}(x) > \mathsf{rank}(y)]$
T.17	$\forall x, y\ [\mathsf{obse}(x) > \mathsf{obse}(y) \ \rightarrow\ \mathsf{rank}(x) > \mathsf{rank}(y)]$
T.18	$\forall x, y\ [\mathsf{infl}(x) > \mathsf{infl}(y) \ \rightarrow\ \mathsf{cent}(x) > \mathsf{cent}(y)]$
T.19	$\forall x, y\ [\mathsf{conf}(x) > \mathsf{conf}(y) \ \rightarrow\ \mathsf{cent}(x) > \mathsf{cent}(y)]$
T.20	$\forall x, y\ [\mathsf{obse}(x) > \mathsf{obse}(y) \ \rightarrow\ \mathsf{cent}(x) > \mathsf{cent}(y)]$

Table 2.14: Further theorems.

consists of all the statements that are logical consequences of its set of axioms (including the axioms themselves, which are trivial consequences of themselves). If we accept the axioms of a theory, then we must also accept all statements that are derivable from the premises, i.e., all statements that are necessarily true if the premises are true. A theorizer is not allowed to include only some of the theorems in the theory, and not to include other theorems (unless the premises are modified accordingly). Of course, in any exposition of a theory, we will only explicitly spell out those theorems that are the most salient consequences of our theory. Since Hopkins assured us that "they are not particularly interesting," they need not be explicitly spelled out in the text.[9] Nevertheless, they are theorems of the set of axioms, and therefore part of Hopkins' theory.

2.3.7 Examining the Axioms

The theory uses the nine axioms of Table 2.4. Rather surprisingly, Hopkins states: "It has perhaps been noticed that there are nine basic propositions, whereas the number logically required to generate the statements is less" (pages 51–52). Apparently, the theory contains redundant axioms, that is, the same set of theorems can be derived from a subset of axioms. We can distinguish between two ways in which axioms may be redundant: 1) an axiom may turn out to be unnecessary for deriving a given set of theorems, or 2) an axiom may turn out to be derivable from the other axioms. In the second case, we make the axiom set more parsimonious by making the derivable axiom into a theorem.[10] In the first case, we may even relax the axiom altogether, making the theory more general. Hopkins' explanation for the use of redundant axioms is even more surprising: "This is because in the theory proper there are nine substantive important relations among the properties" (page 52). It is unclear why he does not want to have theorems that are "substantive important relations among the properties." First, having substantive important theorems is generally regarded as the very purpose of a theory. Second, there will be substantive important theorems anyway, because the axioms are, of course, also theorems themselves since each of them can be trivially derived from the set of axioms.

The first type of superfluous axioms we want to consider are those axioms that are not used for deriving the singled-out theorems. We will first examine the proofs of the theorems as stated above:

- Axiom **A.1** is used in derivation of theorems **T.3**, **T.5**, and **T.13**;
- Axiom **A.2** is used in derivation of theorems **T.3**, **T.10**, and **T.15**;

[9] Of course, if they were undesirable or unwanted consequences they should be spelled out. Note that each of the 'missing' propositions is the converse of one that is discussed in the text (**T.16** is the converse of **A.5**; **T.17** of **T.3**; **T.18** of **T.10**; **T.19** of **A.4**; and **T.20** of **A.2**). It is unclear how the converse of a statement of "theoretical or empirical interest" can be "not particularly interesting."

[10] As Tarski (1946, p.131) puts it: "Fundamentally, we strive to arrive at an axiom system which does not contain a single superfluous statement, that is, a statement that can be derived from the remaining axioms and which, therefore, might be counted among the theorems of the theory under construction."

- Axiom **A.4** is used in derivation of theorem **T.5**;
- Axiom **A.7** is used in derivation of theorem **T.11**;
- Axiom **A.8** is used in derivation of theorems **T.10** and **T.15**;
- Axiom **A.9** is used in derivation of theorem **T.13**;
- Axiom **A.12** is used in derivation of theorem **T.11**; and
- Axiom **A.14** is used in derivation of theorem **T.15**.

The remaining axiom, **A.6**, is not used in any of the proofs.[11] This is also the case in the informal derivations in which Proposition 6 is never used (Hopkins 1964, pp.97–98)—Is this the redundancy that Hopkins is referring to? We can remove axiom **A.6** from the axiom set and still have the same set of singled-out theorems. It is important to note that changing the axiom set does, in general, change the theory. Although the singled-out theorems remain derivable there may be other theorems that cease to belong to the theory. For example, if we decide to remove **A.6** from the axioms, then the statement of **A.6**—which was a substantive important relation—may no longer belong to the theory. Unless of course **A.6** is also a case of the second type of redundancy, i.e., that **A.6** can be derived from the remaining set of axioms. This turns out to be the case.

Axiom 6 The greater his observability, the greater his conformity.

$$\forall x, y\ [\mathsf{obse}(x) > \mathsf{obse}(y)\ \rightarrow\ \mathsf{conf}(x) > \mathsf{conf}(y)]$$

Proof: OTTER can derive **A.6** from **A.8** and **A.12**. A different proof uses **A.8**, **A.14**, **A.1**, and **A.4**.[12]

We can derive **A.6** from the other axioms. As a result, if we decide to remove **A.6** from the axioms set, then the theory does not change. That is to say the same set of statements will be derivable from the smaller set of axioms, in particular, the statement of **A.6** will now become a theorem. This proves that axiom **A.6** is not independent of the other axioms. The other axioms were used for deriving some of the axioms, nevertheless, they might still be redundant if they can be derived from some of the other axioms.

We can try to prove whether the other axioms are independent. Let us consider axiom **A.1**. If we want to prove that **A.1** is independent of the other axioms, we have to show that the other axioms do not imply the truth (nor the falsity) of **A.1**. That is to say, we have to prove that both

[11] This does, of course, not prove that the other axioms are necessary, since there may be different proofs of the theorems that use even less of the axioms.

[12] We can instruct OTTER not to terminate after finding a first proof, but to continue searching for other proofs by setting the '`max_proofs`' parameter, see (McCune 1994b) for details.

1. $\{\mathbf{A.2}, \mathbf{A.4}, \mathbf{A.6}, \mathbf{A.7}, \mathbf{A.8}, \mathbf{A.9}, \mathbf{A.12}, \mathbf{A.14}\} \not\vdash \mathbf{A.1}$ and that

2. $\{\mathbf{A.2}, \mathbf{A.4}, \mathbf{A.6}, \mathbf{A.7}, \mathbf{A.8}, \mathbf{A.9}, \mathbf{A.12}, \mathbf{A.14}\} \not\vdash \overline{\mathbf{A.1}}$.[13]

Proving underivability is, in general, very hard. We will here try to find small counterexamples for these derivability relations. We can prove (2) by finding a model of the other axioms in which **A.1** is true. This amounts to finding a model of the total set of axioms. We already did this in Table 2.6 when we proved consistency. Similarly, we can prove (1) if we can find a model of the other axioms in which **A.1** is false. MACE generated the model with universe $\{0, 1\}$ reprinted in Table 2.15. Finding these two models proves that the axiom **A.1** is

	rank	cent	obse	conf	infl
0	0	0	0	0	0
1	1	0	0	0	0

>	0	1
0	F	F
1	T	F

Table 2.15: A model in which **A.1** is false, and the other axioms hold.

independent of the other axioms—neither its truth, nor its falsity is implied by the other axioms.

Next, we consider **A.2**. Again, the model in Table 2.6 proves that the falsity of **A.2** is not implied by the other axioms. However, we do not manage to find a small model of the other axioms in which **A.2** is false. Maybe **A.2** is not independent of the other axioms? As it turns out, we can derive **A.2** from the other axioms.

Axiom 2 The greater his centrality, the greater his observability.

$$\forall x, y\ [\mathsf{cent}(x) > \mathsf{cent}(y)\ \rightarrow\ \mathsf{obse}(x) > \mathsf{obse}(y)]$$

Proof: OTTER can derive **A.2** from **A.4** and **A.7**.

This proves that axiom **A.2** is not independent of the other axioms. We can also prove that axioms **A.4**, **A.7**, **A.8**, **A.9**, and **A.12** are not independent.

Axiom 4 The greater his centrality, the greater his conformity.

$$\forall x, y\ [\mathsf{cent}(x) > \mathsf{cent}(y)\ \rightarrow\ \mathsf{conf}(x) > \mathsf{conf}(y)]$$

Proof: OTTER can derive **A.4** from **A.2** and **A.6**. A different proof uses **A.2**, **A.8**, and **A.12**.

[13] We use here $\overline{\cdots}$ to indicate the negation of the formula, i.e., if **A.1** is a formula φ then $\overline{\mathbf{A.1}}$ represents the formula $\neg(\varphi)$.

Axiom 7 The greater his conformity, the greater his observability.

$$\forall x, y\ [\mathsf{conf}(x) > \mathsf{conf}(y) \ \rightarrow\ \mathsf{obse}(x) > \mathsf{obse}(y)]$$

Proof: OTTER can derive **A.7** from **A.9**, **A.14**, **A.1**, and **A.2**.

Axiom 8 The greater his observability, the greater his influence.

$$\forall x, y\ [\mathsf{obse}(x) > \mathsf{obse}(y) \ \rightarrow\ \mathsf{infl}(x) > \mathsf{infl}(y)]$$

Proof: OTTER can derive **A.8** from **A.6** and **A.9**.

Axiom 9 The greater his conformity, the greater his influence.

$$\forall x, y\ [\mathsf{conf}(x) > \mathsf{conf}(y) \ \rightarrow\ \mathsf{infl}(x) > \mathsf{infl}(y)]$$

Proof: OTTER can derive **A.9** from **A.7** and **A.8**.

Axiom 12 The greater his influence, the greater his conformity.

$$\forall x, y\ [\mathsf{infl}(x) > \mathsf{infl}(y) \ \rightarrow\ \mathsf{conf}(x) > \mathsf{conf}(y)]$$

Proof: OTTER can derive **A.12** from **A.14**, **A.1**, and **A.4**. A different proof uses **A.14**, **A.1**, **A.2**, and **A.6**.

Finally, let us consider axiom **A.14**. The model in Table 2.6 still proves that the falsity of **A.14** is not implied by the other axioms. MACE generated the model in Table 2.16 (having universe $\{0, 1\}$). This model proves that the truth of

	rank	cent	obse	conf	infl
0	0	0	0	0	0
1	0	0	1	1	1

>	0	1
0	F	F
1	T	F

Table 2.16: A model in which **A.14** is false, and the other axioms hold.

A.14 is not implied by the other axioms. Together with the model in Table 2.6, this proves that axiom **A.14** is independent of the other axioms.

We have proved that the axioms of the theory are not independent—several of the axioms can be derived from the other axioms. We have proven that **A.1** and **A.14** are independent of the other axioms, that is, a smaller axiom set must include these two axioms.

2.3.8 Minimal Sets of Axioms

In the previous section, we proved that seven of the nine axioms are not independent of the other axioms. In this section, we will try to prune the axioms. That is, we will try to make the axiom set more parsimonious by removing superfluous axioms. Each of these seven axioms can be removed from the premise set, since each of them can be derived from the other axioms. Of course, this does not imply that we can remove all seven axioms together, since the removal of one axiom may make another axioms independent relative to the smaller set of axioms. That is, if we remove one of the derivable axioms, we have to test again whether there are still non-independent axioms. If there are none, then the axiom set is minimal, otherwise, we can further reduce the set of axioms and test again for non-independence.

A minimal subset of these nine axioms is the set of five axioms **A.1**, **A.4**, **A.7**, **A.8**, and **A.14** (see Table 2.17). We can prove that this is an axiom set

A.1	$\forall x, y\ [\mathsf{rank}(x) > \mathsf{rank}(y)\ \rightarrow\ \mathsf{cent}(x) > \mathsf{cent}(y)]$
A.4	$\forall x, y\ [\mathsf{cent}(x) > \mathsf{cent}(y)\ \rightarrow\ \mathsf{conf}(x) > \mathsf{conf}(y)]$
A.7	$\forall x, y\ [\mathsf{conf}(x) > \mathsf{conf}(y)\ \rightarrow\ \mathsf{obse}(x) > \mathsf{obse}(y)]$
A.8	$\forall x, y\ [\mathsf{obse}(x) > \mathsf{obse}(y)\ \rightarrow\ \mathsf{infl}(x) > \mathsf{infl}(y)]$
A.14	$\forall x, y\ [\mathsf{infl}(x) > \mathsf{infl}(y)\ \rightarrow\ \mathsf{rank}(x) > \mathsf{rank}(y)]$

Table 2.17: A minimal axiom set for Hopkins' theory.

that will generate all propositions (i.e., all the theorems and other axioms) by proving the 'missing' axioms from this new, smaller, set of axioms. We can use the proofs outlined in the previous section:

$$\begin{aligned}
\{\mathbf{A.4}, \mathbf{A.7}\} &\ \vdash\ \mathbf{A.2} \\
\{\mathbf{A.1}, \mathbf{A.4}, \mathbf{A.8}, \mathbf{A.14}\} &\ \vdash\ \mathbf{A.6} \\
\{\mathbf{A.7}, \mathbf{A.8}\} &\ \vdash\ \mathbf{A.9} \\
\{\mathbf{A.1}, \mathbf{A.4}, \mathbf{A.14}\} &\ \vdash\ \mathbf{A.12}
\end{aligned}$$

These four derivable axioms **A.2**, **A.6**, **A.9**, and **A.12**, which were all "substantive important relations among the properties" (page 52), thus become further theorems of the theory. We can prove that the smaller axiom set, consisting of **A.1**, **A.4**, **A.7**, **A.8**, and **A.14**, is minimal by proving that all axioms in this smaller set are independent of the other axioms in this set. We did this above for axioms **A.1** and **A.14**. We now also succeed in proving the independence of the other three axioms (relative to this smaller set of five axioms, see the models in Tables 2.18–2.20).

Another minimal subset of these nine axioms is the set of five axioms **A.1**, **A.2**, **A.6**, **A.9**, and **A.14** (see Table 2.21). Again, this is an axiom set of the

	rank	cent	obse	conf	infl
0	0	0	0	0	0
1	0	1	0	0	0

$>$	0	1
0	F	F
1	T	F

Table 2.18: A model in which **A.4** is false, and the axioms **A.1**, **A.7**, **A.8**, and **A.14** hold.

	rank	cent	obse	conf	infl
0	0	0	0	0	0
1	0	0	0	1	0

$>$	0	1
0	F	F
1	T	F

Table 2.19: A model in which **A.7** is false, and the axioms **A.1**, **A.4**, **A.8**, and **A.14** hold.

	rank	cent	obse	conf	infl
0	0	0	0	0	0
1	0	0	1	0	0

$>$	0	1
0	F	F
1	T	F

Table 2.20: A model in which **A.8** is false, and the axioms **A.1**, **A.4**, **A.7**, and **A.14** hold.

A.1	$\forall x, y\ [\mathsf{rank}(x) > \mathsf{rank}(y) \ \rightarrow\ \mathsf{cent}(x) > \mathsf{cent}(y)]$
A.2	$\forall x, y\ [\mathsf{cent}(x) > \mathsf{cent}(y) \ \rightarrow\ \mathsf{obse}(x) > \mathsf{obse}(y)]$
A.6	$\forall x, y\ [\mathsf{obse}(x) > \mathsf{obse}(y) \ \rightarrow\ \mathsf{conf}(x) > \mathsf{conf}(y)]$
A.9	$\forall x, y\ [\mathsf{conf}(x) > \mathsf{conf}(y) \ \rightarrow\ \mathsf{infl}(x) > \mathsf{infl}(y)]$
A.14	$\forall x, y\ [\mathsf{infl}(x) > \mathsf{infl}(y) \ \rightarrow\ \mathsf{rank}(x) > \mathsf{rank}(y)]$

Table 2.21: A second minimal axiom set for Hopkins' theory.

theory because:

$$\begin{aligned}\{\mathbf{A.2}, \mathbf{A.6}\} &\vdash \mathbf{A.4}\\ \{\mathbf{A.1}, \mathbf{A.2}, \mathbf{A.9}, \mathbf{A.14}\} &\vdash \mathbf{A.7}\\ \{\mathbf{A.6}, \mathbf{A.9}\} &\vdash \mathbf{A.8}\\ \{\mathbf{A.1}, \mathbf{A.2}, \mathbf{A.6}, \mathbf{A.14}\} &\vdash \mathbf{A.12}\end{aligned}$$

It is a minimal axiom set, because we can prove it is an independent set of axioms (models not reproduced here).

These two axiom sets

$$\{\mathbf{A.1}, \mathbf{A.4}, \mathbf{A.7}, \mathbf{A.8}, \mathbf{A.14}\}$$

$$\{\mathbf{A.1}, \mathbf{A.2}, \mathbf{A.6}, \mathbf{A.9}, \mathbf{A.14}\}$$

have cardinality five and all the other propositions (both theorems and the removed axioms) can be derived from them. There are other minimal axioms sets, but these two are the smallest subsets of the listed axioms that generate all twenty propositions.[14]

2.3.9 Recapitulating

In this section, we presented a formal version Hopkins' theory of 'The Exercise of Influence in Small Groups' by representing his propositions in first-order logic. We evaluated this first-order logic rendition using the logical criteria. In this way, we could prove several properties of the theory. First, we proved that the informal argumentation of Hopkins corresponds to rigorous proofs of theorems in the formal rendition of the theory. Second, we proved that the (formal) theory is non-trivial in the sense that it is consistent or contradiction-free. Third, we proved that the theorems are logically contingent, that is, the claims they make are are both logically falsifiable and satisfiable. Fourth, we proved that the five 'missing' propositions are also derivable from the same axiom set—i.e., they are also theorems of Hopkins' theory. Fifth, we proved that the suggested axioms of the theory are not independent. And sixth, we provided two subsets of the axioms that are independent. Notice that, although we made various changes in the exposition of the theory, the theory itself did not change. The initial formal version of the theory and the final axioms sets can be used to generate precisely the same class of theorems and the same class of models.

[14] All propositions relate two different status properties, therefore, we need at least four axioms to make a coherent theory (if we use less than four axioms, the theory will consist of independent parts). If we want to derive all twenty propositions, we need at least one further axiom that relates the fifth status property back to the first.

2.4 Discussion and Conclusions

In this chapter we discussed formal theory building based on the use of standard first-order logic, and of existing automated reasoning tools. The logic provides us with a number of criteria that can be tested for using computational tools. In principle, each criterion can be tested for by both theorem proving and model generation strategies, for example, a theorem is also sound if it holds in all models of the premise set, or a theory is consistent if the deductive closure of the premise set does not contain a contradiction. In practice, the use of automated theorem provers and model generators is complementary: generating all models or the complete deductive closure of a premise set is impossible. A theorem prover is suitable for proving the inconsistency of the theory, or the soundness of a derivation (requiring only a single proof), and a model generator can prove the consistency of the theory, or the unsoundness of a conjecture (requiring only a single model). In short, much is to be gained by using the right tool for the specific proof/disproof attempt at hand, and even more than just computational differences. Consider, for example, a situation in which the prover fails to prove a conjecture. Determining what caused this failure typically requires a thorough examination of the search-traces—an arduous, time-consuming activity. If the model generator can construct a counterexample to the conjecture, it will become apparent immediately why the proof attempt failed.

As always, there are principal and practical limitations to use of automated reasoning tools: first-order logic is not decidable (although it is semi-decidable: it may detect a consequent eventually); current automated model generators can only find finite models (even only very small ones, cardinalities beyond a dozen seem impractical); and the common practical limitations such as memory, CPU, time.[15] However, none of the proofs and models searches for the case study in section 2.3 requires more than five seconds. Admittedly, this case study concerns a relatively simple theory fragment. Larger theories have been formalized in some of the other case studies (for example [Péli and Masuch 1997] where proofs required up to 30 minutes). Current implementations of automated theorem provers, including OTTER, are very powerful. Automated model generators are of recent incarnation, and are yet far less sophisticated. MACE chokes on deeply nested terms or clauses with many literals (beyond 10 distinct terms). We might end up in a situation in which we cannot prove a conjecture, nor find small counterexamples to it (for example, when all counterexamples have infinite

[15]The input format of OTTER and MACE is, strictly speaking, only accepting sentences—formulas in which no variables occur freely. However, since freely occurring variables are interpreted as (new) constants, this amounts to precisely the same: the assignments of the free variables are precisely the interpretations of the constants added to the vocabulary. Of course, the internal procedures immediately translate the first-order sentences to clauses, which does preserve the satisfiability or unsatisfiability of the original sentences, but not necessarily the logical meaning (due to the skolemization of existential quantifiers).

cardinality). Still, current automated model generators are powerful enough to have solved several open problems in (finite) mathematics (Slaney 1994a).

Chapter 3

Reducing Uncertainty: A Formal Theory of *Organizations in Action*

This chapter presents a formal reconstruction of James D. Thompson's classic contribution to organization theory, *Organizations in Action.* The reconstruction explicates the underlying argumentation structure for Thompson's propositions—literally, theorems or problems to be demonstrated. This allows Thompson's propositions to be derived as theorems in a deductive theory. As it turns out, the formal theory is based on general assumptions using only few primitive concepts. In addition, this theory explains why Thompson's propositions do not hold for noncomplex or "atomic" organizations (a restriction on the domain of application). Furthermore, this study reveals that organizations attempt to reduce constraints in their environment—a heretofore unknown implication of the theory.

3.1 Introduction

Thompson's *Organizations in Action* was published more than three decades ago but is still one of the classics of organization theory. The book provides a unifying perspective on open- and closed-systems thinking in organization theory that has been recognized as an important contribution in its own right (Scott 1998). The environment is a key source of uncertainty for an organization, and Thompson argued that much of organizational action can be explained by the need to reduce uncertainty. Consider a specific action such as "buffering" (e.g., building warehouses or storages), aiming to seal off the organization's technical or operational core from environmental uncertainty. The main lines of his arguments always are couched in explicitly formulated propositions but also are brought to life by examples such as the typologies of technologies (long-linked, mediating, or intensive), interdependencies (pooled, sequential, or reciprocal), and coordination (by standardization, by plan, or by mutual adjustment). His typologies have inspired much research in organizational design (Galbraith 1977) and contingency theory (Mintzberg 1979). However, despite the impressive number of citations, the book itself is not read much anymore—most students of organization science will only know the book through references.[1] Because of this, the typologies and examples of Thompson are getting more attention than his core ideas, which are captured in the propositions. The aim of this chapter is to bring Thompson's main ideas back into focus by presenting a logical formalization of his propositions. That is, we will formalize the propositions of *Organizations in Action* in first-order logic and reconstruct the underlying argumentation for them.

Thompson's *Organizations in Action* is a suitable candidate for a formalization attempt in organization and management theory. First, Thompson's theory is formulated using abstract concepts that transcend individual organizations and particular organization types (Zald 1996). Second, the influential role of *Organizations in Action* ensures that the thinking of quite a number of contemporary scholars is based, at least in part, on ideas presented in this famous book. As a result, the basic assumptions of *Organizations in Action* are likely to turn out to be common assumptions of several organization theories.

In this chapter, we give an axiomatic reconstruction of the first chapters of *Organizations in Action*. Our main focus is on the second chapter of this book. Arguably, this is the most important chapter of the book since it provides a unifying perspective on rational, closed-systems strategies in an open-systems environment. The chapter is structured as follows: First, in §3.2, we will introduce the research methodology of logical formalization; then, in §3.3 we will apply this method to Thompson's *Organizations in Action*; and finally, in §3.4, we will review the formal theory and discuss issues related to our work.

3.2 Research Methodology: Logical Formalization

There is a broad consensus among social scientists on the importance of theory. However, this consensus breaks down as soon as one tries to discuss what a 'theory' is (Sutton and Staw 1995). As Masuch, Bruggeman, Kamps, Péli, and Pólos (1996) argued, any discussion about 'theory' is subject to confusion, unless it is rooted on a clear conception of what 'theory' is. The disciplines of Logic and Philosophy of Science have provided a clear notion of formal, deductive theory.[2] In this chapter, we will use the classical, axiomatic-deductive notion of

[1] The Social Sciences Citation Index lists 145 citations in 1995, 154 in 1996, and 135 in 1997 (Social Sciences Citation Index 1995, 1996, 1997).

[2] Some readers may argue that the social sciences require different scientific techniques (for example of empirical inquiry) than other sciences. This may very well be the case, but the context of justification and the structure of theories are common for all sciences (Rudner 1966;

a theory (Tarski 1956; Popper 1959). The premises of a formal theory consist of general statements (universal laws or empirical generalizations, supplemented with definitions) of which the validity is known or assumed, and the theorems of a formal theory are those statements that are logical consequences of the set of premises, i.e., statements that are necessarily true whenever the premises are true. The formal theory itself is the set containing the premises and all theorems, and theoretical explanations and predictions correspond to deductions of theorems from the set of premises (Popper 1959). The main problem for the application of this notion of theory in the social sciences is that it presupposes an exact notion of logical consequence (how the truth of the premises carries over to theorems). Most social science theories are formulated in natural language, which is inherently ambiguous, and therefore makes it impossible to say of a given conjecture whether or not it is indeed a theorem.[3] Most textbooks on social science research methodology contain chapters on the scientific or logical foundations of social research. Typically, these chapters are used indirectly to provide the background context for discussing social scientific inquiry (Singleton and Straits 1999). Logical formalization is an attempt to address such issues directly, by rationally reconstructing scientific theories and representing them in formal logic.

3.2.1 The Product of Logical Formalization

The result of a logical formalization is a theory in a strict sense. The theory consists of statements in a restricted language. This language contains special symbols from the used formal logic[4], names of variables, and names of constants, predicates, and functions (from the theory's domain of application).

The resulting theory contains an explicit set of premises. The premises form the axioms or basis of the theory. We distinguish between a *definition*, which introduces a name for a meaningful concept that is expressible in terms of other concepts and an *assumption*, which contains a claim that is known or assumed to hold for the theory's (intended) domain of application.

Definitions are used to give a name for meaningful concepts that are expressible in terms of other concepts. In a formal theory there is a clear distinction between primitive concepts, and concepts that are defined in terms of the primitive concepts. A defined concept can be eliminated from the theory by replacing the concept with the expression it stands for. We can only use defined concepts

---

Popper 1969).

[3]Notice however, that there are intuitive notions of what constitutes a sound inference or argumentation in natural language. Constructs like 'if ... then ...,' '... then it follows ...,' or 'Therefore, ... ' frequently appear in articles.

[4]In case of first-order logic these are: quantifiers $\forall$ (for all), $\exists$ (there exists); connectives $\neg$ (not), $\wedge$ (and), $\vee$ (or), $\rightarrow$ (implies), and $\leftrightarrow$ (if and only if); and the equality symbol $=$.

in the definition of other concepts if this does not lead to circularity.[5] In this way, it is possible to write out an equivalent definition that uses only primitive concepts, and it remains possible to eliminate all defined concepts.

The assumptions are general statements, such as universal laws or law-like statements like empirical generalizations.[6] The theory's domain of application is characterized by the assumptions: the theory applies to all domains that satisfy the assumptions (after using the definitions to eliminate defined concepts that may occur in the assumptions). The more general the assumptions of the theory are, the larger is its domain of application. For the empirical testing of the theory it suffices to test the assumptions (if they turn out to be valid then theorems that are logically derivable from them must necessarily be true as well). If assumptions are not directly testable, we can try to falsify them by testing consequences of them (if any of these consequences turns out to be false, then some of the assumptions must be false).

The theory consists of all the statements that are logical consequences of the set of premises (including the premises because they are trivial consequences of themselves). If we accept the premises of a theory, then we must also accept all statements that are logical consequences of the premises, i.e., that are necessarily true if the premises are true. A theorizer is not allowed to include only some of the theorems in the theory, and not to include other theorems (unless the premises are modified accordingly). In any exposition of a theory, there are, apart from the premises, only a small number of salient consequences singled out. We use the following labels: a *theorem*, which is an interesting consequence of a theory because it is a surprising result, or a new testable implication of the theory; a *lemma*, which is a minor theorem—an intermediate result that is used to derive further theorems or lemmas but is of some interest in its own right; and a *corollary*, which is an immediate consequence of a theorem.

The consequences of the premises are the predictions or claims of the theory. The proof of a theorem in a formal theory corresponds to an argument for, or explanation of the claim it makes (giving a causal explanation, i.e., answering *why* the claim holds). That is, if we consider cases in which the premises hold, then the theorem must also hold (i.e., the theorem is a prediction). Conversely, if we consider cases where the theorem holds, then the premises give an explanation for why the theorem holds. If a conjecture is not a consequence of the premises, empirical support for the premises and the conjecture's claim can be independent: even if the premises are valid, the conjecture can be false, and even if we refute the conjecture, all the theory's premises may still be valid.

---

[5]It is impossible to define all the concepts of a theory: if all concepts were defined in terms of the other concepts, then some of the definitions have to be circular.

[6]These statements are also called laws or hypotheses depending on the range of evidence for them. We prefer to use the neutral term assumptions here.

3.2.2 The Process of Logical Formalization

Most social science theories are stated in natural language. The process of logical formalization takes as input a discursive theory (a theory stated in natural language) and outputs a theory in formal logic. This process concerns the interlinked activities of rational reconstruction (reconstructing the claims, premises, and argumentation of the theory) and formal modeling (capturing the claims as theorems that are provable from explicit definitions and assumptions). The process of logical formalization does not obey to strict rules, but involves creativity and insight.[7]

A problem is that not all of the discursive text is theory. A text usually contains examples, discussion, citations, figures, etc.; parts which are not theory (Sutton and Staw 1995). We can reconstruct the theory by identifying sentences that are definitions, or assumptions about the intended application area of the theory, or sentences that are the claims or propositions that the author makes. If we manage to identify definitions, assumptions, and propositions in the text (and formulate these in logic) then we can test for the theorem-hood of the propositions.

However, it is highly unlikely that we can immediately prove the propositions, because not all of the theory is in the discursive text. The authors of a text typically assume a basis of common background knowledge. For example, if a causes b and b causes c then a reader will immediately infer that a also causes c using his background knowledge about the transitivity of causality. In a formal theory, this assumption has to be added explicitly. Making common background knowledge explicit is important, because any ambiguity about their specific content may cause confusion. Moreover, the background knowledge may have unanticipated desirable or unwanted consequences (which are also part of the theory).

There is an even worse problem when there turns out to be a real hiatus in the argumentation of the theory. Sutton and Staw (1995) observe that authors routinely use non-theoretical statements *in lieu of* theory. This indicates that many theories are missing some crucial assumptions. From the viewpoint of justification, we should stop at the discovery of such a hiatus and conclude that (part of) the argumentation, predictions, and explanations of the theory are unsound and must be discarded: the chain of argumentation is as strong as its weakest link. Authors also make unnecessary assumptions or assumptions that can be relaxed, increasing the parsimony of the theory and making it more generally applicable.

Another problem is that organization theory is known to make conflicting requirements and to contain paradoxes (Quinn and Cameron 1988). During the

[7] The same holds for the formalization of theories in traditional mathematics (Blalock 1969). There is always a certain distance to bridge between a partial and ambiguous formulation in natural language and a rigorous formal exposition of the theory in logic or mathematics. A noticeable difference between the formalization in logic and formalization in traditional mathematics is that logic does not require metric precision (which only few theories can provide), but can closely follow the original qualitative argumentation of the natural language theories (Hannan 1997).

formalization of organization theory, such requirements and paradoxes can manifest themselves as contradictions. The theory may contain assumptions that are reasonable when considered independently, but nevertheless lead to a contradiction when combined. These contradictions typically arise if some of the theory's assumptions are formulated too generally (for example when exceptions are not taken into account). Moreover, a discursive theory might be interpreted differently in different contexts. Two interpretations of a theory may each be consistent, but lead to inconsistency when combined. Again, from the viewpoint of justification, we should discard an inconsistent theory.

However, our working hypothesis is that sociological texts have real theoretical content, although it may be obscured by the use of natural language.[8] That is, if we discover a contradiction or hiatus in a theory, we always attempt to reconstruct the theory such that it is contradiction-free and that its argumentation is sound. As a result, we gradually move into the context of discovery and do some theorizing ourselves: We are forced to move away from the explicit statements of the text, and have to rely more and more on our reconstruction of it. Consider the case in which a part of the text strongly suggests an implicit assumption of the theory. Our approach in this is theory-friendly: the goal is to find a favorable interpretation that is consistent and allows us to derive the propositions as theorems.[9] Examining the formal theory is useful during this process of reconstruction. The formal theory may suggest which subset of the assumptions is inconsistent, or suggest which additional (background) assumptions would make a proposition derivable. Logical formalization is a constructive methodology for theory building, because future discussion can be pointed down to the specific premises responsible for it and thereby facilitate further development of the theory.

3.3 Thompson's *Organizations in Action*

Organizations in Action provides a unifying framework for both the classical, closed-systems theories and the emerging open-systems theories of organizations. As Thompson writes, "A central purpose of this book is to identify a framework which might link at important points several of the now independent approaches to the understanding of complex organizations" (Thompson 1967, p.viii).

[8]In fact, there is some controversy over the theoretical status of social sciences. On the one hand, most social scientists do not seem to worry much about the theoretical status of their science, for example the definite phrase 'the theory of ... ' frequently appears in titles. On the other hand, philosophers of science tend to be skeptical about social science theories, for example Rudner (1966, p.40) argues that the social sciences are poor on theory and Balzer et al. (1987, p.xix) state that contemporary sociological theory is in its formative (or pre-scientific) period.

[9]The ambiguities of natural language makes it generally easy to find inconsistent interpretations of a discursive theory, or interpretations in which propositions are not derivable.

This framework has survived the test of time remarkably well, that is, the material is part of mainstream organization theory and citations are frequent. In this chapter, we will mainly focus on the second chapter of *Organizations in Action*. Arguably, the second chapter is the most important one because it provides the crucial link between closed-systems and open-systems strategies. The second chapter starts with one of Thompson's famous typologies, a typology of technologies: *Long-linked technologies* are serial interdependent. Several actions have to be performed in sequence. The prototypical example is an assembly line; *Mediating technologies* are concerned with linking of interdependent customers. Typical examples are commercial banks, insurance firms, and employment agencies; and *Intensive technologies* are crucially dependent upon feedback on their actions. Typical examples are hospitals, construction industry, and military combat teams (Thompson 1967, pp.15–18).

These three types of technology are increasingly susceptible to environmental influences and are therefore decreasingly faithful approximations of closed-systems strategies. The three variations in technology are introduced merely "to illustrate the propositions we wish to develop" (Thompson 1967, p.15). These carefully formulated propositions will be the main focus of study in this chapter, since they capture Thompson's unifying framework. The propositions of the second chapter are listed in Table 3.1.

Label	Content
Proposition 2.1	Under norms of rationality, organizations seek to seal off their core technologies from environmental influences.
Proposition 2.2	Under norms of rationality, organizations seek to buffer environmental influences by surrounding their technical cores with input and output components.
Proposition 2.3	Under norms of rationality, organizations seek to smooth out input and output transactions.
Proposition 2.4	Under norms of rationality, organizations seek to anticipate and adapt to environmental changes which cannot be buffered or leveled.
Proposition 2.5	When buffering, leveling, and forecasting do not protect their technical cores from environmental fluctuations, organizations under norms of rationality resort to rationing.

Note: All propositions are from (Thompson 1967, pp.14–24).

Table 3.1: Propositions of Thompson's Chapter 2.

There has been some debate on the theoretical status of Thompson's propositions. As the editors of a collection of his essays write:

> Unlike some sociologists, Thompson cannot be accused of being "wordy" and of using unnecessary jargon. He uses simple language

> and got to the point quickly. This is especially true for his major work, *Organizations in Action.* Indeed, the reader sometimes feels Thompson gets to the point too quickly; his parsimonious style sometimes leaves the reader behind. Also, although his work contains numerous propositions, all are not derived from a common set of explicit theoretical concepts and assumptions. In this sense Thompson was not a rigorous deductive theorist. Instead, he introduced concepts and formulated propositions appropriate to the organizational realities he was examining, with little effort to show that different sets of propositions derived logically from still more general propositions, assumptions, and concepts. (Rushing and Zald 1976, pp.ix–x)

However, we observed that Thompson's informal argumentation for the propositions suggests an underlying explanatory structure. Therefore, we decided to reconstruct this argumentation in formal logic and build a rigorous deductive theory of *Organizations in Action.* In the rest of this section, we will try to derive Thompson's propositions as theorems in a formal, deductive theory of *Organizations in Action.*

3.3.1 Complex Organizations

Before we can prove the propositions, we will have to introduce a number of general concepts from Thompson's chapter 1. We will only introduce those concepts from chapter 1 that are necessary for deriving the propositions in further sections. The main ingredient of all organization theories is, obviously, organizations. We use the predicate O for "organizations". For example, $\mathsf{O}(o)$ expresses that o is an organization. We also use a predicate SO for a "suborganization" of an organization. For example, $\mathsf{SO}(o_1, o_2)$ expresses that o_2 is a suborganization of (organization) o_1. A suborganization is a part of the organization (but not all arbitrary parts of an organization are suborganizations).

Organizations in Action is explicitly dealing with "complex organizations". Thompson states that complex organizations are ubiquitous in modern societies and gives several examples: manufacturing firms, hospitals, schools, armies, community agencies (Thompson 1967, p.3). However, Thompson does not give a definition of complex organizations, nor does he dwell upon differences between complex and other, noncomplex organizations. We will define complex organizations by some of their characteristics, and will discuss in §3.3.3, below, what noncomplex organizations would look like and why, if they exist, they are exempted from Thompson's theory.

We will use CO for complex organizations. For example, $\mathsf{CO}(o)$ says that o is a complex organization. Thompson gives some structural characteristics of complex organization. He adopts the suggestion of Parsons (1960) that organizations exhibit three distinct levels of responsibility and control: technical, managerial, and institutional. According to Thompson, "every formal organization contains a suborganization whose "problems" are focused around effective performance of the technical function" (Thompson 1967, p.10). We introduce the predicate TC for the technical or operational core of an organization. For example, $\mathsf{TC}(o, tc)$ says that tc is the technical core of o. We will define complex organizations as exactly those organizations that have a technological suborganization:

Definition 1 *Complex organizations.*
$\forall x\ [\mathsf{CO}(x)\ \leftrightarrow\ \mathsf{O}(x) \wedge \exists y\ [\mathsf{SO}(x,y) \wedge \mathsf{TC}(x,y)]]$
(Read: x is a complex organization if and only if x is an organization and there exists a y such that y is a suborganization of x and y is the technical core of x.)

Thompson uses the notion of a technical core (the transformational or production process) in a general sense that applies to all three types of technologies: long-linked, mediating, and intensive technology (pp.15–18). The core technologies of assembly lines are the processing of material and supervision of these operations. In case of mediating technologies like commercial banks, the core activities are the linking of depositors and borrowers. And in case of intensive technologies like hospitals, core activities are the performance of some specific combination of various skills depending, on the patient's state.

Definition 1 still allows complex organizations to have more than one technical core, although Thompson consistently uses the definite article "the" when referring to an organization's core technologies (p.10). We will explicitly assume that the technical core of an organization can be uniquely determined—complex organizations can only have one technical core.

Assumption 1 *The technical core is unique.*
$\forall x, y, z\ [\mathsf{TC}(x,y) \wedge \mathsf{TC}(x,z)\ \rightarrow\ y = z]$
(Read: if both y and z are the technical core of x, then y is equal to z.)

Assumption 1 ensures that we can talk about *the* technical core of a complex organization. This technical core is the technical suborganization of definition 1, as becomes clear from the following lemma:[10]

Lemma 1 *The technical core of a complex organization is a suborganization.*
$\forall x, y\ [\mathsf{CO}(x) \wedge \mathsf{TC}(x,y)\ \rightarrow\ \mathsf{SO}(x,y)]$
(Read: if x is a complex organization with core technologies y, then y is a suborganization of x.)

[10] All derived statements are proved using the automated theorem prover OTTER (Organized Techniques for Theorem-proving and Effective Research; McCune 1994b). We will give here only the outline of those proofs: the first lines of the proofs are by assumption, and the meta-implication symbol, $\Rightarrow$, indicates steps in the proof. Most steps involve *modus ponens*: if ϕ holds and ϕ implies ψ, then ψ holds as well.

Proof: (Using definition 1 and assumption 1.) By Def.1, a complex organization has a technical core that is a suborganization. By Ass.1, an organization can have only one technical core. Therefore, the technical core of a complex organization is a suborganization.

$$
\begin{array}{rcl}
 & \mathsf{CO}(o_1) \wedge \mathsf{TC}(o_1, tc_1) & \text{(by assumption)} \\
\Rightarrow & (\exists tc_2)\ \mathsf{SO}(o_1, tc_2) \wedge \mathsf{TC}(o_1, tc_2) & \text{(by Def.1)} \\
\Rightarrow & tc_1 = tc_2 & \text{(by Ass.1)} \\
\Rightarrow & \mathsf{SO}(o_1, tc_1) &
\end{array}
$$

Lemma 1 is a technicality that will be used in some of the other proofs. It allows us to talk about the technical core of a complex organization by ensuring that the technical core is the specific suborganization of definition 1.

The performance of their core technologies is crucial for organizations. According to Thompson (p.11), "it would therefore be advantageous for an organization subject to criteria of rationality to remove as much uncertainty as possible from its *technical core*." We call this "rational evaluation" and capture this by a predicate, REVA. For example, $\mathsf{REVA}(o, tc)$ expresses that o evaluates tc in terms of technical rationality. As stated above, we assume that core technologies are rationally evaluated:

Assumption 2 *The technical core is rationally evaluated.*
$\forall x, y\, [\mathsf{TC}(x, y) \ \rightarrow\ \mathsf{REVA}(x, y)]$
(Read: if y is the technical core of x, then x will rationally evaluate y.)

Furthermore, if an organization rationally evaluates a particular suborganization, then it will attempt to reduce uncertainty for that suborganization. We introduce a predicate, UC, for uncertainty of a (sub)organization. For example, $\mathsf{UC}(o, u)$ says that o has uncertainty u. We further introduce a predicate, RED, for the reduction. For example, $\mathsf{RED}(o, u, tc)$ expresses that o attempts to reduce u for tc. We can now formulate the assumption that organizations attempt to reduce the uncertainty for suborganizations that are rationally evaluated:

Assumption 3 *Organizations attempt to reduce uncertainty for rationally evaluated suborganizations.*
$\forall x, y, z\, [\mathsf{SO}(x, y) \wedge \mathsf{REVA}(x, y) \wedge \mathsf{UC}(y, z) \ \rightarrow\ \mathsf{RED}(x, z, y)]$
(Read: if y is a suborganization of x, and x rationally evaluates y, and z is the uncertainty of y, then x attempts to reduce uncertainty z for y.)

Using assumptions 2 and 3, we can now derive the following lemma for complex organizations:

Lemma 2 *Complex organizations attempt to reduce uncertainty for their technical cores.*
$\forall x, y, z\ [\mathsf{CO}(x) \land \mathsf{TC}(x, y) \land \mathsf{UC}(y, z)\ \rightarrow\ \mathsf{RED}(x, z, y)]$
(Read: if x is a complex organization with core technologies y, and z is the uncertainty of y, then x attempts to reduce the uncertainty z for y.)

Proof: (Using lemma 1, and assumptions 2, and 3.) By Lem.1, the technical core is a suborganization, and by Ass.2, it is rationally evaluated by the organization. Therefore, by Ass.3, the organization attempts to reduce uncertainty for the technical core.

$$\begin{array}{ccl} & \mathsf{CO}(o_1) \land \mathsf{TC}(o_1, tc_1) \land \mathsf{UC}(tc_1, u_1) & \text{(by assumption)} \\ \Rightarrow & \mathsf{SO}(o_1, tc_1) & \text{(by Lem.1)} \\ \Rightarrow & \mathsf{REVA}(o_1, tc_1) & \text{(by Ass.2)} \\ \Rightarrow & \mathsf{RED}(o_1, u_1, tc_1) & \text{(by Ass.3)} \end{array}$$

Lemma 2 will be an important part of the argumentation in the rest of this chapter, because "coping with uncertainty [appears] as the essence of the administrative process" (p.159).

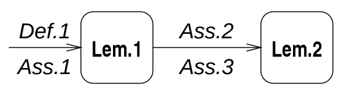

Figure 3.1: Reducing uncertainty (structure of the theory 1).

In this section, we introduced the setting for Thompson's proposition by introducing predicates for organizations, O; suborganizations, SO; and core technologies, TC. This led to the defined notion of complex organizations, CO (Def.1). We assumed that organizations have only one technological suborganization (Ass.1). We derived that the technical core of a complex organization is a suborganization (Lem.1; see Fig.3.1). We furthermore introduced predicates for uncertainty, UC; for rational evaluation, REVA; and for reduction, RED. We assumed that the performance of their core technologies is crucial for organizations (Ass.2), and that an organization will attempt to remove as much uncertainty as possible from suborganizations whose performance is rationally evaluated (Ass.3). We derived that complex organizations attempt to reduce uncertainty for their technical cores (Lem.2).

3.3.2 Sealing Off

In this section, we will now try to prove Thompson's first proposition: "organizations seek to seal off their core technologies from environmental influences" (p.19).

Before we can give a formal version of this proposition, we have to introduce some more predicates. According to Thompson, organizations are subject to environmental influences. These environmental influences consist of both environmental fluctuations and constraints. Environmental fluctuations are dynamic, they reflect the change of market conditions (such as seasonal demand). Environmental constraints are static restrictions on the organization (such as new legislation). We introduce two predicates, FL and CS, for fluctuations and constraints, respectively. For example, $\mathsf{FL}(tc, f, o)$ says that tc is exposed to a fluctuation f from o, and $\mathsf{CS}(tc, c, o)$ expresses that tc is exposed to a constraint c from o. We define environmental influences, ENVI, to be the general term used for both fluctuations and constraints. For example, $\mathsf{ENVI}(tc, i, o)$ says that tc is exposed to environmental influence i from o.

Definition 2 *Environmental influence.*
$\forall x, y, z\ [\mathsf{ENVI}(x, y, z)\ \leftrightarrow\ \mathsf{FL}(x, y, z) \vee \mathsf{CS}(x, y, z)]$
(Read: x is exposed to an influence y from z if and only if x is exposed to a fluctuation y from z or x is exposed to a constraint y from z.)

Thompson argues that "technologies and environments are major sources of uncertainty for organizations" (p.13). To express that environments cause uncertainty, we use a predicate, C, for causality. For example, $\mathsf{C}(i, u)$ expresses that i causes u.[11] Environmental influences cause uncertainty in the organization:

Assumption 4 *Environmental influences cause uncertainty.*
$\forall x, y, z\ [\mathsf{ENVI}(x, y, z)\ \rightarrow\ \exists v\ [\mathsf{UC}(x, v) \wedge \mathsf{C}(y, v)]]$
(Read: if x is exposed to an environmental influence y from z, then there exists a v such that v is the uncertainty of x, and influence y causes uncertainty v.)

Thompson's use of the term "sealing off" (p.19) corresponds closely to the reduction of uncertainty that is due to an environmental influence. An organization seals a suborganization off from an environmental influence if it attempts to reduce the uncertainty caused by this influence. We define a predicate, SEFF, in precisely this way. For example, $\mathsf{SEFF}(o, i, tc)$ expresses that o seals off tc from influence i.

Definition 3 *Sealing off.*
$\forall x, y, z\ [\mathsf{SEFF}(x, y, z)\ \leftrightarrow\ \mathsf{SO}(x, z) \wedge \exists v, w\ [\mathsf{ENVI}(z, y, v) \wedge \mathsf{UC}(z, w) \wedge \mathsf{C}(y, w) \wedge \mathsf{RED}(x, w, z)]]$
(Read: x seals z off from y if and only if z is a suborganization of x, and there exists v and w such that z is exposed to an influence y from v, and y causes uncertainty w of z, and x attempts to reduce w for z.)

[11] Causality is a complex notion with intricate properties. In this chapter, we do not use any formal property of causality, although we would (at least) consider causality to be transitive (if x causes y and y causes, in turn, z then x causes z: $\forall x, y, z\ [\mathsf{C}(x, y) \wedge \mathsf{C}(y, z)\ \rightarrow\ \mathsf{C}(x, z)]$).

We can now prove the following theorem:

Theorem 3 *Complex organizations seal off their core technologies from environmental influences.*
$\forall x, y, z, v\, [\mathsf{CO}(x) \wedge \mathsf{TC}(x, y) \wedge \mathsf{ENVI}(y, z, v) \;\rightarrow\; \mathsf{SEFF}(x, z, y)]$
(Read: if x is a complex organization, and y is the technical core of x, and y is exposed to an influence z from v, then x seals y off from z.)

Proof: (Using lemmas 1 and 2, definition 3, and assumption 4.) By Lem.1, the technical core is a suborganization. By Ass.4, an environmental influence causes uncertainty for the technical core. By Lem.2, complex organizations attempt to reduce this uncertainty for their technical core. Therefore, by Def.3, the organization is sealing its technical suborganization off from the environmental influence.

$$
\begin{array}{rcl}
 & \mathsf{CO}(o_1) \wedge \mathsf{TC}(o_1, tc_1) \wedge \mathsf{ENVI}(tc_1, e_1, o_2) & \text{(by assumption)} \\
\Rightarrow & \mathsf{SO}(o_1, tc_1) & \text{(by Lem.1)} \\
\Rightarrow & (\exists u_1)\ \mathsf{UC}(tc_1, u_1) \wedge \mathsf{C}(e_1, u_1) & \text{(by Ass.4)} \\
\Rightarrow & \mathsf{RED}(o_1, u_1, tc_1) & \text{(by Lem.2)} \\
\Rightarrow & \mathsf{SEFF}(o_1, e_1, tc_1) & \text{(by Def.3)}
\end{array}
$$

This theorem is a formal version of Thompson's proposition 2.1: "Under norms of rationality, organizations seek to seal off their core technologies from environmental influences" (p.19). The phrase "Under norms of rationality" is not explicitly mentioned in the antecedent of theorem 3. These "norms of rationality" seem to underly all propositions, and are reflected in the "rational evaluation" of assumptions 2 and 3, which, by lemma 2, play a role in the argument. The phrase "seek to ..." is captured by the intentional interpretation of the RED predicate: attempt to reduce. Since sealing off is defined in terms of the reduce predicate, it inherits the intentional interpretation.

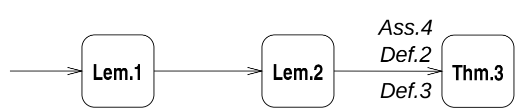

Figure 3.2: Sealing off (structure of the theory 2).

We introduced predicates for environmental fluctuations, FL, and environmental constraints, CS. This led to the defined notion of environmental influence, ENVI (Def.2). We introduced a predicate for causes, C. We assumed that environmental influences cause uncertainty (Ass.4). We defined a predicate for sealing off, SEFF (Def.3), and learned that complex organizations are sealing off their core technologies from environmental influences (Thm.3; see Fig.3.2).

3.3.3 Beyond Thompson: Atomic Organizations

Thompson does not discuss why he restricts his theory to complex organizations. In this section, we will step away from Thompson's book and try to answer questions like: What would noncomplex organizations look like? Do they exist in the real world? If so, why are they excluded from Thompson's theory? How do they fare in the real world?

We will simply define noncomplex organizations as organizations that are not complex organizations. We baptize such organizations as *atomic organizations* and define a predicate ATO for them: $\mathsf{ATO}(o)$ says that o is an atomic organization.

Definition 4 *Atomic organizations.*
$\forall x\ [\mathsf{ATO}(x)\ \leftrightarrow\ \mathsf{O}(x) \wedge \neg\mathsf{CO}(x)]$
(Read: x is an atomic organization if and only if x is an organization and x is not a complex organization.)

Note that we define ATO in terms of another defined predicate, CO. This is only allowed if it does not lead to circularity. In other words, we must be able to trace back the underlying primitive concepts. This can be done by unfolding the definition of complex organizations (definition 1):

Lemma 4 *Atomic organizations are organizations that have no suborganization as technical core.*
$\forall x\ [\mathsf{ATO}(x)\ \leftrightarrow\ \mathsf{O}(x) \wedge \neg\exists y\ [\mathsf{SO}(x,y) \wedge \mathsf{TC}(x,y)]]$
(Read: x is an atomic organization if and only if x is an organization, and there is no y such that y is a suborganization of x and y is the technical core of x.)

Proof: (Using definitions 1 and 4.) By Def.4 atomic organizations are organizations that are not complex organizations. Using Def.1, they are organizations that are either no organizations (which would be a contradiction), or do not have a technical suborganization.

$$\begin{array}{lcl} & \mathsf{ATO}(o_1) & \text{(by assumption)} \\ \Leftrightarrow & \mathsf{O}(o_1) \wedge \neg\mathsf{CO}(o_1) & \text{(by Def.4)} \\ \Leftrightarrow & \mathsf{O}(o_1) \wedge \neg[\mathsf{O}(o_1) \wedge \exists y\ [\mathsf{SO}(o_1,y) \wedge \mathsf{TC}(o_1,y)]] & \text{(by Def.1)} \\ \Leftrightarrow & \mathsf{O}(o_1) \wedge [\neg\mathsf{O}(o_1) \vee \neg\exists y\ [\mathsf{SO}(o_1,y) \wedge \mathsf{TC}(o_1,y)]] & \\ \Leftrightarrow & \mathsf{O}(o_1) \wedge \neg\exists y\ [\mathsf{SO}(o_1,y) \wedge \mathsf{TC}(o_1,y)] & \end{array}$$

Lemma 4 gives an abstract characterization of atomic organizations: atomic organizations do not have their core technologies grouped in a special suborganization—there is no clear separation between their core technologies and other activities. We can use this characterization to try to identify organization that would be atomic in the sense of Def.4. There are some examples of organizations that correspond to this abstract characterization. Examples of atomic organizations are small organizations, especially organizations such as family firms that have only very few employees. These small organizations do not have a clear separation between technological and other activities; for example, they are managed by their owners, who are also involved in the technical operations. Although small-sized organizations have not received much attention in the literature, the majority of organizations has only a very limited number of employees, and small organizations occupy a substantial part of the market (Granovetter 1984). As Granovetter (1984, p.333) notes, "the study of organizations is often taken to be synonymous with the study of "complex organizations"" and "much of what has been done in some otherwise splendid work on the sociology of economic life and complex organizations has proceeded as if the entire waterfront has been covered, when in fact work has concentrated in one important receding pool." Together with small organizations, another example of atomic organizations is new organizations. New enterprise startups usually do not have a fully crystallized management structure; technical and management activities are all performed by the entrepreneur. New organizations involve new roles, which have to be learned (at the cost of inefficiency) (Stinchcombe 1965). Obviously, new and small organizations show considerable overlap.

Lemma 4, stating that atomic organizations do not have a technological suborganization, still leaves us two possibilities in discussing the technical core of atomic organizations. The first option is to assume that the technical core of atomic organizations does not exist at all.[12] This option has as a result that all statements that involve the technical core do not apply to atomic organizations. The second option is to argue that all organizations, including atomic ones, have certain technologies that constitute the core of their activities. Therefore, it makes sense to talk about the core technologies of any organization. If, as in the case of atomic organizations, these technological activities are not grouped together in a suborganization, we treat the entire organization as its own technical core.

Since Thompson does not discuss noncomplex organizations, we can only speculate on which option to choose. We will pursue the second option because it allows us to further investigate characteristics of atomic organizations. We can implement this option by assuming that if the technical core of an organization is not a suborganization, then it is identical with the organization itself:

Assumption 5 *If the technical core of an organization is not a suborganization, then we treat the whole organization as its technical core.*
$\forall x, y\ [\mathsf{O}(x) \wedge \mathsf{TC}(x,y) \wedge \neg\mathsf{SO}(x,y)\ \rightarrow\ x = y]$
(Read: if x is an organization, and y is the technical core of x, and y is not a suborganization of x, then x and y are identical.)

[12] Formally, $\forall x\ [\mathsf{ATO}(x)\ \rightarrow\ \neg\exists y\ [\mathsf{TC}(x,y)]]$.

Assumption 5 gives us indeed the result of the second option:

Lemma 5 *Atomic organizations are identical with their technical core.*
$\forall x, y\, [\mathsf{ATO}(x) \wedge \mathsf{TC}(x, y) \;\rightarrow\; x = y]$
(Read: if x is an atomic organization, and y is the technical core of x, then x and y are identical.)

Proof: (Using lemma 4 and assumption 5.) By Lem.4, atomic organizations do not have a technical core that is a suborganization, that is, it is either not a suborganization or not a technical core (which would be a contradiction). Therefore, the technical core of an atomic organization is not a suborganization, and by Ass.5, it follows that the technical core is identical with the atomic organization.

$$\begin{array}{rcl}
 & \mathsf{ATO}(o_1) \wedge \mathsf{TC}(o_1, tc_1) & \text{(by assumption)} \\
\Rightarrow & \mathsf{O}(o_1) \wedge \neg\exists y\, [\mathsf{SO}(o_1, y) \wedge \mathsf{TC}(o_1, y)] & \text{(By Lem.4)} \\
\Rightarrow & \forall y\, [\neg\mathsf{SO}(o_1, y) \vee \neg\mathsf{TC}(o_1, y)] & \\
\Rightarrow & \neg\mathsf{SO}(o_1, tc_1) \vee \neg\mathsf{TC}(o_1, tc_1) & \\
\Rightarrow & \neg\mathsf{SO}(o_1, tc_1) & \\
\Rightarrow & o_1 = tc_1 & \text{(By Ass.5)}
\end{array}$$

Lemma 5 is the companion of lemma 1, which stated that the technical core of a complex organization is a suborganization. Lemma 5 reiterates that atomic organizations cannot differentiate between their core technologies and other activities: there is no distinction between levels of technical and managerial responsibilities and control.

Now that we have identified the technical core of atomic organizations, we can try to discuss issues relating to its uncertainty. Unfortunately, assumption 4 puts hardly any restriction on uncertainties caused by the environment. Thompson recognized that an organization may be subject to a number of different uncertainties, for example, technological uncertainty and the uncertainty due to influences (Thompson 1967, p.1). As a result, we cannot represent the uncertainty as a function of the organization because a single organization may be exposed to different influences that cause different uncertainties for it. However, we would want to require something weaker, namely that a single environmental influence will only cause a single uncertainty:

Assumption 6 *A single influence causes a single uncertainty.*
$\forall x, y, z, v, w\, [\mathsf{ENVI}(x, y, z) \wedge \mathsf{UC}(x, v) \wedge \mathsf{C}(y, v) \wedge \mathsf{UC}(x, w) \wedge \mathsf{C}(y, w) \;\rightarrow\; v = w]$
(Read: if x is exposed to an influence y from z, and y causes uncertainty v of x, and y also causes uncertainty w of x, then uncertainty v is equal to uncertainty w.)

This assumption is a restriction on assumption 4, which stated that that an influence will cause uncertainty. This restriction is consistent with Thompson's text. Using assumption 6 we can derive a theorem about atomic organizations:

Theorem 6 *The technical core of an atomic organization faces the same environmental uncertainty as the organization.*
$\forall x, y, z, v, w\ [\mathsf{ATO}(x) \wedge \mathsf{TC}(x,y) \wedge \mathsf{ENVI}(x,v,z) \wedge \mathsf{UC}(x,w) \wedge \mathsf{C}(v,w)\ \rightarrow\ \exists u\ [\mathsf{UC}(y,u) \wedge \mathsf{C}(v,u) \wedge u = w]]$
(Read: if x is an atomic organization with core technologies y, and if x is exposed to an influence v from z, and v causes uncertainty w of x, then there exist uncertainty u of the technical core y, caused by v, and the uncertainty u is equal to the uncertainty w.)

Proof: (Using lemma 5 and assumptions 4 and 6.) By Lem.5 atomic organizations are identical with their technical cores. Therefore, the technical core faces the same environmental influence as the organization. By Ass.4, this causes uncertainty of the technical core. Then by Ass.6, the uncertainty of the technical core is the same as the uncertainty of the organization.

$$
\begin{array}{llr}
 & \mathsf{ATO}(o_1) \wedge \mathsf{TC}(o_1, tc_1) \wedge \mathsf{ENVI}(o_1, e_1, o_2) \wedge \mathsf{UC}(o_1, u_1) \wedge \mathsf{C}(e_1, u_1) & \text{(by assumption)} \\
\Rightarrow & o_1 = tc_1 & \text{(by Lem.5)} \\
\Rightarrow & \mathsf{ENVI}(tc_1, e_1, o_2) & \\
\Rightarrow & (\exists u_2)\ \mathsf{UC}(tc_1, u_2) \wedge \mathsf{C}(e_1, u_2) & \text{(by Ass.4)} \\
\Rightarrow & u_2 = u_1 & \text{(by Ass.6)}
\end{array}
$$

Theorem 3 stated that complex organizations seal off their core technologies from environmental influences. Theorem 6, in contrast, states that atomic organizations and their core technologies face the same uncertainty—noncomplex organizations cannot reduce the uncertainty for their technical core.[13] Theorem 6 may help explain the massive failure rates of small organizations (U.S. Small Business Administration 1985), and of new enterprises—the liability of newness (Freeman, Carroll, and Hannan 1983; Brüderl, Preisendörfer, and Ziegler 1992; Hannan 1998). In the case of atomic organizations, an environmental influence causes the same uncertainty on the organizational level as it will on the technical core. As a result, atomic organizations cannot reduce uncertainty for their core technologies. Unlike complex organizations, they have no separate managerial level that can mediate between the technical core and the environment. All the environmental influences that atomic organizations face are faced at equal strength by their core technical activities. As argued before (Perrow 1986), small

[13] Theorem 6 states that the uncertainties on the organizational level and on the technological level are really identical. It would not be rational of an atomic organization to attempt to reduce the uncertainties within the organization because such an attempt must fail.

organizations are trivial organizations, but nevertheless they do occur in great numbers.

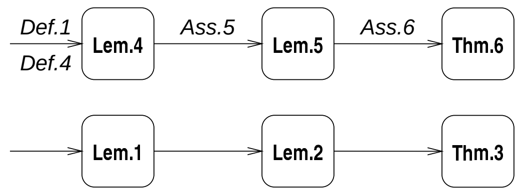

Figure 3.3: Atomic organizations (structure of the theory 3).

In this section, we investigated the possibility of noncomplex organizations. We defined a predicate, ATO, for atomic or noncomplex organizations (Def.4). Unfolding the definitions of complex (Def.1) and atomic organizations (Def.4) characterizes noncomplex organizations as those having no technical suborganization (Lem.4; see Fig.3.3). We assumed that if the technical core of an organization is not a suborganization, then we treat the entire organization as the technical core (Ass.5). We derived that this is the case for noncomplex organizations (Lem.5). We assumed that one environmental influence can only cause one type of uncertainty for an organization (Ass.6), and derived that atomic organizations cannot reduce the uncertainty caused by environmental influences (Thm.6). This indicates that Thompson's restriction to complex organizations is not an arbitrary one: his propositions do not hold for noncomplex organizations.

3.3.4 Buffering and Anticipating

After the *intermezzo* of the §3.3.3, we will continue with our formalization of Thompson's arguments. In this section, we try to prove a formal version of Thompson's second proposition: "organizations seek to buffer environmental influences" (Thompson 1967, p.20), and his fourth proposition: "organization seek to anticipate and adapt to environmental changes" (p.21).

Theorem 3 in §3.3.2 stated that complex organizations attempt to seal off their core technologies from environmental influences. This suggests that complex organizations have some sort of control over influences directed at their (technical) suborganizations. We introduce a predicate, HC, for "having control." For example, $\mathsf{HC}(o, i)$ says that o has control over i. We assume that organizations have some control over environmental influences directed at their suborganizations:

Assumption 7 *Organizations have control over environmental influences on their suborganizations.*
$\forall x, y, z, v\ [\mathsf{O}(x) \wedge \mathsf{SO}(x, y) \wedge \mathsf{ENVI}(y, z, v)\ \rightarrow\ \mathsf{HC}(x, z)]$
(Read: if x is an organization, and y is a suborganization of x, and y is exposed to an influence z from v, then x has control over z.)

The idea is that organizations mediate between the environment and suborganizations like the technical core (Thompson 1967, p.11). Note an organization does not necessarily have *complete* control over such an environmental influence; it just means that the organization can undertake some actions that will reduce the influence (but which may not eliminate it completely). In the rest of this section, we will discuss some concrete actions for reducing uncertainty.

Next, we assume that if organizations attempt to reduce something (such as uncertainty), and they have some control over one of its causes, then they will also attempt to reduce this cause:

Assumption 8 *If an organization attempts to reduce something, and has control over a cause of it, the organization will attempt to reduce the cause.*
$\forall x, y, z, v\ [\mathsf{RED}(x, y, z) \wedge \mathsf{C}(v, y) \wedge \mathsf{HC}(x, v) \ \rightarrow\ \mathsf{RED}(x, v, z)]$
(Read: if x attempts to reduce y for z, and v causes y, and x has control over v, then x attempts to reduce v for z.)

This assumption presupposes the organizational rationality that a reduction of the cause will result in a reduction of the effect. As Thompson writes, "Instrumental action is rooted on the one hand in desired outcomes and on the other hand in beliefs about cause/effect relationships" (p.14). We can now derive the following theorem:

Theorem 7 *Complex organizations attempt to reduce environmental influences for their core technologies.*
$\forall x, y, z, v\ [\mathsf{CO}(x) \wedge \mathsf{TC}(x, y) \wedge \mathsf{ENVI}(y, z, v) \ \rightarrow\ \mathsf{RED}(x, z, y)]$
(Read: if x is a complex organization, and y is the technical core of x, and y is exposed to an influence z from v, then x attempts to reduce z for y.)

Proof: (Using lemmas 1 and 2, definition 1, and assumptions 4, 7, and 8.) By Def.1, complex organizations are organizations, and by Lem.1 the technical core is a suborganization. By Ass.4, an environmental influence causes uncertainty for the technical core. By Lem.2, complex organizations attempt to reduce uncertainty for their technical core. Thus, by Ass.7, organizations have control over the influences on their suborganizations. Therefore, by Ass.8, organizations attempt to reduce this environmental influence for their technical core.

$$\begin{array}{clr}
 & \mathsf{CO}(o_1) \wedge \mathsf{TC}(o_1, tc_1) \wedge \mathsf{ENVI}(tc_1, e_1, o_2) & \text{(by assumption)} \\
\Rightarrow & \mathsf{O}(o_1) & \text{(by Def.1)} \\
\Rightarrow & \mathsf{SO}(o_1, tc_1) & \text{(by Lem.1)} \\
\Rightarrow & (\exists u_1)\ \mathsf{UC}(tc_1, u_1) \wedge \mathsf{C}(e_1, u_1) & \text{(by Ass.4)} \\
\Rightarrow & \mathsf{RED}(o_1, u_1, tc_1) & \text{(by Lem.2)} \\
\Rightarrow & \mathsf{HC}(o_1, e_1) & \text{(by Ass.7)} \\
\Rightarrow & \mathsf{RED}(o_1, e_1, tc_1) & \text{(by Ass.8)}
\end{array}$$

The impact of this theorem becomes more clear in its specific predictions for the two types of environmental influences: fluctuations and constraints. The reduction of environmental fluctuations by an organization is called "buffering" in Thompson: "buffering absorbs environmental fluctuations" (p.21). We define a predicate, BUF, for buffering. For example, $\mathsf{BUF}(o, f, tc)$ says that o is buffering fluctuation f for tc.

Definition 5 *Buffering.*
$\forall x, y, z\,[\mathsf{BUF}(x, y, z) \leftrightarrow \mathsf{SO}(x, z) \wedge \mathsf{FL}(z, y, x) \wedge \mathsf{RED}(x, y, z)]$
(Read: x buffers y for z if and only if z is a suborganization of x, and z is exposed to a fluctuation y from x, and x attempts to reduce y for z.)

Typical examples of buffering are the stockpiling of materials and supplies, and the maintaining of warehouse inventories (p.20). We now have the following corollary:

Corollary 8 (of theorem 7) *Complex organizations buffer environmental fluctuations for their core technologies.*
$\forall x, y, z\,[\mathsf{CO}(x) \wedge \mathsf{TC}(x, y) \wedge \mathsf{FL}(y, z, x) \rightarrow \mathsf{BUF}(x, z, y)]$
(Read: if x is a complex organization, and y is the technical core of x, and y is exposed to a fluctuation z from x, then x buffers z for y.)

Proof: Using theorem 7, lemma 1, and definitions 2, and 5.

	$\mathsf{CO}(o_1) \wedge \mathsf{TC}(o_1, tc_1) \wedge \mathsf{FL}(tc_1, e_1, o_1)$	(by assumption)
$\Rightarrow$	$\mathsf{SO}(o_1, tc_1)$	(by Lem.1)
$\Rightarrow$	$\mathsf{ENVI}(tc_1, e_1, o_1)$	(by Def.2)
$\Rightarrow$	$\mathsf{RED}(o_1, e_1, tc_1)$	(by Thm.7)
$\Rightarrow$	$\mathsf{BUF}(o_1, e_1, tc_1)$	(by Def.5)

This corollary is a formal version of Thompson's proposition 2.2: "Under norms of rationality, organizations seek to buffer environmental influences by surrounding their technical cores with input and output components" (p.20).

The theorem also makes a specific prediction in case of environmental constraints. The reduction of environmental constraints by an organization is called "anticipating and adapting" or "forecasting" in Thompson: "To the extent that environmental fluctuations can be anticipated, however, they can be treated as **constraints** on the technical core" (p.22). We define a predicate, ANA, for anticipating and adapting. For example, $\mathsf{ANA}(o, c, tc)$ says that o is anticipating and adapting to constraint c for tc.

Definition 6 *Anticipating and adapting.*
$\forall x, y, z\,[\mathsf{ANA}(x, y, z) \leftrightarrow \mathsf{SO}(x, z) \wedge \mathsf{CS}(z, y, x) \wedge \mathsf{RED}(x, y, z)]$

(Read: x anticipates and adapts to y for z if and only if z is a suborganization of x, and z is exposed to a constraint y from x, and x attempts to reduce y for z.)

Anticipation and adaptation typically involves the reallocation of resources according to the forecasted market demand or supply constraints. This gives rise to another corollary:

Corollary 9 (of theorem 7) *Complex organizations anticipate and adapt to an environmental constraint for their core technologies.*
$\forall x, y, z\, [\mathsf{CO}(x) \wedge \mathsf{TC}(x, y) \wedge \mathsf{CS}(y, z, x) \;\rightarrow\; \mathsf{ANA}(x, z, y)]$
(Read: if x is a complex organization, and y is the technical core of x, and y is exposed to a constraint z from x, then x anticipates and adapts to z for y.)

Proof: Using theorem 7, lemma 1, and definitions 2, and 6.

$$
\begin{array}{clc}
 & \mathsf{CO}(o_1) \wedge \mathsf{TC}(o_1, tc_1) \wedge \mathsf{CS}(tc_1, e_1, o_1) & \text{(by assumption)} \\
\Rightarrow & \mathsf{SO}(o_1, tc_1) & \text{(by Lem.1)} \\
\Rightarrow & \mathsf{ENVI}(tc_1, e_1, o_1) & \text{(by Def.2)} \\
\Rightarrow & \mathsf{RED}(o_1, e_1, tc_1) & \text{(by Thm.7)} \\
\Rightarrow & \mathsf{ANA}(o_1, e_1, tc_1) & \text{(by Def.6)}
\end{array}
$$

This corollary is a formal version of Thompson's proposition 2.4: "Under norms of rationality, organizations seek to anticipate and adapt to environmental changes which cannot be buffered or leveled" (p.21).

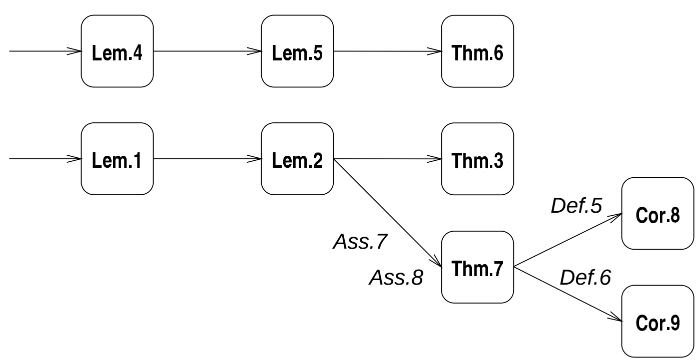

Figure 3.4: Buffering and anticipating (structure of the theory 4).

We have introduced a predicate for "having control," HC, and assumed that organizations have (some) control over environmental influences on their suborganizations (Ass.7). We further assumed that if organizations have control over the cause of something they want to reduce, they will attempt to reduce this cause (Ass.8), and we derived that complex organizations attempt to reduce environmental influences for their technical cores (Thm.7; see Fig.3.4). We defined buffering as the reduction of fluctuations, BUF (Def.5), defined anticipating and adapting as the reduction of constraints, ANA (Def.6), and derived that complex organizations will buffer environmental fluctuations (Cor.8) and anticipate and adapt to environmental constraints (Cor.9).

3.3.5 Smoothing or Leveling

In this section, we attempt to derive a formal version of Thompson's third proposition: "organizations seek to smooth out input and output transactions" (p.21). According to Thompson, organizations also attempt to reduce fluctuations in the environment (p.21). Apparently, organization have some control over specific elements of their environment. We define a predicate, CEE, for controlled elements in the environment. For example, $\mathsf{CEE}(o_1, o_2)$ says that o_2 is an element in the environment of o_1 over which o_1 has some control:[14]

Definition 7 *Controlled environmental elements.*
$\forall x, y\ [\mathsf{CEE}(x, y)\ \leftrightarrow\ \mathsf{O}(x) \wedge \forall z\ [\mathsf{ENVI}(x, z, y)\ \rightarrow\ \mathsf{HC}(x, z)]]$
(Read: y is an element in x's controlled environment if and only if x is an organization, and for all z such that x is exposed to an influence z from y it is the case that x has control over z.)

This definition, again, does not imply that organizations have unilateral control over other elements in their environment, the amount of control may be limited. Using this definition, we can now derive:

Theorem 10 *Complex organizations attempt to reduce environmental influences in their controlled environment for their technical core.*
$\forall x, y, z, v, w\ [\mathsf{CO}(x) \wedge \mathsf{TC}(x, y) \wedge \mathsf{ENVI}(y, z, x) \wedge \mathsf{CEE}(x, v) \wedge \mathsf{ENVI}(x, w, v) \wedge \mathsf{C}(w, z)\ \rightarrow\ \mathsf{RED}(x, w, y)]$
(Read: if x is a complex organization, and y is the technical core of x, and y is exposed to an influence z from x, and v is in the controlled environment of x, and x is exposed to an influence w from v, and w causes z, then x attempts to reduce w for y.)

Proof: (Using theorem 7, definition 7, and assumption 8.) By Thm.7, complex organizations reduce environmental influences for their technical core. If this environmental influence is caused by another environmental influence on an element of the controlled environment of the organization, then, by Def.7, the organization has some control over the second influence. Therefore, by Ass.8 the

[14] We wish to thank the reviewer who pointed out a deficiency in an earlier version of this definition.

organization will attempt to reduce the second influence for the organization.

$$\begin{array}{lcl} & \mathsf{CO}(o_1) \wedge \mathsf{TC}(o_1, tc_1) \wedge \mathsf{ENVI}(tc_1, e_1, o_1) \wedge \mathsf{CEE}(o_1, o_2) & \\ & \wedge \mathsf{ENVI}(o_1, e_2, o_2) \wedge \mathsf{C}(e_2, e_1) & \text{(by assumption)} \\ \Rightarrow & \mathsf{RED}(o_1, e_1, tc_1) & \text{(by Thm.7)} \\ \Rightarrow & \mathsf{HC}(o_1, e_2) & \text{(by Def.7)} \\ \Rightarrow & \mathsf{RED}(o_1, e_2, tc_1) & \text{(by Ass.8)} \end{array}$$

We will investigate the impact of this theorem by considering the specific predictions it makes for environmental fluctuations. Thompson uses the term "smoothing" for the reduction of fluctuations in the environment: "smoothing or leveling involves attempts to reduce fluctuations in the environment" (p.21). We define a predicate, SM, for smoothing. For example, $\mathsf{SM}(o, f, tc)$ says that o smoothes f for tc:

Definition 8 *Smoothing.*
$\forall x, y, z\ [\mathsf{SM}(x, y, z) \ \leftrightarrow\ \mathsf{SO}(x, z) \wedge \exists v\ [\mathsf{FL}(x, y, v) \wedge \mathsf{RED}(x, y, z)]]$
(Read: x smoothes y for z if and only if z is a suborganization of x, and there exists a v such that x is exposed to a fluctuation y from v, and x attempts to reduce y for z.)

A typical example of smoothing is price mechanisms: by charging premiums during peak periods and inducements during slow periods (p.21). Note the difference between buffering and smoothing: buffering concerns the reduction of fluctuations within an organization, whereas smoothing concerns the reduction of fluctuation in the environment. We now have the following corollary:

Corollary 11 (of theorem 10) *Complex organizations smooth environmental fluctuations in their controlled environment for their technical core.*
$\forall x, y, z, v, w\ [\mathsf{CO}(x) \wedge \mathsf{TC}(x, y) \wedge \mathsf{FL}(y, z, x) \wedge \mathsf{CEE}(x, v) \wedge \mathsf{FL}(x, w, v) \wedge \mathsf{C}(w, z) \ \rightarrow \mathsf{SM}(x, w, y)]$
(Read: if x is a complex organization, and y is the technical core of x, and y is exposed to a fluctuation z from x, and v is in the controlled environment of x, and x is exposed to a fluctuation w from v, and w causes z, then x smoothes w for y.)

Proof: Using theorem 10, lemma 1 and definitions 2 and 8.

$$\begin{array}{lcl} & \mathsf{CO}(o_1) \wedge \mathsf{TC}(o_1, tc_1) \wedge \mathsf{FL}(tc_1, e_1, o_1) \wedge \mathsf{CEE}(o_1, o_2) & \\ & \wedge \mathsf{FL}(o_1, e_2, o_2) \wedge \mathsf{C}(e_2, e_1) & \text{(by assumption)} \\ \Rightarrow & \mathsf{ENVI}(tc_1, e_1, o_1) \wedge \mathsf{ENVI}(o_1, e_2, o_2) & \text{(by Def.2)} \\ \Rightarrow & \mathsf{RED}(o_1, e_2, tc_1) & \text{(by Thm.10)} \\ \Rightarrow & \mathsf{SO}(o_1, tc_1) & \text{(by Lem.1)} \\ \Rightarrow & \mathsf{SM}(o_1, e_2, tc_1) & \text{(by Def.8)} \end{array}$$

This corollary is a formal version of proposition 2.3 in Thompson: "Under norms of rationality, organizations seek to smooth out input and output transactions" (p.21).

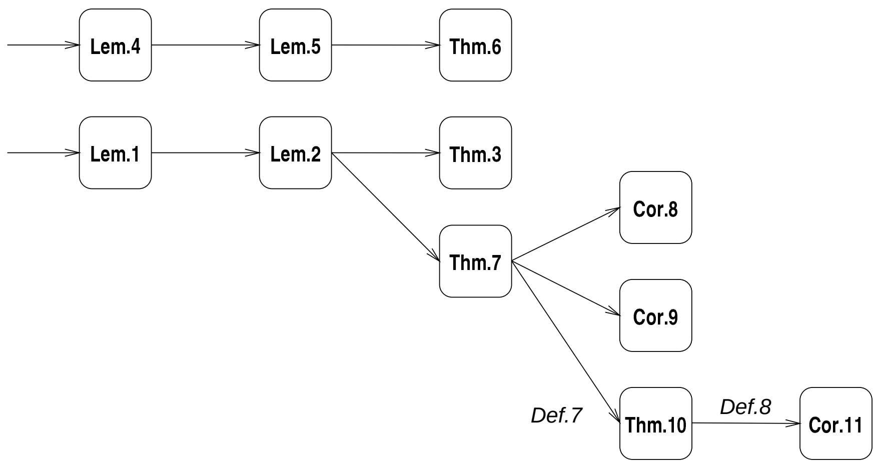

Figure 3.5: Smoothing (structure of the theory 5).

In the above section, we have defined predicates for "controlled environmental element," CEE (Def.7), and for smoothing, SM (Def.8). We derived that complex organizations attempt to reduce environmental influences in their controlled environment (Thm.10; see Fig.3.5). In particular, we derived that complex organizations attempt to smooth out fluctuations in their controlled environment (Cor.11).

3.3.6 Beyond Thompson: Negotiating

After proving a corollary on the reduction of fluctuations in the environment (Cor.11), we come across an interesting question: What about the reduction of constraints in the environment? Thompson discusses propositions for both the reduction of fluctuations and the reduction of constraints in the organization but only discusses one proposition for the reduction of fluctuations in the environment (see Table 3.2). Can complex organizations, analogous to the reduction of fluctuations, also reduce constraints in the environment? The general theorem about the reduction of influences in the environment, theorem 10, suggests that this is the case. Since there is no corresponding proposition in Thompson, we will define a new concept that treats the reduction of constraints in the environment. This term is related to the anticipation and adaptation of definition 6. The difference is that in case of anticipating and adapting the reduction of constraints takes place within the organization, whereas here the reduction of constraints

	In the Organization	In the Environment
Fluctuations ..	Buffering (Cor.8)	Smoothing (Cor.11)
Constraints ...	Anticipating and adapting (Cor.9)	...

Table 3.2: Reduction of Constraints in the Environment?

takes place in the environment. For want of a better term, we call the reduction of constraints in the environment "negotiating" and define a predicate, NEG, for it. For example, $\mathsf{NEG}(o, c, tc)$ says that o negotiates c for tc:

Definition 9 *Negotiating.*
$\forall x, y, z\ [\mathsf{NEG}(x, y, z)\ \leftrightarrow\ \mathsf{SO}(x, z) \wedge \exists v\ [\mathsf{CS}(x, y, v) \wedge \mathsf{RED}(x, y, z)]]$
(Read: x negotiates y for z if and only if z is a suborganization of x, and there exists a v such that x is exposed to a constraint y from v, and x attempts to reduce y for z.)

The definition of negotiating does not restrict the domain of the theory, it only defines NEG as shorthand for the reduction of constraints in the environment, allowing us to formulate statements more concisely.

The corollaries about buffering, smoothing, and anticipating and adapting are claims of Thompson (propositions 2.2, 2.3, and 2.4 respectively). A corollary about negotiating is not mentioned in Thompson but would complete the four logical possibilities to reduce fluctuations and constraints within organizations and within the environment (see Table 3.2). Using the same set of assumptions used to derive versions of the other propositions, we can derive the following corollary:

Corollary 12 (of theorem 10) *Complex organizations negotiate environmental constraints in their controlled environment for their technical core.*
$\forall x, y, z, v, w\ [\mathsf{CO}(x) \wedge \mathsf{TC}(x, y) \wedge \mathsf{CS}(y, z, x) \wedge \mathsf{CEE}(x, v) \wedge \mathsf{CS}(x, w, v) \wedge \mathsf{C}(w, z)\ \rightarrow\ \mathsf{NEG}(x, w, y)]$
(Read: if x is a complex organization, and y is the technical core of x, and y is exposed to a constraint z from x, and v is in the controlled environment of x, and x is exposed to a constraint w from v, and w causes z, then x negotiates w for y.)

Proof: Using theorem 10, lemma 1 and definitions 2 and 9.

$$\begin{array}{ccl}
 & \mathsf{CO}(o_1) \wedge \mathsf{TC}(o_1, tc_1) \wedge \mathsf{CS}(tc_1, e_1, o_1) \wedge \mathsf{CEE}(o_1, o_2) & \\
 & \wedge \mathsf{CS}(o_1, e_2, o_2) \wedge \mathsf{C}(e_2, e_1) & \text{(by assumption)} \\
\Rightarrow & \mathsf{ENVI}(tc_1, e_1, o_1) \wedge \mathsf{ENVI}(o_1, e_2, o_2) & \text{(by Def.2)} \\
\Rightarrow & \mathsf{RED}(o_1, e_2, tc_1) & \text{(by Thm.10)} \\
\Rightarrow & \mathsf{SO}(o_1, tc_1) & \text{(by Lem.1)}
\end{array}$$

$$\Rightarrow \qquad \mathsf{NEG}(o_1, e_2, tc_1) \qquad \text{(by Def.9)}$$

Although there is no corresponding proposition in Thompson (1967), this corollary follows from exactly the same assumptions that we used to derive the other theorems. Negotiation is a hitherto unknown implication of the theory: the theory predicts that organizations negotiate constraints in their environment.

The discovery of a new prediction of the theory gives us a new possibility for the empirical testing of the theory. If the empirical evidence supports the prediction of the theory, our confidence in the theory is strengthened. If, on the other hand, the prediction is not conform to the empirical evidence, we have falsified the theory. That is, we have at least falsified our formal reconstruction of it. We will discuss the new prediction in the light of some recent findings reported in the literature.

Can we be more specific about what "negotiating" in the sense of corollary 12 means? Negotiating is defined as attempts to reduce constraints in the environment. As a result, the reduction of constraints in the environment has an effect for all organizations that are subject to this constraint. All these organizations share the benefits of the reduction and therefore have a collective interest in reducing the constraint in the environment. For the prototypical example of a constraint, new legislation, this would mean that organizations will attempt to reduce the effects of legislation. We actually find support for such a claim in recent empirical findings on legalization (Edelman 1992; Sutton, Dobbin, Meyer, and Scott 1994; Sutton and Dobbin 1996). These studies investigate the introduction of equal employment opportunity and affirmative action (EEO/AA) laws. Their main finding is that organizations can collectively mediate the impact of these laws. "Such laws set in motion a process of definition during which organizations test and collectively construct the form and boundaries of compliance in a way that meets legal demands yet preserves managerial interests" (Edelman 1992, p.1532). This presents a clear case in which organizations successfully reduce environmental uncertainty by what we termed negotiating. Although negotiating is an implicit consequence of Thompson, it is not an unknown topic in organization theory. Finding this new prediction increases our confidence in Thompson's theory and in our formal interpretation of it.

Note that negotiating is defined as the action of individual organizations. However, consider what happens if more than one organization is facing the same environmental constraint. In that case, corollary 12 predicts that all these organizations will attempt to reduce this constraint in the environment.[15] The NEG

[15] For example, we can derive for the case of two organizations subject to the same constraint: $\forall x_1, x_2, y_1, y_2, z_1, z_2, v, w\ [\mathsf{CO}(x_1) \wedge \mathsf{TC}(x_1, y_1) \wedge \mathsf{CS}(y_1, z_1, x_1) \wedge \mathsf{CEE}(x_1, v) \wedge \mathsf{CS}(x_1, w, v) \wedge \mathsf{C}(w, z_1) \wedge \mathsf{CO}(x_2) \wedge \mathsf{TC}(x_2, y_2) \wedge \mathsf{CS}(y_2, z_2, x_2) \wedge \mathsf{CEE}(x_2, v) \wedge \mathsf{CS}(x_2, w, v) \wedge \mathsf{C}(w, z_2) \rightarrow \mathsf{NEG}(x_1, w, y_1) \wedge \mathsf{NEG}(x_2, w, y_2)]$

predicate captures the rationality of individual organizations to engage in a (collective) attempt to reduce a constraint in the environment. Of course, whether such an attempt is successful may very well depend on the proportion of organizations participating in these attempts.

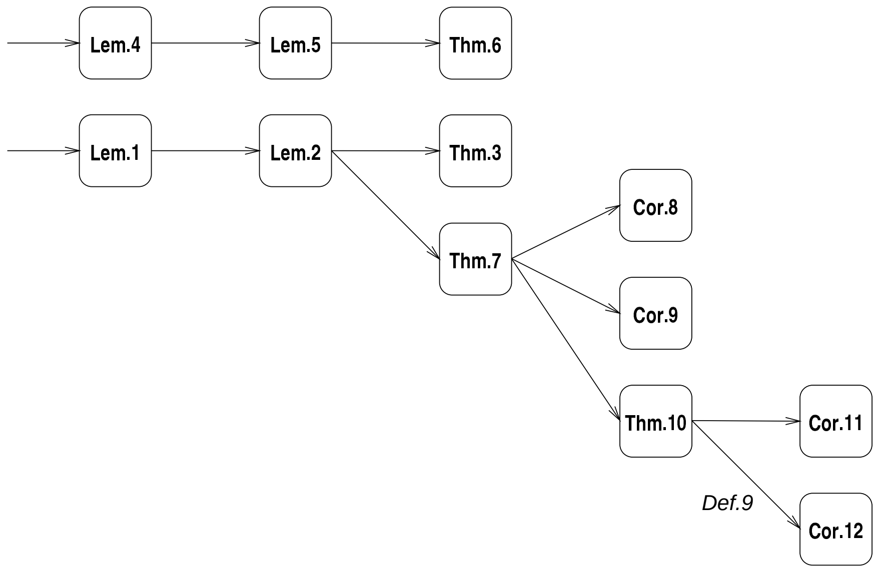

Figure 3.6: Negotiating (structure of the theory 6).

In this section, we defined NEG, a predicate for negotiating (Def.9). Negotiating is not mentioned in Thompson but can be defined in analogy with smoothing (Def.8). A corollary on negotiation (Cor.12; see Fig.3.6) can be derived from the same set of assumptions used for deriving the other theorems. This concludes our formal theory of *Organizations in Action*. We derived formal versions of four of the five propositions in chapter 2 from general axioms of organization theory.

3.4 Discussion and Conclusions

The main benefit of formal theories is that they provide clarity, both about the theory's propositions and about its argumentation structure. We have disambiguated the theory's natural language formulation by coding it in a formal language. The formal theory contains explicit definitions of complex organizations, sealing off, buffering, anticipating and adapting, and smoothing. Apart from giving a formal and precise formulation of the propositions, we also managed to trace back the argumentation for these propositions by finding reasonable and sufficient

by two applications of corollary 12, and so on.

underlying assumptions. The formal theory is consistent.[16] All lemma's, theorems, and corollaries in §3.3 are sound consequences of the premises. The proofs give sound explanations for the claims and show that the claims are unavoidable consequences of the premises.

In the formal theory of *Organizations in Action*, we used the predicates listed in Table 3.3. The theory is parsimonious because it uses only ten primitive predi-

Predicate	Meaning
Primitive:	
$\mathsf{O}(o)$	o is an organization
$\mathsf{SO}(o, so)$	so is a suborganization of o
$\mathsf{TC}(o, tc)$	tc is the technical core of o
$\mathsf{REVA}(o, tc)$	o rationally evaluates tc
$\mathsf{UC}(o, u)$	o has uncertainty u
$\mathsf{RED}(o, i, tc)$	o attempts to reduce i for tc
$\mathsf{FL}(tc, f, o)$	tc is exposed to a fluctuation f from o
$\mathsf{CS}(tc, c, o)$	tc is exposed to a constraint c from o
$\mathsf{C}(i, u)$	i causes u
$\mathsf{HC}(o, i)$	o has control over i
Defined:	
$\mathsf{CO}(o)$	o is a complex organization
$\mathsf{ENVI}(tc, i, o)$	tc is exposed to an influence i from o
$\mathsf{SEFF}(o, i, tc)$	o seals off tc from i
$\mathsf{ATO}(o)$	o is an atomic organization
$\mathsf{BUF}(o, f, tc)$	o buffers f for tc
$\mathsf{ANA}(o, c, tc)$	o anticipates and adapts to c for tc
$\mathsf{CEE}(o_1, o_2)$	o_2 is in o_1's controlled environment
$\mathsf{SM}(o, f, tc)$	o smoothes f for tc
$\mathsf{NEG}(o, c, tc)$	o negotiates c for tc

Table 3.3: Notation Used in the Formal Theory.

cates or concepts. In a formal theory, there is a clear distinction between primitive concepts and concepts that are defined in terms of the primitive concepts. A defined concept can be eliminated from the theory by replacing the concept with the expression it stands for. Note that we may only use defined concepts in the definition of other concepts if this does not lead to circularity.[17] The nine defined

[16] A formal theory is consistent if it has a model. There are computer programs currently available that generate models of a formal theory. Formally, such a model is an interpretation function that assigns objects of the domain to the constants, functions, and predicates of the theory. We used MACE (Models And Counter-Examples; McCune 1994a) to generate such a model (and thereby proved the consistency of the theory). Technical details are available upon request from the authors.

[17] It is impossible to define all the concepts of a theory: if all concepts were defined in terms

predicates do not affect the theory's parsimony because they can be substituted for by their definiens (and the same theorems would still be derivable). The ten primitive predicates are general notions of organization theory (see Table 3.3).

The assumptions characterize the application domain of the theory: the theory applies to all domains that satisfy the assumptions.[18] Our reconstruction revealed that Thompson's argumentation is largely based on fairly basic assumptions of organization theory. We used only eight assumptions and two of them (Ass.5 and 6) are only used for our discussion of noncomplex organizations. The underlying assumptions are general assumptions about organizations. As a result, the formal theory of *Organizations in Action* is a general theory, and the axiomatic structure of the theory facilitates further extension.

The argumentation structure of the theory was shown in Figure 3.6 in §3.3.6. The theory consists of two parts: a part about noncomplex organizations (lemmas 4 and 5, and theorem 6) and a part about complex organizations (roughly comparable to Thompson's propositions). Both parts are related by definition 4, which defines the notion of atomic organizations as the complement of complex organizations. There are four possible actions that organizations can perform in order to seal off their technical cores from environmental influences (see Table 3.2 in §3.3.6). Furthermore, we showed that buffering (corollary 8) and anticipating and adapting (corollary 9) are corollaries of a more general theorem about the reduction of environmental influences in the organization (theorem 7). Similarly, smoothing (corollary 11) and negotiating (corollary 12) are corollaries of a more general theorem about the reduction of influences in the environment (theorem 10). As a result, the (formal) theory of *Organizations in Action* is a coherent theory.

When comparing the theorems of our formal theory with the propositions in the original text, there are some notable differences. We did provide formal versions of Thompson's propositions 2.1, 2.2, 2.3, and 2.4. However, we did not give a formal version of proposition 2.5: "When buffering, leveling, and forecasting do not protect their technical cores from environmental fluctuations, organizations under norms of rationality resort to rationing" (Thompson 1967, p.23). Thompson's propositions 2.1 through 2.4 concern only environmental uncertainty—they all treat the reduction of environmental influences. Proposition 2.5 about rationing, in contrast, treats the reduction of the *effects* of environmental influences. In chapter 2 of *Organizations in Action*, Thompson only tells *when* organizations resort to rationing: namely, when buffering, leveling (smoothing), and forecasting (anticipating and adapting) do not protect the technical core. However, he does not explain *why* organizations resort to rationing. This explanation requires more detailed knowledge about the technological dependencies and the

of the other concepts, then some of the definitions have to be circular.

[18]The definitions are only important for determining the meaning of the defined predicates used in the assumptions (in our case only definition 2, the definition of environmental influences).

uncertainty caused by these dependencies. That is, explaining *why* organizations resort to rationing requires material properly contained in the other chapters of *Organizations in Action* (notably chapter 5 on the interdependence of components). Without including the further chapters, we are unable to explain why organizations resort to rationing, since: "Rationing is an unhappy solution, for its use signifies that the technology is not operating at its maximum" (p.23).

The formal theory identifies a number of unknown consequences of Thompson's theory. Thompson (1967) does not discuss noncomplex or atomic organizations explicitly, although he formulates his propositions for complex organizations. Atomic organizations form an interesting special case because they are more vulnerable to environmental influences. We proved that atomic organizations face the same uncertainty as their technical cores (theorem 6). Consequently, atomic organizations cannot comply with one of the main principles of "organizational rationality" as advocated by Thompson. The explicit treatment of atomic organizations is important because it gives an explanation for the fact that organizations embed their core technologies in managerial activities. These managerial activities negotiate between the technical suborganization and those who use its products. Without them, the organization would be an atomic organization and therefore open and unprotected to any environmental influence. Having a managerial level allows complex organizations to seal off their core technologies from environmental influences (theorem 3 and proposition 2.1 in Thompson). We also proved that organizations can attempt to reduce environmental uncertainty by reducing constraints in the environment, which we termed negotiating (corollary 12). Negotiating is not mentioned in Thompson but completes the four logical possibilities to reduce fluctuations and constraints within the organization and in the environment. Negotiating is a hitherto unknown implication of Thompson's theory. In other words, the theory predicts that organizations attempt reduce constraints in their environment. This negotiating is not an unknown topic in organization theory: recent empirical findings on legalization (Edelman 1992; Sutton et al. 1994; Sutton and Dobbin 1996) describe how organizations can (collectively) negotiate the impact of constraints in the environment.

3.4.1 Related and further research

We started this chapter by arguing that, although Thompson's theory remains influential, it seems to have lost much of its cachet. Recent literature focuses on alternative theories for explaining the complex relation between organizations and their environment, such as new institutionalism (Powell and DiMaggio 1991) and organizational ecology (Hannan and Freeman 1989). The feeling that Thompson's theory still has much to offer to contemporary scholars inspired us to conduct the formal analysis reported in this chapter.

The formal theory of *Organizations in Action* presented in this chapter is an axiomatic theory in which the underlying assumptions are made explicit.[19] Discussing these underlying assumptions allows us to make comparisons with other theories in organization theory. We used eight assumptions. Assumption 1 postulates the uniqueness of the technical core. This captures some of the background knowledge that had to be made explicit in order to derive the theorems. The assumption is a technicality that makes the notion of technical core more clear. The first major premises, assumptions 2 and 3, capture some the rationality principles underlying Thompson's theory. Assumption 2 states that the performance of the technical core is rationally evaluated, and assumption 3 states that organizations attempt to reduce uncertainty for rationally evaluated suborganizations. It is the engine of the theory: all the explanations of the formal theory depend on it. As Thompson (1967, p.159) states: "Uncertainty appears as the fundamental problem for complex organizations, and coping with uncertainty, as the essence of the administrative process." These two assumptions capture the core of Thompson's argument and seem typical for rational adaptation theories, such as contingency theory (Lawrence and Lorsch 1967) and resource-dependence theory (Pfeffer and Salancik 1978). Of course, institutional and ecological theorists will not readily agree with them. The (new) institutional theory is explicitly based "on the rejection of rational-actor models" (DiMaggio and Powell 1991, p.8) and organizational ecology is constructed as "an alternative to the dominant adaptation perspective" (Hannan and Freeman 1977, p.929).

One of the key assumptions of an open-systems perspective on organizations is captured by assumption 4, which states that environmental influences cause uncertainty. Although much changed since Thompson's book appeared in the sixties, this assumption seems as pertinent as ever. And even though one could argue that, on the one hand, environmental influences seem to generate less uncertainty due to current information technology and flexible work practices, one might just as well argue that, on the other hand, this is compensated for by the increased volatility of the environment. The next two assumptions, assumptions 5 and 6, are only used for the discussion of non-complex or atomic organizations.

[19] Masuch and Huang (1996) give a different formalization of Thompson in a multi-agent, action logic. The objectives of their formalization are different from ours: our objective was to recover the underlying axioms on which Thompson's argumentation is based, whereas their primary objective is to experiment with a new formal logic that is specially designed for representing actions. They argue: "actions presuppose *attitudes* and engender *change*, and both are notoriously hard to express in the extentional context of standard logics, e.g., First-Order Logic" (Masuch and Huang 1996, p.72). This may be the case, in the sense that modeling of actual actions happening in an organizational domain may be overly elaborate or complicated in first-order logic. However, modeling a theory *about* actions can very well be expressed in first-order logic, as we showed in this chapter. This corresponds with the findings of Masuch and Huang (1996) since, as it turned out, they did not need to use either the multi-agent or the action features of their logic. Due to the different objectives, the formal theory of *Organizations in Action* presented in this chapter makes the underlying argumentation structure of Thompson explicit, whereas Masuch and Huang use a series of abstract goal definitions (characterizing particular rational agents) to explain the propositions.

Assumption 5 postulates that we treat the whole organization as its technical core if it has no suborganizations. This assumption is not based on Thompson's text but expresses a convention that allowed us to discuss noncomplex or atomic organizations. Assumption 6 is a technicality that states that single influence causes a single uncertainty (a minor restriction on assumption 4). This assumption makes explicit the background knowledge that allows for the discussion of atomic organizations.

The next assumption, assumption 7, states that organizations have (some) control over influences directed at their suborganizations. This is one of the core assumptions of rational adaptation theories but will certainly be challenged by institutional and especially ecological theorists. However, these different views may be less orthogonal than they appear to be at first glance. Assumption 7 by no means implies that organizations can control their environments. First, the amount of control that organizations have may be very limited. Second, even if organizations have control, in principle, over influences, their capability to effectively use this control may depend on their capability to predict such influences. Third, even if an organization can foresee an influence, it may lack the ability to react decisively, due to complex internal decision procedures, rendering their control useless.

Finally, assumption 8 states another rationality principle underlying the theory: if an organization attempts to reduce something, and has some control over the cause of it, the organization will attempt to reduce the cause. As Thompson (1967, p.14) writes "Instrumental action is rooted on the one hand in desired outcomes and on the other hand in beliefs about cause/effect relationships." This assumption is an important part of the explanation for theorems 7 and 10 and the corollaries about buffering, smoothing, and anticipating and adapting (as well as negotiating). Assumption 8 is a general principle of rationality that seems generally acceptable. Of course, ecologists will de-emphasize the organization's capability to entertain beliefs about cause/effect relationships or, even more so, question the capacity to have control over causes. However, institutionalists and ecologists would agree that in the (according to them, improbable) event that the antecedent is satisfied, the consequent should hold as well.

The formal reconstruction revealed that Thompson's theory can be related to several alternative theories such as organizational ecology (Hannan and Freeman 1989) and new institutionalism (Powell and DiMaggio 1991). Although organizational ecology was originally presented as an alternative to rational adaptation theory, their position turns out to be more qualified. Organizational ecologists carefully distinguish between the individual intentions of organizations and the organizational outcomes. The part of *Organizations in Action* we analyzed in this chapter is treating organizational intentions (or desired outcomes in Thompson's parlance). Organizational ecologists do not necessarily reject Thompson's assumptions about the rational intentions of individual organizations but would argue that the relation between these intentions and organizational outcomes is weak—resulting in organizations being relatively inert and unable to change their structure to better match the environment (Hannan and Freeman 1984). Also Thompson does not believe that intentions and organizational outcomes are in perfect relation, for example, when he argues that "the basic threat to organizational success lies in interdependence with an uncooperative environment" (Thompson 1967, p.160). And ecologists, on the other hand, are "acknowledging that organizational changes of some kinds occur frequently and that organizations sometimes even manage to make radical changes in strategies and structures" (Hannan and Freeman 1984, p.149). The views of organizational ecologists and Thompson are not in contradiction, as can be evaluated by comparing this chapter with the formalization of organizational ecology's inertia theory (Péli et al. 1994). Although not in contradiction, there is a noticeable difference in the degree in which organizations are believed to be able to successfully realize their intentions in organizational outcomes. Several authors have made proposals for reconciling adaptation and evolutionary selection perspectives (Tushman and Romanelli 1985; Levinthal 1991; Amburgey, Kelly, and Barnett 1993).

Institutionalists have been more radical in their rejection of the rational adaptation perspective and explicitly focus on "properties of supra-individual units of analysis that cannot be reduced to aggregations or direct consequences of individuals' attributes or motives" (DiMaggio and Powell 1991, p.8). Rather surprisingly, the formal reconstruction revealed that Thompson's theory can be directly related to new institutionalism. We proved corollary 12, stating that organizations attempt to reduce constraints in the environment, which corresponds to findings reported in institutional theory (Edelman 1992; Sutton et al. 1994; Sutton and Dobbin 1996). The formal theory suggests that adaptation theories and institutional theories are not mutually inconsistent. Moreover, adaptation theories can offer explanations for phenomena that are usually conceived as requiring institutional argumentation. This affirms that adaptation theories may have been discarded too soon: they can even offer explanations beyond the domains with which they are traditionally associated.

Deriving the corollary on negotiating in the environment strengthens the connection of *Organizations in Action* with related adaptation theories such as the resource-dependence theory. Pfeffer and Salancik (1978, pp.154–157) analyze interorganizational behaviors from the perspective of uncertainty reduction. The fact that traditional adaptation theories and neoinstitutional theories can offer alternative explanations raises some interesting research questions on the limits of both approaches to organization theory. For example, Sutton and Dobbin (1996, p.794) observe two types of responses to legal uncertainty, which "sustain the neoinstitutional argument, but offer new support for efficiency and labor-control hypothesis." There seem to be no *a priori* reasons to reject adaptation-based explanations (nor to reject institutional explanations). Kraatz and Zajac (1996, p.812) explored the limits of neoinstitutional theory, and their "findings reveal surprisingly little support for neoinstitutional predictions." This leads them to conclude that "current research on organization-environment relations may underestimate the power of traditional adaptation-based explanations in organizational sociology" (p.812). Further research is needed to elucidate the relation between (rational) adaptation theories and (neo)institutional theories and to investigate how adaptation theories might be embedded in institutional theory and *vice versa*. The formalization of one dominant adaptation theory is only a small step toward clarifying the relation between adaptation theories and institutional theories.

One of our future research directions is to extend the material by incorporating further chapters of Thompson. Our main focus has been on the propositions in chapter 2. This chapter provides the crucial link between organizations that strive to use the rationality of closed-systems strategies in an open-systems environment by attempting to seal off their core technology from influences of the environment. Thompson (1967, p.1) argues that the two basic sources for uncertainty for organizations are *technologies* and *environments*. The part of Thompson we formalized in this chapter concerns uncertainty that is strictly caused by the environment. We intend to incorporate also those uncertainties that are caused by technologies or by a combination of technologies and environments. This would then allow for explaining why organizations have to resort to unhappy solutions like rationing (the missing proposition 2.5). The axiomatic structure of the theory facilitates such further extension.

Building a deductive theory of *Organizations in Action* explicated the underlying assumptions of the framework that Thompson proposes. Making the underlying assumptions explicit implies that they too can become the subject of discussion and criticism. Although we made particular efforts to motivate these assumptions by the text itself, we do not claim that we have always chosen the single best solution. However, by being explicit about our assumptions, any debate about the theory and its assumptions can be unambiguously pointed down. This allows for a constructive mode of theory building in which alternatives can be generated, evaluated on their merits, and well-argued choices can be made.

Chapter 4

Criteria for Formal Theory Building

This chapter provides practical operationalizations of criteria for evaluating scientific theories, such as the consistency and falsifiability of theories and the soundness of inferences, that take into account definitions. The precise formulation of these criteria is tailored to the use of automated theorem provers and automated model generators—generic tools from the field of automated reasoning. The use of these criteria is illustrated by applying them to a first-order logic representation of a classic organization theory, Thompson's Organizations in Action.

4.1 Introduction

Philosophy of science's classical conception of scientific theories is based on the axiomatization of theories in (first-order) logic. In such an axiomatization, the theory's predictions can be derived as theorems by the inference rules of the logic. In practice, only very few theories from the empirical sciences have been formalized in first-order logic. One of the reasons is that the calculations involved in formalizing scientific theories quickly defy manual processing. The availability of automated reasoning tools allows us to transcend these limitations. In the social sciences, this has led to renewed interest in the axiomatization of scientific theories (Péli et al. 1994; Péli and Masuch 1997; Péli 1997; Bruggeman 1997; Hannan 1998; Kamps and Pólos 1999). These authors present first-order logic versions of heretofore non-formal scientific theories.

The social sciences are renowned for the richness of their vocabulary (one of the most noticeable differences with theories in other sciences). Social science theories are usually stated using many related concepts that have subtle differences in meaning. As a result, a formal rendition of a social science theory will use a large vocabulary. We recently started to experiment with the use of definitions as a means to combine a rich vocabulary with a small number of primitive terms.

Now definitions are unlike theorems and unlike axioms. Unlike the-

> orems, definitions are not things we prove. We just declare them by fiat. But unlike axioms, we do not expect definitions to add substantive information. A definition is expected to add to our convenience, not to our knowledge. (Enderton 1972, p.154)

If dependencies between different concepts are made explicit, we may be able to define some concepts in terms of other concepts, or in terms of a smaller number of primitive concepts. If the theory contains definitions, the defined concepts can be eliminated from the theory by expansion of the definitions. Eliminating the defined concepts does not affect the theory, in the sense that the models and theorems of the theory remain the same.

This chapter provides practical operationalizations of criteria for evaluating scientific theories, such as the consistency and falsifiability of theories and the soundness of inferences. In earlier discussion of the criteria for evaluating theories, we did not distinguish between different types of premises (Kamps 1998a). Here, we want to consider cases in which some of the theory's premises are definitions.[1] In this chapter, we will provide practical operationalizations of these criteria that take definitions into account, and illustrate their use on a formal fragment of organization theory.

4.2 Logical formalization

Most social science theories are stated in ordinary language (except, of course, for mathematical theories in economics). The main obstacle for the formalization of such a discursive theory is their rational reconstruction: interpreting the text, distinguishing important claims and argumentation from other parts of the text, and reconstructing the argumentation. This reconstruction is seldom a straightforward process, although there are some useful guidelines (Fisher 1988). When the theoretical statements are singled-out, they can be formulated in first-order logic. The main benefit of the formalization of theories in logic is that it provides clarity by providing an unambiguous exposition of the theory (Suppes 1968).

[1] Following (Enderton 1972), predicate definitions are formulas of the form:

$$\forall x_1, \ldots, x_n \, [\mathsf{P}(x_1, \ldots, x_n) \;\leftrightarrow\; \varphi(x_1, \ldots, x_n)]$$

where P is a (new) n-ary relation symbol, and φ a formula in the original language (i.e., not containing P) in which only $x_1,\ldots,x_n$ may occur free. Function definitions are formulas of the form:

$$\forall x_1, \ldots, x_{n+1} \, [\mathsf{f}(x_1, \ldots, x_n) = x_{n+1} \;\leftrightarrow\; \psi(x_1, \ldots, x_{n+1})]$$

where f is a (new) n-ary function symbol, and ψ a formula in the original language (i.e., not containing f) in which only $x_1,\ldots,x_{n+1}$ may occur free. A function f is well defined if and only if the theory contains

$$\forall x_1, \ldots, x_n \exists! x_{n+1} \, [\psi(x_1, \ldots, x_{n+1})]$$

Defined symbols are only allowed in the defining formulas if this does not lead to circularity.

Moreover, the fields of logic and philosophy of science have provided a number of criteria for evaluating formal theories, such as the consistency and falsifiability of theories and the soundness of inferences. Our aim is to develop support for the axiomatization of theories in first-order logic by giving specific operationalizations of these criteria. These specific formulations are chosen such that the criteria can be established in practice with relative ease, i.e., such that existing automated reasoning tools can be used for this purpose.[2]

4.2.1 Criteria for Evaluating Theories

We will use the following notation. Let Σ denote the set of premises of a theory. A formula φ is a theorem of this theory if and only if it is a logical consequence, i.e., if and only if $\Sigma \models \varphi$. The theory itself is the set of all theorems, in symbols, $\{\varphi \mid \Sigma \models \varphi\}$.

Consistency The first and foremost criterion is the *consistency* of a theory: we can tell whether a theory in logic is contradiction-free. We can prove the consistency of a theory by generating a model of the premises. This model will be a model of the entire theory, i.e., all derivable theorems will necessarily hold in it. If a theory is inconsistent, it cannot correspond to its intended domain of application. Therefore, empirical testing should focus on identifying those premises that do not hold in its domain. The formal theory can suggest which assumptions are problematic by identifying (minimal) inconsistent subsets of the premises.

The theory is *consistent* if we can find a model $\mathcal{A}$ such that the premises are satisfied: $\mathcal{A} \models \Sigma$. A theory is *inconsistent* if we can derive a contradiction, $\perp$, from the premises: $\Sigma \vdash \perp$.

Soundness Another criterion is the *soundness* of arguments in a theory: we can tell whether a claim undeniably follows from the given premises. Many of our basic propositions are inaccessible for direct empirical testing. Such propositions can be indirectly tested by their testable implications (Hempel 1966). In case of unsound argumentation, examining the counterexamples provides useful guidance for revision of the theory.

A theorem φ is *sound* if it can be derived from the premises $\Sigma \vdash \varphi$. A theorem φ is *unsound* (i.e., φ is no theorem) if we can construct a counterexample, that is, a model $\mathcal{A}$ in which the premises hold, and the theorem is false: $\mathcal{A} \models \Sigma$ and $\mathcal{A} \not\models \varphi$.

[2] Of course, we hope that this will be regarded as an original contribution, but the claim to originality is a difficult one to establish. The novelty is in the combination of ideas from various fields and our debts to the fields of logic and philosophy of science fan out much further than specific citations indicate.

Falsifiability Falsifiability of a theorem means that it is possible to refute the theorem. Self-contained or tautological statements are unfalsifiable—their truth does not depend on the empirical assumptions of the theory. Falsifiability is an essential property of scientific theories (Popper 1959). If no state of affairs can falsify a theory, empirical testing can only reassert its trivial validity. A theory is falsifiable if it contains at least one falsifiable theorem.

An initial operationalization of falsifiability is: a theorem φ is *unfalsifiable* if it can be derived from an empty set of premises: $\vdash \varphi$ and *falsifiable* if we can construct a model $\mathcal{A}$ (of the language) in which the theorem is false: $\mathcal{A} \not\models \varphi$. Note that we cannot require this model to be a model of the theory. A theorem is necessarily true in all models of the theory (otherwise it would not be a theorem). We should therefore ignore the axioms of the theory, and consider arbitrary models of the language. This initial formulation works for some unfalsifiable statements like tautologies, but may fail in the context of definitions. Consider the following simple example: a theory that contains a definition of a Bachelor predicate.

$$\forall x\ [\mathsf{Bachelor}(x)\ \leftrightarrow\ \mathsf{Man}(x) \wedge \mathsf{Unmarried}(x)]$$

In such a theory, we will have the following theorem.

$$\forall x\ [\mathsf{Bachelor}(x)\ \rightarrow\ \mathsf{Man}(x)]$$

Using falsifiability as formulated above, we would conclude that this statement is falsifiable. It is easy to construct models (of the language) in which the theorem is false, that is, models in which an object a occurs such that $\mathsf{Bachelor}(a)$ is assigned *true* and $\mathsf{Man}(a)$ is *false*. However, if we expand the definition, then the theorem becomes

$$\forall x\ [\mathsf{Man}(x) \wedge \mathsf{Unmarried}(x)\ \rightarrow\ \mathsf{Man}(x)]$$

This expanded version of the theorem is tautologically true and therefore unfalsifiable by the above formulation. This is problematic, since "we do not expect definitions to add substantive information" (Enderton 1972, see the above quotation). Therefore, the elimination of defined concepts should also not affect any of the criteria for evaluating theories. Although we have to ignore the axioms, we must take the definitions into account when establishing falsifiability.

In the context of definitions $\Sigma_{\mathsf{def}} \subseteq \Sigma$, a theorem φ is *falsifiable* if there exists a model $\mathcal{A}$ in which the definitions hold, and the theorem is false: $\mathcal{A} \models \Sigma_{\mathsf{def}}$ and $\mathcal{A} \not\models \varphi$. A theorem φ is *unfalsifiable* if the theorem can be derived from only the set of definitions: $\Sigma_{\mathsf{def}} \vdash \varphi$.

Satisfiability Satisfiability is the counterpart of falsifiability. Satisfiability of a theorem ensures that it can be fulfilled. Self-contradictory statements are unsatisfiable. It makes no sense to subject an unsatisfiable theorem to empirical testing, since it is impossible to find instances that corroborate the theorem.

In the context of definitions $\Sigma_{\mathsf{def}} \subseteq \Sigma$, a theorem φ is *satisfiable* if there exists a model $\mathcal{A}$ in which the definitions hold, and the theorem is true: $\mathcal{A} \models \Sigma_{\mathsf{def}}$ and $\mathcal{A} \models \varphi$. A theorem φ is *unsatisfiable* if we can derive a contradiction from only the set of definitions and the theorem: $\Sigma_{\mathsf{def}} \cup \{\varphi\} \vdash \perp$.

Contingency Theorems that can both be fulfilled and refuted are called *contingent*—their validity strictly depends on the axioms, they are neither tautologically true, nor self-contradictory. The empirical investigation of non-contingent theorems does not make any sense because the outcome is predetermined.

A theorem is *contingent* if it is both satisfiable and falsifiable. A theorem is *non-contingent* if it is unsatisfiable or unfalsifiable (or both).[3]

Independence The set of premises should contain no superfluous premises, that is to say each premise should be *independent* of the other premises. An set of premises is independent if all premises are independent of the other premises.

A premise σ is independent of the other premises if neither its truth, nor its falsity is implied by the other premises: if $\Sigma \setminus \{\sigma\} \nvdash \sigma$ and $\Sigma \setminus \{\sigma\} \nvdash \neg\sigma$. The first part amounts to finding a model $\mathcal{A}$ such that $\mathcal{A} \models \Sigma \setminus \{\sigma\} \cup \{\sigma\}$, which is of course a model $\mathcal{A} \models \Sigma$, equivalent to our operationalization of consistency. The second part amounts to finding a model $\mathcal{B}$ such that $\mathcal{B} \models \Sigma \setminus \{\sigma\} \cup \{\neg\sigma\}$.

Further advantages Making the inference structure of a theory explicit will make it possible to assess the theory's *explanatory and predictive power* (by looking at the set of theorems because these are the predictions of the theory, and the proofs give explanations for them); its *domain* or scope (by investigating the models of the theory); its *coherence* (for example, a theory may turn out to have unrelated or independent parts); its *parsimony* (for example, it may turn out that some assumptions are not necessary, or can be relaxed); and other properties.

This list of criteria should not be considered as final, and may be extended with various other criteria if they appear useful. For example, one of the criteria that is frequently discussed is *completeness*. A theory is complete if of any two contradictory sentences formulated in terms of theory under consideration at least one sentence can be proved in that theory (Tarski 1946, p.135). One attempts to formulate a set of axioms that would allow for the derivation of the largest possible set of statements that can be established as valid (and not a single false one). Due to a logical law, the law of the excluded middle, one of every two contradictory sentences must be true. Hence, this would ultimately lead to a complete theory (provided that the vocabulary is finite). It is clear that none of the sociological axiomatizations we discuss realizes this ultimate ideal, it is

[3] We can also apply the above operationalizations to the definitions. All definitions turn out to be non-contingent, since any definition will be unfalsifiable. This neatly corresponds to the fact that definitions should not add to our knowledge (Enderton 1972, p.154).

even unclear whether there will ever be a complete axiomatization of a significant empirical science theory.

4.2.2 Computational tools

The above operationalizations require particular proof or model searches for establishing the criteria. The field of automated reasoning has provided us with automated theorem provers and model generators—generic tools that can directly be used for computational testing of the criteria. Automated theorem provers, such as OTTER (McCune 1994b), are programs that are designed to find proofs of theorems. Typical theorem provers use *reductio ad absurdum*, that is, the program attempts to derive a contradiction from the premises and the negation of the theorem. A theorem prover can also be used to prove that a theory is inconsistent if it can derive a contradiction from the set of premises of the theory. Automated model generators, such as MACE (McCune 1994a), are programs that can find (small) models of sets of sentences. A model generator can prove the consistency of a theory, if it can generate a model of the premises. It can also be used to prove underivability of a conjecture, by attempting to generate a model of the premises in which the conjecture is false. Table 4.1 summarizes how to test

CRITERION	OTTER	MACE
Consistency		$(\exists\mathcal{A})\mathcal{A} \models \Sigma$
Inconsistency	$\Sigma \vdash \bot$	
Soundness	$\Sigma \cup \{\neg\varphi\} \vdash \bot$	
Unsoundness		$(\exists\mathcal{A})\mathcal{A} \models \Sigma \cup \{\neg\varphi\}$
Falsifiability		$(\exists\mathcal{A})\mathcal{A} \models \Sigma_{\mathsf{def}} \cup \{\neg\varphi\}$
Unfalsifiability	$\Sigma_{\mathsf{def}} \cup \{\neg\varphi\} \vdash \bot$	
Satisfiability		$(\exists\mathcal{A})\mathcal{A} \models \Sigma_{\mathsf{def}} \cup \{\varphi\}$
Unsatisfiability	$\Sigma_{\mathsf{def}} \cup \{\varphi\} \vdash \bot$	
Independence		$(\exists\mathcal{A})\ \mathcal{A} \models \Sigma$ and $(\forall\sigma \in \Sigma)(\exists\mathcal{B})\ \mathcal{B} \models \Sigma \setminus \{\sigma\} \cup \{\neg\sigma\}$

Note: Σ denotes a premise set, with $\Sigma_{\mathsf{def}} \subseteq \Sigma$ the definitions in this set and $\sigma \in \Sigma$ an individual premise, and φ a conjecture or theorem.

Table 4.1: Criteria and Automated Reasoning Tools.

for the criteria.[4]

The decision to use either automated theorem provers or model generators is not an arbitrary one. Although (syntactic) proof-theoretic and (semantic) model-

[4] OTTER and MACE are companion programs that can read the same input format. This facilitates switching between theorem proving and model searching, depending on which type of tool is most suitable for the specific proof or disproof attempt at hand. Kamps (1998a) discusses the use of these tools for formal theory building in detail.

theoretical characterizations are logically equivalent, determining the criteria is a different matter. First, there is a fundamental restriction on what we can hope to achieve because first-order logic is undecidable. Although undecidable, first-order logic is semi-decidable—we can prove $\Sigma \vdash \bot$ if it is true. We suggest the use of theorem provers only for cases in which a contradiction can be derived. The other cases are, in general, undecidable (causing theorem provers to run on for ever). However, finding finite models is, again, decidable—we can find a finite model $\mathcal{A}$ such that $\mathcal{A} \models \Sigma$ if there exists such a finite model. Current model generators can only find finite models—even only models of small cardinalities. We have no solution for those cases in which only infinite models exist (or only models that are too large for current programs). Second, the tests we suggest to evaluate the criteria do not only prove a criterion, but also present a specific proof or model that is available for further inspection. In simple cases theorem provers may terminate after exhausting their search space without finding a contradiction, proving indirectly that the problem set is consistent, or that a conjecture is not derivable. Even in these cases a direct proof is far more informative, for example, if we can find specific counterexamples to a conjecture, it is immediately clear why our proof attempt has failed.

4.3 Case Study: A Formal Theory of Organizations in Action

We will illustrate the criteria outlined above by applying them to the formal theory of Organizations in Action (Kamps and Pólos 1999), a formal rendition of (Thompson 1967). Thompson (1967) is one of the classic contributions to organization theory: it provides a framework that unifies the perspective treating organizations as closed systems, with the perspective that focuses on the dependencies between organizations and their environment. This framework has influenced much of the subsequent research in organization theory. Thompson (1967) is an ordinary language text, in which only the main propositions are clearly outlined. Kamps and Pólos (1999) provide a formal rendition of the first chapters of the book, by reconstructing the argumentation used in the text.

The formal theory uses the predicate symbols reprinted in Table 4.2. Although the text of (Thompson 1967) does not contain explicit definitions, the use of terminology in the text strongly suggests strict dependencies between several important concepts. This allowed for the definition of these concepts in terms of a small number of primitive notions of organization theory. Table 4.3 lists the premises (both definitions and assumptions) and theorems of the formal theory. In the formal theory, the key propositions of (Thompson 1967) can be derived as theorems (theorem 3, corollaries 8, 9, and 11). The proofs of these theorems are based on a reconstruction of the argumentation in the text (assumptions 1–8).

Predicates	
Primitive:	
$\mathsf{O}(o)$	o is an organization
$\mathsf{SO}(o, so)$	so is a suborganization of o
$\mathsf{TC}(o, tc)$	tc is the technical core of o
$\mathsf{REVA}(o, tc)$	tc is rationally evaluated by o
$\mathsf{UC}(o, u)$	o has uncertainty u
$\mathsf{RED}(o, i, tc)$	o attempts to reduce i for tc
$\mathsf{FL}(tc, f, o)$	tc is exposed to a fluctuation f from o
$\mathsf{CS}(tc, c, o)$	tc is exposed to a constraint c from o
$\mathsf{C}(i, u)$	i causes u
$\mathsf{HC}(o, i)$	o has control over i
Defined:	
$\mathsf{CO}(o)$	o is a complex organization
$\mathsf{ENVI}(tc, i, o)$	tc is exposed to an influence i from o
$\mathsf{SEFF}(o, i, tc)$	o seals off tc from i
$\mathsf{ATO}(o)$	o is an atomic organization
$\mathsf{BUF}(o, f, tc)$	o buffers f for tc
$\mathsf{ANA}(o, c, tc)$	o anticipates and adapts to c for tc
$\mathsf{CEE}(o_1, o_2)$	o_2 is in o_1's controlled environment
$\mathsf{SM}(o, f, tc)$	o smoothes f for tc
$\mathsf{NEG}(o, c, tc)$	o negotiates c for tc

Table 4.2: Predicates (Kamps and Pólos 1999).

Premises and Theorems	
Premises:	
Def.1	$\forall x\,[\mathsf{CO}(x) \leftrightarrow \mathsf{O}(x) \wedge \exists y\,[\mathsf{SO}(x,y) \wedge \mathsf{TC}(x,y)]]$
Def.2	$\forall x,y,z\,[\mathsf{ENVI}(x,y,z) \leftrightarrow \mathsf{FL}(x,y,z) \vee \mathsf{CS}(x,y,z)]$
Def.3	$\forall x,y,z\,[\mathsf{SEFF}(x,y,z) \leftrightarrow \mathsf{SO}(x,z) \wedge \exists v,w\,[\mathsf{ENVI}(z,y,v) \wedge \mathsf{UC}(z,w) \wedge \mathsf{C}(y,w) \wedge \mathsf{RED}(x,w,z)]]$
Def.4	$\forall x\,[\mathsf{ATO}(x) \leftrightarrow \mathsf{O}(x) \wedge \neg\mathsf{CO}(x)]$
Def.5	$\forall x,y,z\,[\mathsf{BUF}(x,y,z) \leftrightarrow \mathsf{SO}(x,z) \wedge \mathsf{FL}(z,y,x) \wedge \mathsf{RED}(x,y,z)]$
Def.6	$\forall x,y,z\,[\mathsf{ANA}(x,y,z) \leftrightarrow \mathsf{SO}(x,z) \wedge \mathsf{CS}(z,y,x) \wedge \mathsf{RED}(x,y,z)]$
Def.7	$\forall x,y\,[\mathsf{CEE}(x,y) \leftrightarrow \mathsf{O}(x) \wedge \forall z\,[\mathsf{ENVI}(x,z,y) \rightarrow \mathsf{HC}(x,z)]]$
Def.8	$\forall x,y,z\,[\mathsf{SM}(x,y,z) \leftrightarrow \mathsf{SO}(x,z) \wedge \exists v\,[\mathsf{FL}(x,y,v) \wedge \mathsf{RED}(x,y,z)]]$
Def.9	$\forall x,y,z\,[\mathsf{NEG}(x,y,z) \leftrightarrow \mathsf{SO}(x,z) \wedge \exists v\,[\mathsf{CS}(x,y,v) \wedge \mathsf{RED}(x,y,z)]]$
Ass.1	$\forall x,y,z\,[\mathsf{TC}(x,y) \wedge \mathsf{TC}(x,z) \rightarrow y = z]$
Ass.2	$\forall x,y\,[\mathsf{TC}(x,y) \rightarrow \mathsf{REVA}(x,y)]$
Ass.3	$\forall x,y,z\,[\mathsf{SO}(x,y) \wedge \mathsf{REVA}(x,y) \wedge \mathsf{UC}(y,z) \rightarrow \mathsf{RED}(x,z,y)]$
Ass.4	$\forall x,y,z\,[\mathsf{ENVI}(x,y,z) \rightarrow \exists v\,[\mathsf{UC}(x,v) \wedge \mathsf{C}(y,v)]]$
Ass.5	$\forall x,y\,[\mathsf{O}(x) \wedge \mathsf{TC}(x,y) \wedge \neg\mathsf{SO}(x,y) \rightarrow x = y]$
Ass.6	$\forall x,y,z,v,w\,[\mathsf{ENVI}(x,y,z) \wedge \mathsf{UC}(x,v) \wedge \mathsf{C}(y,v) \wedge \mathsf{UC}(x,w) \wedge \mathsf{C}(y,w) \rightarrow v = w]$
Ass.7	$\forall x,y,z,v\,[\mathsf{O}(x) \wedge \mathsf{SO}(x,y) \wedge \mathsf{ENVI}(y,z,v) \rightarrow \mathsf{HC}(x,z)]$
Ass.8	$\forall x,y,z,v\,[\mathsf{RED}(x,y,z) \wedge \mathsf{C}(v,y) \wedge \mathsf{HC}(x,v) \rightarrow \mathsf{RED}(x,v,z)]$
Theorems:	
Lem.1	$\forall x,y\,[\mathsf{CO}(x) \wedge \mathsf{TC}(x,y) \rightarrow \mathsf{SO}(x,y)]$
Lem.2	$\forall x,y,z\,[\mathsf{CO}(x) \wedge \mathsf{TC}(x,y) \wedge \mathsf{UC}(y,z) \rightarrow \mathsf{RED}(x,z,y)]$
Thm.3	$\forall x,y,z,v\,[\mathsf{CO}(x) \wedge \mathsf{TC}(x,y) \wedge \mathsf{ENVI}(y,z,v) \rightarrow \mathsf{SEFF}(x,z,y)]$
Lem.4	$\forall x\,[\mathsf{ATO}(x) \leftrightarrow \mathsf{O}(x) \wedge \neg\exists y\,[\mathsf{SO}(x,y) \wedge \mathsf{TC}(x,y)]]$
Lem.5	$\forall x,y\,[\mathsf{ATO}(x) \wedge \mathsf{TC}(x,y) \rightarrow x = y]$
Thm.6	$\forall x,y,z,v,w\,[\mathsf{ATO}(x) \wedge \mathsf{TC}(x,y) \wedge \mathsf{ENVI}(y,v,z) \wedge \mathsf{UC}(y,w) \wedge \mathsf{C}(v,w) \rightarrow \exists u\,[\mathsf{UC}(x,u) \wedge \mathsf{C}(v,u) \wedge w = u]]$
Thm.7	$\forall x,y,z,v\,[\mathsf{CO}(x) \wedge \mathsf{TC}(x,y) \wedge \mathsf{ENVI}(y,z,v) \rightarrow \mathsf{RED}(x,z,y)]$
Cor.8	$\forall x,y,z\,[\mathsf{CO}(x) \wedge \mathsf{TC}(x,y) \wedge \mathsf{FL}(y,z,x) \rightarrow \mathsf{BUF}(x,z,y)]$
Cor.9	$\forall x,y,z\,[\mathsf{CO}(x) \wedge \mathsf{TC}(x,y) \wedge \mathsf{CS}(y,z,x) \rightarrow \mathsf{ANA}(x,z,y)]$
Thm.10	$\forall x,y,z,v,w\,[\mathsf{CO}(x) \wedge \mathsf{TC}(x,y) \wedge \mathsf{ENVI}(y,z,x) \wedge \mathsf{CEE}(x,v) \wedge \mathsf{ENVI}(x,w,v) \wedge \mathsf{C}(w,z) \rightarrow \mathsf{RED}(x,w,y)]$
Cor.11	$\forall x,y,z,v,w\,[\mathsf{CO}(x) \wedge \mathsf{TC}(x,y) \wedge \mathsf{FL}(y,z,x) \wedge \mathsf{CEE}(x,v) \wedge \mathsf{FL}(x,w,v) \wedge \mathsf{C}(w,z) \rightarrow \mathsf{SM}(x,w,y)]$
Cor.12	$\forall x,y,z,v,w\,[\mathsf{CO}(x) \wedge \mathsf{TC}(x,y) \wedge \mathsf{CS}(y,z,x) \wedge \mathsf{CEE}(x,v) \wedge \mathsf{CS}(x,w,v) \wedge \mathsf{C}(w,z) \rightarrow \mathsf{NEG}(x,w,y)]$

Table 4.3: A Formal Theory of Organizations in Action (Kamps and Pólos 1999).

	Consistent	Sound	Falsifiable	Satisfiable	Contingent
Lem.1	yes	yes	yes	yes	yes
Lem.2		yes	yes	yes	yes
Thm.3		yes	yes	yes	yes
Lem.4		yes	no	yes	no
Lem.5		yes	yes	yes	yes
Thm.6		yes	yes	yes	yes
Thm.7		yes	yes	yes	yes
Cor.8		yes	yes	yes	yes
Cor.9		yes	yes	yes	yes
Thm.10		yes	yes	yes	yes
Cor.11		yes	yes	yes	yes
Cor.12		yes	yes	yes	yes

Table 4.4: Evaluating the Theory.

Additionally, the formal theory explains why the theory is restricted to a particular type of organizations (theorem 6). Moreover, it derives a heretofore unknown implication of the theory (corollary 12) that relates Thompson's theory to recent empirical findings and current developments in organization theory. For detailed discussion we refer the reader to (Kamps and Pólos 1999).

The criteria of the previous section played an important role during the construction of the formal theory. Table 4.4 give an assessment of the final version of the theory in terms of the criteria.

Consistency Using an automated model generator it is easy to find models of the theory. Mace produced a model of cardinality 4 within a second (a model having universe $\{0, 1, 2, 3\}$, reprinted in tables 4.5 and 4.6). This is a prototypical model of the theory corresponding to the claims of theorem 3, theorem 7, and corollary 8. It represents an organization that seals its core technologies off from environmental fluctuations, by the use of buffering (for example the stockpiling of materials and supplies). It is easy to verify that all premises (and theorems) hold in the model—the model proves that the theory is consistent.

Finding any arbitrary model of the theory is, formally speaking, sufficient to prove its consistency. We can find models on smaller cardinalities. For example, on cardinality 1 there exists a trivial model that assigns *false* to all predicates. In practice, we try to find more natural models of the theory. That is, we can examine the models and see if they conform to our mental models of the theory. This is an easy safeguard against hidden inconsistencies—theories that are only

O	
0	T
1	F
2	F
3	F

SO	0	1	2	3
0	F	T	F	F
1	F	F	F	F
2	F	F	F	F
3	F	F	F	F

TC	0	1	2	3
0	F	T	F	F
1	F	F	F	F
2	F	F	F	F
3	F	F	F	F

REVA	0	1	2	3
0	F	T	F	F
1	F	F	F	F
2	F	F	F	F
3	F	F	F	F

UC	0	1	2	3
0	F	F	F	F
1	F	F	F	T
2	F	F	F	F
3	F	F	F	F

$\mathsf{RED}(0,3,1) = \text{T}$ and $\mathsf{RED}(0,2,1) = \text{T}$ and $\mathsf{RED}(x,y,z) = \text{F}$ otherwise.

$\mathsf{FL}(1,2,0) = \text{T}$ and $\mathsf{FL}(x,y,z) = \text{F}$ otherwise.

$\mathsf{CS}(x,y,z) = \text{F}$ for all x, y, z in $\{0,1,2,3\}$.

C	0	1	2	3
0	F	F	F	F
1	F	F	F	F
2	F	F	F	T
3	F	F	F	F

HC	0	1	2	3
0	F	F	T	F
1	F	F	F	F
2	F	F	F	F
3	F	F	F	F

Table 4.5: A Model of the Theory (only primitives).

CO	
0	T
1	F
2	F
3	F

$\mathsf{ENVI}(1,2,0) = \text{T}$ and $\mathsf{ENVI}(x,y,z) = \text{F}$ otherwise.

$\mathsf{SEFF}(0,2,1) = \text{T}$ and $\mathsf{SEFF}(x,y,z) = \text{F}$ otherwise.

$\mathsf{BUFF}(0,2,1) = \text{T}$ and $\mathsf{BUFF}(x,y,z) = \text{F}$ otherwise.

Table 4.6: Selected Defined Predicates (extending Table 4.5).

consistent because background knowledge has remained implicit.[5] We can look for prototypical models of the theory directly by adding premises that express appropriate initial conditions (typically existential statements). If we find a model of this enlarged set of premises, it is obviously also a model of the theory.

Soundness In the final formal theory (as reprinted in Table 4.3), all theorems are derivable. Figure 4.1 shows the inference structure of the formal theory. None

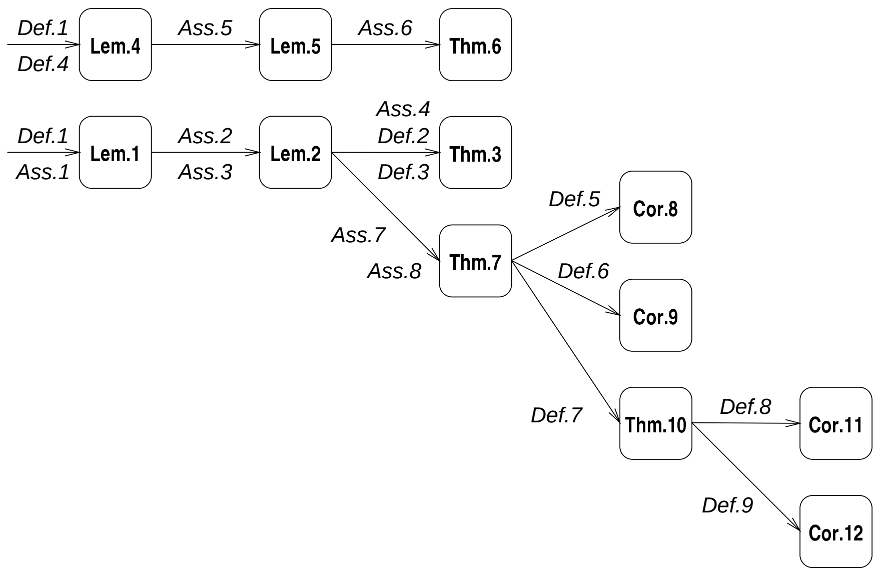

Figure 4.1: Inferences in the Formal Theory

of the proofs is very complex (the automated theorem prover OTTER required only 12 seconds for the longest proof).

The original text of the theory presupposes common background knowledge—assumptions taken for granted in the substantive field. In order for the theorems

[5]Consider the following simple example:

$$\forall x\,[\mathsf{Dog}(x) \;\rightarrow\; \mathsf{Bark}(x)]$$
$$\mathsf{Rottweiler}(\mathsf{Johnny})$$
$$\neg\mathsf{Bark}(\mathsf{Johnny})$$

Can we find a model of this theory? Yes, although inspection of the models will reveal that in every model of the theory, Johnny the rottweiler is not a dog. These models are nonintended models because of the (implicit) background knowledge that rottweilers are a particular breed of dogs.

$$\forall x\,[\mathsf{Rottweiler}(x) \;\rightarrow\; \mathsf{Dog}(x)]$$

Adding this assumption to the theory will make it inconsistent, that is, we can then derive a contradiction from it.

to be deductively derivable, several implicit assumptions had to be added to the theory (notably assumptions 1, 6 and 7). Which precise background assumptions to add is one of the thorniest problems in the formalization of a theory, requiring a deep understanding of the substantive field under consideration. Fortunately, the formal tools can help: suppose we cannot derive a theorem due to a missing background assumption. We can prove that the theorem is underivable by generating counterexamples, that is, models of the premises in which the theorem is false. If the unsoundness of the theorem is due to missing background knowledge, inspection of the counterexamples will reveal that they are nonintended models—models that violate our common sense, or implicit background assumptions from the substantive domain. For example, we found models of organizations having more than one operational core (conflicting with assumption 1, which is implicit in the original text). We can make the theorem derivable if we add sufficient assumptions to exclude these nonintended models from the theory.

Falsifiability We tried to prove the falsifiability of the theorems as discussed above: by finding a model in which the definitions hold and the theorem is false. We failed to find such a model for lemma 4. As it turns out, lemma 4 can be derived from just definitions 1 and 4—proving that this lemma is unfalsifiable. Lemma 4 is true by definition, and therefore does not make an empirical claim. If we would subject lemma 4 to empirical testing, we will be unable to refute it, but can at best reassert the trivial validity of the statement.

Fortunately, the other theorems of the formal theory are falsifiable. For each of these theorems, we can find models of the definitions in which the theorem is false (not reproduced here). Mace generated these models in a matter of seconds.

Satisfiability/Contingency For proving the satisfiability of the theorems, we need to find models that make both the definitions and the theorem true. The model of tables 4.5 and 4.6 also proves the satisfiability of all theorems. As a result, we can conclude that only lemma 4 is non-contingent—it is not an empirical statement, but its truth is determined by virtue of the definitions only. The other theorems make empirical claims that can, in principle, be corroborated or refuted by empirical testing.

Independence For proving independence of the premises, we need to show that neither the truth, nor the falsity of a premise is implied by the other premises. The model of tables 4.5 and 4.6 also proves that the falsity of each premise cannot be implied by the other premises. We can show that neither the truth is implied by, for each premise, generating models in which the premise is false, and the other premises hold. Mace has no problems generating these models (not reproduced here), thereby proving that the premises are independent—none of them is superfluous.

4.4 Discussion and Conclusions

This chapter discussed the axiomatization of scientific theories in first-order logic. We provided practical operationalizations of criteria for evaluating scientific theories, such as the consistency and falsifiability of theories and the soundness of inferences. The precise formulation of these criteria is tailored to the use of computational support. The tests for the criteria, in practice amounting to particular proof or model searches, can be directly performed by existing automated reasoning tools.

The efficient treatment of definitions is one of the basic research problems in automated reasoning (Wos 1988, Problem 30). A naive approach is to eliminate all defined predicates and functions from the problem set by expansion of the definitions. This, however, also eliminates useful ways chunking information and as a result it "increases the likelihood that a program will get lost" (Wos 1988, p.62).[6] Since only few problems are provable without the expansion of (some of) the definitions, dealing with definitions is a difficult challenge for automated reasoning tools. As a result, most automated theorem provers treat definitions and axioms alike (a notable exception is [Giunchiglia and Walsh 1989]). Interestingly, the above argument does not seem to apply to automated model generators.[7] The search space of automated model generators is the set of all possible models, i.e., all possible interpretation functions of the vocabulary. Reducing the vocabulary of the formal language by eliminating the defined concepts will proportionally reduce this search space. Moreover, after eliminating the defined concepts, the definitions themselves can be removed from the problem set, which reduces the number of "constraints" that need to be taken into account when deciding whether a particular interpretation is a model of the problem set.

We used the criteria to evaluate a formal rendition of a classic organization theory (Kamps and Pólos 1999). Assessing the criteria allows for an exact evaluation of the merits of a theory. In some cases, this may reveal important facts about the theory, for example, the case study showed that one of the derived statements is unfalsifiable—empirical investigation of it is futile. However, we do not view the criteria as rigid, final tests. Quite the opposite, in our experience the criteria are especially useful during the process of formalizing a theory. During the construction of a formal theory, the criteria can provide useful feedback on how to revise the theory in case of a deficiency. For example, examining

[6] This is substantiated by the theory in our case study: after eliminating all defined predicates, OTTER proved several theorems slightly faster, but some others slower. These are preliminary observations without taking into account the time needed expand the definitions (a preprocessing step that is done once for any number of queries). This expansion was done manually but is of no great complexity: since definitions are not allowed to be circular, we only need to expand each definition once.

[7] For the theory in our case study, eliminating defined concepts gave significantly better performance on all model searches (roughly halving MACE's processor and memory usage). Again, these are preliminary observations that do not consider the preprocessing of definitions.

counterexamples can reveal which implicit (background) assumptions need to be added to the theory. There are, of course, important principled and practical limitations to the axiomatization of theories in first-order logic: the undecidability of first-order logic, the scientific knowledge available in the substantive domains, or the availability of resources like processor power, memory, and time. There is yet no equivocal answer to the question whether it is possible, or even desirable, to axiomatize large parts of substantive domains. Axiomatization is often viewed as the ultimate step in the lifetime of a scientific theory—the axioms are frozen in their final form, and active research moves on to areas where still progress can be made. The main motivation for the research reported in this chapter is that the formalization of theories can play a broader role: it need not end the life of a theory, but rather contribute to its further development.

Chapter 5

The Process of Axiomatizing Scientific Theories

This chapter discusses the axiomatization of scientific theories in formal logic. Such an axiomatization is traditionally viewed as the ultimate step in the justification of a theory. A first-order logic rendition of a theory gives an explicit, unambiguous exposition of the theory. This allows, in turn, for testing for a number of criteria (such as consistency, soundness of derivations, satisfiability and falsifiability of theorems) that can be evaluated using generic tools from automated reasoning. We can make a rigorous evaluation of a scientific theory by assessing these criteria. However, as it turns out, these criteria exceed their use as rigid, final tests and are especially useful during the process of formalizing a theory. The criteria can provide useful feedback on how to revise the theory in case of a deficiency. For example, they can identify implicit (background) assumptions of the theory. As a result, the tools of the context of justification can also play an important role in the revision of the theory—an activity belonging to the context of discovery.

We will give a detailed illustration by the formalization of a theory fragment from Zetterberg's "On Theory and Verification in Sociology."

5.1 Introduction

In recent years, several authors started working on the formal reconstruction of ordinary language, social science theories. The main focus is on organization theory, and especially on the branch based on a natural selection perspective on organizations (Hannan and Freeman 1989). This resulted in the logical formalization or axiomatization of several parts of "organizational ecology": The inertia theory was formalized in (Péli et al. 1994); the life-history theory in (Péli and Masuch 1997); the niche width theory in (Péli 1997; Bruggeman 1997); and the age-dependence theory in (Hannan 1998). Further efforts were directed at a classical version of contingency theory, Thompson (1967) "Organizations in Action," as formalized in (Kamps and Pólos 1999) and at several relatively simple theories, such as Hage (1965) "An axiomatic theory of organizations" as formalized in (Kamps 1998a) and a theory from Zetterberg (1965) "On Theory and Verification in Sociology" that we treat as a case study in this chapter. These efforts resulted in first-order logic versions of previously non-formal scientific theories, which reconstruct the textual argumentation of the original texts by rendering it in formal logic. A notable exception is Hannan (1998), who uses the logical formalization to develop new theory "to clarify an area of research characterized by conflicting claims and divergent empirical findings" (p.126).

Following Reichenbach (1938, pp.6–7), scientific activities are traditionally dichotomized into the context of justification and the context of discovery. The axiomatization of theories in (first-order) logic is traditionally viewed as the final step in the justification of a theory. A first-order logic rendition of a theory gives an explicit, unambiguous exposition of the theory. This allows, in turn, for testing for a number of criteria (such as consistency, soundness of derivations, satisfiability and falsifiability of theorems) that can be evaluated using generic tools from automated reasoning. We can make a rigorous evaluation of a scientific theory by assessing these criteria. However, as it turns out, these criteria exceed their use as rigid, final tests and are especially useful during the process of formalizing a theory. The criteria can provide useful feedback on how to revise the theory in case of a deficiency—an activity belonging to the context of discovery. For example, they can identify implicit (background) assumptions of the theory. As a result, the tools of the context of justification can also play an important role in the context of discovery.

This chapter is structured as follows: First, in §5.2, we will discuss the axiomatization of theories; next, in §5.3, we will illustrate this by axiomatizing the theory of (Zetterberg 1965); in §5.4 we will critically review and repeatedly revise the formal theory; and finally, in §5.5, we will draw conclusions and discuss issues related to our work.

5.2 Axiomatizing Theories

5.2.1 The Product of Formalization

The axiomatization or logical formalization of non-formal theories consists of the interlinked activities of rational reconstruction (reconstructing the claims, premises, and argumentation of a theory) and formal modeling (capturing the claims as theorems that are provable from explicit assumptions). The main benefit of the formalization of theories in logic is that it provides clarity (Suppes 1968). It will provide an unambiguous exposition of the theory, containing explicit axioms and theorems. Moreover, the logic allows us to formulate a number of criteria for evaluating theories. In the previous chapter, we discussed various criteria for evaluating theories, including the consistency of the theory and the soundness of derivations. These criteria can be used for assessing relevant properties of the theory.

5.2.2 The Process of Formalization

Most social science theories are stated in ordinary language (for example in articles in social science journals). As a result, the main obstacle for logical formalizing such a discursive theory is their rational reconstruction: interpreting the text, distinguishing important claims and argumentation from other parts of the text, and reconstructing the argumentation. This reconstruction is seldom a straightforward process, although there are some useful guidelines (see, for example, the method in Fisher 1988). When the theoretical statements are singled-out, they can be formulated in first-order logic. This initial formalization can be evaluated by the criteria such as consistency, and soundness of arguments.

A strict justification point of view would stop after evaluating the theory by these criteria. However, since theories stated in ordinary language are typically partial and incomplete, it is highly unlikely that our initial formalization of the theory is completely satisfactory. For example, the initial formalization may turn out to be inconsistent, or some of the claims may not be derivable. Finding such undesirable properties does reveal deficiencies of our initial formal rendition of the theory. Is it justified to pass this verdict on to the original theory? This is not necessarily the case and, instead, we may attempt to revise our initial formalization such that it is a better reconstruction of the original theory. Fortunately, analyzing the criteria can provide useful feedback for the revision of our initial formalization.

Recover from Inconsistency

Although theories in natural language rarely contain conspicuous contradictions, the ambiguities of ordinary language can easily obscure them. As a result, our initial formalization of a theory may turn out to be inconsistent.

The proof of the inconsistency of the theory (as provided by an automated theorem prover) is a derivation of a contradiction from a specific subset of the premises. Examining this proof will clarify what caused the inconsistency, which may, in turn, suggest how to resolve the contradiction. For example, by changing a definition, or by making some assumption weaker. After revising the premises we have to repeat the test for consistency, to ensure that the modifications are sufficient. Also, a premise set may contain several different contradictions. In these cases, repeated testing for inconsistencies allows for a piecemeal revision of the theory.

Recover from Unsound Argumentation

Since authors typically assume a body of common background knowledge, it is unlikely that all needed premises are mentioned in the source text. One of the thorniest problems during the logical formalization of a theory, is to make these implicit background assumptions explicit. As a result, some of the theory's claims may turn out to be underivable in our initial formalization of the theory. Of course, it can also be the case that a claim turns out to be a false conjecture.

The proof that the derivation of a claim is unsound (as provided by an automated model generator) is a counterexample, that is, a model of the premises in which the claim is false. Examining this model gives important feedback for possible revision of the theory. On the one hand, it may be the case that, although a model of the premises, it is a non-intended model in the sense that it seems highly unlikely that the author intended this model to belong to his or her theory. For example, the model may conflict with common sense, or with common background knowledge in the domain of the theory. In this case we have to add this common sense or background knowledge to the premises of the theory, and see if we can now derive the claim. There may be several implicit underlying assumptions, which give rise to different counterexamples. On the other hand, the model may be an intended, faithful model of the theory and in this model the claim is false. In this case, the original claim is too strong to be supported by the theory, and we have to modify the claim. A typical example is the case in which the claim is overstated, and the model presents a known exception that should be taken into account. After revising the claim we can test whether we can derive this weaker claim. There may be other exceptions to the claim and repeated tests for unsoundness allows for a piecemeal treatment of them. In some cases, the claim may have to be retracted altogether, or we are forced to add further assumptions that will restrict the theory's domain of application such that the claim will hold on this restricted domain of the theory.

As a result, the process of formalizing theories proceeds through several iterations—it is a cyclic process in which the formal theory is repeatedly revised. Moreover, these modifications may have an impact on the original theory. Consider the case in which the original theory is inconsistent. If we can resolve the inconsistency in the formal rendition, we can translate this revision back to the original theory. Consider the case in which the original theory contains a hiatus. If we can find reasonable assumptions that make the claims derivable in the formal theory, we can, again, translate these added assumptions back to the original theory. It may also be the case that the formal theory reveals that certain restricting assumptions of the original theory are not necessary or can be relaxed. In short, in these cases the formal theory and the original theory evolve in parallel.

In the next section, we will give a detailed illustration of this process by discussing the formalization of a theory fragment from Zetterberg (1965)'s '*On*

Theory and Verification in Sociology.'

5.3 Case Study: Zetterberg's Theory

This section contains a logical formalization of Zetterberg's *On Theory and Verification in Sociology* (1965). Zetterberg's theory is stated in ordinary language—it is not a formal theory—but the main propositions are clearly outlined. It is an 'axiomatic theory' consisting of 10 propositions.[1] The ten propositions are about five 'variables' of social groups (p.159):

1. the number of associates per member in the group;
2. the solidarity of the group;
3. the consensus of the beliefs, values, and norms in the group;
4. the division of labor in the group; and
5. the extent to which persons are rejected (excluded) from the group when they violate norms.

Unfortunately, Zetterberg does not elaborate on the sociological meaning of these concepts, and leaves their further interpretation to the reader.[2] Presumably, the number of associates per member is indicating the (average) number of members with whom a particular member interacts. In some groups with a well-defined role structure, members may only interact with few other members, while in some small groups virtually every pair of members interacts. The solidarity of the group typically indicates the community of interests and objectives (usually viewed as the counterpart of competition between members). The distinction between solidarity and consensus of the beliefs, values, and norms in the group is rather elusive, although it is of course possible that there is consensus about the differences of interests or objectives of individual members (they are, so to say, 'agreeing to disagree'). The division of labor in a group refers to the structural necessity of members to interact in order to function as a group. Finally, the number of rejections per deviant indicates the extent to which persons are rejected (excluded) from the group when they violate norms. In closely knitted

[1]Note that it is exceptional for a sociological theory to be stated in an axiomatic form. This perspicuous structure makes Zetterberg (1965) a suitable candidate for a case study for we can focus on the formal modeling of the theory without doing an extensive reconstruction of the argumentation.

[2]This can be explained in part by the fact that Zetterberg is mainly interested in methodology. An earlier version of this theory was introduced as a "fictitious example" to illustrate the advantages of axiomatic theory for sociological research (Zetterberg 1955, p.534). He also refers to it as "a somewhat distorted version of Durkheim's theory of division of labor" (Zetterberg 1965, p.160).

groups, minor deviations may result in exclusion of members, whereas in loosely affiliated groups major deviations are permissible.

Table 5.1 lists ten propositions relating these five variables. Zetterberg regards

P.1	The greater the division of labor, the greater the consensus.
P.2	The greater the solidarity, the greater the number of associates per member.
P.3	The greater the number of associates per member, the greater the consensus.
P.4	The greater the consensus, the smaller the number of rejections per deviant.
P.5	The greater the division of labor, the smaller the number of rejections per deviant.
P.6	The greater the number of associates per member, the smaller the number of rejections per deviant.
P.7	The greater the division of labor, the greater the solidarity.
P.8	The greater the solidarity, the greater the consensus.
P.9	The greater the number of associates per member, the greater the division of labor.
P.10	The greater the solidarity, the smaller the number of rejections per deviant.

Note: All propositions are from Zetterberg (1965, pp.159–160).

Table 5.1: The propositions of Zetterberg's theory.

the last four propositions (labeled 7 through 10) as axioms of the theory. These last four propositions show, according to Zetterberg, some relation to Durkheim's classical work on the division of labor (Durkheim 1893). The first six propositions (labeled 1 through 6) are claimed to be derivable from the four axioms using "the deduction rules of ordinary language" (p.163). According to Zetterberg, we can make the following derivations (p.161): Proposition **P.1** can be inferred from **P.7** and **P.8**; Proposition **P.2** from **P.7** and **P.9**; Proposition **P.3** from **P.8** and derived proposition **P.2**; Proposition **P.4** from **P.8** and **P.10**; Proposition **P.5** from **P.7** and **P.10**; and Proposition **P.6** from **P.9** and derived proposition **P.5**.

5.3.1 Formalization

We use first-order logic to formalize Zetterberg's propositions. In the formalization we use unary functions to represent the five variables (see Table 5.2).

$\mathsf{napm}(x)$	number of associates per member in group x
$\mathsf{soli}(x)$	solidarity of group x
$\mathsf{cons}(x)$	consensus of the beliefs, values, and norms in group x
$\mathsf{dlab}(x)$	division of labor in group x
$\mathsf{nrpd}(x)$	number of rejections per deviant in group x
$x > y$	x is greater than y (or y is smaller than x)

Table 5.2: Functions and predicates.

5.3.2 Axioms

Zetterberg (1965, p.160) regards Propositions 7 through 10 as axioms of the theory. A rendering of these propositions in first-order logic is presented in Table 5.3.

A.7	$\forall x, y\ [\mathsf{dlab}(x) > \mathsf{dlab}(y)\ \rightarrow\ \mathsf{soli}(x) > \mathsf{soli}(y)]$
A.8	$\forall x, y\ [\mathsf{soli}(x) > \mathsf{soli}(y)\ \rightarrow\ \mathsf{cons}(x) > \mathsf{cons}(y)]$
A.9	$\forall x, y\ [\mathsf{napm}(x) > \mathsf{napm}(y)\ \rightarrow\ \mathsf{dlab}(x) > \mathsf{dlab}(y)]$
A.10	$\forall x, y\ [\mathsf{soli}(x) > \mathsf{soli}(y)\ \rightarrow\ \mathsf{nrpd}(y) > \mathsf{nrpd}(x)]$

Table 5.3: A formalization of Propositions 7–10.

Is the (formal) theory consistent? The theory is consistent if it has a model. We used the automated model generator MACE in an attempt to generate models of the axioms. Table 5.4 shows a model of Propositions 7–10.[3] We can easily verify

	napm	soli	cons	dlab	nrpd
0	0	1	1	1	0
1	0	0	0	0	1

>	0	1
0	F	F
1	T	F

Table 5.4: A model of Axioms 7–10.

that the axioms hold in the model of Table 5.4: Axiom 7 holds because the group with greater division of labor (0) has also a greater solidarity; Axiom 8 holds because the group with greater solidarity (0) also has greater consensus; Axiom 9 holds vacuously because there is no group with greater number of associates per member; and finally, Axiom 10 holds because the group with greater solidarity (0) has a smaller number of rejections per deviant. Zetterberg's theory has a model, therefore it is consistent.

[3] On domain size 2, MACE generates 1024 models of the axioms.

5.3.3 Theorems

We will now investigate the explanatory or predictive power of the theory. Zetterberg claims that "these four propositions [7–10] can be used to derive the other findings which thus become theorems" (p.161). In these derivation he uses "the deduction rules of natural language" (p.163). A first-order logic rendering of these (intended) theorems is presented in Table 5.5.

T.1	$\forall x, y\ [\mathsf{dlab}(x) > \mathsf{dlab}(y) \ \rightarrow\ \mathsf{cons}(x) > \mathsf{cons}(y)]$
T.2	$\forall x, y\ [\mathsf{soli}(x) > \mathsf{soli}(y) \ \rightarrow\ \mathsf{napm}(x) > \mathsf{napm}(y)]$
T.3	$\forall x, y\ [\mathsf{napm}(x) > \mathsf{napm}(y) \ \rightarrow\ \mathsf{cons}(x) > \mathsf{cons}(y)]$
T.4	$\forall x, y\ [\mathsf{cons}(x) > \mathsf{cons}(y) \ \rightarrow\ \mathsf{nrpd}(y) > \mathsf{nrpd}(x)]$
T.5	$\forall x, y\ [\mathsf{dlab}(x) > \mathsf{dlab}(y) \ \rightarrow\ \mathsf{nrpd}(y) > \mathsf{nrpd}(x)]$
T.6	$\forall x, y\ [\mathsf{napm}(x) > \mathsf{napm}(y) \ \rightarrow\ \mathsf{nrpd}(y) > \mathsf{nrpd}(x)]$

Table 5.5: A formalization of Propositions 1–6.

We used the automated theorem prover OTTER to attempt proving Propositions 1–6 from the four axioms. We can find proofs of Proposition 1, 3, 5, and 6.

Theorem 1 *The greater the division of labor, the greater the consensus*:

$$\forall x, y\ [\mathsf{dlab}(x) > \mathsf{dlab}(y) \ \rightarrow\ \mathsf{cons}(x) > \mathsf{cons}(y)]$$

Proof: OTTER can derive **T.1** from **A.7** and **A.8**.

Theorem 3 *The greater the number of associates per member, the greater the consensus*:

$$\forall x, y\ [\mathsf{napm}(x) > \mathsf{napm}(y) \ \rightarrow\ \mathsf{cons}(x) > \mathsf{cons}(y)]$$

Proof: OTTER can derive **T.3** from **A.9**, **A.7**, and **A.8**.

Theorem 5 *The greater the division of labor, the smaller the number of rejections per deviant*:

$$\forall x, y\ [\mathsf{dlab}(x) > \mathsf{dlab}(y) \ \rightarrow\ \mathsf{nrpd}(y) > \mathsf{nrpd}(x)]$$

Proof: OTTER can derive **T.5** from **A.7** and **A.10**.

Theorem 6 *The greater the number of associates per member, the smaller the number of rejections per deviant*:

$$\forall x, y\ [\mathsf{napm}(x) > \mathsf{napm}(y) \ \rightarrow\ \mathsf{nrpd}(y) > \mathsf{nrpd}(x)]$$

Proof: OTTER can derive **T.6** from **A.9**, **A.7**, and **A.10**.

However, we cannot derive Propositions 2, the greater the solidarity, the greater the number of associates per member:

$$\forall x, y\ [\mathsf{soli}(x) > \mathsf{soli}(y)\ \rightarrow\ \mathsf{napm}(x) > \mathsf{napm}(y)]$$

Nor can we derive Proposition 4, the greater the consensus, the smaller the number of rejections per deviant:

$$\forall x, y\ [\mathsf{cons}(x) > \mathsf{cons}(y)\ \rightarrow\ \mathsf{nrpd}(y) > \mathsf{nrpd}(x)]$$

According to Zetterberg is Proposition 2 derivable from Proposition 7 and 9 (p.161). We can prove that the derivation of **T.2** is unsound if we can find a counterexample, that is, if we can find a model of the axioms in which the intended theorem does not hold. As it turns out, we can construct a counterexample to this claim (see Table 5.6). Finding this counterexample proves that **T.2** is not

	napm	soli	cons	dlab	nrpd
0	0	0	0	0	0
1	1	0	0	0	0

>	0	1
0	T	F
1	F	F

Table 5.6: Counterexample to Propositions 2.

a consequence of **A.7–10**. Zetterberg also claims that Proposition 4 is derivable from Proposition 8 and 10 (p.161). Again, we can construct a counterexample to this claim (see Table 5.7). These counterexamples prove that Propositions 2

	napm	soli	cons	dlab	nrpd
0	0	0	0	0	0
1	0	0	1	0	0

>	0	1
0	F	T
1	F	F

Table 5.7: Counterexample to Propositions 4.

and 4 are no theorems of the axioms. On the positive side, we can derive a new proposition as theorem:

Theorem 11 *The greater the number of associates per member, the greater the solidarity*:

$$\forall x, y\ [\mathsf{napm}(x) > \mathsf{napm}(y)\ \rightarrow\ \mathsf{soli}(x) > \mathsf{soli}(y)]$$

Proof: OTTER can derive **T.11** from **A.9** and **A.7**.

Theorem 11 is the converse of the (underivable) Proposition 2. Of course, there are also many trivial theorems that can be derived, such as all tautologies or theorems already subsumed by the spelled-out theorems.

5.3.4 Recapitulating

We formalized Zetterberg's propositions as **T.1–6** and **A.7–10** in Table 5.5 and 5.3 respectively. The resulting theory contains propositions **T.1**, **T.3**, **T.5**, and **T.6** as theorems that can be derived from the four axioms, **A.7–10**. However, we proved that propositions **T.2** and **T.4** cannot be derived and are no theorems of the theory. We are able to prove theorem **T.3** despite the suggestion that its proof depends upon theorem **T.2**. Fortunately, we could find a (different) proof of **T.3** using axioms **A.9**, **A.7**, and **A.10** (or, equivalently, using axiom **A.9** and theorem **T.1**).[4] The proofs of **T.1**, **T.5**, and **T.6** are according to the suggested inferences. Table 5.8 compares the verbal theory with our formal version. Our

Verbal theory:	Formal theory:	
P.1 from **P.7** and **P.8**	**T.1** from **A.7** and **A.8**	$\surd$
P.2 from **P.7** and **P.9**	No theorem **T.2**	
P.3 from **P.8** and derived **P.2**	**T.3** from **A.9**, **A.7** and **A.8**	$\surd$ (!)
P.4 from **P.8** and **P.10**	No theorem **T.4**	
P.5 from **P.7** and **P.10**	**T.5** from **A.7** and **A.10**	$\surd$
P.6 from **P.9** and derived **P.5**	**T.6** from **A.9**, **A.7** and **A.10**	$\surd$

Table 5.8: The Verbal and Formal Theory

formalization attempt seems moderately successful: we could derive four of the six intended theorems.

5.4 Revision

Our formalization of Zetterberg's theory prompts a number of questions: How do these results in the formal version of the theory relate to the original version? Have we uncovered a deficiency in the original theory? Went something wrong in our reconstruction of his arguments? Can we come up with a different interpretation in which all the intended theorems are derivable? We will try to answer these questions in this section.

5.4.1 Limit Explanatory/Predictive Power

One option is to do nothing: We have given, arguably, the most natural first-order rendition of the propositions. If two of the conjectures are not derivable in the formal version of the theory, then we have an important argument to discard these propositions as false conjectures. That is, we keep Axioms 7–10 as stated

[4] In fact, the situation would not change much if we could prove **T.2**, since the derivation of **T.3** from **T.2** and **A.8** is unsound!

formalized in Table 5.3, and reduce the set of theorems to Theorem 1, 3, and 5–6 as stated in Table 5.5. As Zetterberg (1965, p.163) remarks, "our deductions are not too precise, so long as our concepts are defined in normal prose, and the deduction rules of ordinary language are used." It may be not unreasonable to assume that now that we use formal logic, having a precisely defined language and a strict notion of deduction, we have to discard two of the intended, ordinary language theorems as false conjectures.

5.4.2 Nonintended Models and Real Counterexamples

Before passing such a severe verdict on the theory, it seems more reasonable to first make a detailed examination of the evidence. It is important to note that we did more than proving that the two conjectures are underivable, since we produced the counterexamples that prove the underivability. These counterexamples are available for inspection.

When analyzing the counterexample of Table 5.6, we immediate find a strange feature. The model gives an unnatural interpretation of the ">"-relation: $(0 > 0)$ is true, whereas $(0 > 1)$, $(1 > 0)$, and $(1 > 1)$ are false. This model is not one of the models we intended to be models of the theory. It seems unreasonable to discard a conjecture because of the existence of such a nonintended model—a model that violates our commonsense or background knowledge of the substantive domain. Any exposition of a theory presupposes a set of common background knowledge. In a formal exposition of a theory, relevant parts of this implicit background knowledge have to be added explicitly to the theory. Finding counterexamples that are nonintended models clearly indicates what background knowledge should be added to the formal theory. We would assume that the ">"-relation denotes a strictly larger relation (Meaning Postulate 1) and that on the domain $\{0, 1\}$ it holds that $(1 > 0)$ (Meaning Postulate 2). We decide to add these axioms to the theory (see Table 5.9).

MP.1	$\forall x, y\, \neg[(x > y) \wedge (y > x)]$
MP.2	$(1 > 0)$

Table 5.9: Background assumptions.

Adding these two background assumptions to the theory will substantially reduce the number of models of the theory. Specifically, and more importantly, the models in Tables 5.6 and 5.7 (that were counterexamples to **T.2** and **T.4**) do no longer belong to the theory. This means that we can retry to prove **T.2** and **T.4**. Unfortunately, OTTER is still unable to prove either of them: there must exist other counterexamples to Propositions 2 and 4.

MACE proves that Proposition 2 is still underivable by generating the counterexamples in Table 5.10 (without loss of generality, we use only **A.7**, **A.9**,

MP.1, and **MP.2**).[5] Similarly, for Proposition 4 we find the counterexamples in

	soli	dlab	napm
0	0	0	0
1	1	0	0
0	0	0	1
1	1	0	1
0	0	0	0
1	1	1	0
0	0	0	1
1	1	1	1
0	0	1	0
1	1	1	0
0	0	1	0
1	1	1	0

>	0	1
0	F	F
1	T	F

Table 5.10: Counterexamples to Propositions 2.

Table 5.11 (without loss of generality, we use only **A.8**, **A.10**, **MP.1**, and **MP.2**).[6] The models in Tables 5.10 and 5.11 cannot easily be discarded as nonintended

	cons	soli	nrpd
0	0	0	0
1	1	0	0
0	0	0	0
1	1	0	1
0	0	0	0
1	1	0	1
0	0	1	0
1	1	1	0
0	0	1	0
1	1	1	1
0	0	1	1
1	1	1	1

>	0	1
0	F	F
1	T	F

Table 5.11: Counterexamples to Propositions 4.

models, they seem to be genuine counterexamples to propositions 2 and 4. We will now explore different ways of dealing with them.

[5] Mace finds twelve models on domain size 2. Table 5.10 lists six of them, the other six are copies with arguments 0 and 1 interchanged.

[6] Mace finds twelve models on domain size 2, those listed in Table 5.11 and the six copies with arguments 0 and 1 interchanged.

5.4.3 Weaken Theorems

We will now regard the counterexamples as faithful models of the theory. Therefore, if we want to retain Propositions 2 and 4, we must reformulate them such that they hold in the models that are counterexamples to their initial formulation, e.g., in models such as in Tables 5.10 and 5.11 respectively. In order to hold in a larger set of models, the reformulated propositions must be weaker than the original formulations. We will attempt to find reformulated propositions that are provable from the axioms, yet still close to the original formulation of the theorems. Note that this does not change the theory: we will only change our exposition of the theory by singling out more consequences of the axioms explicitly. Since all consequences of the axiom set are, by definition, part of the theory, the theory does not change if we single out more of them.

In case of Proposition 2, '*the greater the solidarity, the greater the number of associates per member*', the new formulation must hold in (at least) the models of Table 5.10. Let us analyze these models: they all have the form $\mathsf{soli}(1) > \mathsf{soli}(0)$ and $\mathsf{napm}(0) = \mathsf{napm}(1)$. A weaker version of Proposition 2 that also holds in these models is: '*The greater the solidarity, the greater* ***or equal*** *the number of associates per member*':

$$\textbf{T.2}^- \qquad \forall x, y\,[\mathsf{soli}(x) > \mathsf{soli}(y) \;\rightarrow\; \neg(\mathsf{napm}(y) > \mathsf{napm}(x))]$$

Now that we have reformulated Proposition 2, we can retry to prove it. (We have not changed the axioms, so the model in Table 5.4 still proves that the theory is consistent.) Although we have dealt with the (type of) counterexamples in Table 5.10, there may still be other counterexamples. There turn out to be none, since we can prove the reformulated theorem.

Theorem 2$^-$ *The greater the solidarity, the greater or equal the number of associates per member*:

$$\forall x, y\,[\mathsf{soli}(x) > \mathsf{soli}(y) \;\rightarrow\; \neg(\mathsf{napm}(y) > \mathsf{napm}(x))]$$

Proof: OTTER can derive **T.2$^-$** from **A.7**, **A.9**, and **MP.1**.[7]

We can try to apply the same strategy to Proposition 4. Let us analyze the models in Table 5.11. There are counterexamples of the form $\mathsf{cons}(1) > \mathsf{cons}(0)$ and $\mathsf{nrpd}(0) = \mathsf{nrpd}(1)$ and counterexamples of the form $\mathsf{cons}(1) > \mathsf{cons}(0)$ and $\mathsf{nrpd}(0) > \mathsf{nrpd}(1)$. Moreover, there are also models in which Proposition 4 holds (so, these are not counterexamples), and these have the form $\mathsf{cons}(1) > \mathsf{cons}(0)$ and $\mathsf{nrpd}(1) > \mathsf{nrpd}(0)$ (for example, the model in Table 5.4).

[7] Note that the proof of **T.2$^-$** requires the background assumption **MP.1**. If we had not discovered the relevance of this assumption earlier, then all counterexamples would be non-intended models violating **MP.1**. In other words, we would have discovered this background assumption now by investigating counterexamples to **T.2$^-$**.

It seems like the axioms put hardly any constraint on the relation between consensus and number of rejections per deviant! A weaker version of Proposition 4 that holds in all these models must be very weak—so weak that is a tautology. For example: '*The greater the consensus, the smaller,* ***or equal, or higher*** *the number of rejections per deviant.*' Theorem 4^- is derivable from an empty premise set.[8]

Theorem 4^- *The greater the consensus, the smaller,* ***or equal, or higher*** *the number of rejections per deviant*:

$$\begin{aligned}\forall x, y\,[\mathsf{cons}(x) > \mathsf{cons}(y) \quad \rightarrow \quad & (\mathsf{nrpd}(y) > \mathsf{nrpd}(x) \\ & \vee (\neg(\mathsf{nrpd}(x) > \mathsf{nrpd}(y)) \wedge \neg(\mathsf{nrpd}(y) > \mathsf{nrpd}(x))) \\ & \vee \mathsf{nrpd}(x) > \mathsf{nrpd}(y))]\end{aligned}$$

Proof: OTTER can derive **T.4**$^-$ from an empty premise set.

As a result, the statement of **T.4**$^-$ makes no empirical claim and is unfalsifiable. Summarizing: we can derive a (weaker) version of Proposition 2 (Theorem 2^-), but this does not help in deriving Proposition 4.

5.4.4 Strengthen Axioms

A final option is to regard the counterexamples as models that are outside the intended domain of the theory. That is, the theorems do hold but on a smaller domain. Therefore, we must reformulate the axioms such that models such as those in Tables 5.10 and 5.11 are no longer models of the revised axioms. In order to hold in a smaller set of models, the revised axioms must be stronger than the original axioms. There are several ways to make Proposition 2 derivable. We choose a way that follows the original argumentation as closely as possible.

Zetterberg argues that Proposition 2 is derivable from Proposition 7 and 9. As stated above, the counterexamples of Table 5.10 have the form $\mathsf{soli}(1) > \mathsf{soli}(0)$ and $\mathsf{napm}(0) = \mathsf{napm}(1)$. A natural way to exclude these counterexamples is to add as an axiom that *the greater the solidarity, the greater the number of associates per member*:

$$\forall x, y\,[\mathsf{soli}(x) > \mathsf{soli}(y) \;\rightarrow\; \mathsf{napm}(x) > \mathsf{napm}(y)]$$

However, that would mean that we add Proposition 2 as an axiom (which means that we can trivially derive it).

A second way to exclude the counterexamples is to add as axioms that *the greater the solidarity, the greater the division of labor* (the converse of Proposition 7):

[8] We interpret here equal as 'not smaller than' and 'not higher than.'

$$\textbf{A.12} \qquad \forall x, y\, [\mathsf{soli}(x) > \mathsf{soli}(y) \;\rightarrow\; \mathsf{dlab}(x) > \mathsf{dlab}(y)]$$

and *the greater the division of labor, the greater the number of associates per member* (the converse of Proposition 9):

$$\textbf{A.13} \qquad \forall x, y\, [\mathsf{dlab}(x) > \mathsf{dlab}(y) \;\rightarrow\; \mathsf{napm}(x) > \mathsf{napm}(y)]$$

Adding these two axioms takes care of (the type of) counterexamples in Table 5.10 (Axiom 12 removes models 1–2, and 5–6; Axiom 13 removes models 3–4). We can now make a new attempt at proving Proposition 2 using the revised set of axioms. If this attempt succeeds, the revision has removed all counterexamples. As it turns out, we can indeed prove Proposition 2.

Theorem 2 *The greater the solidarity, the greater the number of associates per member*:

$$\forall x, y\, [\mathsf{soli}(x) > \mathsf{soli}(y) \;\rightarrow\; \mathsf{napm}(x) > \mathsf{napm}(y)]$$

Proof: OTTER can derive **T.2** from **A.12** and **A.13**.

In case of Proposition 4 we must take care of the counterexamples in Table 5.11 (after adding Axioms 12 and 13, Proposition 4 is still not derivable, and the same counterexamples remain). All counterexamples in Table 5.11 are of the form $\mathsf{cons}(1) > \mathsf{cons}(0)$ and $\mathsf{soli}(0) = \mathsf{soli}(1)$,

To restore the argumentation for Proposition 4 we only need to add an axiom stating that *the greater the consensus, the greater the solidarity* (the converse of Axiom 8):

$$\textbf{A.14} \qquad \forall x, y\, [\mathsf{cons}(x) > \mathsf{cons}(y) \;\rightarrow\; \mathsf{soli}(x) > \mathsf{soli}(y)]$$

This takes care of all counterexamples, because a new proof attempt of Proposition 4 succeeds.

Theorem 4 *The greater the consensus, the smaller the number of rejections per deviant*:

$$\forall x, y\, [\mathsf{cons}(x) > \mathsf{cons}(y) \;\rightarrow\; \mathsf{nrpd}(y) > \mathsf{nrpd}(x)]$$

Proof: OTTER can derive **T.4** from **A.14** and **A.10**.

One way to view the revisions above, is to regard this as adding three extra axioms, **A.12–14**. But there is another way: the initial formal version of Propositions 7, 8, and 9 can be combined with their converses, axioms **A.12**, **A.14**, and **A.13** respectively, into a revised version of these propositions as presented in Table 5.12. Viewed in this way, we have revised our formalization of Propositions 7–9, by interpreting the natural language statements like '*the greater the solidarity, the greater the consensus*' as a logical biconditional. That is, as 'the solidarity is greater if and only if the consensus is greater.' This interpretation is

A.7*	$\forall x, y\ [\mathsf{dlab}(x) > \mathsf{dlab}(y) \leftrightarrow \mathsf{soli}(x) > \mathsf{soli}(y)]$
A.8*	$\forall x, y\ [\mathsf{soli}(x) > \mathsf{soli}(y) \leftrightarrow \mathsf{cons}(x) > \mathsf{cons}(y)]$
A.9*	$\forall x, y\ [\mathsf{napm}(x) > \mathsf{napm}(y) \leftrightarrow \mathsf{dlab}(x) > \mathsf{dlab}(y)]$

Table 5.12: Revised formalization of Propositions 7–9.

justifiable by considering the inherent ambiguity of the ordinary language statements. Moreover, it is warranted by the author's explicit remark to "let us assume that the research suggests that the link between determinant and result in these propositions is necessary and reversible" Zetterberg (1965, p.160).

We have now revised the axioms of the theory, so we must again pose the question: Is the formal theory (still) consistent? Our earlier model in Table 5.4 is no longer a model of the theory, because the stronger version of Proposition 7 (axiom **A.7***) does not hold in it. Fortunately, there are still models of our revised versions of Zetterberg's Propositions 7–10, for example the model shown in Table 5.13.[9]

	napm	soli	cons	dlab	nrpd
0	1	1	1	1	0
1	0	0	0	0	1

$>$	0	1
0	F	F
1	T	F

Table 5.13: A model of Axioms **A.7***, **A.8***, **A.9***, and **A.10**.

We can easily verify that the revised axioms hold in the model of Table 5.13: Axiom **A.7*** holds because the group with greater division of labor (0) has also a greater solidarity; Axiom **A.8*** holds because the group with greater solidarity (0) also has greater consensus; Axiom **A.9*** holds because the group with greater number of associates per member has a greater division of labor; and finally, Axiom **A.10** holds because the group with greater solidarity (0) has a smaller number of rejections per deviant. This revised version of Zetterberg's theory has a model, therefore it is consistent.

Now all (intended) theorems are provable: the proofs of Proposition 1, 3, 5, and 6 (and 11) are still valid because we have only added axioms (or strengthened them); Proposition 2 is derivable from 7 and 8 (both in their revised form);

[9]On domain size 2, MACE generates 66 models of the revised axioms (and the background assumptions). The model in Table 5.13 is a prototypical model of the theory. There are two "ideal types" of groups (Zetterberg 1955, p.539):

> *Mechanical groups* marked by: 1) low division of labor; 2) low solidarity; 3) small membership; and 4) strong rejection of deviates from group norms.
> *Organic groups* marked by: 1) high division of labor; 2) high solidarity; 3) large membership; and 4) little rejection of deviates.

Group 0 in the model is an organic group, and group 1 is a mechanical group.

Proposition 4 is derivable from 8 (in its revised form) and the original Proposition 10.

Moreover, there are two more theorems derivable.

Theorem 15 *the greater the consensus, the greater the division of labor*:

$$\textbf{T.15} \qquad \forall x, y\ [\mathsf{cons}(x) > \mathsf{cons}(y) \ \rightarrow\ \mathsf{dlab}(x) > \mathsf{dlab}(y)]$$

Proof: OTTER can derive **T.15** from **A.14** and **A.12**.

Theorem 16 *the greater the consensus, the greater the number of associates per member*:

$$\textbf{T.16} \qquad \forall x, y\ [\mathsf{cons}(x) > \mathsf{cons}(y) \ \rightarrow\ \mathsf{napm}(x) > \mathsf{napm}(y)]$$

Proof: OTTER can derive **T.16** from **A.14**, **A.12**, and **A.13**.

Similar to the axioms, we can combine the Theorems 1, 2, and 3, with (new) Theorems 15, 11, and 16 respectively, as Theorems **T.1***, **T.2***, and **T.3*** in Table 5.14.

T.1*	$\forall x, y\ [\mathsf{dlab}(x) > \mathsf{dlab}(y) \ \leftrightarrow\ \mathsf{cons}(x) > \mathsf{cons}(y)]$
T.2*	$\forall x, y\ [\mathsf{soli}(x) > \mathsf{soli}(y) \ \leftrightarrow\ \mathsf{napm}(x) > \mathsf{napm}(y)]$
T.3*	$\forall x, y\ [\mathsf{napm}(x) > \mathsf{napm}(y) \ \leftrightarrow\ \mathsf{cons}(x) > \mathsf{cons}(y)]$

Table 5.14: A revised formalization of Theorems 1–3.

We have now re-formalized three of the four axioms as biconditionals, i.e., **A.7***, **A.8***, and **A.9***, which allows us to derive all intended theorems, **T.1**–**6**. Moreover, we can derive three of these theorems as stronger biconditionals, i.e., **T.1***, **T.2***, and **T.3***.

Adding axioms allows us to derive all propositions, including the missing Propositions 2 and 4. Moreover, the stronger formulations of Axioms 7–9, i.e., **A.7***, **A.8***, and **A.9***, also restore the intended inference patterns. For example, unlike in our initial formalization, Proposition 3 (**T.3***) can now be inferred from Proposition 2 (**T.2***) and Proposition 8 (**A.8***).

5.4.5 Proposition 10

We have now formalized three of the four axioms as biconditionals (**A.7***, **A.8***, and **A.9***), and left the remaining axiom in its original version (**A.10**). Although strictly speaking not necessary for proving the theorems, one could argue that it is more natural to formalize the fourth axiom, Proposition 10, in a similar way as the other axioms. Let us explore this option.

A.10*	$\forall x, y\ [\mathsf{soli}(x) > \mathsf{soli}(y) \leftrightarrow \mathsf{nrpd}(y) > \mathsf{nrpd}(x)]$

Table 5.15: Revised formalization of Propositions 10.

Assume Proposition 10 is reformalized as **A.10*** in Table 5.15: *the solidarity is greater if and only if the number of rejections per deviant is smaller.*

The stronger version of Proposition 10 allows for the derivations of three more theorems derivable.

Theorem 17 *the greater the number of rejections per deviant, the smaller the consensus*:

$$\textbf{T.17} \qquad \forall x, y\ [\mathsf{nprd}(x) > \mathsf{nprd}(y) \rightarrow \mathsf{cons}(y) > \mathsf{cons}(x)]$$

Proof: OTTER can derive **T.17** from **A.10*** and **A.8**.

Theorem 18 *the greater the number of rejections per deviant, the smaller the division of labor*:

$$\textbf{T.18} \qquad \forall x, y\ [\mathsf{nprd}(x) > \mathsf{nprd}(y) \rightarrow \mathsf{dlab}(y) > \mathsf{dlab}(x)]$$

Proof: OTTER can derive **T.18** from **A.10*** and **A.12**.

Theorem 19 *the greater the number of rejections per deviant, the smaller the number of associates per member*:

$$\textbf{T.19} \qquad \forall x, y\ [\mathsf{nprd}(x) > \mathsf{nprd}(y) \rightarrow \mathsf{napm}(y) > \mathsf{napm}(x)]$$

Proof: OTTER can derive **T.19** from **A.10***, **A.12**, and **A.13**.

Again, we can combine the (new) Theorems 17, 18, and 19 with Theorems 4, 5, and 6 respectively as Theorems 4*, 5*, and 6* in Table 5.16. We now have for-

T.4*	$\forall x, y\ [\mathsf{cons}(x) > \mathsf{cons}(y) \leftrightarrow \mathsf{nrpd}(y) > \mathsf{nrpd}(x)]$
T.5*	$\forall x, y\ [\mathsf{dlab}(x) > \mathsf{dlab}(y) \leftrightarrow \mathsf{nrpd}(y) > \mathsf{nrpd}(x)]$
T.6*	$\forall x, y\ [\mathsf{napm}(x) > \mathsf{napm}(y) \leftrightarrow \mathsf{nrpd}(y) > \mathsf{nrpd}(x)]$

Table 5.16: A revised formalization of Theorems 4–6.

malized all 10 propositions in the same way (both axioms **A.7***–**10*** and theorems **T.1***–**6*** as biconditionals). Table 5.17 compares the verbal theory with our last formal version. In this version, all suggested theorems are derivable. Moreover, they are derivable by the suggested inferences.

Verbal theory:	Formal theory:	
P.1 from **P.7** and **P.8**	**T.1*** from **A.7*** and **A.8***	$\surd$
P.2 from **P.7** and **P.9**	**T.2*** from **A.7*** and **A.9***	$\surd$
P.3 from **P.8** and derived **P.2**	**T.3*** from **A.9***, **A.7*** and **A.8***	$\surd$
P.4 from **P.8** and **P.10**	**T.4*** from **A.8*** and **A.10***	$\surd$
P.5 from **P.7** and **P.10**	**T.5*** from **A.7*** and **A.10***	$\surd$
P.6 from **P.9** and derived **P.5**	**T.6*** from **A.9***, **A.7*** and **A.10***	$\surd$

Table 5.17: The Verbal and Formal Theory

Is the (formal) theory still consistent? The theory is consistent if it has a model. The model in Table 5.13 is also a model of Axioms **A.7***, **A.8***, **A.9***, and **A.10***.[10]

5.4.6 Recapitulating

We have now presented four formalizations of Zetterberg's natural language theory as summarized in Table 5.18.

	Premises:	Theorems:
(1)	**A.7**, **A.8**, **A.9**, **A.10**	**T.1**, **T.3**, **T.5**, **T.6**
(2)	**A.7**, **A.8**, **A.9**, **A.10**, **MP.1**	**T.1**, **T.2**$^-$, **T.3**, **T.5**, **T.6**
(3)	**A.7***, **A.8***, **A.9***, **A.10**	**T.1***, **T.2***, **T.3***, **T.4**, **T.5**, **T.6**
(4)	**A.7***, **A.8***, **A.9***, **A.10***	**T.1***, **T.2***, **T.3***, **T.4***, **T.5***, **T.6***

Table 5.18: Four Formal Version of the Theory.

Versions 1 and 2 use the same axioms, therefore they characterize the same theory.[11] If we would choose between these first two versions, then the exposition in version 2 is closer to the natural language exposition of the theory because it presents a version of theorem 2.

Version 3 allows us to derive all propositions, including Proposition 4, and is therefore a more natural representation of the natural language theory. The price for deriving all propositions is a set of stronger axioms, which narrows down the theory's domain of application. Version 3 allows use to derive the propositions using the suggested inferences. In version 3 of the theory, some propositions (**T.1***, **T.2***, **T.3***, **A.7***, **A.8***, and **A.9***) are represented using biconditionals, whereas other propositions (**T.4**, **T.5**, **T.6**, and **A.10**) are represented as normal conditionals. The difference in the translation of propositions is admittedly *ad*

[10] On domain size 2, MACE generates 34 models of the revised axioms (and the background assumptions).

[11] In the table, we only list the background assumptions if they are needed in the derivations.

hoc: we have reinterpreted those axioms that were necessary for deriving at least a conditional version of the theorems.

Version 4 reinterprets all axioms as biconditionals, and as a result allows for also deriving biconditional statements. Version 4 gives a natural reconstruction of Zetterberg's theory in first-order logic. It gives a uniform, formal interpretation of all propositions (both axioms and theorems), it is consistent, the derivations of theorems are sound, and all theorems are satisfiable and falsifiable.[12] Our goal is to formalize Zetterberg's theory, and, arguably, version 4 is our best candidate: it seems to model the natural language of the original theory closely, and it satisfies all the logical criteria we formulated.

The explanatory power of version 4 comes at a price: we had to reformalize the axioms as biconditionals. These versions of the axioms are strong. Consequently, they restrict the domain of the theory. We can investigate the domain of the theory by looking at its models. Table 5.19 shows the number of models of

	Premises:	Models on $\{0,1\}$:
(1)	**A.7**, **A.8**, **A.9**, **A.10**	1024
(2)	**A.7**, **A.8**, **A.9**, **A.10**, **MP.1**, **MP.2**	142
(3)	**A.7***, **A.8***, **A.9***, **A.10**, **MP.1**, **MP.2**	66
(4)	**A.7***, **A.8***, **A.9***, **A.10***, **MP.1**, **MP.2**	34

Table 5.19: Number of Models on a Simple Domain.

universe $\{0,1\}$. On domain size two, version 4 of the theory has models similar to the model presented in Table 5.13, and there are models in which all variables are equal for both groups (i.e., for group 0 and group 1).[13] The axioms are so strong that all five variables become virtually identical, trivializing the theory. Although version 4 of the theory is a more natural translation of the natural language wording, its strong axioms seem unrealistic. To a lesser extent, the same holds for version 3 in which four of the five variables are virtually identical. Although, version 2 of the theory does not derive a version of proposition 4, it is perhaps the most promising candidate.

[12] For proving the satisfiability of a theorem, we have to find a model (ignoring the axioms) where it holds. All theorems are satisfied in the model in Table 5.13. For proving falsifiability of a theorem, we have to construct a model (ignoring the axioms) in which the theorem is does not hold. For example, the models in Table 5.10 still prove the falsifiability of proposition 2 (note that these models do no longer belong to the revised theory).

[13] The model presented in Table 5.13 is one of the 34 models on domain size 2 that are produced by the automated model generator MACE. No less than 32 of these models make all axioms (and theorems) vacuously true! In these 32 models all variables are equal for both groups (both 0 or both 1 for the five variables yield $2^5 = 32$ models). In the model in Table 5.13 all five variables are unequal, and the copy with arguments 0 and 1 interchanged is the last remaining model.

5.5 Discussion and Conclusions

The axiomatization of scientific theories in formal logic dates back, at least, to the logical positivists (Ayer 1959). A formal theory is defined as the deductive closure of its set of axioms and theoretical explanations and predictions correspond to deductions from the set of premises (Popper 1959). Formal logic provides several criteria for evaluating the theory, such as the consistency of the theory and soundness of deductions—criteria traditionally associated with the context of justification. We argued that these criteria can also play an important role in the revision of a theory—an activity traditionally associated with the context of discovery. The tests we suggested do not only prove the criteria, but also give a particular derivation or model that explains why a certain criterion holds or fails to hold. Examining these proofs or models provides crucial information for revising the formal theory. Moreover, this revision may have an impact on the original theory. This can be of great importance since even a minor modification of the original theory may avoid the costs involved in the empirical testing of incorrect or irrelevant hypotheses. The criteria facilitate a piecemeal revision of the theory, resulting in a cyclic process of theory development.[14]

The detection of nonintended models deserves special attention: the implications of this simple result are quite fundamental. The mechanization or automatization of human intelligence is an age-old problem. Human intelligence has a bewildering richness of latent knowledge available that plays a crucial role in human reasoning (Boden 1998). Two of the most fundamental problems in the formalization of human intelligence are how to unearth this background knowledge, and how to determine which parts of it are relevant for the problem at hand. Finding counterexamples that are nonintended models provides a partial solution: in a sense, the formal theory can tell us what it is missing. Moreover, the underlying human knowledge is often tacit, making its articulation beyond the owner's control (Polanyi 1958). Notice that even a model violating a theorist's tacit domain knowledge can still be recognized by him/her as a nonintended model. Furthermore, confronting a theorist with such a nonintended model can make him/her aware of the underlying tacit understanding and can provide crucial help in its articulation.

The process of revision is essentially interactive. We attempt to use computational support for those tasks for which computers are better equipped. For example, we use automated reasoning tools for finding proofs or models. Notice that humans theorizers have often difficulty in finding counterexamples that are non-intended models. Theorizers tend to ignore these models since they conflict with their common-sense or with their understanding of the substantive domain. Fortunately, an automated model generator does not have such a bias. On the

[14] Much like the essay on the *polyhedra conjecture* in (Lakatos 1976) to which our case study in §5.3 shows some remarkable resemblance.

other hand, a human theorizer can use this knowledge to distinguish between non-intended models and genuine counterexamples. This decision is crucial because it determines whether we need to revise the premises (in case of a non-intended model) or whether we need to revise the conjecture (in case of a genuine counterexample). This decision is difficult to make for automated systems because it would require a full axiomatization of all relevant common, background knowledge.

The product of an axiomatization attempt, a first-order logic rendition of a theory, is a deductive theory. Although we advocate deductive theories, we do not want to de-emphasize other modes of reasoning. Quite the contrary. Consider, for example, the step to revise the theory to account for a counterexample. Such an attempt to revise the theory is *abductive*. In fact, this step is using an extended form of abduction, since we may either decide to change the premises to explain the claim (traditional abduction, see for example Aliseda-LLera 1997), or decide to change the claim such that it can be explained by the original premises.[15] Although the product of an axiomatization is a deductive theory, the process of axiomatizing a theory is essentially non-deductive.

[15] Note that only in the first case we really modify the formal theory—being all logical consequences of the premise set. In the second case only the exposition of the theory changes.

Chapter 6

Partial Deductive Closure

In formal logic, the deductive closure of a theory is taken for granted—in fact, the formal definition of a "theory" is a set of sentences closed under the rules of inference. In reality, however, it is impossible to generate the complete deductive closure of a premise set, for the resulting set is infinite (since it contains all tautologies, i.e., expressions that are always true and hence follow from any set). In consequence, the complete deductive closure of a premise set is neither realizable nor desirable—only *partial* deductive closures provide useful results. In this chapter, we present an algorithm that performs an efficient partial deductive closure for an important class of formulas, i.e., conditional formulas that relate two relevant domain properties. Statements of this form provide the backbone of any empirical social science; arguably, it is the most important class of empirical statements in the social sciences.

The implemented program is applied to an important organization theory, Organizational Ecology. The algorithm generates more theorems than the original discursive theory made explicit, some of which are of theoretical interest.

6.1 Introduction

Computer simulation has become a standard tool in Management Science. A domain is represented in a formal language, and the properties of the domain are investigated through the properties of the formal representation (the "simulation model"). Traditionally, such representations used equational mathematics, *e.g.*, numerical differential equations, as the formal language (Forrester 1961; Cohen, March, and Olsen 1972; Burton and Obel 1984). Recent years, however, have seen various efforts to use "qualitative" or "declarative" representation languages instead (Baligh, Burton, and Obel 1990; Glorie et al. 1990; Blanning 1992; Carley and Prietula 1994; Péli et al. 1994; Kamps and Péli 1995). These efforts are usually inspired by progress in Artificial Intelligence—and by domains so complex that they defy numerical representation. The researcher may not know the numerical value of a variable, but still want to incorporate this variable in the model. For this reason, qualitative languages are an attractive alternative for numerical languages.

With the advent of expert systems, a large variety of qualitative languages became available. The common denominator of these languages is usually (a fragment of) First-Order Logic (FOL), the best known formal logic. This commonality has focussed attention on the use of formal logic as a representation language, both in Management Science and elsewhere (Kimbrough and Lee 1988a,b; Kimbrough 1990; Masuch 1992; Bhargava and Kimbrough 1994). As a *logic*, FOL has considerable expressive power and has useful features for developing better theories. For example, FOL provides precise criteria for theoretical *consistency* (is a theory contradiction-free?), *soundness* (are the explanations of a theory logically correct?) and *contingency* (is the theory falsifiable?).

By testing a theory for these properties, the researcher can develop better theories with the help of logic. The researcher formulates his theory in a logical language, and the computer investigates the logical properties of this representation. The logical representation itself could then play the role of a simulation model, where the computation of the "outcome" is done through logical inferencing.

Now, if one wants to spell out the logical consequences of a set of assumptions, then one is, technically speaking, working on the "deductive closure" of this set "under the rules of inference." In formal logic, the deductive closure of a theory is taken for granted—in fact, the formal definition of a "theory" is a set of assumptions closed under the rules of inference (Tarski 1956). In reality, however, it is impossible to generate the complete deductive closure of a premise set, for the resulting set is infinite (since it contains all *tautologies*, *i.e.*, expressions that are always true and hence follow from any set). In consequence, the complete deductive closure of a premise set is neither realizable nor desirable—only *partial* deductive closures provide useful results.

In this chapter, we present an algorithm that performs an efficient partial deductive closure for an important class of formulas. We call this class SPtSP, ("single property to single property"); it comprises conditional formulas that link one property of an object to another property of an object. Examples include: *if the size of an organization increases, then its inertia increases*; and *if the inertia of an organization increases, then its survival chance increases.* Statements of this form provide the backbone of any empirical social science; arguably, it is the most important class of empirical statements in the social sciences.

Section 6.2 discusses the role of formal logic in theory-building; in analogy to the *empirical cycle* in social science research (De Groot 1961, 1969), we describe a *logical cycle* of theory development. Section 6.3 describes the algorithm. The next two sections (6.4 and 6.5) show the algorithm in action, working on the "inertia" fragment of an important organization theory, *Organizational Ecology*

(Hannan and Freeman 1984, 1989). As it turns out, our algorithm generates more theorems than the original discursive theory of organizational inertia made explicit, some of which are of theoretical interest. The last section discusses some limits to our approach, regarding both the algorithm and the use of FOL as a representation language.

The research reported in this chapter is part of a larger effort at the Center for Computer Science in Organization and Management (CCSOM) of the University of Amsterdam to develop a formal methodology of theory analysis and theory building. This effort includes the application of standard logics to existing theories in organization and management (Péli et al. 1994; Péli and Masuch 1994; Bruggeman 1997), the development of "non-standard" logics especially suited for the representation of *action* theories (Pólos and Masuch 1995; Huang, Masuch, and Pólos 1996; Masuch and Huang 1996), and the development of software that supports theory building using formal logic (Ó Nualláin 1993).

6.2 Developing Theories with Logical Tools

Theories can be seen as propositional systems, and logic is traditionally used to order such systems. Logic provides the rationale for theoretical explanations; explanations, in turn, provide the justification for a theory. In this view, the "implications" of a theory are crucial for its justification: the better a theory, the better its "predictions." The formal definition of a theory is a set of statements "closed" under the rules of inference (Tarski 1956), which would suggest that the logic itself does the job of providing the implications. In practice, however, the logical closure of a theory is not a given; some agent is needed to make logic happen, and to carry out the deductions. As a consequence, we need to distinguish between a *premise set* (an explicitly stated set of premises), a *complete theory* (the premise set closed under the rules of inference), and *intermediate* theories that represent the premise set plus some, but not necessarily all, of its logical consequences. In addition, we must recognize that theories are not always conceived as explicitly stated sets of formulas; they can also be perceived as some kind of knowledge of a theoretician regarding a domain, regardless how this knowledge is represented. For want of a better term, we call this kind of knowledge *theoretical expectation*. Obviously, theoretical expectations depend on their bearer and may change as the theory evolves (Lakatos and Musgrave 1970).

Formal inferencing and theoretical expectations interact. Conclusions that confirm earlier expectations strengthen confidence in the theory and unearth new ways to test it. Unexpected conclusions invite the theoretician to revise either his expectations or the original theory. If the theory does not make the right predictions, it must change, so finding out which conclusions are implied by a theory is an important part of theory building. Much like the empirical cycle in social science research (problem identification; hypothesis formulation; research design; data collection; data analysis; hypothesis testing), there is a *logical cycle* in the interaction between theoretical expectations and formal inferencing.

The logical cycle can be described using the terminology introduced above (Figure 6.1): Theoretical expectations (TE) are formalized, yielding an interme-

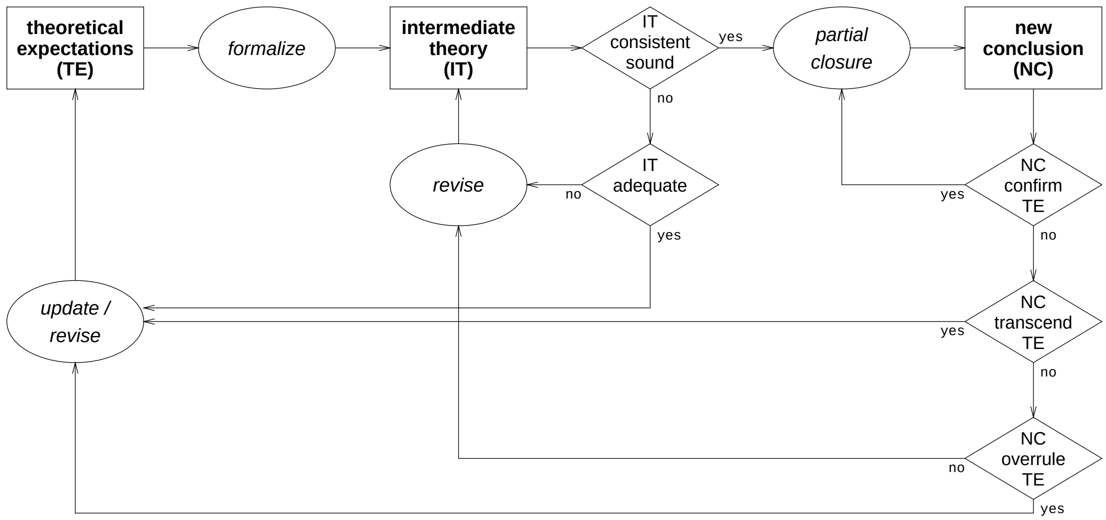

Figure 6.1: The logical cycle.

diate theory (IT). The intermediate theory may or may not be consistent. If it is inconsistent, either the formalization or the theoretical expectations require revisions. If it is consistent, continued theorizing yields a partial closure of the original IT. New conclusions (NC) can (1) confirm the theoretical expectations, (2) transcend the theoretical expectations, or (3) contradict the theoretical expectation. In the first case, the partial closure can continue; in the second case, the theoretical expectations need to be updated; in the third case, either the formalization of the original expectations or the expectations themselves need to be revised. The cycle is never truly complete, since the deductive closure of any set is infinite. But not all conclusions that are technically derivable from a given set are of interest. For example, tautologies are derivable from any set of premises. But because they are always true, they tell us nothing about the domain—they are true regardless of the structure of the domain.

Although everybody agrees that theoretical labor is an important part of a researcher's work—thinking a theory through, taking it to its logical conclusions, ascertaining its consistency and coherence, and so on—the logical cycle has received very little attention in the literature. In the past, there was no method to guide the researcher more systematically through a maze of potential conclusions. With the advent of logical programming, however, we can now automate a partial deductive closure of a given set of assumptions.

6.3 Deduction of Theorems

We can get a partial deductive closure of a set of premises by applying inference rules to deduce new expressions (in technical terms "well-formed formulas," or "formulas" for short). In this section, we describe an algorithm that performs a partial deductive closure of an important class of theorems. Our algorithm focuses on theorems that relate two properties of the domain. We call this class of statements "single property to single property" (SPtSP); it comprises conditional formulas that link one property of an object to another property of an object, as noted above. The algorithm is called PDC-1.

6.3.1 Informal Description of the Algorithm

The algorithm uses premises of the form SPtSP. An SPtSP expression relates one quantifiable property to another, such as *higher inertia yields to higher survival chances*. Additionally, such an expression may have constraints that restrict it to certain types of objects, *e.g.*, to *reorganization-free organizations*: *reorganization-free organizations with higher inertia have higher survival chances*. Assume that a domain is characterized by the following premises:

$$\mathsf{Constraints}_1 \wedge \mathsf{Property}_1 \rightarrow \mathsf{Property}_2$$
$$\mathsf{Constraints}_2 \wedge \mathsf{Property}_2 \rightarrow \mathsf{Property}_3$$
$$\mathsf{Constraints}_3 \wedge \mathsf{Property}_3 \rightarrow \mathsf{Property}_4$$

A theorem that relates $\mathsf{Property}_1$ with $\mathsf{Property}_4$ can be deduced from these three premises by "cutting out" $\mathsf{Property}_2$ and $\mathsf{Property}_3$:

$$\mathsf{Constraints}_1 \wedge \mathsf{Constraints}_2 \wedge \mathsf{Constraints}_3 \wedge \mathsf{Property}_1 \rightarrow \mathsf{Property}_4$$

As cases in point, we take three premises from the domain of Organizational Ecology. The following notation is used to represent SPtSP class expressions: $[\mathsf{Constraints}] \wedge \mathsf{Property}_1 \rightarrow \mathsf{Property}_2$. The $\mathsf{Property}_1$ and $\mathsf{Property}_2$ are referring to two quantifiable properties of the domain. The $[\mathsf{Constraints}]$ are the restricting conjuncts (either one or more conjuncts). The square brackets "$[\cdots]$" are used to differentiate the conjuncts of the $\mathsf{Constraints}$ from $\mathsf{Property}_1$ in the antecedent. These brackets are added only to improve readability, and are ignored by the formal machinery.

Assume we have the following premises (assumption 5, 3b and 2a in section 6.4 respectively):

- Larger organizations have higher inertia than smaller organizations of the same class:
$\forall c, i_1, i_2, s_1, s_2, t_1, t_2, x, y([\mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Class}(x, c, t_1) \wedge \mathsf{Class}(y, c, t_2)$
$\wedge\, \mathsf{Size}(x, s_1, t_1) \wedge \mathsf{Size}(y, s_2, t_2) \wedge \mathsf{Iner}(x, i_1, t_1) \wedge \mathsf{Iner}(y, i_2, t_2)]$
$\wedge\, (s_2 > s_1) \rightarrow (i_2 > i_1))$

(Read: for all $c, i_1, i_2, s_1, s_2, t_1, t_2, x, y$ if x and y are organizations of the same class c at time t_1 and t_2, and s_1 and i_1 are, respectively, the size and inertia of x at t_1 and s_2 and i_2 are, respectively, the size and inertia of y at t_2, and s_2 exceeds s_1, then i_2 exceeds i_1.)

- Reorganization-free organizations with higher inertia have higher reproducibility:
  $\forall i_1, i_2, rp_1, rp_2, t_1, t_2, x, y([\mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Reorg\_free}(x, t_1, t_1)$
  $\wedge\, \mathsf{Reorg\_free}(y, t_2, t_2) \wedge \mathsf{Iner}(x, i_1, t_1) \wedge \mathsf{Iner}(y, i_2, t_2) \wedge \mathsf{Repr}(x, rp_1, t_1)$
  $\wedge\, \mathsf{Repr}(y, rp_2, t_2)] \wedge (i_2 > i_1) \rightarrow (rp_2 > rp_1))$
  (Read: for all $i_1, i_2, rp_1, rp_2, t_1, t_2, x, y$ if x and y are organizations not in reorganization at time t_1 and t_2 respectively, and i_1 and rp_1 are, respectively, the inertia and reproducibility of x at t_1, and i_2 and rp_2 are, respectively, the inertia and reproducibility of y at t_2, and i_2 exceeds i_1 then rp_2 exceeds rp_1.)

- Organizations with higher reproducibility have higher reliability/accountability:
  $\forall ra_1, ra_2, rp_1, rp_2, t_1, t_2, x, y([\mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Repr}(x, rp_1, t_1) \wedge \mathsf{Repr}(y, rp_2, t_2)$
  $\wedge\, \mathsf{Relacc}(x, ra_1, t_1) \wedge \mathsf{Relacc}(y, ra_2, t_2)] \wedge (rp_2 > rp_1) \rightarrow (ra_2 > ra_1))$
  (Read: for all $ra_1, ra_2, rp_1, rp_2, t_1, t_2, x, y$ if x and y are organizations at time t_1 and t_2 respectively, and rp_1 and ra_1 are, respectively, the reproducibility and reliability/accountability of x at t_1, and rp_2 and ra_2 are, respectively, the reproducibility and reliability/accountability of y at t_2, and rp_2 exceeds rp_1 then ra_2 exceeds ra_1.)

These premises are used to create a new theorem relating *size* and *reliability/accountability*, by following the same steps as in the abstract example. The set of constraints on this theorem is the superset of the constraints on the individual premises. Identical conjuncts in this set can be removed (since $(p \wedge p) \leftrightarrow p)$. This leads to the following new theorem:

- Larger reorganization-free organizations have higher reliability/accountability than smaller organizations of the same class:
  $\forall c, ra_1, ra_2, s_1, s_2, t_1, t_2, x, y([\mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Reorg\_free}(x, t_1, t_1)$
  $\wedge\, \mathsf{Reorg\_free}(y, t_2, t_2) \wedge \mathsf{Class}(x, c, t_1) \wedge \mathsf{Class}(y, c, t_2) \wedge \mathsf{Size}(x, s_1, t_1) \wedge \mathsf{Size}(y, s_2, t_2)$
  $\wedge\, \mathsf{Relacc}(x, ra_1, t_1) \wedge \mathsf{Relacc}(y, ra_2, t_2)] \wedge (s_2 > s_1) \rightarrow (ra_2 > ra_1))$
  (Read: for all $c, ra_1, ra_2, s_1, s_2, t_1, t_2, x, y$ if x and y are organizations of the same class c and not in reorganizational period at time points t_1 and t_2 respectively, and s_1 and ra_1 are, respectively, the size and reliability/accountability of x at t_1, and s_2 and ra_2 are, respectively, the size and reliability/accountability of y at t_2, and s_2 exceeds s_1 then ra_2 exceeds ra_1.)

6.3.2 Formal Specification

The partial deductive closure of a premise set is generated in three steps:

Filter premises The premises of the domain are filtered for SPtSP premises. Only these premises are used to construct new theorems.

Deduce new theorems The construction of new theorems. The algorithm uses the SPtSP premises to derive new SPtSP theorems.

Filter new theorems The set of constructed theorems is refined by removing: 1) vacuously true theorems, 2) weaker versions of other theorems in the set, and 3) superfluous conjuncts.

Table 6.1 introduces the notation we use to refer to the premise set and the derived theorems:

Symbol	Description
Σ	the original premise set used as a starting point for the algorithm
$\Sigma^\star$	the subset of SPtSP premises in the original premise set Σ
Σ'	the set of SPtSP theorems derived from $\Sigma^\star$
Σ^+	the resulting set of premises and theorems $\Sigma \cup \Sigma'$, *i.e.*, a partial closure of Σ

Table 6.1: Premise sets and theorems.

Filter Premises

The algorithm PDC-1 first identifies the premises that are in SPtSP form, that is, of the form:

$$[\mathsf{Constraints}_i] \wedge \mathsf{Property}_i \rightarrow \mathsf{Property}_j$$

Let $\Sigma^\star$ denote the premise set Σ restricted to SPtSP premises. Define $\Sigma^\star$ as follows:

1. For every premise $\sigma \in \Sigma$:
   (a) IF σ belongs to the SPtSP class
   THEN $\sigma \in \Sigma^\star$
   ELSE $\sigma \notin \Sigma^\star$
2. Nothing else is in $\Sigma^\star$

Deduce New Theorems

PDC-1 constructs new theorems for each pair of properties described in the SPtSP premises. The algorithm uses a depth-first, forward-chaining strategy.

Let $\Sigma^\star$ denote the set of SPtSP premises and Φ the set of properties described in $\Sigma^\star$. Let σ_{ij} denote an SPtSP formula $([\mathsf{Constr}_i] \wedge \mathsf{Prop}_i \rightarrow \mathsf{Prop}_j)$. The deduced theorems are denoted by Σ', the set of SPtSP theorems derived from $\Sigma^\star$.

Step 1 The algorithm constructs new theorems for each pair $\{\mathsf{Prop}_a, \mathsf{Prop}_b\}$ of properties in Φ. Pairs of the same properties $\{\mathsf{Prop}_a, \mathsf{Prop}_a\}$ are excluded, since they would yield tautologies like *organizations with higher inertia have higher inertia*.

Step 2 For each pair of properties $\{\mathsf{Prop}_a, \mathsf{Prop}_b\}$ the algorithm uses a premise having Prop_a in the antecedent:

$$[\mathsf{Constr}_a] \wedge \mathsf{Prop}_a \rightarrow \mathsf{Prop}_1$$

to construct an initial formula σ_{a1}, which is refined step by step. If there are no (more) premises that relate Prop_a to another property, there are no (more) theorems that relate Prop_a and Prop_b. The algorithm terminates for $\{\mathsf{Prop}_a, \mathsf{Prop}_b\}$ and continues with the next pair of properties.

Step 3 The formula σ_{ai} that relates Prop_a and Prop_i

$$[\mathsf{Constr}_{ai}] \wedge \mathsf{Prop}_a \rightarrow \mathsf{Prop}_i$$

is refined if the algorithm can find a premise that relates Prop_i and Prop_{i+1}:

$$[\mathsf{Constr}_i] \wedge \mathsf{Prop}_i \rightarrow \mathsf{Prop}_{i+1}$$

This results in a formula $\sigma_{a(i+1)}$ that relates Prop_a and Prop_{i+1}:

$$[\mathsf{Constr}_{ai} \wedge \mathsf{Constr}_i] \wedge \mathsf{Prop}_a \rightarrow \mathsf{Prop}_{i+1}$$

The antecedent of this formula contains the combined constraints of σ_{ai} and $\sigma_{i(i+1)}$, we refer to these constraints of $\sigma_{a(i+1)}$ as simply $\mathsf{Constr}_{a(i+1)}$. To avoid cycles in this refinement process, the algorithm only considers premises that introduce a new property. For example, if an initial formula relating *inertia* and *reproducibility* is found in Step 2, then a premise relating *reproducibility* and *inertia* is not applied in Step 3, since the property *inertia* was used before. This prevents the construction of some tautologies (like *organizations with higher inertia have higher inertia*).

If the formula σ_{ai} cannot be further refined (there is no (further) premise that relates Prop_i and Prop_{i+1}), the algorithm has reached a dead end. The algorithm retracts the last refinement σ_{ai} and attempts to construct other refinements of $\sigma_{a(i-1)}$ (the previous version of σ_{ai}).

Step 4 The refinement of formula σ_{ai} is completed if a theorem is constructed that relates Prop_a and Prop_b (in other words Prop_{i+1} in Step 3 is the desired Prop_b). The new theorem σ_{ab} is added to Σ', the set of deduced SPtSP theorems. Since there may be more than one theorem that relates Prop_a and Prop_b, the algorithm retracts the last refinement of σ_{ai} to σ_{ab}

after a successful theorem construction in Step 4 in order to find formulas constructed using different combinations of premises. This allows for the construction of theorems relating to different contexts, like *the survival chance of reorganizing organizations decreases with time* and *the survival chance of reorganization-free organizations increases with time.*

Otherwise, if Prop_{i+1} in Step 3 is not the desired Prop_b, Step 3 is repeated.

PDC-1 uses a standard algorithm for variable unification (Robinson 1965) in Step 3. This variable unification is necessary to determine that a formula, such as $\forall o_1, t_a(\mathsf{O}(o_1, t_a))$, can be made equal to another formula, such as $\forall x, y(\mathsf{O}(x, y))$, because the variables can be unified, in this case x with o_1 and y with t_a.

Filter Theorems

We have three filters to refine the set of constructed theorems. The first filter removes vacuously true theorems, the second removes weaker or identical versions of theorems, and the last filter removes superfluous conjuncts in theorems. The filters create a more concise set of theorems by removing non-interesting theorems.

Vacuously true theorems Sometimes premises cannot be combined directly, since different premises may have incompatible constraints. If we construct a new candidate theorem with incompatible premises, *i.e.*, premises having contradictory constraints. For example, we create a vacuously true theorem by combining premises for reorganizing organizations with premises for reorganization-free organizations. A vacuously true theorem has an antecedent that can never be fulfilled in the context of the premises (there is no model that satisfies both the antecedent and the premises). Recall that the theorems have the form of an *if-then* statement: if the antecedent holds then the consequent must hold. Therefore, we can prove any consequent from the theorem's antecedent if the antecedent never holds. In the truth table for the logical conditional, if p is false then the conditional $p \to q$ is always true, regardless of q's truth value.

Suppose that the premise set contains the following premise:

> An organization cannot be reorganization-free and reorganizing at the same time:
> $\forall x, t_1, t_2(\neg(\mathsf{Reorg\_free}(x, t_1, t_2) \wedge \mathsf{Reorg}(x, t_1, t_2)))$

Assume that we have constructed a theorem with the following constraints:

> For all reorganization-free organizations under reorganization ...
> $\forall x, t_1, t_2, \ldots([\mathsf{O}(x, t_1) \wedge \mathsf{Reorg\_free}(x, t_1, t_2) \wedge \mathsf{Reorg}(x, t_1, t_2) \wedge \ldots] \wedge \ldots \to \ldots)$

In this case, the antecedent of the theorem is inconsistent with the premise set, and the theorem is true regardless of the consequent (which may be even the *falsum*).

In sum, if we can substitute the *falsum* for the consequent in a theorem, and this theorem still holds, then the theorem is vacuously true. Such a theorem only holds because of the (hidden) contradiction in its antecedent. Vacuously true theorems do not provide any theoretical insights, therefore, their removal does not affect the theory.

Filter out weaker versions of a theorem Different premises can lead to different theorems that relate the same pair of properties. Sometimes these theorems are complementary: for example, one theorem is restricted to reorganizations, and another to reorganization-free periods, like *the survival chance of reorganizing organizations decreases with time* and *the survival chance of reorganization-free organizations increases with time.* But in other cases, one of them may subsume the other: for example, if one theorem states that *the inertia of organization x is larger than the inertia of organization y* and another theorem that *the inertia of organization x is larger than* **or equal to** *the inertia of organization y*, the latter theorem is weaker. In this case weaker versions of the same theorem are removed. The filter evaluates for every theorem whether it is the unique or strongest version of the formula. Identical theorems and theorems with weaker antecedents or stronger consequents are removed. Identical or weaker versions of a theorem are uninteresting because they can be derived from a stronger version. For example,

$$[\mathsf{O}(o_1) \wedge \mathsf{O}(o_2)] \wedge (\mathsf{Prop}_1(o_1) \geq \mathsf{Prop}_1(o_2)) \rightarrow (\mathsf{Prop}_2(o_1) > \mathsf{Prop}_2(o_2))$$

is preferable to

$$[\mathsf{O}(o_1) \wedge \mathsf{O}(o_2) \wedge \mathsf{Reorg}(o_1) \wedge \mathsf{Reorg}(o_2)] \wedge (\mathsf{Prop}_1(o_1) > \mathsf{Prop}_1(o_2)) \\ \rightarrow (\mathsf{Prop}_2(o_1) \geq \mathsf{Prop}_2(o_2))$$

since (i) $[\mathsf{O}(o_1) \wedge \mathsf{O}(o_2)]$ is implied by $[\mathsf{O}(o_1) \wedge \mathsf{O}(o_2) \wedge \mathsf{Reorg}(o_1) \wedge \mathsf{Reorg}(o_2)]$; (ii) $(\mathsf{Prop}_1(o_1) \geq \mathsf{Prop}_1(o_2))$ is implied by $(\mathsf{Prop}_1(o_1) > \mathsf{Prop}_1(o_2))$; and finally (iii) $(\mathsf{Prop}_2(o_1) > \mathsf{Prop}_2(o_2))$ implies $(\mathsf{Prop}_2(o_1) \geq \mathsf{Prop}_2(o_2))$.

Simplify theorems A constructed theorem may require the existence of "intermediate" conjuncts. The formula relating *size* and *reliability/accountability* derived in the example of section 6.3.1 would actually read:

$$\forall c, i_1, i_2, ra_1, ra_2, rp_1, rp_2, s_1, s_2, t_1, t_2, x, y([\mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Reorg\_free}(x, t_1, t_1) \\ \wedge \mathsf{Reorg\_free}(y, t_2, t_2) \wedge \mathsf{Class}(x, c, t_1) \wedge \mathsf{Class}(y, c, t_2) \wedge \mathsf{Size}(x, s_1, t_1) \wedge \mathsf{Size}(y, s_2, t_2) \\ \wedge \mathbf{Iner}(\mathbf{x}, \mathbf{i_1}, \mathbf{t_1}) \wedge \mathbf{Iner}(\mathbf{y}, \mathbf{i_2}, \mathbf{t_2}) \wedge \mathbf{Repr}(\mathbf{x}, \mathbf{rp_1}, \mathbf{t_1}) \wedge \mathbf{Repr}(\mathbf{y}, \mathbf{rp_2}, \mathbf{t_2}) \\ \wedge \mathsf{Relacc}(x, ra_1, t_1) \wedge \mathsf{Relacc}(y, ra_2, t_2)] \wedge (s_2 > s_1) \rightarrow (ra_2 > ra_1))$$

The *inertia* and *reproducibility* conjuncts must exist for the conditions of the premises to be fulfilled. If their existence is postulated (as is the case in the formal inertia theory), these conjuncts can be removed from the constraints.

- For every organization, there is some inertia that it has:
  $\forall x, t(\mathsf{O}(x, t) \rightarrow \exists i(\mathsf{Iner}(x, i, t)))$

- For every organization, there is some reproducibility that it has:
  $\forall x, t(\mathsf{O}(x,t) \rightarrow \exists rp(\mathsf{Repr}(x, rp, t)))$

The *inertia* and *reproducibility* conjuncts can now be derived from the *organization* conjuncts. The set of constraints is simplified (this was tacitly done in section 6.3.1):

$$\forall c, ra_1, ra_2, s_1, s_2, t_1, t_2, x, y([\mathsf{O}(x,t_1) \wedge \mathsf{O}(y,t_2) \wedge \mathsf{Reorg\_free}(x,t_1,t_1) \\ \wedge \mathsf{Reorg\_free}(y,t_2,t_2) \wedge \mathsf{Class}(x,c,t_1) \wedge \mathsf{Class}(y,c,t_2) \wedge \mathsf{Size}(x,s_1,t_1) \wedge \mathsf{Size}(y,s_2,t_2) \\ \wedge \mathsf{Relacc}(x,ra_1,t_1) \wedge \mathsf{Relacc}(y,ra_2,t_2)] \wedge (s_2 > s_1) \rightarrow (ra_2 > ra_1))$$

This concludes the description of our algorithm for the partial closure of SPtSP formulas. In the next two sections, we show the algorithm in action; it is applied to a fragment of an important organization theory, the "inertia" part of Organizational Ecology (OE). We first give a brief account of OE, then provide a formalization of the inertia part of OE, and finally show how the algorithm performs the partial deductive closure of this part.

6.4 Inertia Fragment of Organizational Ecology

Most organizational theories regard organizations as agents that adapt rationally to changing environments (Thompson 1967; Mintzberg 1979). Organizational Ecology, in contrast, sees organizational structures evolving through environmental selection. When environmental conditions change, new organizations emerge, and maladapted organizations die.

Organizational ecology employs analogies from biology. Genes determine the action repertoire of organisms, whereas organizations' repertoires are fixed by their core features. Organizations of the same form make up a population (just as organisms of the same form make up a species). Several factors inhibit the flexibility and adaptation of organizations, such as sunk costs, political coalitions, or the hazards of lost legitimacy.

Organizational ecology considers changes in the environment to be largely unpredictable. Organizations are characterized by structural inertia—if they adapt, they do so slowly. Contrary to the rational adaptation approach, however, successful organizations are inert, not flexible. Organizations must produce their products or services reliably and account for their actions rationally. To do so, they must be able to reproduce their structures smoothly. But the factors that facilitate their reproducibility make organizations resistant to change. Thus, inertia is a byproduct of reproducibility.

Organizational ecology does recognize the possibility of rational adaptation. To adapt, organizations must reorganize. Organizations can change their structures to a certain degree, but reorganizations typically involve changes in core features. Such changes are dangerous; they involve large resources, and organizational learning must start anew and higher up on the learning curve. Even if an organization survives a major reorganization, its environment may change in unexpected ways and the reorganization might be in vain. So if organizations attempt to adapt, they are not likely to succeed. Inert organizations—those that resist the temptation to reorganize—may be less at risk than flexible organizations.

The inertia part of OE is originally described in (Hannan and Freeman 1984). Table 6.2 reprints the core of the theory of organizational inertia as published

Label	Content
Assumption 1	Selection in populations of organizations in modern societies favors forms with high reliability of performance and high levels of accountability.
Assumption 2	Reliability and accountability require that organizational structures be highly reproducible.
Assumption 3	High levels of reproducibility of structure generate strong inertial pressures.
Theorem 1 (from Ass.1–3)	Selection within populations of organizations in modern societies favors organizations whose structures have high inertia.
Assumption 4	Reproducibility of structure increases monotonically with age.
Theorem 2 (from Ass.2,4)	Structural inertia increases monotonically with age.
Theorem 3 (from Ass.4,Thm.1)	Organizational death rates decrease with age.
Assumption 5	The level of structural inertia increases with size for each class of organizations.
Assumption 6	The process of attempting reorganization lowers reliability of performance.
Theorem 4 (from Ass.1,6)	Attempts at reorganization increase death rates.
Assumption 7	Organizational death rates decrease with size.
Assumption 8	Structural reorganization produces a liability of newness.
Assumption 9	The death rate of organizations attempting structural change rises with the duration of the reorganization.
Assumption 10	Complexity increases the expected duration of reorganization.
Theorem 5 (from Ass.9,10)	Complexity increases the risk of death due to reorganization.

Note: All propositions are from (Hannan and Freeman 1984).

Table 6.2: The Inertia Fragment of Organizational Ecology.

there. The authors explicitly list the assumptions on which the theorems are based, and the intended inference structure of the theory.[1] A formalization of this theory in FOL has been published in (Péli et al. 1994). We use the formulas representing the premises of (Péli et al. 1994) as input, and let our algorithm derive the theorems. The premises will be discussed in this section, and the derivable theorems in section 6.5. Table 6.3 characterizes the relation symbols

Symbol	Description
$\mathsf{Class}(x, c, t)$	object x is a member of class c at time t
$\mathsf{Compl}(x, cp, t)$	object x is characterized by complexity cp at time t
$\mathsf{Iner}(x, i, t)$	object x has a value of inertia i at time t
$\mathsf{O}(x, t)$	object x is an organization at time t
$\mathsf{Relacc}(x, ra, t)$	object x has a value of reliability/accountability ra at time t
$\mathsf{Reorg}(x, t_1, t_2)$	object x reorganizes between times t_1 and t_2
$\mathsf{Reorg\_free}(x, t_1, t_2)$	object x has a reorganization-free period between times t_1 and t_2
$\mathsf{Reorg\_type}(x, rt, t)$	object x is in a reorganization of type rt at time t
$\mathsf{Repr}(x, rp, t)$	object x has a value of reproducibility rp at time t
$\mathsf{Sc}(x, p, t)$	the chance of survival of object x is p at time t
$\mathsf{Size}(x, s, t)$	object x has a size s at time t
$\mathsf{Time}(t)$	object t is a time-point
$x > y$	value x is larger than value y

Table 6.3: The meaning of the relation symbols.

that are used in the formulas of the inertia fragment.

As noted above, OE stipulates that inertia—not flexibility—helps organizations to survive (Theorem 1), the reason being that inertia is associated with reliability and other features that help organizations to survive. Assumptions 1–3 are put forward to justify Theorem 1 (the original justification in [Hannan and Freeman 1984] is not sound, but one can derive the theorem by strengthening the assumptions, as shown in [Péli et al. 1994]).

Assumption 1 Organizations with higher reliability and accountability have higher survival chance:
$\forall p_1, p_2, ra_1, ra_2, t_1, t_2, x, y([\mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Relacc}(x, ra_1, t_1) \wedge \mathsf{Relacc}(y, ra_2, t_2)$
$\wedge\, \mathsf{Sc}(x, p_1, t_1) \wedge \mathsf{Sc}(y, p_2, t_2)] \wedge (ra_2 > ra_1) \rightarrow (p_2 > p_1))$

Assumption 2a Organizations with higher reproducibility have higher reliability and accountability:
$\forall ra_1, ra_2, rp_1, rp_2, t_1, t_2, x, y([\mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Relacc}(x, ra_1, t_1) \wedge \mathsf{Relacc}(y, ra_2, t_2)$
$\wedge\, \mathsf{Repr}(x, rp_1, t_1) \wedge \mathsf{Repr}(y, rp_2, t_2)] \wedge (rp_2 > rp_1) \rightarrow (ra_2 > ra_1))$

[1] Note that the theory is not very parsimonious: assumption 5, 7 and 8 are not used for any derivation according to the theory.

Assumption 2b Organizations with higher reliability and accountability have higher reproducibility:
$\forall ra_1, ra_2, rp_1, rp_2, t_1, t_2, x, y([\mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Relacc}(x, ra_1, t_1) \wedge \mathsf{Relacc}(y, ra_2, t_2)$
$\wedge \mathsf{Repr}(x, rp_1, t_1) \wedge \mathsf{Repr}(y, rp_2, t_2)] \wedge (ra_2 > ra_1) \rightarrow (rp_2 > rp_1))$

Assumption 3a Reorganization-free organizations with higher reproducibility have higher inertia:
$\forall i_1, i_2, rp_1, rp_2, t_1, t_2, x, y([\mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Reorg\_free}(x, t_1, t_1) \wedge \mathsf{Reorg\_free}(y, t_2, t_2)$
$\wedge \mathsf{Repr}(x, rp_1, t_1) \wedge \mathsf{Repr}(y, rp_2, t_2) \wedge \mathsf{Iner}(x, i_1, t_1) \wedge \mathsf{Iner}(y, i_2, t_2)] \wedge (rp_2 > rp_1) \rightarrow (i_2 > i_1))$

Assumption 3b Reorganization-free organizations with higher inertia have higher reproducibility:
$\forall i_1, i_2, rp_1, rp_2, t_1, t_2, x, y([\mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Reorg\_free}(x, t_1, t_1) \wedge \mathsf{Reorg\_free}(y, t_2, t_2)$
$\wedge \mathsf{Repr}(x, rp_1, t_1) \wedge \mathsf{Repr}(y, rp_2, t_2) \wedge \mathsf{Iner}(x, i_1, t_1) \wedge \mathsf{Iner}(y, i_2, t_2)] \wedge (i_2 > i_1) \rightarrow (rp_2 > rp_1))$

The next two theorems spell out the consequences of environmental selection through time: (surviving) organizations will tend to become increasingly inert (Theorem 2), so their survival chances increase (Theorem 3). Justifying these theorems requires the assumption that "reproducibility of structure increases monotonically with age" (Assumption 4).

Assumption 4 The reproducibility of reorganization-free organizations increases with time:
$\forall rp_1, rp_2, t_1, t_2, x([\mathsf{Time}(t_1) \wedge \mathsf{Time}(t_2) \wedge \mathsf{O}(x, t_1) \wedge \mathsf{O}(x, t_2) \wedge \mathsf{Reorg\_free}(x, t_1, t_2)$
$\wedge \mathsf{Repr}(x, rp_1, t_1) \wedge \mathsf{Repr}(x, rp_2, t_2)] \wedge (t_2 > t_1) \rightarrow (rp_2 > rp_1))$

Under the heading of 'reorganization,' Theorem 4 covers organizational change: organizations may attempt structural change, but such change puts the organization more at risk than inertia. Hannan and Freeman (1984) claim that Theorem 4 relies on the assumptions that reorganization lowers the reliability of organizational performance (Assumption 6), and that the structural inertia of an organization increases with size (for organizations belonging to the same class; Assumption 5).

Assumption 5 Larger organizations of the same class have higher inertia:
$\forall c, i_1, i_2, s_1, s_2, t_1, t_2, x, y([\mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Class}(x, c, t_1) \wedge \mathsf{Class}(y, c, t_2) \wedge \mathsf{Size}(x, s_1, t_1)$
$\wedge \mathsf{Size}(y, s_2, t_2) \wedge \mathsf{Iner}(x, i_1, t_1) \wedge \mathsf{Iner}(y, i_2, t_2)] \wedge (s_2 > s_1) \rightarrow (i_2 > i_1))$

Assumption 6 The reliability and accountability of reorganizing organizations decreases with time:
$\forall ra_1, ra_2, t_1, t_2, x([\mathsf{Time}(t_1) \wedge \mathsf{Time}(t_2) \wedge \mathsf{O}(x, t_1) \wedge \mathsf{O}(x, t_2) \wedge \mathsf{Reorg}(x, t_1, t_2)$
$\wedge \mathsf{Relacc}(x, ra_1, t_1) \wedge \mathsf{Relacc}(x, ra_2, t_2)] \wedge (t_2 > t_1) \rightarrow (ra_1 > ra_2))$

The last and fifth theorem of the inertia fragment states that "complexity increases the risk of death due to reorganization." To simplify the setup, the fifth theorem is added as a premise (Assumption 7); this gives a total of seven assumptions.[2]

[2]The fifth theorem is a kind of meta-theorem. It takes as an assumption that *complexity*

Assumption 7 (Theorem 5) More complex organizations of the same class have lower survival chances after reorganizations of the same type:
$\forall c, c_1, c_2, p, p_1, p_2, re, t_a, t_b, t_c, x, y([\mathsf{O}(x, t_a) \wedge \mathsf{O}(y, t_a) \wedge \mathsf{O}(x, t_c) \wedge \mathsf{O}(y, t_c) \wedge \mathsf{Class}(x, c, t_a)$
$\wedge\ \mathsf{Class}(y, c, t_a) \wedge \mathsf{Sc}(x, p, t_a) \wedge \mathsf{Sc}(y, p, t_a) \wedge \mathsf{Reorg}(x, t_a, t_b) \wedge \mathsf{Reorg}(y, t_a, t_c)$
$\wedge\ \mathsf{Reorg\_type}(x, re, t_a) \wedge \mathsf{Reorg\_type}(y, re, t_a) \wedge \mathsf{Reorg\_free}(x, t_b, t_c) \wedge \mathsf{Sc}(x, p_1, t_c)$
$\wedge\ \mathsf{Sc}(y, p_2, t_c) \wedge \mathsf{Compl}(x, c_1, t_a) \wedge \mathsf{Compl}(y, c_2, t_a)] \wedge (c_2 > c_1) \rightarrow (p_1 > p_2))$

In addition, the premise set contains twelve premises that formulate the necessary background knowledge, *e.g.*, organizations are either reorganizing or not reorganizing. These premises are implicitly used in the original text (Meaning Postulates 1–12).

Meaning Postulate 1 Reorganization-free from t_1 to t_2 means reorganization-free at t_1 and at t_2:
$\forall x, t_1, t_2(\mathsf{Reorg\_free}(x, t_1, t_2) \rightarrow \mathsf{Reorg\_free}(x, t_1, t_1) \wedge \mathsf{Reorg\_free}(x, t_2, t_2))$

Meaning Postulate 2 Something which is equal cannot be larger, and something that is larger cannot be smaller:
$\forall x, y(\neg((x > y) \wedge (x = y)) \wedge \neg((x > y) \wedge (y > x)))$

Meaning Postulate 3 If x is larger than y and y is larger than z then x is larger than z:
$\forall x, y, z((x > y) \wedge (y > z) \rightarrow (x > z))$

Meaning Postulate 4 A reorganization takes time:
$\forall x, t_a, t_b(\mathsf{Reorg}(x, t_a, t_b) \rightarrow (t_b > t_a))$

Meaning Postulate 5 An organization cannot be reorganization-free and reorganizing at the same time:
$\forall x, t_1, t_2(\neg(\mathsf{Reorg\_free}(x, t_1, t_2) \wedge \mathsf{Reorg}(x, t_1, t_2)))$

Meaning Postulate 6 An organization cannot change its class without reorganizing:
$\forall x, t_1, t_2, c_1, c_2(\mathsf{O}(x, t_1) \wedge \mathsf{O}(x, t_2) \wedge \mathsf{Reorg\_free}(x, t_1, t_2) \wedge \mathsf{Class}(x, c_1, t_1) \wedge \mathsf{Class}(x, c_2, t_2)$
$\rightarrow (c_1 = c_2))$

Meaning Postulate 7 If an organization exists at t_1 and at t_2 then this organization exists between t_1 and t_2:
$\forall x, t, t_1, t_2(\mathsf{O}(x, t_1) \wedge \mathsf{O}(x, t_2) \wedge (t > t_1) \wedge (t_2 > t) \rightarrow \mathsf{O}(x, t))$

Meaning Postulate 8 For every organization, there is some reliability and accountability that it has:
$\forall x, t(\mathsf{O}(x, t) \rightarrow \exists ra(\mathsf{Relacc}(x, ra, t)))$

---

causes longer reorganization periods. Due to the fact that survival chance increases during reorganization-free periods (Theorems 3) and decreases during reorganization (Theorem 4), it concludes that complexity decreases the survival chance due to reorganization. This meta-reasoning requires a slightly more complex form of the algorithm than the version presented in this chapter. We treat the original fifth theorem as an assumption (Assumption 7), which enables us to use the simpler version of the PDC-1 algorithm and still derive the same number of theorems.

Meaning Postulate 9 For every organization, there is some reproducibility that it has:
$\forall x, t(\mathsf{O}(x,t) \rightarrow \exists rp(\mathsf{Repr}(x,rp,t)))$

Meaning Postulate 10 For every organization, there is some survival chance that it has:
$\forall x, t(\mathsf{O}(x,t) \rightarrow \exists p(\mathsf{Sc}(x,p,t)))$

Meaning Postulate 11 For every organization, there is some inertia that it has:
$\forall x, t(\mathsf{O}(x,t) \rightarrow \exists i(\mathsf{Iner}(x,i,t)))$

Meaning Postulate 12 For every organization, there is some class to which it belongs:
$\forall x, t(\mathsf{O}(x,t) \rightarrow \exists c(\mathsf{Class}(x,c,t)))$

6.5 Application of PDC-1 to the Inertia Fragment

This section exemplifies the results of the PDC-1 when applied to the inertia part of OE. Starting with the seven assumptions and twelve background assumptions listed in section 6.4, PDC-1 generates a total of seventeen theorems—twelve more than are presented in the original text. Several of the new theorems are theoretically important. The first five theorems coincide with the theorems of the original text; their theoretical importance has been justified in (Hannan and Freeman 1984).

Theorem 1 Reorganization-free organizations with higher inertia have higher survival chances:
$\forall i_1, i_2, p_1, p_2, t_1, t_2, x, y([\mathsf{Reorg\_free}(x,t_1,t_1) \wedge \mathsf{Reorg\_free}(y,t_2,t_2) \wedge \mathsf{Iner}(x,i_1,t_1)$
$\wedge\, \mathsf{Iner}(y,i_2,t_2) \wedge \mathsf{O}(x,t_1) \wedge \mathsf{O}(y,t_2) \wedge \mathsf{Sc}(x,p_1,t_1) \wedge \mathsf{Sc}(y,p_2,t_2)] \wedge (i_2 > i_1) \rightarrow (p_2 > p_1))$

Theorem 2 The inertia of reorganization-free organizations increases with time:
$\forall i_1, i_2, t_1, t_2, x([\mathsf{Time}(t_1) \wedge \mathsf{Time}(t_2) \wedge \mathsf{Reorg\_free}(x,t_1,t_2) \wedge \mathsf{O}(x,t_1) \wedge \mathsf{O}(x,t_2)$
$\wedge\, \mathsf{Reorg\_free}(x,t_1,t_1) \wedge \mathsf{Reorg\_free}(x,t_2,t_2) \wedge \mathsf{Iner}(x,i_1,t_1) \wedge \mathsf{Iner}(x,i_2,t_2)]$
$\wedge\, (t_2 > t_1) \rightarrow (i_2 > i_1))$

Theorem 3 The survival chance of reorganization-free organizations increases with time:
$\forall p_1, p_2, t_1, t_2, y([\mathsf{Sc}(y,p_1,t_1) \wedge \mathsf{Sc}(y,p_2,t_2) \wedge \mathsf{Time}(t_1) \wedge \mathsf{Time}(t_2) \wedge \mathsf{O}(y,t_1) \wedge \mathsf{O}(y,t_2)$
$\wedge\, \mathsf{Reorg\_free}(y,t_1,t_2)] \wedge (t_2 > t_1) \rightarrow (p_2 > p_1))$

Theorem 4 The survival chance of reorganizing organizations decreases with time:
$\forall p_1, p_2, t_1, t_2, y([\mathsf{Sc}(y,p_1,t_1) \wedge \mathsf{Sc}(y,p_2,t_2) \wedge \mathsf{Time}(t_2) \wedge \mathsf{Time}(t_1) \wedge \mathsf{O}(y,t_2) \wedge \mathsf{O}(y,t_1)$
$\wedge\, \mathsf{Reorg}(y,t_2,t_1)] \wedge (t_1 > t_2) \rightarrow (p_2 > p_1))$

Theorem 5 (Assumption 7) More complex organizations of the same class have lower survival chances after reorganizations of the same type:
$\forall c, c_1, c_2, p, p_1, p_2, re, t_a, t_b, t_c, x, y([\mathsf{O}(x,t_a) \wedge \mathsf{O}(y,t_a) \wedge \mathsf{O}(x,t_c) \wedge \mathsf{O}(y,t_c) \wedge \mathsf{Class}(x,c,t_a)$
$\wedge\, \mathsf{Class}(y,c,t_a) \wedge \mathsf{Sc}(x,p,t_a) \wedge \mathsf{Sc}(y,p,t_a) \wedge \mathsf{Reorg}(x,t_a,t_b) \wedge \mathsf{Reorg}(y,t_a,t_c)$

$\wedge\,\mathsf{Reorg\_type}(x, re, t_a) \wedge \mathsf{Reorg\_type}(y, re, t_a) \wedge \mathsf{Reorg\_free}(x, t_b, t_c) \wedge \mathsf{Sc}(x, p_1, t_c)$
$\wedge\,\mathsf{Sc}(y, p_2, t_c) \wedge \mathsf{Compl}(x, c_1, t_a) \wedge \mathsf{Compl}(y, c_2, t_a)] \wedge (c_2 > c_1) \rightarrow (p_1 > p_2))$

Theorems 6 and 7 are straightforward extensions of Assumptions 4 and 6.

Theorem 6 The reliability and accountability of reorganization-free organizations increases with time:
$\forall ra_1, ra_2, t_1, t_2, y([\mathsf{Relacc}(y, ra_1, t_1) \wedge \mathsf{Relacc}(y, ra_2, t_2) \wedge \mathsf{Time}(t_1) \wedge \mathsf{Time}(t_2) \wedge \mathsf{O}(y, t_1)$
$\wedge\,\mathsf{O}(y, t_2) \wedge \mathsf{Reorg\_free}(y, t_1, t_2)] \wedge (t_2 > t_1) \rightarrow (ra_2 > ra_1))$

Theorem 7 The reproducibility of reorganizing organizations decreases with time:
$\forall rp_1, rp_2, t_1, t_2, x([\mathsf{Time}(t_1) \wedge \mathsf{Time}(t_2) \wedge \mathsf{Reorg}(x, t_1, t_2) \wedge \mathsf{O}(x, t_2) \wedge \mathsf{O}(x, t_1) \wedge \mathsf{Repr}(x, rp_1, t_2)$
$\wedge\,\mathsf{Repr}(x, rp_2, t_1)] \wedge (t_2 > t_1) \rightarrow (rp_2 > rp_1))$

The first important new theorem, Theorem 8, says that organizational size has a positive impact on survival chance. It hinges on Assumption 5, and confirms the theoretical expectation regarding the context of environmental selection. It helps build confidence in the premise set as an adequate representation of OE.

Theorem 8 Larger reorganization-free organizations of the same class have higher survival chances:
$\forall c, p_1, p_2, s_1, s_2, t_1, t_2, x, y([\mathsf{Class}(x, c, t_1) \wedge \mathsf{Class}(y, c, t_2) \wedge \mathsf{Size}(x, s_1, t_1) \wedge \mathsf{Size}(y, s_2, t_2)$
$\wedge\,\mathsf{Reorg\_free}(x, t_1, t_1) \wedge \mathsf{Reorg\_free}(y, t_2, t_2) \wedge \mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Sc}(x, p_1, t_1) \wedge \mathsf{Sc}(y, p_2, t_2)]$
$\wedge\,(s_2 > s_1) \rightarrow (p_2 > p_1))$

Theorems 9 and 10 follow from Theorem 8 on the basis of the reasoning leading to Theorem 1. They are extensions of Theorem 8.

Theorem 9 Larger reorganization-free organizations of the same class have higher reproducibility:
$\forall c, rp_1, rp_2, s_1, s_2, t_1, t_2, x, y([\mathsf{Reorg\_free}(x, t_1, t_1) \wedge \mathsf{Reorg\_free}(y, t_2, t_2) \wedge \mathsf{Repr}(x, rp_1, t_1)$
$\wedge\,\mathsf{Repr}(y, rp_2, t_2) \wedge \mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Class}(x, c, t_1) \wedge \mathsf{Class}(y, c, t_2) \wedge \mathsf{Size}(x, s_1, t_1)$
$\wedge\,\mathsf{Size}(y, s_2, t_2)] \wedge (s_2 > s_1) \rightarrow (rp_2 > rp_1))$

Theorem 10 Larger reorganization-free organizations of the same class have higher reliability and accountability:
$\forall c, ra_1, ra_2, s_1, s_2, t_1, t_2, x, y([\mathsf{Class}(x, c, t_1) \wedge \mathsf{Class}(y, c, t_2) \wedge \mathsf{Size}(x, s_1, t_1) \wedge \mathsf{Size}(y, s_2, t_2)$
$\wedge\,\mathsf{Reorg\_free}(x, t_1, t_1) \wedge \mathsf{Reorg\_free}(y, t_2, t_2) \wedge \mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Relacc}(x, ra_1, t_1)$
$\wedge\,\mathsf{Relacc}(y, ra_2, t_2)] \wedge (s_2 > s_1) \rightarrow (ra_2 > ra_1))$

Theorem 11, however, is unexpected: the normal expectation is that organizations can decrease in size without reorganizing. Theorem 11 points either to a weakness in the premise set as the formal representation of OE, or to a limitation of OE itself. On closer inspection, the premise set appears to provide an adequate representation of the theory's assumptions (fortunately, the original text [Hannan and Freeman 1984] gives an explicit list of both assumptions and theorems). So we may conclude that Theorem 11 points to a limitation of OE. OE appears less general than expected, or, to put it more formally, its apparent class of models is smaller than expected. In fact, OE appears to imply a dichotomy between (1) organizations under "normal" conditions and (2) organizations under reorganization. OE's theoretical setup dictates that any decrease in size requires reorganization, so Theorem 11 gives a different meaning to the term "reorganization," or, rather points out how general the meaning of this term is in the theory of OE.

Theorem 11 The size of reorganization-free organizations of the same class does not decrease with time:
$\forall c, s_1, s_2, t_1, t_2, x([\mathsf{Time}(t_1) \wedge \mathsf{Time}(t_2) \wedge \mathsf{Reorg\_free}(x, t_1, t_2) \wedge \mathsf{Reorg\_free}(x, t_2, t_2) \wedge \mathsf{Reorg\_free}(x, t_1, t_1) \wedge \mathsf{O}(x, t_2) \wedge \mathsf{O}(x, t_1) \wedge \mathsf{Class}(x, c, t_2) \wedge \mathsf{Class}(x, c, t_1) \wedge \mathsf{Size}(x, s_2, t_2) \wedge \mathsf{Size}(x, s_1, t_1)] \wedge (t_2 > t_1) \rightarrow \neg(s_1 > s_2))$

Theorems 12 through 13 are expected, but they, too, show in subtle ways the limits of OE by demonstrating the equivalence of inertia, reliability, and reproducibility. This equivalence is not intended by the original text, but is required to establish the soundness of Theorem 1 (as argued in [Péli et al. 1994]). By implication, Theorems 12–13 reiterate a problem in the explanatory structure of the original theory.

Theorem 12 Organizations with higher reproducibility have higher survival chance:
$\forall p_1, p_2, rp_1, rp_2, t_1, t_2, x, y([\mathsf{Repr}(x, rp_1, t_1) \wedge \mathsf{Repr}(y, rp_2, t_2) \wedge \mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Sc}(x, p_1, t_1) \wedge \mathsf{Sc}(y, p_2, t_2)] \wedge (rp_2 > rp_1) \rightarrow (p_2 > p_1))$

Theorem 13a Reorganization-free organizations with higher inertia have higher reliability and accountability:
$\forall i_1, i_2, ra_1, ra_2, t_1, t_2, x, y([\mathsf{Reorg\_free}(x, t_1, t_1) \wedge \mathsf{Reorg\_free}(y, t_2, t_2) \wedge \mathsf{Iner}(x, i_1, t_1) \wedge \mathsf{Iner}(y, i_2, t_2) \wedge \mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Relacc}(x, ra_1, t_1) \wedge \mathsf{Relacc}(y, ra_2, t_2)] \wedge (i_2 > i_1) \rightarrow (ra_2 > ra_1))$

Theorem 13b Reorganization-free organizations with higher reliability and accountability have higher inertia:
$\forall i_1, i_2, ra_1, ra_2, t_1, t_2, x, y([\mathsf{Reorg\_free}(x, t_1, t_1) \wedge \mathsf{Reorg\_free}(y, t_2, t_2) \wedge \mathsf{Iner}(x, i_1, t_1) \wedge \mathsf{Iner}(y, i_2, t_2) \wedge \mathsf{O}(x, t_1) \wedge \mathsf{O}(y, t_2) \wedge \mathsf{Relacc}(x, ra_1, t_1) \wedge \mathsf{Relacc}(y, ra_2, t_2)] \wedge (ra_2 > ra_1) \rightarrow (i_2 > i_1))$

The next three theorems, Theorems 14–16, point out some implications of Theorem 5, and so does Theorem 17, but Theorem 17 unexpectedly points to a specific relationship between complexity and size under conditions of reorganization.

Theorem 14 More complex organizations of the same class have lower or equal reliability and accountability after reorganizations of the same type:
$\forall c, c_1, c_2, p, ra_1, ra_2, re, t_a, t_b, t_c, x, y([\mathsf{O}(x, t_a) \wedge \mathsf{O}(y, t_a) \wedge \mathsf{Class}(x, c, t_a) \wedge \mathsf{Class}(y, c, t_a) \wedge \mathsf{Sc}(x, p, t_a) \wedge \mathsf{Sc}(y, p, t_a) \wedge \mathsf{Reorg}(x, t_a, t_b) \wedge \mathsf{Reorg}(y, t_a, t_c) \wedge \mathsf{Reorg\_type}(x, re, t_a) \wedge \mathsf{Reorg\_type}(y, re, t_a) \wedge \mathsf{Reorg\_free}(x, t_b, t_c) \wedge \mathsf{Compl}(x, c_1, t_a) \wedge \mathsf{Compl}(y, c_2, t_a) \wedge \mathsf{O}(x, t_c) \wedge \mathsf{O}(y, t_c) \wedge \mathsf{Relacc}(x, ra_1, t_c) \wedge \mathsf{Relacc}(y, ra_2, t_c)] \wedge (c_2 > c_1) \rightarrow \neg(ra_2 > ra_1))$

Theorem 15 More complex organizations of the same class have lower or equal reproducibility after reorganizations of the same type:

$\forall c, c_1, c_2, p, re, rp_1, rp_2, t_a, t_b, t_c, x, y([\mathsf{O}(x, t_a) \wedge \mathsf{O}(y, t_a) \wedge \mathsf{Class}(x, c, t_a) \wedge \mathsf{Class}(y, c, t_a)$
$\wedge\ \mathsf{Sc}(x, p, t_a) \wedge \mathsf{Sc}(y, p, t_a) \wedge \mathsf{Reorg}(x, t_a, t_b) \wedge \mathsf{Reorg}(y, t_a, t_c) \wedge \mathsf{Reorg\_type}(x, re, t_a)$
$\wedge\ \mathsf{Reorg\_type}(y, re, t_a) \wedge \mathsf{Reorg\_free}(x, t_b, t_c) \wedge \mathsf{Compl}(x, c_1, t_a) \wedge \mathsf{Compl}(y, c_2, t_a) \wedge \mathsf{O}(x, t_c)$
$\wedge\ \mathsf{O}(y, t_c) \wedge \mathsf{Repr}(x, rp_1, t_c) \wedge \mathsf{Repr}(y, rp_2, t_c)] \wedge (c_2 > c_1) \rightarrow \neg(rp_2 > rp_1))$

Theorem 16 More complex organizations of the same class have lower or equal inertia after reorganizations of the same type:
$\forall c, c_1, c_2, i_1, i_2, p, re, t_a, t_b, t_c, x, y([\mathsf{O}(x, t_a) \wedge \mathsf{O}(y, t_a) \wedge \mathsf{Class}(x, c, t_a) \wedge \mathsf{Class}(y, c, t_a)$
$\wedge\ \mathsf{Sc}(x, p, t_a) \wedge \mathsf{Sc}(y, p, t_a) \wedge \mathsf{Reorg}(x, t_a, t_b) \wedge \mathsf{Reorg}(y, t_a, t_c) \wedge \mathsf{Reorg\_type}(x, re, t_a)$
$\wedge\ \mathsf{Reorg\_type}(y, re, t_a) \wedge \mathsf{Reorg\_free}(x, t_b, t_c) \wedge \mathsf{Compl}(x, c_1, t_a) \wedge \mathsf{Compl}(y, c_2, t_a) \wedge \mathsf{O}(x, t_c)$
$\wedge\ \mathsf{O}(y, t_c) \wedge \mathsf{Reorg\_free}(x, t_c, t_c) \wedge \mathsf{Reorg\_free}(y, t_c, t_c) \wedge \mathsf{Iner}(x, i_1, t_c) \wedge \mathsf{Iner}(y, i_2, t_c)]$
$\wedge\ (c_2 > c_1) \rightarrow \neg(i_2 > i_1))$

Theorem 17 More complex organizations of the same class have lower or equal size after reorganizations of the same type:
$\forall c, c_1, c_2, new\_c, p, re, s_1, s_2, t_a, t_b, t_c, x, y([\mathsf{O}(x, t_a) \wedge \mathsf{O}(y, t_a) \wedge \mathsf{Class}(x, c, t_a) \wedge \mathsf{Class}(y, c, t_a)$
$\wedge\ \mathsf{Sc}(x, p, t_a) \wedge \mathsf{Sc}(y, p, t_a) \wedge \mathsf{Reorg}(x, t_a, t_b) \wedge \mathsf{Reorg}(y, t_a, t_c) \wedge \mathsf{Reorg\_type}(x, re, t_a)$
$\wedge\ \mathsf{Reorg\_type}(y, re, t_a) \wedge \mathsf{Reorg\_free}(x, t_b, t_c) \wedge \mathsf{Compl}(x, c_1, t_a) \wedge \mathsf{Compl}(y, c_2, t_a)$
$\wedge\ \mathsf{Reorg\_free}(x, t_c, t_c) \wedge \mathsf{Reorg\_free}(y, t_c, t_c) \wedge \mathsf{O}(x, t_c) \wedge \mathsf{O}(y, t_c) \wedge \mathsf{Class}(x, new\_c, t_c)$
$\wedge\ \mathsf{Class}(y, new\_c, t_c) \wedge \mathsf{Size}(x, s_1, t_c) \wedge \mathsf{Size}(y, s_2, t_c)] \wedge (c_2 > c_1) \rightarrow \neg(s_2 > s_1))$

In sum, the partial closure through PDC-1 has improved the theory. Theorems 6 through 10 confirm the theoretical expectations. Theorem 11 is unexpected, and contradicts the theoretical expectation. Since the intermediate theory appeared to be adequate, we revised the theoretical expectations instead. The consequence is that the apparent class of models of the theory is reduced. Theorems 12–13 are expected (given the revised theoretical expectations). The next three theorems, Theorems 14–16, strengthen the original theoretical expectations, but the last, Theorem 17, is also unexpected, and forces an update of the theoretical expectations. The premises of the inertia fragment and the theorems derived by PDC-1 are shown in Figure 6.2. The nodes represent the theorems (a theorem relates the top-node with the node); the arrows denote the premises that yield the theorems.

6.6 Discussion and Conclusions

We have argued that formal logic helps researchers to improve a theory in various ways. In a static perspective, logic can help them to answer questions of consistency and explanatory soundness. In a dynamic perspective, logic can help to discover hidden implications of a given theory, or, more precisely, implications of a formal representation of the theory in logical terms. As the logical cycle demonstrates, the partial closure can advance a theory in various ways, either by reinforcing original theoretical expectations about a domain, or by suggesting a modification of those expectations. Conversely, if there is no reason to modify the expectations, the partial closure can point out weaknesses in the formal

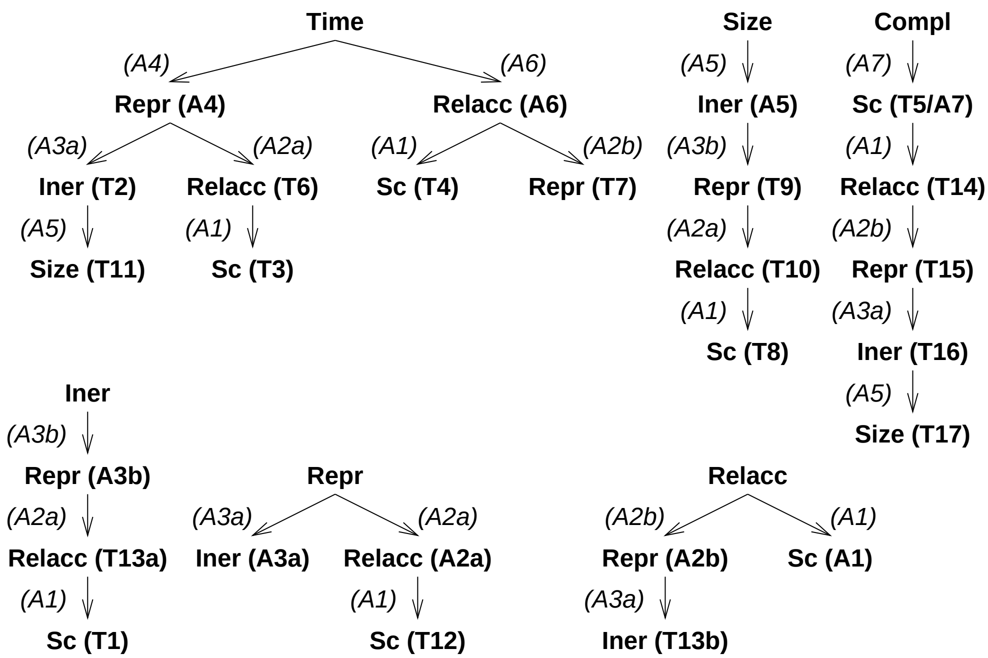

Figure 6.2: The assumptions (An) and theorems (Tn) of the inertia fragment.

representation of the theory.

PDC-1 has its limits, of course.

First, the algorithm works only on a fragment of FOL, namely on formulas that we called "single property to single property." This fragment is arguably important—important enough to allow for a formalization of the inertia part of OE—but it does not have the full expressive power of FOL.

Second, FOL itself has its limits. There has been a lively debate in philosophy about the use of FOL as a tool for formalizing scientific theories (Ayer 1959). Many scientific theories involve notions that FOL cannot handle directly, such as counterfactual conditions and intensional constructs. One could even argue that scientific laws are not material implications as provided by FOL. Because of this, there is a broad agreement that FOL on its own is too weak to formalize all scientific theories. However, our goals are more modest; we focus on theories that do not transcend the expressive boundaries of FOL. Even if scientific laws are not material implications they do imply material implications (as was kindly pointed out to us by one reviewer), and it remains of interest to generate implicit consequences thereof. *Organizational ecology*, the domain theory examined in this chapter, is restricted to object-related statements about properties and relations.

Third, FOL is not necessarily the most elegant or efficient language. Once a representation in FOL has been generated for a specific domain, simpler, or more parsimonious representations may suggest themselves (Newell and Simon 1972; Brachman and Levesque 1985). But the general experience in natural language representation points to a trade off between specificity and flexibility. For specific domains, specialized languages may appear more appropriate, but such languages are not easily generalizable to other domains. Conversely, a general language may not give the most efficient representation for a particular domain, but it carries over more easily to other domains; because of this, a general language is more appropriate for a generic application for theorem-finding. Using FOL has one additional, very important advantage: its formal properties are well-understood. The formal properties of specialized *ad hoc* languages are, as a rule, not known. For example, without a formal semantics, one has no criteria for soundness; without a proof theory, one has no machinery for derivations.

Fourth, PDC-1 is restricted to finding the logical implications of a given set of premises. It cannot generate new conclusions in a logical sense. In fact, once the logic plus a set of premises are fixed, no deductive procedure can generate logically new conclusions; logically new conclusions require new premises or a new logic. The motivation for this research was to generate theorems that are "new" in an *empirical* sense: implied by the logic but (perhaps) unknown to the researcher. Such conclusions may or may not be of particular interest—in this sense the choice to focus on SPtSP formulas is of a heuristic nature. There is no guarantee that PDC-1 (or more general algorithms, for that matter) will always generate interesting (empirically) new theorems. In our case study the algorithm does generate interesting theorems, but more cases are needed to settle this empirical matter.

The usefulness of PDC-1 was demonstrated for an important organization theory, Organizational Ecology; the algorithm generated a set of new theorems, including some of real theoretical importance, notably theorem 11. The fact that the theory implies that organizations cannot decrease in size under normal conditions is clearly important for gauging the OE's scope and setup. Of course, the algorithm's job could also have been done "by hand." In fact, some of the new theorems had already been detected by hand, as reported in (Péli et al. 1994); but then, some had not.

Our tool can also be used during the original formalization of the theory. Recall that the algorithm also identified all previously known theorems from the premises, so it could have been used to check the soundness of the original theory. In fact, the investigation of the inertia fragment by our method would have revealed several more or less serious flaws in the explanatory structure of the original presentation of the theory. In particular, it would have pointed out that the derivation of Theorem 1 (selection favors inertia), arguably the most important theorem of the inertia part, is unsound, and that a sound derivation requires additional qualifications that reduce the scope of the original theory quite considerably, see (Péli et al. 1994). Furthermore, the algorithm can help to enlarge the scope of the theory by helping the theoretician to find out "what would happen if" he would add new assumptions to the original premise set. In this way, PDC-1 makes an important step towards an application for logical simulation.

As a direct follow-up of the reported research, we want to extend the PDC-1 algorithm to other classes of theorems. Taking into account that a complete deductive closure will comprise infinitely many theorems (most of them completely uninteresting), extending the PDC-1 algorithm should be done with care. The job of the algorithm is not only to derive a particular class of theorems, but also to ensure that nothing else is derived.

Chapter 7

Qualitative Reasoning beyond the Physics Domain

The ability to reason qualitatively about physical systems is important to understanding and interacting with the physical world for both humans and intelligent machines. Accordingly, its study has become an important subject of research in artificial intelligence. Qualitative reasoning techniques are traditionally associated with the domain of physics, although the domain of application is, in fact, much broader. This chapter investigates the application of qualitative reasoning techniques beyond the domain of physics. It presents a case study of application in the social sciences: the density dependence theory of organizational ecology. It discusses how the different nature of soft science domains will complicate the process of model building. Furthermore, it shows that the "model building" process can also help making theoretically important decisions, and, as a result, have an impact on the original theory. This will require a shift in focus from the "model simulation" process towards the "model building" process.

7.1 Introduction

During the last decades, Qualitative Reasoning (QR) has been an active area of research (Weld and de Kleer 1990; Kuipers 1994). The field has reached a consensus on the main issues. As a consequence of this, the time has come to think about extending its domain of application. QR is traditionally associated with the physics domain. This domain of application has been so dominant that qualitative reasoning is often called qualitative physics (for example, Forbus 1988). Extending the domain of QR prompts an interesting question: is the dominant relation between QR and physics based on ontological arguments?

The application of QR outside the traditional physics domain seems, indeed, possible. Kuipers (1994) lists applications in biology (irreversible population change, predator-prey ecology), chemistry (chemical engineering), economics (supply and demand, micro-economics), and medicine (glaucoma, drug metabolism).[1] The answer to our question appears to be negative, there are no ontological reasons to explain why the majority of QR-research deals with physics applications. Does the lack of ontological arguments mean that the relation between QR and physics is purely accidental? Probably not, there may be other, pragmatical arguments to explain this relation. A plausible explanation for the historical choice to reason about physical systems is the formal, well-understood nature of the physics domain. This indicates that QR outside physics, albeit possible, may still not be the very same as QR inside physics. The different nature of the domain may require different emphases. This prompts another question: does the application of QR outside physics require a change in methodology?

It is with these questions in mind that we performed the case study reported in this chapter. The intention of the chapter is to investigate the differences between physics and other domains in the context of QR. Therefore, a natural choice of domain for this case study is a "soft" science domain. The soft sciences are in many respects the opposite of physics, in being highly non-formal, less well-understood. We have chosen to build a QR-application in the social sciences.

This chapter is organized in the following way. Section 7.2 gives a short introduction to the framework for qualitative reasoning that we used for our research. Section 7.3 and 7.4 introduce the domain of our case study, i.e., the density dependence theory of organizational ecology. Section 7.5 describes the qualitative density dependence model, and summarizes its qualitative behavior prediction. Section 7.6 lists the results that were obtained during the construction and use of the qualitative model. And finally, in section 7.7 we evaluate our case study in retrospect, trying to answer the questions posed above.

7.2 Representational Context

The model of this chapter is implemented in a domain-independent qualitative reasoning shell called GARP (Bredeweg 1992). Garp incorporates many features of the component-centered (De Kleer and Brown 1984), and the process-centered (Forbus 1984) approaches in QR. Initial conditions are described in *case models*, the theory itself in *model fragments* (consisting of *conditions* and *givens*). Case models and model fragments can be expressed in terms of: *entities* (like liquid), *quantities* (like amount), *values* and *derivatives* (like $\{-, 0, +\}$), and *dependencies* (like (in)equalities, proportionalities, influences, etc.); or in terms of other model fragments.

The behavior of a system during a particular time period is described by the set of applicable model fragments. The behavior over time periods is determined by the application of transition rules between states of behavior.

[1] Several other authors have also reported applications of QR outside physics.

7.3 The Density Dependence Theory

Mainstream organizational theories regard organizations as agents that adapt rationally to changing environments (Thompson 1967; Mintzberg 1979). These theories describe organizations from an *individual* viewpoint. Complementarily, a change in environmental resource conditions affects the whole *population* of organizations. For example, if resource conditions deteriorate, the total population of organizations will decline (despite the efforts of individuals to avoid this fate). Organizational ecology (Hannan and Freeman 1989) describes the process by which organizational populations grow and decline due to changing environmental conditions. Organizational ecology abstracts from the rational behavior of individuals, populations are solely dependent on the environmental conditions.

The density dependence theory (Hannan and Carroll 1992) is at the heart of organizational ecology: it describes the dynamics that underlie the growth of an isolated population as a function of the population's density. It serves as a base model for other parts of organizational ecology that investigate the demographic behavior of different (sub)populations under changing environmental conditions (e.g., niche strategists, life history strategists) or during reorganization (e.g., the inertia-fragment [Hannan and Freeman 1989; Péli et al. 1994]). The density dependence theory assumes that the founding and mortality rates of a population are affected by two opposing forces: by the degree of legitimation that the population enjoys and by the intensity of competition between the members of the population.

Legitimation reflects the institutional standing of the population. A high level of legitimation means that the organizational population has the status of a taken-for-granted solution to given problems. Organizations of high legitimation are desirable partners for other organizations when making exchange relations. Moreover, founding new organizations in a highly legitimated population is also easier. The theory assumes that the founding rate is directly proportional to the legitimation of the population, and that the mortality rate is inversely proportional to it. Legitimation increases monotonically with density. The beneficial effect of growing density is especially important when there are only a few organizations in the environment. If organization density is high already, then the founding of an additional organization does not improve the population's institutional standing significantly. The higher the density of a population is, the smaller is the legitimating effect of additional organizations.

An intensifying *competition* between the member organizations of the population increases the mortality. It also decreases the founding rate of the population: managers are reluctant to initiate new organizations if the chance of success is low. The theory assumes that the intensity of competition is directly proportional to the mortality rate and inversely proportional to the founding rate. Since competition is about resources, the intensity of competition increases with density. The density dependence theory claims that increasing density intensifies competition at an increasing rate.

The beneficial effects of legitimation prevail at low densities, while the negative effects of competition dominate if density is high. As a result, the demographic rates change with density in a *non-monotonic* way. The founding rate increases over the lower density range and decreases above a certain value. On the other hand, the mortality rate decreases at low densities, and increases later. When the two rates becomes equal, the population reaches its equilibrium size: this value is called the *carrying capacity* of the given resource environment.

Hannan and Carroll (1992, chap.2) give the following description of the intuitive theory specified above:

Competition The intensity of competition, C, increases with density, N, at an increasing rate. That is, $C = \varphi(N)$; and $\varphi' > 0$ and $\varphi'' > 0$.

Legitimation The intensity of legitimation, L, increases with density at a decreasing rate. That is, $L = \upsilon(N)$; and $\upsilon' > 0$ and $\upsilon'' < 0$.

Founding Rate The founding rate of an organizational population, λ, is inversely proportional to the intensity of competition within the population, and directly proportional to the legitimation. That is, $\lambda \propto 1/C$ and $\lambda \propto L$.

Mortality Rate The mortality rate of an organizational population, μ, is directly proportional to the intensity of competition within the population, and inversely proportional to the legitimation. That is, $\mu \propto C$ and $\mu \propto 1/L$.

For environments with a positive carrying capacity, it is assumed that legitimation exceeds competition at low densities, that is, $L_i \geq C_i$ for $i < N$. To avoid negative founding and mortality rates, the range of the legitimation and competition functions also has to be confined to non-negative numbers. Since legitimation and competition occur in the denominator of the mortality and the founding rate, respectively, their value cannot be zero either. The theory assumes that $L > 0$ and $C > 0$.

The legitimation and competition functions are depicted in Figure 7.1. Competition increases with density at an increasing rate, and legitimation increases with density at a decreasing rate. In this figure, N_0 denotes the initial point, and N_{cc} denotes the point where legitimation and competition are equal.

7.4 Applying the Theory

Our aim is to simulate the growth pattern of populations, in other words, how the density of a population changes over time. Hannan and Freeman (1989) use the Lotka-Volterra definition of growth rate, ρ, as the difference between founding and mortality rates of the population. That is, $\rho = \lambda - \mu$. The growth rate can be

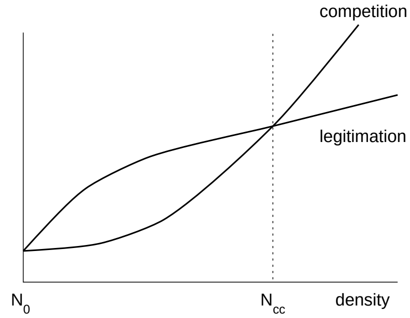

Figure 7.1: Legitimation and competition as functions of density.

calculated from legitimation and competition directly: if $\lambda = L/C$ and $\mu = C/L$ then $\rho = L/C - C/L$.[2]

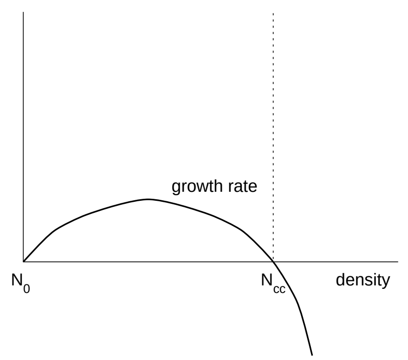

Figure 7.2: Growth rate as a function of density.

Figure 7.2 depicts the growth rate of an organizational population, based on the legitimation and competition functions in Figure 7.1.

We can now investigate the theory's predictions about the change of population size. If the environment is initially empty (Figure 7.3), then the density dependence theory predicts a sigmoid (or S-shaped) population growth. This growth pattern is the well-known *logistic curve* of the Lotka-Volterra model.

[2]In fact, founding and mortality rates may differ by a factor. That is, $\lambda = a * L/C$ and $\mu = b * C/L$. We ignore these factors to allow for the direct calculation of the growth rates, because the use of a simple model helps to convey the main point of our argument. Moreover, if the factors a and b are given we can rescale the C and L functions to $C^\star$ and $L^\star$, so that $\lambda = L^\star/C^\star$ and $\mu = C^\star/L^\star$. The rescaled $C^\star$ ($L^\star$) still satisfies the criteria of increasing at an increasing (decreasing) rate.

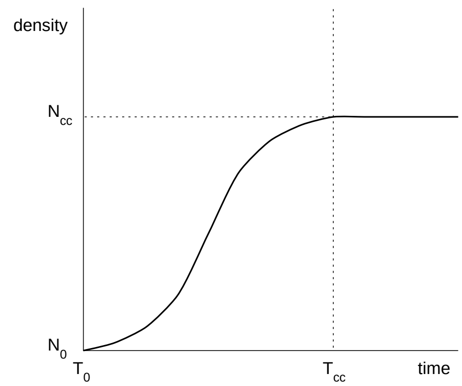

Figure 7.3: Expected behavior for an empty environment.

7.5 Qualitative Density Dependence

In section 7.4 we have established that growth rate $\rho = L/C - C/L$. Having only positive legitimation and competition values, this means that $\rho > 0$ if $L > C$, $\rho = 0$ if $L = C$, and $\rho < 0$ if $C > L$. The sign of $L - C$ equals the sign of ρ. The qualitative model of growth rate therefore is $\rho =_Q L - C$.[3]

Section 7.3 also gives constraints on legitimation and competition. Legitimation is increasing with density at a decreasing rate, $\delta L/\delta N > 0$ and $\delta^2 L/\delta N^2 < 0$. Competition is increasing with density at an increasing rate, $\delta C/\delta N > 0$ and $\delta^2 C/\delta N^2 > 0$. But instead of derivatives to density, derivatives to time are needed. If legitimation is greater than competition (a growing population) density will increase with time. This means that we can use the derivatives listed above. This qualitative behavior can be modeled as follows: $L \propto_Q N$, $\delta L \propto_Q 1/N$, $C \propto_Q N$, and $\delta C \propto_Q N$.

If legitimation is smaller than competition (a declining population) density will decrease with time. This means that we have to read Figure 7.1 from the right to the left. In this part, competition is decreasing with decreasing density at an *increasing* rate (i.e, the derivative is negative, but the second-order derivative is positive), and legitimation is decreasing with decreasing density at a *decreasing* rate. This means that the signs of the second-order derivatives of L and C change. This qualitative behavior can be modeled as follows: $L \propto_Q N$, $\delta L \propto_Q N$, $C \propto_Q N$, and $\delta C \propto_Q 1/N$.

If density is increasing, it has a positive effect on the second order derivative of competition and a negative effect on the second order derivative of legitimation. If density is decreasing these effects are reversed. In the equilibrium points (le-

[3] To denote the difference between qualitative dependencies and their mathematical counterparts, we use "$=_Q$" instead of "$=$" for equalities, "$\propto_Q$" instead of "$\propto$" for proportionalities, and "I" for influences. To be able to deal with higher order derivatives, we introduce new quantities (named $\delta \ldots$) that correspond to the derivatives of the original quantities (using influences).

gitimation equals competition, i.e., N_0 and N_{cc}), density is not changing, making this difference disappear.

$$
\begin{array}{rcl}
N & I^+ & \rho \\
L & I^+ & \delta L \\
C & I^+ & \delta C \\
L & \propto_{Q^+} & N \\
\delta L & \propto_{Q^-} & N \\
(\delta L & \propto_{Q^+} & N) \\
C & \propto_{Q^+} & N \\
\delta C & \propto_{Q^+} & N \\
(\delta C & \propto_{Q^-} & N) \\
\rho & =_Q & L - C
\end{array}
\begin{array}{r}
(7.1) \\ (7.2) \\ (7.3) \\ (7.4) \\ (7.5) \\ (7.6) \\ (7.7) \\ (7.8) \\ (7.9) \\ (7.10)
\end{array}
$$

Figure 7.4: Dependencies of qualitative density dependence.

Figure 7.4 summarizes the model. Dependencies 7.1 to 7.3 are included for technical reasons, they allow the use of higher-order derivatives by modeling them as normal values. Dependencies 7.4 and 7.5 ensure that legitimation is increasing with density at a decreasing rate (if legitimation is smaller than competition, dependency 7.6 replaces 7.5). Dependencies 7.7 and 7.8 ensure that competition is increasing with density at an increasing rate (if legitimation is smaller than competition, dependency 7.9 replaces 7.8). Dependency 7.10 calculates the growth rate from the legitimation and competition values.

We have now modeled the causal chain of the density dependence theory: i) the trade-off between competition and legitimation causes a certain growth rate, ii) the growth rate will affect the density of the population, iii) the change in density, in its turn, will affect competition and legitimation, iv) etcetera.

Using this qualitative density dependence model, the qualitative reasoning shell GARP can make behavior predictions of the following case models:

Case A: An empty environment Scenario A describes an environment that contains no members, but has some positive amount of resources. In this scenario, the initial values are $N = 0$ (the initial density is zero), $L = C$, and $\delta L > \delta C$ (indicating the presence of resources). The behavior prediction is as follows (see Figure 7.5): State 1 is (just before) the point T_0, state 2 corresponds to T_0, state 3 to the interval between T_0 and T_1, state 4 to the point T_1, state 5 to the interval between T_1 and T_2, and finally, state 6 to point T_2.

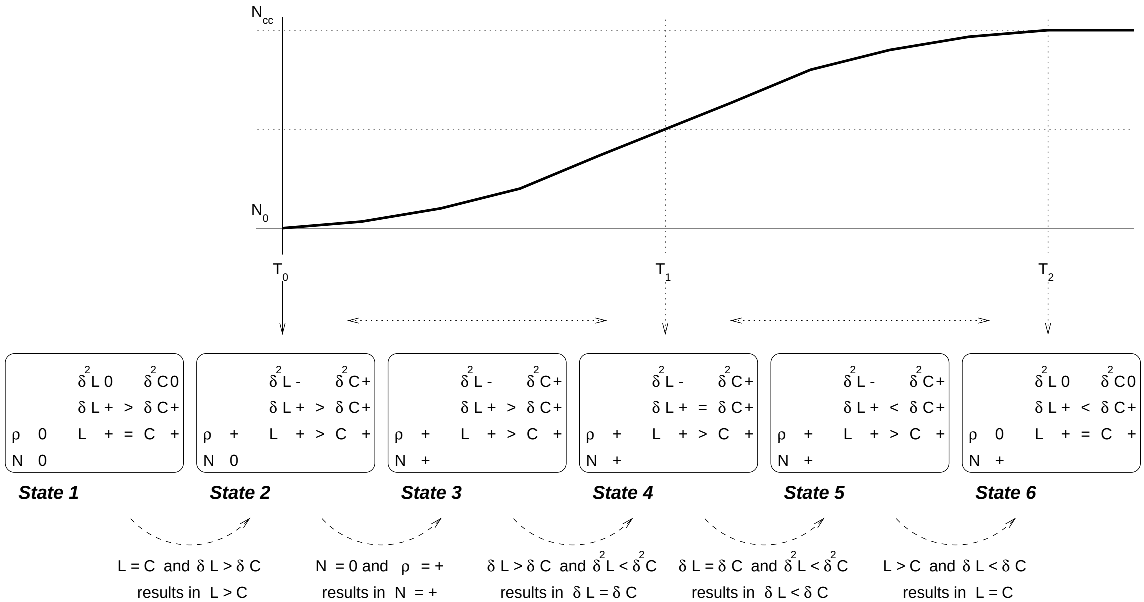

Figure 7.5: Case A: An empty environment.

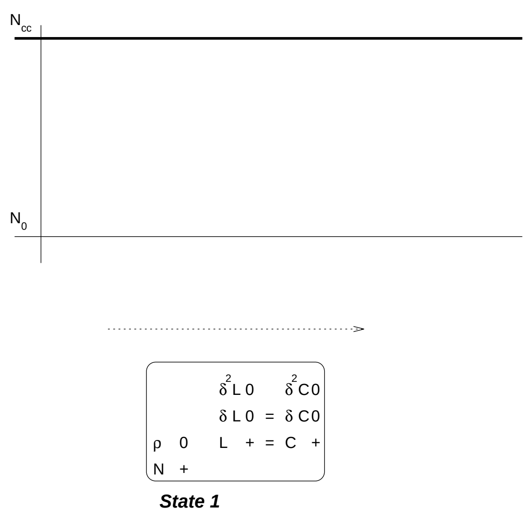

Figure 7.6: Case B: A population in equilibrium.

Case B: A population in equilibrium Scenario B describes an organizational population that is at its carrying capacity: there is a positive density, and the legitimation and competition are in balance. In this scenario, the initial values are $N > 0$ (the initial density is positive), $L = C$, $\delta L = 0$, and $\delta C = 0$ (the competition and legitimation are in equilibrium). The behavior prediction of Case B results in a single steady state (see Figure 7.6).

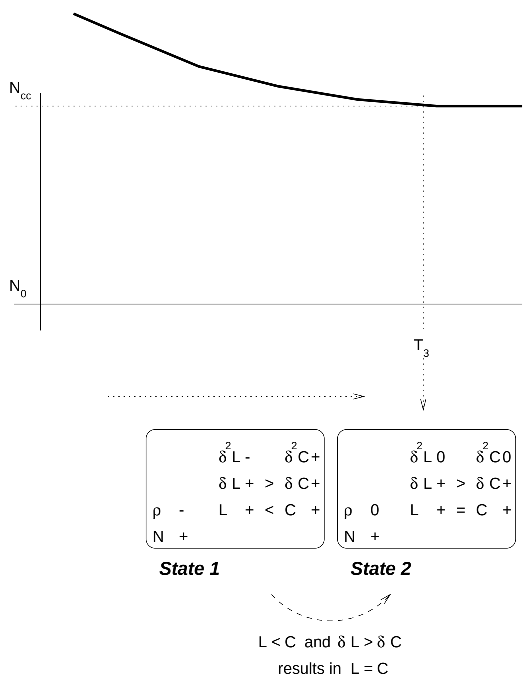

Figure 7.7: Case C: An overcrowded environment.

Case C: An overcrowded environment Scenario C describes an organizational population that is above its carrying capacity: there is a positive density, but the competition exceeds the legitimation of the population. In this scenario, the initial values are $N > 0$ (the initial density is positive), $L < C$, and $\delta L > \delta C$ (the competition exceeds the legitimation). The behavior prediction is as follows (see Figure 7.7): State 1 corresponds to the interval before T_3, and state 2 to the point T_3.

7.6 Results

Hannan and Carroll (1992) give a formal mathematical model of density dependence, as well as an explicit qualitative description using the proportionalities and the signs of the derivatives as summarized in section 7.3. The qualitative description provides the intuitions underlying the mathematical description of the theory. For building the qualitative model, we only used the qualitative description of the theory. This resulted in a model that closely follows the theoretical intuitions. It describes the core elements of the theory and abstracts from unnecessary detail. The model can be used to become familiar with the theory and the theory's predictions. The resulting qualitative density dependence model is able to derive the behavior predicted by the theory. The qualitative model is more general than the parametrical model (and the resulting quantitative simulations). A quantitative, parametrical model makes, out of necessity, various non-trivial assumptions about parameters, functions, etcetera.

During the modeling process we had to make several decisions of theoretical importance. These decisions reveal implicit assumptions underlying the theory. Making these assumptions explicit is an important contribution to the original theory. Apart from identifying hidden assumptions underlying the theory, the qualitative simulator was also able to predict unidentified consequences of the theory. We will discuss some of these implicit assumptions and consequences in detail:

First, the theory gives no information about the derivatives of legitimation and competition at zero density. Therefore, it is possible that the derivative of competition is initially higher than the derivative of legitimation, i.e., $\delta C_0 > \delta L_0$ (as depicted in Figure 7.8). That is, the derivative of growth rate is initially

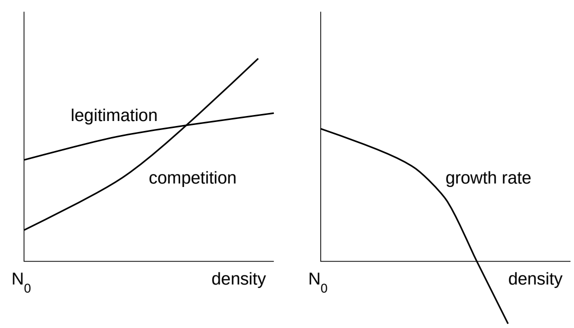

Figure 7.8: Monotonically growing population.

negative, resulting in a *monotonic* population growth. This is in conflict with the theory's claim that the demographic rates are *non-monotonic*. A constraint is needed on the initial derivatives of competition and legitimation: at zero density, the derivative of legitimation is greater than the derivative of competition ($\delta L_0 > \delta C_0$). This constraint is an additional premise that needs to be added to the theory.

Second, in state 3 of case A the simulator predicts a state transition from unequal derivatives of legitimation and competition to equal derivatives. Although this transition is likely to occur, it is not guaranteed to take place. In state 3, $L > C$, $\delta L > \delta C$, $\delta^2 L < 0$, and $\delta^2 C > 0$. There is no guarantee that δL will become equal to δC. If legitimation and competition behave as in Figure 7.9, the

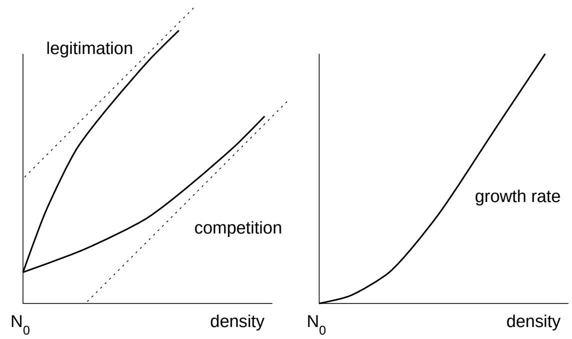

Figure 7.9: Forever expanding population.

population will stay in state 3, that is, it grows exponentially into infinity. This is clearly unintended: real-world's resources are always finite in number. This case can be avoided by assuming that competition exceeds legitimation after a certain density value, that is, there exists a certain density $N^\star$ beyond which $C > L$ (recall that legitimation exceeds competition at low densities).[4] This is another additional premise that needs to be added to the theory.

Third, the decline of overcrowded populations is also captured by the density dependence theory.[5] If there are more organizations around than the carrying capacity of the environment, for instance due to migration, the density falls until the population reaches the carrying capacity (see Figure 7.10). The theory claims

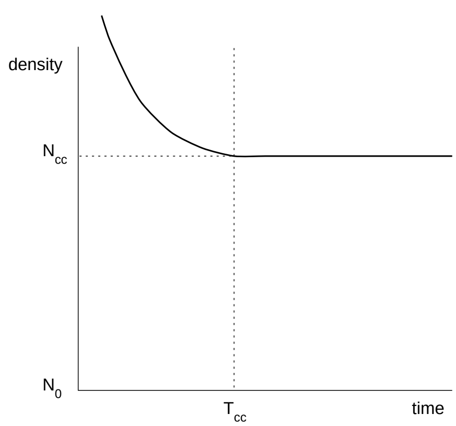

Figure 7.10: Expected behavior for an overcrowded environment.

[4]A different remedy, namely to impose an upper-bound on legitimation, is suggested in (Péli 1993).

[5]This additional case model was found by hand, although it could have been discovered using a *total-envisionment* of states and transitions.

that the density dependence of organizational populations is non-monotonic. This is, indeed, the case for increasing populations. The decrease of populations manifests, in contrast, a monotonic pattern. This is an unknown consequence of the theory that sheds light on the relative speeds of growth and decline processes: it predicts that the decline of a population will be significantly more rapid that its growth.

7.7 Discussion

As discussed in the previous sections, the density dependence theory of organizational ecology was successfully modeled as a QR-application. In accordance with existing research, our case study did not reveal any ontological arguments that would prevent the application of QR outside physics.

A full-fetched discussion of the metaphysics of QR is outside the scope of this chapter. This discussion should focus on the level of *abstraction*. Physics has reached a high level of abstraction using terms such as *energy* or *gravity* as abstractions of underlying forces. These terms have an advanced nature (anyone who tried to explain to a child what *gravity* is will have noticed this). The non-physics domain of this chapter and those of (Kuipers 1994) have also reached a level of abstraction that is sufficient for building QR-applications.

Although feasible, the application of QR in a soft science domain is somewhat different from applying QR in physics. This difference seems to emerge from the different natures of both domains. There is one important difference between the domain of physics and a domain in the soft sciences. Not surprisingly, soft science domains are less well-understood, less formalized. In short, soft sciences lack the deep understanding of domain knowledge that characterizes physics. This is a subtle difference, but it has important methodological consequences.[6]

7.7.1 Finding the Right Model for the Job

A first consequence is that building models on less well-understood, less formal domains will require more effort, and therefore will be more time consuming. Although the value of the "model simulation" process (Forbus 1988) remains, the "model building" process must obtain a more prominent status.

There is another, deeper, consequence. In physics, terms such as *liquid flow* have a clearly established meaning (Hayes 1985a). In the soft sciences, the nomenclature is less developed; there is hardly any consensus. Terms like *legitimation*

[6]In the following, we will exaggerate this difference in understanding of the domains in order to make our arguments more clear. In reality, insufficient knowledge about the domain can play a role in all domains, including physics (as one reviewer put it: "Modeling in the physical sciences and in engineering is still very much an art"). Thus, insights in the way to handle insufficient domain knowledge are useful for QR in general.

vary in their precise meaning among different theories. For example, the term legitimation denotes, in the density dependence theory, the taken-for-grantedness of an organizational population. On many occasions, the precise nature of a variable is only defined by the behavior it manifests in the context of a theory, and not by its label. As a consequence, the constructed domain models will not be very generic, they will only be reusable to a very limited extent.

In the traditional case (Falkenhainer and Forbus 1991), the use of a particular term, say *liquid flow*, induces the use of a particular model fragment specifying the behavior related to this term. But when terminology has no exact meaning this process is reversed: a needed model fragment determines the use of a certain term. The prediction of certain desired behavior, e.g., a sigmoid density function in the density dependence theory, requires the opposing force of two underlying causes. After constructing the model fragments with unlabeled variables, we start to look for sociological meaningful names, for example the terms legitimation and competition, for the two underlying causes.

Summarizing, the application of QR in soft science domains will require more effort in the model building process. Moreover, the constructed model fragments will only be reusable to a limited extent. Thus, it is less likely that the modeling process can be facilitated by existing libraries of generic model fragments.

7.7.2 Finding the Right Job for the Model

The same fact that causes the model building process to be more troublesome, also makes it more worthwhile. Let us analyze what is disturbing the modeling process. Is it the "incompetence" of the model builder? Although the bounded rationality of a human model builder is certainly an important factor that tampers with the modeling process, there is no reason why the same model builder should be less competent when modeling a soft science theory. Is it the "incompetence" of the theory? We think so: during the modeling process many decisions have to be made, that have an impact on the original theory.

The theory can be ambiguous and allow for various interpretations. If these ambiguities are solved in the qualitative model, this solution corresponds also to an improvement of the original theory (see section 7.6 for examples).[7] The explication of the underlying structure of a theory will provide new theoretical insights. The model can reveal underlying assumptions, and thereby shed light on the theory's domain of application. Furthermore, the simulator can identify unforeseen (and even counterintuitive) consequences of a theory, and thereby clarify the theory's predictive and explanatory power. In short, the original theory will evolve in parallel with its qualitative model during the modeling process.[8]

---

[7] Note that handling incomplete knowledge is one of the strong points of QR.

[8] Contemporary philosophy of science argues that theory development follows a cyclic pattern (Kuhn 1962; Lakatos 1976; Balzer et al. 1987). After the initial formation of a theory, it is repeatedly revised to account for anomalous observations (and may, in the end, be abandoned

Suppose we attempt to build the qualitative model of a premature, possibly imperfect theory. Instead of a straightforward translation, the model building will require decisions of theoretical importance. These decisions can be facilitated by simulation runs: various alternatives can be implemented and evaluated for their impact on the behavior prediction. The predicted behavior may not be in accordance with intuitions, logic, or empirical knowledge. These discrepancies will guide the model builder in the revision of the qualitative model (or of the expectations).[9] Moreover, these revisions will also apply to the original theory. In this way, the tedious "debugging" steps in traditional QR acquire a new character. They become experiments at the frontier of a science: every successful and every unsuccessful revision of the model may extend our knowledge about the theory.

---

if the revision would require too radical changes, resulting in a paradigm shift). QR can play an important role in this diachronic structure of a theory, it allows us to recreate the theory-evolution process at a miniature scale.

[9]Several machine learning tools to support the evolutionary model building process (Falkenhainer and Rajamoney 1988), and the diagnose/repair step (Bredeweg and Schut 1993) have been reported.

Chapter 8

Discussion & Related Work

> But to come very near to a true theory, and to grasp its precise application, are two very different things, as the history of science teaches us. Everything of importance has been said before by somebody who did not discover it.
> —Alfred North Whitehead, *The Organisation of Thought*

An exhaustive discussion of relevant related work is beyond the scope of this thesis, since it would need to consider a seemingly endless list of various issues from diverse fields as organization theory, sociology, methodology, philosophy, logic, automated reasoning, and artificial intelligence. Instead, we will address some relatively isolated issues, mainly focusing on those not already discussed in earlier chapters.

8.1 Computational Theory Building

The main line of this thesis has been the practical application of logic to scientific theories, with a special focus on the role that computational support can play in formal theory building. Of course, any of the computational derivations we made in our case studies could have been made 'by hand' by anyone with sufficient proficiency in logic. In fact, each of these derivations, when considered in isolation, would be regarded as rather simple by logicians.[1] However, performing large numbers of these derivations in practice is a different matter and here computational support can play a crucial role. The crucial importance of support tools was recognized long before the availability of computers. One of the earliest formalizations, outside mathematics, of a scientific theory using first-order logic addressed biology (Woodger 1937, 1939). Woodger (1939) deliberates in an epilogue about what a modern visitor to the Solomon's House from Bacon's *New Atlantis* would encounter. The main change he envisions is

> a Calculator—a category of worker not provided for by Bacon. The Calculator operates a gigantic machine, like our machines but of very much wider scope, being capable of working out the consequences of any hypothesis which can be formulated in the universal notion. (Woodger 1939, p.79)

In a sense, we have been trying to provide the type of machine that Woodger envisioned, by investigating the ways in which modern computers can support the axiomatization of, and computational reasoning with, scientific theories. Although the role of computational support should not be underestimated—it played a crucial role and, either directly or indirectly, motivated many decisions we made—the number of pages that explicitly discuss the computer-based implementation has remained fairly limited. This can be immediately explained by our decision to reuse generic computational tools from the field of automated reasoning for the search for proofs and models. As a result of this, the quest for computational support amounts to finding appropriate operationalizations from the various logically equivalent ways to characterize various criteria, such as the consistency of the theory and the soundness of theoretical derivations. It is important to note that, due to the undecidability of first order logic, there are fundamental restrictions on what we can ultimately achieve (Beth 1958b). Well within these limitations, the developments of modern software and, especially, hardware have caused, and continue to cause, a spectacular increase in the amount of problems that can be solved by computational tools (Wos 1998). For example, automated reasoning tools did solve various open problems in mathematics (Slaney 1994a). The relative simplicity of the derivations gives us, perhaps, some confidence that the queries we pose can be solved in practice by computational tools.

The computational component is also one of the main distinctions with earlier efforts to formalize theories in logic, such as the formal approach of (Kyburg 1968) and the structuralist approach to theories (Balzer et al. 1987). Our approach closely resembles that of Kyburg (1968), who presents first-order logic renditions

[1]We are here ignoring the fact that human proofs are usually only partially formal, and that human theorem proving can be error prone (even logicians can have a 'bad' day, see also [Hodges 1998]). Moreover, some of the derivations are, albeit technically not very complex, still difficult to perform by human theorem provers; think of the construction of counterexamples that are nonintented models of the theory.

of theory fragments from the fields of mathematics, geometry, physics, psychology, sociology, and biology. There are striking similarities between our case studies and the sociological example used in Kyburg's chapter 2, and the formalization of a psychological theory in his chapter 11. Kyburg's chapter on sociology, chapter 12, is perhaps somewhat different for it mainly relies on the differential equations used in Simon's mathematical model (1952; 1957) of Homans' theory of the human group (1950). Perhaps the main difference with our computational approach is that Kyburg makes all his derivations manually—indeed an admirable effort.

The structuralist approach is representing theories as set-theoretic models and allows for the substantial use of standard mathematics (Balzer et al. 1987).[2] The structuralist approach has been widely applied to a variety of theories from the empirical sciences. Although mainly motivated by the formalization of physics (Sneed 1971), there are various applications of the structuralist approach to other empirical sciences (Stegmüller 1986; Balzer, Moulines, and Sneed forthcoming, see). These applications include structuralist formalizations of sociological theories (for example, Balzer 1990; Manhart 1994), notwithstanding their (initial) disclaimer that contemporary sociological theory may simply not yield the demanded formal structures (Balzer et al. 1987, p.xix). The wide range of applications shows that the structuralist approach provides convenient notions for formally expressing scientific theories.[3] However, it is yet unclear whether these notions are equally attractive for computational reasoning with them.

8.2 Formal Theory

The formalization of scientific theories in formal logic has a long history. The intimate relation between logic and scientific methodology was at the core of the logical positivism of the Vienna Circle (Carnap 1928; Ayer 1959; Neurath et al. 1970).[4] Many of the classic textbooks on logic have even the dual goal of introducing both logic and scientific methodology (for example, Cohen and Nagel 1934; Tarski 1946; Carnap 1958). There are various ways in which formal logic can be applied to scientific theories (see, for example, the discussion of the potential interaction between modern logic and philosophy of science in [Van

[2] The structuralist approach is also known as the 'non-statement view' on theories (Stegmüller 1973), also sometimes called the new structuralism (to distinguish it from the 'old' structuralism which was in fact closely connected with the traditional axiomatic method [Piaget 1971]).

[3] Whether there are fundamental reasons for using set theory is debatable. As Hintikka (1998, p.304) writes: "[W]hen the sharpest philosophers of science realized that a study of 'the logical syntax of the language of science' was not enough, they resorted to set theory for their conceptualizations. Ironically some misguided philosophers of science have continued to seek salvation in set theory long after the development of logical semantics and systematic model theory."

[4] Although logical positivism was for some time out of favor in philosophy, there is a current re-evaluation of logical positivism (Friedman 1988, 1991).

Benthem 1982]). Despite this potential and despite the illustrious history, there are few current interfaces between logic and philosophy of science (Kuipers 1997). Moreover, there is no equivocal advise on which formalism is most suitable to represent particular theories. Therefore, we decided to apply only classical, well-understood formalism and notions. We used classical first-order logic, and viewed theories as deductive systems (Tarski 1956), in which theoretical explanations and prediction correspond to deductions of theorems (Popper 1959).[5] Of course, this should not be interpreted as a claim that first-order logic is the most suitable logic for representing scientific theories, nor even as a claim that such a logic exists.

Throughout the thesis, we have provided formalizations of actual theories from the field of sociology and, especially, its subfield of organization science. This is, perhaps, a controversial choice, because the paradigm examples of theories in the empirical sciences belong to physics. The field of sociology is even renowned for its lack of systematic theory, and consequently, sociology appears as an unlikely subject for an axiomatization attempt using formal logic. Nevertheless, all sciences, including the social sciences, share the same methodological principles. In particular, whether a theoretical argument in sociology is valid should be judged by the same criteria as an argument in physics. In short, the "logic" of the social sciences is not different from that of the other sciences (Nagel 1956; Rudner 1966; Popper 1969; Neurath 1970).

This does not imply that browsing through a sociological journal gives the same impression as browsing those of fields like, say, physics. There are, of course, major differences between the actual theories of sociology and those of physics. One of the most noticeable differences is the rigorous, mathematical discourse of physics and the informal, qualitative, verbal discourse of sociology. Since descriptions of sociological theories are stated in ordinary language, they are usually partial and incomplete. This is not to say that these verbal theories do not contain creative and interesting ideas. Quite the contrary, sociological theories contain as a rule many original ideas, some of which are genuine insights. Unfortunately, these insights are easily lost within the bewildering richness. The logical formalization of theories is an attempt to ferret out these insights.

In physics, the axioms of theories are regarded as fundamental laws of nature. The premises of the formal theories we reconstruct are not necessarily such self-evident, universally true statements. Our goals are much more modest; we want to represent the theories as they appear in the literature. Our formal reconstructions use assumptions that are, in general, more basic statements than the theorems.

[5] The view of a theory as a deductively closed set of sentences is attributed to Alfred Tarski. Although we will use the expressions 'deductive theory' and 'deductive system' synonymously, it is perhaps worth to point out that Tarski (1956, p.343) uses them "in quite distinct senses." A 'deductive theory' is a realization (or model) of an axiom system, whereas every set of sentences which contains all its consequences is a 'deductive (closed) system.' See (Putnam 1962; Hempel 1970; Suppe 1974) for discussion and criticism of standard notions of theory, and see (Balzer et al. 1987) for a set-theoretic representation of scientific theories.

However, it seems unrealistic to expect that any of them will turn out to be a fundamental law of nature. At least, we would not regard these more basic assumptions as 'final' statements. The axioms we postulate are provisional, in the sense that they may turn out to be oversimplifications.[6] In as far as such provisional axioms are insightful and aid us in understanding the theory, this is in itself a valuable contribution to the theory.[7] We want to stimulate further development of the theories by making their underlying assumptions part of the discussion. If some of the assumptions do not withstand the test of criticism, we can immediately show which of the theorems have to be retracted as well, or which alternative assumptions would rescue them. In short, finding the more basic underlying assumptions can be an important step towards finding some of the (more) fundamental laws of the social sciences.

Although we have focused on the axiomatization of (fragments of) social science theories, there seem to be no fundamental reasons why the application should be restricted to theories from this particular field. Although we did not investigate the axiomatization of other substantive disciplines, we would want to argue, with a touch of irony, that the logic of the natural sciences is not very different from the logic of the social sciences. Therefore, the tools we used in the formalization of theories from sociology, may also be of use for the formalization of theories from other fields, including physics and even mathematics. It should however be noted that there are important factors that may impede the logical formalization of theories from natural science domains like physics. Formal logic seems particularly well suited to represent the informal, qualitative discourse of sociology. Perhaps the most important reason for this—and one that is easily overlooked—is simply the use of primitive (or undefined) terms.

[6]This is, of course, a matter of degree rather than a fundamental difference between the sciences: even the celebrated theories of physics give only approximate explanations of empirical phenomena (for example, by the use of various simplifying assumptions). As Popper (1959, p.111) put it "The empirical basis of science has thus nothing 'absolute' about it. Science does not rest upon solid bedrock. The bold structure of its theories rises, as it were, above a swamp. It is like a building erected on piles. The piles are driven down from above into the swamp, but not down to any natural or 'given' base; and if we stop driving the piles deeper, it is not because we have reached firm ground. We simply stop when we are satisfied that the piles are firm enough to support the structure, at least for the time being."

[7]Using the distinction between heuristic and nonheuristic axioms (Suppes 1993), it is clear that the axioms we have in mind should have a heuristic value in understanding the theory under consideration. The notion of heuristic axioms may seem like a contradiction in terms: 'heuristics' are often interpreted as informal rules of thumb, much like the exact opposite of formal axioms. It is in this sense that Pólya (1957, p.113) writes, "What is bad is to mix up heuristic reasoning with rigorous proof." We also do not want to mix up layers of informal and formal reasoning, but, just as the informal reasoning, we want the formal reasoning to be insightful and aid in understanding of the theory. Heuristic axioms are "axioms that seem intuitively to organize and facilitate our thinking about the subject, and in particular our ability to formulate, in an ordinarily self-contained way, problems concerned with the phenomena governed by the theory and their solution" (Suppes 1993, p.58).

> Our knowledge of the things denoted by the primitive terms ... is very comprehensive and is by no means exhausted by the adopted axioms. But this knowledge is, so to speak, our private concern which does not exert the least influence on the construction of our theory. In particular, in deriving theorems from the axioms, we make no use whatsoever of this knowledge, and behave as though we did not understand the content of the concepts involved in our considerations, and if we knew nothing about them that had not been expressly asserted by the axioms. We disregard, as is commonly put, the meaning of the primitive terms adopted by us, and direct our attention exclusively to the form of the axioms in which these terms occur. (Tarski 1946, pp.121–122)

This is of course a particularly useful feature if there is no clear intuition (and only partial consensus) on the meaning of primitive terms. Even if the underlying intuitions on the meaning of primitive terms are incommensurable (Kuhn 1996), researchers may still agree on a number of axioms in which these primitive terms appear. The rigorous, mathematical, quantitative discourse of physics would require the formalization of substantial parts of quite advanced mathematics before substantive results of theoretical interest can be reached. Think of the mathematics presupposed by Einstein's theory of relativity (Friedman 1983), although even in case of relativity theory there are efforts to formalize this theory in formal logic (Rakić 1997a,b; Andréka, Madarász, Németi, Sági, and Sain 1998).

8.3 Logical Criteria

We used only fairly standard, or naive, criteria for evaluating theories. This can be easily explained by the fact that we restricted our methodological discussion to issues that actually arose in the formalization of these theories. The motivation for this restriction is that it ensures that we deal with issues that are practically relevant, and avoid overemphasizing extremely exceptional cases, or even a digression into purely hypothetical cases. The immediate consequence of this restriction is that we extensively dealt with issues that would be regarded as well-known and elementary in other fields.[8] This should, of course, not be

[8] As Lazarsfeld (1962, p.463) observed: "[M]odern philosophers of science are only concerned with the natural sciences [and] do not pay attention to the empirical work in social research that is actually going on today. In a way, this is easily understood. In the correct sense of the term, as yet there is no systematic theory in the social sciences, only research procedures and a number of low-level generalizations. This would force the philosopher to become acquainted with technical details and results which are often rather boring, and would put him on a disadvantage vis-à-vis a colleague who can speculate on the meaning of such basic ideas as the uncertainty principle or the notion of relativity. The reading of empirical social research, which often lacks the big sweep of the more developed natural sciences is not only personally unrewarding but

interpreted to imply that we regard any of our notions as final. We would not hesitate, and do even anticipate, to incorporate various further refinements.

The result of an axiomatization attempt should be a better understanding of the structure of a theory by clarifying the underlying assumptions and requirements that the theory makes. One of the crucial motivations for an attempt to formulate more perspicuous theories is to facilitate their empirical testing. Unearthing the structure of the theory allows us to assess the consequences of the outcome of empirical research. In particular, we can single-out problematic parts of the theory in case of an unfavorable outcome.[9] In practice, empirical testing of theories depends largely on the used operational definitions of measurements (Bridgman 1927). In the social sciences, these operational definitions are seldom straightforward measurement conventions. There is little consensus on the operational definition of crucial concepts in the social science, and empirical researchers take, for understandable reasons, considerable freedom in their choice. As a result of this, empirical testing in the social sciences is frequently inconclusive: empirical confirmation is, at best, restricted to the particular operationalization at hand, and empirical refutation can be even totally due, or at least blamed on, the choice of a particular operationalization. Moreover, it is far from straightforward to compare different empirical tests of the same theory. The gravity of this problem is greatly enhanced by the fact that theories in the social sciences use a large vocabulary of terms that have various subtle differences in meaning.

Although many of the primitive concepts that sociological theories use are clearly related, there are seldom attempts to systematize the vocabulary. In chapter 3, we made such an attempt by formulating definitions of previously undefined concepts.[10] Before the introduction of definitions we would have regarded all terms as observational (conform the original texts). Introducing defined notions (although we would regard these as descriptive rather than stipulative definitions [Hempel 1966]) prompts questions on which terms should be regarded as theoretical and which as observable. Although there is no natural demarcation between observational and theoretical terms (Lakatos 1970), there are, at least intuitive, differences in the degree of observability of primitive terms (or perhaps

also does not confer much prestige. Understandably the general reader is more curious about the philosophical implications of the natural sciences which have so greatly influenced our daily lives."

[9]Despite popular folklore, a theory can of course only be refuted by a falsifying *hypothesis* and not by a single falsifying observation (Popper 1959, pp.86–87). Such a falsifying hypothesis should be a reproducible effect, of some generality, that is in itself falsifiable, and has received sufficient empirical confirmation or corroboration.

[10]In retrospect, this was done roughly along the lines of Carnap's explication of familiar but vague concepts (Carnap 1950, esp. chap.1). This is probably no great surprise, since Carnap (1970, p.59) wrote much earlier, "The last field to be dealt with is *social science*. ... Here we need no detailed analysis because it is easy to see that every term in this field is reducible to terms of the other fields. ... Many terms can even be defined on that basis, and the rest is certainly reducible to it."

rather in the degree of nonobservability). It may be not unreasonable to identify some notions as theoretical terms (for concrete examples, think of notions like 'legitimation' and 'competition' in the theory of chapter 7). Theoretical terms defined by simple definitions are easy to eliminate from a theory. It seems of great importance to investigate ways to define theoretical terms, or if they turn out not to be definable, find other ways to eliminate them (for example, along the lines of [Ramsey 1931]). Introducing a clear distinction between observable and theoretical terms, for example by simply assuming that the vocabulary consists of two disjoint sets (similar to Simon 1977, chap.6.7), will also allow for the further refinement of some of the used criteria (especially of the falsifiability criterion).

We have used various criteria for the (computational) evaluation of theories, such as consistency, soundness, falsifiability, satisfiability, contingency, and independence, that basically amount to the standard criteria from logic and philosophy of science. The criteria of the consistency of a theory and soundness of the derivation of theorems are stable notions from the field of logic. Nevertheless, there can still be various refinements in the way these formal criteria are used in practice. For example, we do not look for arbitrary models when proving consistency, but try to establish whether intended models (corresponding to typical cases of the theory) are models of the theory. Another example is to distinguish between various types of counterexamples, i.e., by distinguishing nonintended models from 'real' counterexamples.

The criterion of falsifiability is not as straightforward and there exist various proposals for it.[11] We used falsifiability as a criterion for theorems, while falsifiability is usually used as a criterion of theories (including Popper 1959). The reason for this is simple: the consistency of a theory directly implies the consistency of every statement of the theory. In case of falsifiability this is reversed, the falsifiability of any single statement of the theory implies the falsifiability of the entire theory. Simply put, a falsifiable theory can still have unfalsifiable theorems (for example, see the theory in chapter 4), and to identify them can be of obvious relevance for empirical research on the theory.

The notion of falsifiability as used in this thesis is formulated in terms of models: in its naive form, a theorem (or theory) is falsifiable, if there exist models in which the theorem (or theory) is false. This seems very different from Popper's original proposal, which is defined in terms of *basic statements*.

> A theory is to be called 'empirical' or 'falsifiable' if it divides the class of basic statements into the following two non-empty subclasses. First, the class of all those basic statements with which it is inconsistent (or which it rules out, or prohibits): we call this the class of *potential falsifiers* of the theory; and secondly, the class of those basic

[11] For a particular strict form of falsifiability, see also the FITness (Finite and Irrevocably Testable) criterion (Simon and Groen 1973; Simon 1983). Theories that satisfy this criterion have the property that all their theoretical terms are eliminable.

> statements which it does not contradict (or which it 'permits'). We can put this more briefly by saying: a theory is falsifiable if the class of its potential falsifiers is not empty. (Popper 1959, p.86)

Popper's famous notion of basic statements is referring to formulas of "the form of singular existential statements" (Popper 1959, p.102) with the additional requirement that these statement must be "asserting that an observable event is occurring in a certain individual region of space and time" (Popper 1959, p.103). This syntactic notion of basic statements is, of course, very different from our semantic notion of models.[12]

Taking into account the elementary logical relations between statements and their models, the resulting notions of falsifiability are not so different. First, falsifiability in terms of statements implies falsifiability in terms of models. Assume that a theory is falsifiable in the Popperian sense, i.e., there exists a basic statement that the theory prohibits. Then, in all models in which this basic statement is true (there are such models since basic statements are presumably satisfiable), the theory must be false. Consequently, there do exist models in which the theory is false, i.e., the theory is falsifiable in terms of models. This immediately ensures that those theories that we would regard as unfalsifiable, are also unfalsifiable in the Popperian sense, i.e., if there are no models in which the theory is false, there cannot exist (consistent) basic statements that the theory prohibits. Second, it is less clear whether falsifiability in terms of models implies falsifiability in terms of basic statements, for we would need to construct a basic statement. Assume that there exists a model in which the theory is false. With some handwaving, one could think of a model as a very long conjunction of literals (atomic formulas or their negations) by writing out the interpretation function (at least for small, finite models). These literals (at least the positive ones) closely resemble (relatively atomic) basic statements. Given the fact that all consistent conjunctions of basic statements are basic statements (Popper 1963, p.386), this long conjunction is, arguably, a basic statement. Recall that the theory was false in this model, therefore, the theory cannot be consistent with this long conjunction.

We did not deal with the material testability requirement so far, but there is

[12]Having said this, one could nevertheless ponder about a direct relation between (the intuitions behind) these two notions, considering the fact that formal semantics was developed much later. It is interesting to note that it makes perfect sense to substitute 'models' for every occurrence of 'basic statements' in the definition of falsifiability (Popper 1959, p.86, quoted above). In fact, we would get an exact description of how models work. Any theory partitions its class of models of the language (or structures) into two disjoint subclasses: a subclass of models in which it is true (i.e., the models of the theory), and a subclass of models in which the theory is false. There are remarkable similarities between the notions of models and the development of basic statements (or 'test statements' as they are later called [Popper 1974, p.1106]). In particular, see the discussion in (Popper 1959, pp.100–103), (Popper 1963, p.386), and especially (Ayer 1974; Popper 1974), where 'test statements' and 'counterexamples' are used on par. Kuipers (1987, pp.82–85) makes similar remarks when providing structuralist explications of various Popperian notions.

an elegant way to incorporate it into the notion of models. We are of course free to choose a universe for the models of a first-order logic language. Specifically, we can choose a (non-empty) set of objects that fulfill these testability requirements, i.e., 'observable' events that are spacio-temporally sufficiently restricted. The resulting models with this universe will be as 'observable' as the primitive terms permit them to be (if observable and nonobservable terms are distinguished, one could further refine this material requirement).

8.4 The Process of Axiomatization

The application of formalization approaches is, of course, not restricted to scientific theories.[13] Although risking to end up in an infinite regress, we can even try to apply our formalization method to the formalization process of chapter 5 itself. That is, we can extend our informal discussion of chapter 5 by trying to formalize the process of formalizing theories. Although this is not the proper place to pursue this subject extensively, we will here make an initial first step by rationally reconstructing the axiomatization process (mainly in order to highlight similarities and differences with related work).

When undertaking an effort to axiomatize a substantive theory, the ultimate goal is to provide a formal theory that is consistent, in which the conjectures are derivable, theorems are falsifiable, and the axioms are independent. There are logical dependencies between these criteria, that suggest the order in which tests for them should be performed. These dependencies include: (i) It seems a natural choice to establish the consistency of the axioms before testing the soundness of derivations of conjectures.[14] The soundness criterion has no discriminating effect on an inconsistent theory; every formula is a theorem of an inconsistent theory. Moreover, we do not need to know the precise theorems, before we can investigate the theory's consistency. Deductive theories are defined as the set of logical consequences of a set of premises. As a result, we only need to consider the theory's premises when testing for consistency—soundly derived theorems cannot make a theory inconsistent. (ii) It seems natural to test for soundness of the derivations before testing for the falsifiability and satisfiability. We use falsifiability and satisfiability as criteria of theorems. Therefore, it is reasonable to first establish the theorem-hood of conjectures. (iii) The satisfiability of theorems can be tested for indirectly: if the theory is consistent, then a soundly derivable

[13] Although it is clear that the formalization of other domains requires some changes, i.e., the restriction to universal premises; the use of classical, monotonic logics; or even bringing in "psychologism" (Beth and Piaget 1966).

[14] Note that these are merely heuristics, one could just as well argue that the underivability of some conjectures is far more likely than that the theory turns out to be inconsistent. This may result in modification of the axioms, which would require establishing consistency anew. Fortunately, the final result seems invariant of the order in which tests are performed, provided that the theorizer makes consistent decisions.

theorem is satisfiable. If a conjecture is not derivable, it is relevant to test whether the conjecture is at least satisfiable. (iv) Independence of axioms should be tested for after establishing consistency. Consistency implies one side of the tests for independence, namely that the negation of each axiom is not implied by the other axioms. Moreover, if a subset of the axioms is inconsistent then all other axioms (and their negations) will be derivable.

Keeping these dependencies in mind, we will give an initial rational reconstruction of the formalization process.[15] Before any of the criteria can be applied, we have to formulate an initial formal version.

1. Reconstruct the theory, and give an initial formalization of the premises (Σ) and conjectures (Γ).

The sociological theories treated in this thesis are based on publications in the sociological literature. Since descriptions of sociological theories are, in general, stated in ordinary language, they are usually partial and incomplete. The first step towards their formalization is to reconstruct the theoretical argumentation in these texts. This crucial rational reconstruction is largely outside the scope of this thesis.[16] In some of the chapters we have, for this reason, selected scientific texts that do not require extensive rational reconstruction. We have mainly dealt with the formal part of the formalization process, i.e., what happens after an initial reconstruction of a theory is completed. The rational reconstruction gives a non-formal description of the theory, which at least contains an explicit set of premises and an explicit set of (intended) theorems. The next step is to formalize the reconstructed premises and theorems in logic. The argumentation of this initial formal theory will closely resemble the argumentation of the original text, in particular, it may still be partial and incomplete. In the next steps, we will try to refine the initial formalization into a formal theory, i.e., a set of formulas that represent the premises, and a set of theorems that can be derived from the premises.

The first criterion to test for is consistency.

[15]Readers should bear in mind that this initial reconstruction is highly provisional and makes several serious simplifications. For example, we will ignore definitions because that would require a far more complicated (and more interesting) reconstruction. Moreover, we will here assume that we can conclusively determine the criteria. Of course, models may be too large, or proofs too complicated for us to find in practice (or in principle, since first-order logic is not decidable). Therefore, it is very well possible that we can neither find a derivation, nor a model.

[16]The rational reconstruction seems to be an art rather than a science (although there are some definite do's and don'ts). We want to distill from this description, at least, 1) the most important concepts and claims, 2) the assumptions that are presupposed by these arguments, and 3) the justification or argumentation for these claims. The excellent (Fisher 1988) gives a general method for analyzing scientific argumentation (a close reading method that is non-formal but strongly motivated by insights from logic).

2. Test for (in)consistency of Σ by generating at least one model of Σ (proving that Σ is consistent), or by deriving a contradiction from Σ (proving that Σ is inconsistent).

   If there exist models then (check model(s) and) proceed with step 3

   If the theory gives rise to a contradiction, then the proof singles out a specific derivation of a contradiction from a subset of the premises. If this contradiction can be resolved, then redo step 2 (there may be other contradictions). Otherwise, abandon the (formal) theory.

In practice, we will generate a number of models corresponding to typical scenarios treated by the theory, by adding appropriate initial conditions (typically existential statements). For example, if the conjectures are conditional statements (like in the theories treated in this thesis), we can try to find models in which the antecedent part of a conditional conjecture is satisfied. Testing for consistency in this sense roughly amounts to testing whether (some of) the intended models of the theory are allowed for by the formal axioms.

The theories we treated consist entirely out of universal conditional formulas and, as a result, they make no existential claims. It is therefore impossible that they are formally inconsistent, because there will be models of the theory in which none (or only few) of the antecedents are satisfied.[17] Of course, something weaker than a formal inconsistency may be the case: we may be unable to find some of the intended models of the theory. In these cases, the combination of the theory with some common initial conditions (existential statements) is formally inconsistent. A typical case of such a hidden inconsistency is when a premise is overstated and does not take particular exceptions into account. To resolve such an 'inconsistency' we simply have to extend the conditional part of the overstated assumption, such that all known exceptions are taken into account. There do exist logics, in particular nonmonotonic logics (Ginsberg 1987; Brewka, Dix, and Konolige 1997), that can deal with some of these 'inconsistencies.' Since the detection of these inconsistencies and the resulting revisions are likely to be of theoretical interest, we prefer to use a logic that does not implicitly resolve such deficiencies.

Our discussion in chapter 5 mainly focused on the soundness of derivations of conjectures.

3. For each conjecture $\gamma_i \in \Gamma$, test whether there exists a proof such that $\Sigma \vdash \gamma_i$ (proving that γ_i is soundly derivable, shortly γ_i is sound) or models $\mathcal{A}$ such that $\mathcal{A} \models \Sigma \cup \{\neg\gamma_i\}$ (proving that $\Sigma \not\models \gamma_i$, i.e., γ_i is not soundly derivable, shortly γ_i is unsound).

[17] Except for very contrived cases: for example, $\{\forall x[\mathsf{A}(x) \rightarrow \mathsf{B}(x)], \forall x[\mathsf{A}(x) \rightarrow \neg\mathsf{B}(x)], \forall x[\neg\mathsf{A}(x) \rightarrow \mathsf{B}(x)], \forall x[\neg\mathsf{A}(x) \rightarrow \neg\mathsf{B}(x)]\}$ is formally inconsistent.

If there exist a derivation then (check proof and) proceed with next conjecture γ_{i+1}.

If there exist counterexample(s) $\mathcal{A}$ then distinguish between:

(a) Is $\mathcal{A}$ a nonintended model (conflicting with common-sense or with implicit background assumptions), then revise Σ by adding background knowledge σ such that $\mathcal{A} \not\models \Sigma \cup \{\sigma\}$.

Then redo step 2 (prove consistency) and retry deriving this γ_i (there may be other counterexamples; earlier proven conjectures will still be theorems).

(b) If step 3a fails, then $\mathcal{A}$ is an exception to γ_i. Try to revise γ_i by weakening it to γ_i' in such a way that γ_i' is true in $\mathcal{A}$, in symbols, $\mathcal{A} \models \gamma_i'$.

Then retry proving this γ_i' (there may be other counterexamples; the theory is still consistent and earlier proven conjectures will still be theorems).

(c) If step 3b fails, then try to restrict the domain of the theory by strengthening Σ to Σ' in such a way that $\mathcal{A}$ is no longer a model of the restricted theory, in symbols, $\mathcal{A} \not\models \Sigma'$.

If there is an acceptable restriction Σ' then redo step 2 (proving consistency) and retry proving this γ_i (there may be other counterexamples; earlier proven conjecture will still be theorems).

If there is no acceptable restriction, then proceed to step 3d.[18]

(d) Discard γ_i as a false conjecture, and proceed with the next conjecture γ_{i+1}.

It is important to note that the revision of axioms to make a conjecture derivable is not a deductive step—it is a form of abduction.[19] The traditional form of abduction in artificial intelligence is: given a background theory K and an observation F, abduction computes an explanation E such that

$$K \cup E \models F$$

where $K \cup E$ is consistent (Poole 1988). Typically, the explanation, E, is a conjunction of initial conditions that allows the knowledge base, K, to entail a

[18] This is a good place to test whether γ_i is at least satisfiable, i.e., whether there exists models $\mathcal{A}$ such that $\mathcal{A} \models \gamma_i$. Recall, that we ignore here the presence of definitions (discussed in chapter 4).

[19] Abductive inferences are widely studied from the point of view of modern logic (Tan 1992; Flach 1995; Aliseda-LLera 1997), see also (Janssen and Tan 1992) on abductive explanations in economics.

particular observation, F. That is, K is a set of general statements, and E and F are (conjunctions of) ground facts.

In case that $K \models F$ an empty explanation suffices (in other words, traditional deductive explanations apply). Most literature on abduction ignores this case, and focuses on what one might call abduction *proper*, that is, the case that $K \not\models F$. Most of the abduction literature assumes $K \not\models F$ as a given, whereas we first have to establish it in step 3. Notice that proving $\Sigma \not\models \gamma$ is non-trivial in first-order logic, even undecidable in general, although in many cases small counterexamples exists. Finding such a counterexample not only proves that γ is underivable, but can also be used to decide between further revisions of the theory. Most abduction systems use a one-step approach. That is to say, they will immediately generate a set of explanations, i.e., a set of E's such that $K \cup E \models F$ and $K \cup E$ consistent (sometimes these explanations can be pruned using various criteria). The revisions we have in mind deal only with a (particular set of) counterexample(s). It is not unlikely that we may have to deal with a number of different (sets of) counterexamples. This results in a piecemeal revision, which may require both revision of the axioms as well as revision of the conjecture.

Our approach was to clearly distinguish between revising the theory (either by adding implicit background knowledge, or a proper revision by making Σ stronger such that the models that were counterexamples are no longer models of the theory), or revising the intended theorem (by making γ weaker such that it will be true in the models that were counterexamples). The abduction literature generally does not make this distinction, and seems to cover only the case in which the theory is revised. However, recall that the explanation typically represents the appropriate initial conditions that allow for a singular fact to be deduced from a general knowledge base. This seems intuitively a revision of the intended theorem and, consequently, we would prefer the alternative notation:[20]

$$K \models (E \to F)$$

where $K \not\models \neg E$. The resulting explanations (i.e., the E's) are the same, but there is of course a fundamental difference between the revision of a theory (or knowledge base), or the reformulation of a conjecture (or 'observation'). Whether this distinction is also relevant for other domains may depend on the specific knowledge base.

A common refinement of the abduction pattern is to study cases in which the observation is an *anomaly*. The observation F is an anomaly if it is contrary to our expectations, that is, in case the theory predicted the opposite $K \models \neg F$. Notice that the 'standard' abduction pattern fails in this case.[21] If we want to

[20]Using the *deduction theorem*: $\Sigma \cup \{\phi\} \vdash \psi$ if and only if $\Sigma \vdash (\phi \to \psi)$ with ϕ having no free variables (Chang and Keisler 1990).

[21]If $K \models \neg F$ then (in monotonic logics) for all explanations E, it is the case that $K \cup E \models \neg F$. If now E also explains F, that is, if $K \cup E \models F$, then we know that $K \cup E$ is *in*consistent (contrary

have a consistent theory that predicts F, we have to retract some axioms first. In other words, we have to find a contraction E_1 and an extension E_2 such that

$$(K \setminus E_1) \cup E_2 \models F$$

and $(K \setminus E_1) \cup E_2$ is consistent. These patterns are widely studied in the literature on belief revision or theory change (Alchourrón, Gärdenfors, and Makinson 1985).

Assume we want to derive an intended theorem and it is the case that the negation of this theorem is already implied by the theory. In this case, we will be unable to find models of the axioms in which the theorem is true. This amount to a test whether an intended theorem γ_i is at least consistent with the premises, i.e., whether there exists a model $\mathcal{A}$ such that $\mathcal{A} \models \Sigma \cup \{\gamma_i\}$, or whether $\Sigma \cup \{\neg\gamma_i\} \vdash \bot$.[22] A good place to test for this is immediately after establishing that a conjecture is not derivable in step 3. Alternatively, we could try to prove this immediate after proving the consistency of the premise in step 2, by trying to find a model $\mathcal{A}$ such that $\mathcal{A} \models \Sigma \cup \Gamma$.

It is unclear whether these patterns do occur in our setup. In our case studies all the conjectures have the form of a universal statement. Now suppose we would be able to derive the negation of such a statement. Because the negation of a universal statement is an existential statement, this implies that we can already derive an existential theorem. We have dealt with strictly universal theories, which do not imply existential statements. Nevertheless, it seems that our earlier method can also be applied in this, hypothetical, case. That is to say, inspection of models and counterexamples can lead to appropriate revision of the theory, including the retraction of some axioms if some of the intended models cannot be constructed.

Finally, we should test for the falsifiability of the theorems, and the independence of the axioms. The falsifiability is tested by showing the logical possibility of a theorem to be false:[23]

4. For each $\gamma_i \in \Gamma$, test whether there exist models $\mathcal{A}$ such that $\mathcal{A} \models \neg\gamma_i$ (proving that γ_i is falsifiable), or a proof such that $\vdash \gamma_i$ (proving that γ_i is unfalsifiable).

to the above requirement). At least, this is the case for monotonic logics—the abduction pattern depends on the used logic. If the underlying logic is *nonmonotonic* (denoted by the '$\mid\!\sim$') then the traditional abduction pattern should be extended (Inoue and Sakama 1995). In nonmonotonic logics we have also negative explanations, since we can also make F derivable by deleting some E from K. That is, a revision from $K \not\mid\!\sim F$ to $K \setminus E \mid\!\sim F$ (called *extended* abduction in [Inoue and Sakama 1995]). Specifically, in nonmonotonic logics there may exist consistent revisions from $K \mid\!\sim \neg F$ to $K \setminus E \mid\!\sim F$ and from $K \mid\!\sim \neg F$ to $K \cup E \mid\!\sim F$.

[22] Notice that this is a stronger requirement than satisfiability as operationalized in chapter 4. Our notion of satisfiability is the counterpart of falsifiability and therefore ignores the axioms of the theory (although definitions are taken into account).

[23] Recall that we assume here that there are no definitions. An operationalization of falsifiability that takes definitions into account was discussed in chaper 4.

> If a theorem γ_i is unfalsifiable, then we may decide to exclude it from the set of singled-out theorems (i.e., from Γ).

If there is at least one falsifiable theorem, then the theory is falsifiable. Notice that even in the extreme case that all the singled-out theorems are unfalsifiable, we cannot immediately conclude that the theory as a whole is unfalsifiable. The theory may still have other falsifiable consequences, and in this extreme case the theory may therefore still be perfectly falsifiable. In such cases, one could decide to test the falsifiability of the axioms themselves, and if at least one of the axioms is falsifiable, the theory is falsifiable. Only if all the axioms are unfalsifiable, we can conclude that the theory is unfalsifiable. We restrict here our attention to the falsifiability of theorems because the falsifiability of axioms is closely related to their independence. If some of the axioms are unfalsifiable, then these axioms will be necessarily not independent of the other axioms. The independence of the axioms is investigated in the following manner.

5. For each $\sigma_i \in \Sigma$, test whether there exist models $\mathcal{A}$ such that $\mathcal{A} \models \Sigma \setminus \{\sigma_i\} \cup \{\neg\sigma_i\}$ (together with consistency proving that σ_i is independent of the other axioms) or a proof such that $\Sigma \setminus \{\sigma_i\} \vdash \sigma_i$ (proving that σ_i is not independent of the other axioms).

   If an axiom σ_i is not independent of $\Sigma \setminus \{\sigma_i\}$ then we can make the axiom set more parsimonious by moving σ_i from the set axioms to the set of theorems, and redo step 5 with the smaller axiom set Σ' (with $\Sigma' = \Sigma \setminus \{\sigma_i\}$).

This concludes our initial rational reconstruction of the process of axiomatization. We will not pursue a formal reconstruction of this process here, but it is clear that this process not deductive. This corresponds to the well-known fact that theory revision is nonmonotonic even if the underlying theory is monotonic (Makinson and Gärdenfors 1991).[24] Rott (1991) is an excellent source for further discussion on nonmonotonic theory revision. Despite the overwhelming number of different nonmonotonic logics (Brewka et al. 1997), it is unclear whether any of these logics can be used for a formalization of the axiomatization process.

8.5 Discovery?

The process of axiomatization as discussed in the previous section shows a clear relation with Lakatos' famous heuristic pattern of mathematical discovery as discussed in his "proofs and refutations":

[24] Of course, a theory *about* the revision process may again turn out to be monotonic.

> There is a simple pattern of mathematical discovery – or of the growth of informal mathematical theories. It consists of the following stages:[a]
> (1) *Primitive conjecture.*
> (2) *Proof (a rough thought-experiment or argument, decomposing the primitive conjecture into subconjectures or lemmas).*
> (3) *'Global' counterexamples (counterexamples to the primitive conjecture) emerge.*
> (4) *Proof re-examined: the 'guilty lemma' to which the global counterexample is a 'local' counterexample is spotted. This guilty lemma may have previously been 'hidden' or may have been misidentified. Now it is made explicit, and built into the primitive conjecture as a condition. The theorem – the improved conjecture – supersedes the primitive conjecture with the new proof-generated concept as its paramount new feature.*
> These four stages constitute the essential kernel of proof analysis. But there are some further stages which frequently occur:
> (5) *Proofs of other theorems are examined to see if the newly found lemma or the new proof generated concept occurs in them: this concept may found lying at cross-roads of different proofs, and thus emerge as of basic importance.*
> (6) *The hitherto accepted consequences of the original and now refuted conjecture are checked.*
> (7) *Counterexamples are turned into new examples – new fields of inquiry open up.*
>
> [a]As I have stressed the actual historical pattern may deviate slightly from this heuristic pattern. Also the fourth stage may sometimes precede the third (even in the heuristic order) – an ingenious proof analysis may suggest the counterexample.
>
> (Lakatos 1976, pp.127–128)

Lakatos' brilliant essay "proofs and refutations," an extensive discussion of the *polyhedra conjecture*, goes far beyond the simple example we discussed in detail in chapter 5. This conjecture, baptized the "Euler's formula" by Pólya (1954a, p.52), states that for polyhedra the number faces plus the number of vertices equals the number edges plus two, in symbols, $F + V = E + 2$ or, equivalently, $V - E + F = 2$.[25] Lakatos (1976) discusses in great detail, based on historical

[25]The Euler's formula is extensively discussed in (Pólya 1954a, chap.III, pp.35–58). Lakatos' essay dramatically extends Pólya's initial discussion:

> The phase of *conjecturing* and *testing* in the case of $V - E + F = 2$ is discussed in Pólya ([1954], vol. I, the first five sections of the third chapter, pp.35–41). Pólya stopped here, and does not deal with the phase of *proving* … Our discussion starts where Pólya stops. (Lakatos 1976, p.7)

Rather surprisingly, and notwithstanding Lakatos' contribution, this seems not to give Pólya adequate credits: First, the discussion does in fact continue for two more sections (Pólya 1954a, material, the intricate process of incessant improvement of guesses by speculation and criticism. His discussion includes dealing with global and local counterexamples (falsifying whole conjecture or an intermediate lemma); rejection of the conjecture; improving the conjecture by exception-barring methods (comparable to what we called weakening of the conjecture); rejection of counterexamples by monster-barring or monster-adjustment (comparable to what we called the removal of nonintended models by adding implicit background knowledge). Most of the revisions in case of the polyhedra conjecture concern the used definitions. The similarities will become even more apparent if we extend our discussion by considering the revision of definitions. Notice, however, that this will make the process more complicated for the revision of a definition will affect all occurrences of the defined term. A defined term may, of course, occur in several of the conjectures, but the same term may also occur in some of the axioms (in which case a revision of the definitions yields a different theory). Moreover, the effect of such a revision also depend on where the defined term occurs in a conjecture or axiom. For example, if we strengthen the definition of a defined term occurring in a conjecture that is a conditional statement, the conjecture will become weaker if the defined term occurs in the antecedent part, and stronger if it occurs in the consequent part of the conjecture.

Lakatos' heuristic pattern of mathematical discovery, influenced by the revival of mathematical heuristics (Pólya 1945, 1954a,b, 1962, 1965), is discussing informal mathematics. Our investigations are based on the logical axiomatization of scientific theories, a pre-eminent part of formal mathematics. As a result, there are, despite the various similarities, also some notable differences. Perhaps the most prominent difference is the trivial but extremely useful observation that, in a strictly formal theory, counterexamples need not mysteriously "emerge" nor be suggested by "an ingenious proof analysis"—they are merely formal models of the theory.

It is important to note that Lakatos is explicitly dealing with mathematical discovery. This immediately suggests links between our computational approach and research in machine discovery or scientific discovery (Langley et al. 1987; Shrager and Langley 1990; Simon 1995).[26] Most of the machine discovery systems deal with the induction of theories from facts or data, and our theory-driven

§III.6/7, pp.41–43), among others discussing critical counterexamples like the picture-frame or doughnut shaped polyhedra (that feature prominently in Lakatos' method [Lakatos 1976, p.19]). Second, the discussion even further continues in the remarks and comments on chapter III (Pólya 1954a, pp.52–58) by refining the statement and eventually leading to exercises 9, 29 and 39 in which versions of the Euler's formula are to be proved. Third, there is an appendix that gives the solutions of the exercises in chapter III (Pólya 1954a, pp.222–227), including the proofs of three version of the Euler's *theorem*.

[26] Lakatos' discussion plays an important role in reasearch on theory change (Darden 1990, 1991, 1992), and has inspired various systems in machine learning (Hayes-Roth 1984; Falkenhainer and Rajamoney 1988). The underlying Popperian methodology features prominently in (Shapiro 1991), which provides an elegant algorithm that, given an unknown model that serves

approach is complementary to these data-driven approaches. Our focus has been entirely on inferences at the theoretical level: addressing how to go from hypotheses and conjectures to axioms and theorems, and for the time being ignoring how a hypothesis or conjecture can be inductively inferred from empirical data. Many of the working systems in machine discovery are based on heuristic search and *cognitive* models of science (Giere 1992) and usually does not worry too much about the underlying principles involved (for which it has been criticized by some [Glymour 1992a]). A striking example is the most famous machine discovery system, AM (Lenat 1976), whose impressive results seem to be difficult to reproduce (Ritchie and Hanna 1984). A further difference is that early machine discovery systems tried to de-emphasize human intervention and were supposed to operate autonomously, whereas our approach is essentially interactive. Recent machine discovery systems have re-evaluated the role of user interaction. Human-computer collaboration has now been identified as the source of generality among successful discovery programs (Valdés-Pérez 1999).[27] De Jong and Rip (1997) give a vivid discussion of how this role might evolve in the future.

8.6 Sociology

Throughout the thesis, our approach has been application-oriented. In all chapters, we provided or discussed formalizations of actual social science theories. The social sciences have rarely been the subject of an axiomatization attempt using formal logic.[28] However, over the years, there has been a lot of effort in the mathematical formalization of the social sciences. The most successful efforts belong to the field of economics, which is dominated by mathematical theories (an excellent

as an oracle deciding the truth of sentences in this model, can inductively infer a theory that implies all (testable) sentences that are true in this model.

[27]Lenat always stated that all results were obtained in "totally unguided runs" (Davis and Lenat 1982, p.129). However, the user could improve AM's performance with "about a factor of 2 or 3 speedup" by guiding AM to focus on specific concepts, and "obtain much nicer results" by adding new concepts and results to AM's knowledge base. Moreover, Lenat already envisioned that machine discovery systems would ultimately be interactive: "There is one important observation to be made: the very best examples of AM in action were brought to full fruition only by a human developer. That is, AM thought of a couple of great concepts, but couldn't develop them well on its own. A human (the author) then took them and worked on them by hand, and interesting results were achieved. These result could be told to AM, who could then go off and look for new concepts to investigate. This interaction is of course at much lower frequency than the kind of rapidfire question/answering talked about above. Yet it seems that such synergy may be the ultimate mode of AM-like systems: creative *assistants* to experts" (Davis and Lenat 1982, p.130).

[28]A notable exception in sociology is (Kyburg 1968, chap.12), which presents a first-order logic version of Simon's mathematical model (1952; 1957) of Homans' theory of the human group (1950).

example is Debreu 1959).[29] Outside economics, however, the situation is different. Consider the field of sociology. The mathematical formalization of theories in sociology was attracting much attention during the sixties and early seventies (for example, Lazarsfeld 1954; Kemeny and Snell 1962; Coleman 1964; Blalock 1969; Fararo 1973).[30] Many sociologists had high hopes that mathematical sociology could provide a fruitful theoretical perspective that could further progress sociology as a scientific discipline, similar to Lewin's remarks on the neighboring discipline of social psychology:

> The greatest handicap of applied psychology has been the fact that, without proper theoretical help, it had to follow the costly, inefficient, and limited method of trial and error. Many psychologists working today in an applied field are keenly aware of the need for close co-operation between theoretical and applied psychology. This can be accomplished in psychology, as it has been in physics, if the theorist does not look toward applied problems with highbrow aversion or with a fear of social problems, and if the applied psychologist realizes that there is nothing so practical as a good theory. (Lewin 1951, p.169)

However, the high hopes were never fulfilled and, despite all efforts, the interest in formal theorizing phased out during the seventies, and formal theory is now situated at the periphery of sociology (Sørensen 1978; Hage 1994a; Hannan 1997). It is unclear what caused the untimely demise of formal theorizing in sociology. Various knowledgeable insiders have tried to explain demise of mathematical theory (see the collection devoted to this [Hage 1994a]). There seems to be no simple explanation and we will here just reiterate three of the causes that seem pertinent. The reasons proposed include the fact that most sociologists seem to be content with the relatively simplistic theories that are available.

> Nearly all sociologists give lip service to the notion that social reality is complex. Yet we seem to prefer relatively simple theoretical explanations of this reality, together with data analyses and measurement-error assumptions that are also highly simplistic. (Blalock 1994, p.121)

A second explanation is the fact that most applications of formal theorizing in sociology try to adopt formal theories from other disciplines, instead of formalizing

[29] In economics, formal logic plays a similar role as in mathematics, i.e., logic is used for addressing foundational issues and regular mathematics is thought to be informal (see the discussion in Dow, Krugman, Weintraub, Backhouse, and Chick 1998).

[30] The relation between sociology and positivism dates back even longer: the introduction of both the terms 'positivism' (or at least, 'positive philosophy') and 'sociology' is commonly attributed to one and the same person, Auguste Comte. Although it is the case that logic and sociology are put at considerable distance in his hierarchy of sciences, which is roughly a linear series starting with logic and mathematics and winding up with biology and sociology (Comte 1830).

theories that originate from sociology.

> The construction of models implementing a theory about a sociological phenomenon would be regarded a central activity of mathematical sociology. It is, however, not the activity that dominates mathematical sociology. Rather, most contributions to the literature apply mathematical models already existing and usually borrowed from other disciplines to sociological phenomena. The sociological interpretation of these models renders them mathematical sociology. (Sørensen 1978, pp.345–346)

A third cause of the demise of formal theorizing in sociology may be a miss match between the vague, qualitative discourse of the verbal theories and the demands of disciplined, mathematical discourse.

> The availability of natural-language theories were—and are—partial and imprecise: use of formal languages, especially mathematics, demanded closure and extreme (usually metric) precision. Effort at formalizing sociological theories with classical tools required the analyst to assume too much—e.g., to supply the missing assumptions about metrics, continuity, and differentiability of functions. As a result the formalizations offered during this period yielded models that failed to resonate closely with the original theories. The formalizers tended to blame the theorists for not building sufficiently precise theories, and theorists increasingly stopped paying attention. (Hannan 1997, pp.145–146)

The unfortunate result of all of this is that the current state of theory in sociology is a confused state of fragmentation and politicization (Hage 1994b). Some of the theories in sociology contain insightful and important ideas, but these can be easily lost in the overall confusion.

In the light of these remarks above, it may be difficult to convince social scientists that the current interest in logical formalization will not share the same fate. Of course, only time can tell, but we believe that there are some important reasons why the current interest in logical formalization need not share the fate of its predecessors.

The first reason is that logical formalization uses a formal language that allows a precise and formal expression of statements that can still be as general as its natural language expression. Mathematical formalization, such as the construction of causal models, typically requires quantification and closure (the assumption that all unmentioned variables and interactions are irrelevant), and various highly technical assumptions like the uncorrelatedness of rest terms with independent variables in the same equation (Blalock 1969, 1971; Hannan 1971a).[31] It important to note that none of these assumptions is required by the logical formalization of a theory. As Hannan (1997, p.146) commented on the formalization of parts of his theory:

> FOL [first-order logic] allows natural language statements to be given an expression that is formal (in the sense that it can be used in calculations by a theorem prover) yet remains close to the natural language. The semantic richness of FOL narrows the gap between the theorist's statement and the formalizer's rendition.

The logical reconstruction of an ordinary language theory allows the analyst to follow the original theory's argumentation closely—making results obtained in the formal version of the theory also apply to the original theory. The aim of an axiomatization attempt is to formalize the theoretical knowledge as captured in an authoritative theory, and we do not expect to write down the basic axioms of sociology or organization science. Formalizing this theoretical knowledge can be, of course, in itself an important, initial step towards the formulation of some of the (more) basic axioms of sociology.

The second reason is the current focus on the formalization of theories that originate from the domain of sociology. We do not attempt to fit a sociological theory into a pre-existing mathematical framework, but attempt to fit a mathematical framework onto a pre-existing sociological theory—as readers who attempt to see through our logical notation may have noticed. Of course, giving a sociological interpretation to formal theories originating from other disciplines can lead to important insights, or even start new research areas in sociology. In fact, some of the most important research is inspired by theories originating from other disciplines. Think for example of how economic game theory (Von Neumann and Morgenstern 1944) inspired much research in decision-making, cooperation, and rational choice (see, for example, Axelrod 1984; Coleman 1990). Nevertheless, we feel that there is a complementary need to formalize theories originating from sociology itself.

The third reason is the availability of the computational support that can deal with calculations that are required for the formalization of theories. If these tools play only a small part of the role computational support played in advancing

[31] As Hannan (1971b, p.473) stated, "such approaches to making causal inferences from nonexperimental data place a number of constraints on the analyst. In particular, he must make explicit most of the assumptions underlying both his model and analysis operations." In as far as these are underlying assumptions of the model, making them explicit is an important contribution to the theory. However, in as far as these are underlying assumptions of the analysis operations, they can be regarded as merely constraints on the applicability of these techniques. These constraints seems to be non-trivial, considering that "[g]iven the state of theory and data in the nonexperimental social sciences, it is highly unlikely that all of the assumptions underlying any one of the techniques will be met in any substantively interesting application" (Hannan 1971b, p.473).

the use of statistics in the social sciences, the prospects of logical formalization are more than promising. Although too early for a full assessment, there are some positive telltale signs: the field of automated reasoning recognized the use of these tools for building and testing theories as one of its most important long-term applications (Loveland 1996).

The fourth reason is that there are few alternatives that address sociological theory head on. Many theoretical texts in sociology are not very explicit on the underlying assumptions that they make. If we want to advance the state of sociological theory, it seems inevitable to make these underlying assumptions part of the discussion. As Blalock (1964, p.171) put it:

> Quite clearly, *a failure to state one's assumptions explicitly does not make them disappear in some magical way.* It *does, however, make it much more difficult to evaluate and reject a given theoretical system.* Theoretical inadequacies are harder to spot, and untestable theories remain to clutter up the literature. Such a state of affairs is hardly desirable.

Sociologists will not dispute the validity of this statement and will broadly agree with it, however, it may remain a different matter whether they will also act accordingly.[32]

Modern statistics provides powerful tools for analyzing our data, logical formalization is an attempt to provide similar tools for analyzing our theories. First-order logic is simple and clear (Kiser 1997), and does in itself not require expensive empirical research (Péli et al. 1994). Moreover, it may actually prevent us from

[32] As an anonymous reviewer confided to us in a revealing remark ending with some disquieting questions: "This is an admirable sentiment, and one that I fully endorse, but I feel it may not adequately take account of the *realpolitik* of our discipline (and perhaps most social science fields). Cynical as it may sound, some major theorists may not welcome techniques that expose logical flaws in their theories and increase the theories' exposure to critical assessment. While most of the major management theories have been subjected to empirical testing, it is very difficult to think of a case in which that testing has been conclusive—either a conclusive verification or a conclusive falsification. The openendedness of these streams of 'testing' research can be attributed, at least in part, to the logical and empirical ambiguity of the theories being tested. This openendedness is frustrating for those who want some empirical resolution of theoretical questions, but it plays to the political interests of the theorists, who accumulate more citations to their work and more reputational advantages the longer an inconclusive stream of research is perpetuated. There is little or no stigma attached to the theorist if her theory is never conclusively verified or falsified; instead, we tend to lay the blame for lack of empirical answers on the empirical researchers and their methods. Thus, I believe there is a disincentive on the part of theorists to make their theories logical and falsifiable, and a resistance to the endeavors of those who would like to force theory in that direction. I do not argue that theorists intentionally set out to make their theories murky or non-falsifiable; but there is an unconscious response to the existing reward structure that reduces effort to rectify those conditions. Assuming there is some validity to my argument, how would you 'sell' your logical formalization approach? How realistic are the prospects that such a technique would ever be adopted by the 'towers of power' in management?"

performing empirical research on inconsistent or unfalsifiable theories. Nevertheless, there are no silver bullets and logical formalization cannot be considered a *panacea* (Péli et al. 1994). The logical criteria are additional and complementary to the standard, empirical criteria for evaluating theories and ignore many dimensions that researchers use in order to discern useful theories, for example their empirical plausibility, their relevance, the insight they provide, their social or political philosophy, or their elegance and intuitive appeal. This is not as disadvantageous as it may appear at first glance. The logical criteria are objective, structural criteria on scientific theories. They are fundamentally different from empirical, subjective, or political criteria. Confounding them can create unnecessary confusion (Masuch et al. 1996).

There has been much debate on what constitutes a sociological theory (Freese 1980; Van de Ven 1989b; Sutton and Staw 1995). The case studies in this thesis present theories in the strict sense of a deductive system (Tarski 1956).[33] One way to view this strict notion is as a theory in its purest form, or as an 'ideal type' of theory. Few sociologists will argue against deductive systems satisfying the criteria of theories (or even of *good* theories [Van de Ven 1989a], or of *strong* theories [Sutton and Staw 1995])—even without agreeing on the exact criteria that theories should meet. Using the strict notion of theories allows us to be on the safe side, as it were. We do not want to de-emphasize the importance of other, less formal notions of theory. Formulating theoretical ideas in a less formal manner can be extremely useful, especially during earlier stages of theory development. The discovery of theories requires creativity and insight—activities with which formal logic is rarely associated. As Weick (1995) remarked: 'what theory is *not*, theorizing is.' The process of theory development may very well resemble some form of 'disciplined imagination' (Weick 1989)—maybe logic can help us understand the 'discipline' involved.

[33] Similar views have been proposed earlier. This notion, although used informally, can be traced in the writings of (for example, Durkheim 1938; Weber 1968; Neurath 1970; Beth 1958a; Kaplan 1964; Homans 1967; Merton 1968; Stinchcombe 1968; Elster 1978).

Bibliography

T. W. Adorno, R. Dahrendorf, H. Pilot, H. Albert, J. Habermas, and K. R. Popper, editors. *Der Positivismusstreit in der deutschen Soziologie*. Hermann Luchterhand Verlag, Darmstadt und Neuwied, 1969.

C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50: 510–530, 1985.

A. Aliseda-LLera. *Seeking Explanations: Abduction in Logic, Philosophy of Science and Artificial Intelligence*. ILLC dissertation series 1997-4, Department of Philosophy, Stanford University, Stanford CA, 1997.

T. L. Amburgey, D. Kelly, and W. P. Barnett. Resetting the clock: The dynamics of organizational change and failure. *Administrative Science Quarterly*, 38:51–73, 1993.

H. Andréka, J. X. Madarász, I. Németi, G. Sági, and I. Sain. Analyzing the logical structure of relativity theory via model theoretic logic. Manuscript, Mathematical Institute of the Hungarian Academy of Sciences, Budapest, Hungary, 1998.

R. Axelrod. *The Evolution of Cooperation*. Basic Books, New York, 1984.

A. J. Ayer, editor. *Logical Positivism*, The library of philosophical movements. The Free Press, New York, 1959.

A. J. Ayer. Truth, verification and verisimilitude. In Schilpp (1974), pages 684–691.

H. H. Baligh, R. M. Burton, and B. Obel. Devising expert systems in organization theory: The organizational consultant. In Masuch (1990), chapter 2, pages 35–58.

W. Balzer. A basic model for social institutions. *Journal of Mathematical Sociology*, 16:1–29, 1990.

W. Balzer, C. U. Moulines, and J. D. Sneed. *An Architectonic for Science: The Structuralist Program*, volume 186 of *Synthese Library*. D. Reidel Publishing Company, Dordrecht, 1987.

W. Balzer, C. U. Moulines, and J. D. Sneed, editors. *Structuralist Knowledge Representation: Paradigmatic Examples*, Poznań Studies in the Philosophy of the Sciences and the Humanities. Rodopi, Amsterdam, forthcoming.

J. B. Barney and W. G. Ouchi, editors. *Organizational Economics: Towards a New Paradigm for Understanding and Studying Organizations*. Jossey-Bass Publishers, San Francisco CA, 1986.

J. Barwise and J. Etchemendy. *The Language of First-Order Logic*. CSLI Publications, Stanford CA, revised and expanded third edition, 1992.

J. Barwise and J. Etchemendy. *Hyperproof*. CSLI Publications, Stanford CA, 1994.

W. Bechtel. *Philosophy of Science: An Overview for Cognitive Science*. Lawrence Erlbaum Associates Publishers Hillsdale NJ, 1988.

E. W. Beth. *De Weg der Wetenschap: Inleiding tot de Methodenleer der Empirische Wetenschappen*. De Erven F. Bohn N.V., Haarlem, 1958a.

E. W. Beth. On machines which prove theorems. *Simon Stevin: Wis- en Natuurkundig Tijdschrift*, 32:49–60, 1958b.

E. W. Beth and J. Piaget. *Mathematical Epistemology and Psychology*. D. Reidel Publishing Company, Dordrecht, 1966.

H. K. Bhargava and S. O. Kimbrough, editors. *Special Issue on Logic Modeling*, volume 11(2) of *Decision Support Systems*. Elsevier Science Publishers, North-Holland, 1994.

H. M. Blalock, Jr. *Causal Inferences in Nonexperimental Research*. The University of North Carolina Press, Chapel Hill NC, 1964.

H. M. Blalock, Jr. *Theory Construction: From Verbal to Mathematical Formulations*. Prentice-Hall, Englewood Cliffs NJ, 1969.

H. M. Blalock, Jr., editor. *Causal Models in the Social Sciences*. Aldine Atherthon, Chicago, 1971.

H. M. Blalock, Jr. Why have we failed to systematize reality's complexities? In Hage (1994a), chapter 7, pages 227–242.

R. W. Blanning. Knowledge, metaknowledge, and explanation in intelligent organizational models. In Masuch and Warglien (1992), chapter 3, pages 65–86.

M. A. Boden. Creativity and artificial intelligence. *Artificial Intelligence*, 103: 347–356, 1998.

R. J. Brachman and H. J. Levesque, editors. *Readings in Knowledge Representation*. Morgan Kaufmann Inc. Los Altos CA, 1985.

B. Bredeweg. *Expertise in Qualitative Prediction of Behaviour.* PhD thesis, University of Amsterdam, 1992.

B. Bredeweg and C. Schut. Reducing ambiguity by learning assembly behaviour. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence '93*, pages 980–985. Chambéry, France, August 1993, 1993.

G. Brewka, J. Dix, and K. Konolige. *Nonmonotonic Reasoning: an overview.* Number 73 in CSLI Lecture Notes. CSLI Publications, Stanford CA, 1997.

P. W. Bridgman. *The Logic of Modern Physics.* Macmillan, New York, 1927.

J. Brüderl, P. Preisendörfer, and R. Ziegler. Survival chances of newly founded business organizations. *American Sociological Review*, 57:227–242, 1992.

J. Bruggeman. Niche width theory reappraised. *Journal of Mathematical Sociology*, 22(2):201–220, 1997.

A. Bundy, editor. *Automated Deduction—CADE-12: Proceedings of the* 12^{th} *International Conference on Automated Deduction, Nancy, France, June 26-July 1, 1994*, volume 814 of *Lecture Notes in Artificial Intelligence.* Springer-Verlag, Berlin, 1994.

R. M. Burton and B. Obel. *Designing Efficient Organizations. Modeling and Experimentation.* North Holland, New York, 1984.

K. M. Carley and M. J. Prietula, editors. *Computational Organization Theory.* Lawrence Erlbaum Associates, Hillsdale New Jersey, 1994.

R. Carnap. *Der logische Aufbau der Welt.* Weltkreis-Verlag, Berlin, 1928.

R. Carnap. *Logical Foundations of Probability.* The University of Chicago Press, Chicago IL, 1950.

R. Carnap. *Introduction to Symbolic Logic and its Applications.* Dover Publications, New York NY, 1958.

R. Carnap. Logical foundations of the unity of science. In Neurath et al. (1970), chapter I.1, pages 42–62.

C. C. Chang and H. J. Keisler. *Model Theory*, volume 73 of *Studies in Logic and the foundations of mathematics.* North-Holland, Amsterdam, third edition, 1990.

M. D. Cohen, J. G. March, and J. P. Olsen. A garbage can model of organizational choice. *Administrative Science Quarterly*, 17:1–25, 1972.

M. R. Cohen and E. Nagel. *An Introduction to Logic and Scientific Method.* Routledge and Kegan Paul, London, 1934.

J. S. Coleman. *Introduction to Mathematical Sociology.* The Free Press, New York, 1964.

J. S. Coleman. *Foundations of Social Theory.* The Belknap Press, Cambridge MA, 1990.

A. Comte. *Cours de Philosophie Positive.* Paris, 1830. Six volumes published from 1830–1842.

H. L. Costner and R. K. Leik. Deductions from "axiomatic theory". *American Sociological Review*, 29:819–835, 1964.

R. M. Cyert and J. G. March. *A behavioral theory of the firm.* Prentice-Hall, Englewood Cliffs NJ, 1963.

L. Darden. Diagnosing and fixing faults in theories. In Shrager and Langley (1990), pages 319–346.

L. Darden. *Theory Change in Science: Strategies from Mendelian Genetics.* Monographs on the History and Philosophy of Biology. Oxford University Press, New York, Oxford, 1991.

L. Darden. Strategies for anomaly resolution. In Giere (1992), pages 251–273.

E. Davis. *Representations of Commonsense Knowledge.* Morgan Kaufmann, San Mateo CA, 1990.

E. Davis. Axiomatizing qualitative process theory. In B. Nebel, C. Rich, and W. Swartout, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference (KR'92)*, pages 177–188. Morgan Kaufmann Publishers, San Mateo CA, 1992.

E. Davis. The naive physics perplex. *AI Magazine*, 19(4):51–79, 1998.

R. Davis and D. B. Lenat. *Knowledge-Based Systems in Artificial Intelligence.* McGraw-Hill advanced computer science series. McGraw Hill, New York, 1982.

A. D. de Groot. *Methodologie: Grondslagen van onderzoek en denken in de gedragswetenschappen.* Mouton & Co, 's Gravenhage, 1961.

A. D. de Groot. *Methodology: Foundations of Inference and Research in the Behavioral Sciences.* Mouton, The Hague, 1969.

H. de Jong and A. Rip. The computer revolution in science: steps towards the realization of computer-supported discovery environments. *Artificial Intelligence*, 91(2):225–256, 1997.

J. de Kleer and J. S. Brown. A qualitative physics based on confluences. *Artificial Intelligence*, 24:7–83, 1984.

H. de Nivelle. *Bliksem 1.10 User Manual.* Max Planck Institut für Informatik, Saarbrücken, Germany, 1999.

G. Debreu. *Theory of Value: An Axiomatic Analysis of Economic Equilibrium.* John Wiley & Sons, New York, 1959.

P. J. DiMaggio and W. W. Powell. Introduction. In Powell and DiMaggio (1991), chapter 1, pages 1–38.

S. C. Dow, P. Krugman, E. R. Weintraub, R. E. Backhouse, and V. Chick. Controversy: Formalism in economics. *The Economic Journal*, 108:1826–1869, 1998.

E. Durkheim. *De la division du travail social.* Felix Alcan, Paris, 1893.

E. Durkheim. *The Rules of Sociological Method.* Free Press, Glencoe IL, 1938.

L. B. Edelman. Legal ambiguity and symbolic structures: Organizational mediation of civil rights law. *American Journal of Sociology*, 97:1531–1576, 1992.

J. Elster. *Logic and Society: Contradictions and Possible Worlds.* John Wiley & Sons, Chichester, 1978.

J. Elster. *Nuts and Bolts for the Social Sciences.* Cambridge University Press, 1989.

H. B. Enderton. *A Mathematical Introduction to Logic.* Academic Press, New York, 1972.

B. Falkenhainer and K. D. Forbus. Compositional modeling: finding the right model for the job. *Artificial Intelligence*, 51:95–143, 1991.

B. Falkenhainer and S. Rajamoney. The interdependencies of theory formation, revision, and experimentation. In *Proceedings of the 5th International Conference on Machine Learning*, pages 353–366. University of Michigan, Ann Arbor MI, 1988.

T. J. Fararo. *Mathematical Sociology: An Introduction to Fundamentals.* Wiley-Interscience, New York, London, 1973.

A. Farquhar. *Automated Modeling of Physical Systems in the Presence of Incomplete Knowledge.* PhD thesis, Department of Computer Science, The University of Texas at Austin, 1993.

E. A. Feigenbaum and J. Feldman, editors. *Computers and Thought.* McGraw-Hill, New York, 1963.

A. Fisher. *The logic of real arguments.* Cambridge University Press, Cambridge UK, 1988.

M. Fitting. *First-Order Logic and Automated Theorem Proving.* Springer-Verlag, New York, second edition, 1996.

P. Flach. *Conjectures: an inquiry concerning the logic of induction.* ITK dissertation series 1995-1, Tilburg University, Tilburg, The Netherlands, 1995.

K. D. Forbus. Qualitative process theory. *Artificial Intelligence*, 24:85–168, 1984.

K. D. Forbus. Qualitative physics: Past, present and future. In H. E. Shrobe, editor, *Exploring Artificial Intelligence: Survey Talks from the National Conference on Artificial Intelligence*, chapter 7, pages 239–298. Morgan-Kaufman CA, 1988.

J. W. Forrester. *Industrial Dynamics.* MIT Press, Cambridge MA, 1961.

J. Freeman, G. R. Carroll, and M. T. Hannan. The liability of newness: Age dependence in organizational death rates. *American Sociological Review*, 48: 692–710, 1983.

L. Freese. Formal theorizing. *Annual Review of Sociology*, 6:187–212, 1980.

M. Friedman. *Foundations of Space-Time Theories: Relativistic Physics and Philosophy of Science.* Princeton University Press, Princeton NJ, 1983.

M. Friedman. Logical truth and analyticity in Carnap's "logical syntax of language". In W. Aspray and P. Kitcher, editors, *History and Philosophy of Modern Mathematics*, volume XI of *Minnesota Studies in the Philosophy of Science*, pages 82–94. University of Minnesota Press, Minneapolis, 1988.

M. Friedman. The re-evaluation of logical positivism. *The Journal of Philosophy*, LXXXVIII:505–519, 1991.

J. R. Galbraith. *Organization Design.* Addison-Wesley, Reading MA, 1977.

M. Genesereth and N. Nilsson. *Logical Foundations of Artificial Intelligence.* Morgan Kaufmann, San Mateo, CA, 1987.

R. N. Giere, editor. *Cognitive Models of Science*, volume XV of *Minnesota Studies in the Philosophy of Science.* University of Minnesota Press, Minneapolis MN, 1992.

M. L. Ginsberg, editor. *Readings in Nonmonotonic Reasoning.* Morgan Kaufmann, Los Altos CA, 1987.

F. Giunchiglia and T. Walsh. Theorem proving with definitions. In A. Cohn and J. Dennison, editors, *AISB-89: Proceedings of the seventh conference of Artificial Intelligence and the Simulation of Behaviour*, pages 175–183. Pitman, London, 1989.

J. C. Glorie, M. Masuch, and M. Marx. Formalizing organizational theory: A knowledge-based approach. In Masuch (1990), chapter 4, pages 79–81.

C. Glymour. Invasion of the mind snatchers. In Giere (1992), pages 465–471.

C. Glymour. *Thinking Things Through: An Introduction to Philosophical Issues and Achievements.* MIT Press, Cambridge MA, 1992b.

N. Goodman. *The Structure of Appearance*, volume 107 of *Synthese Library.* D. Reidel Publishing Company, Dordrecht NL, third edition, 1977.

A. Grandori. *Perspectives on Organization Theory.* Ballinger Publishing Company, Cambridge MA, 1987.

M. Granovetter. Small is bountiful: Labor markets and establishment size. *American Sociological Review*, 49:323–334, 1984.

J. Hage. An axiomatic theory of organizations. *Administrative Science Quarterly*, 10:289–320, 1965.

J. Hage, editor. *Formal Theory in Sociology.* State University of New York Press, Albany, 1994a.

J. Hage. Sociological theory: Complex, fragmented, and politicized. In *Formal Theory in Sociology* Hage (1994a), chapter 3, pages 52–65.

M. T. Hannan. *Aggregation and Disaggregation in Sociology.* Lexington Books, Lexington MA, 1971a.

M. T. Hannan. Problems of aggregation. In Blalock (1971), chapter 27, pages 473–508.

M. T. Hannan. Comment: On logical formalization of theories from organizational ecology. In Raftery (1997), pages 145–150.

M. T. Hannan. Rethinking age dependence in organizational mortality: Logical formalizations. *American Journal of Sociology*, 104:126–164, 1998.

M. T. Hannan and G. R. Carroll. *Dynamics of Organizational Populations: density, legitimation, and competition.* Oxford University Press, New York, Oxford, 1992.

M. T. Hannan and J. Freeman. The population ecology of organizations. *American Journal of Sociology*, 82:929–964, 1977.

M. T. Hannan and J. Freeman. Structural inertia and organizational change. *American Sociological Review*, 49:149–164, 1984.

M. T. Hannan and J. Freeman. *Organizational Ecology.* Harvard University Press, Cambridge MA, 1989.

P. J. Hayes. The naive physics manifesto. In D. Michie, editor, *Expert Systems in the Micro-Electronic Age*, pages 242–270. Edinburgh University Press, Edinburgh Scotland, 1979.

P. J. Hayes. Naive physics I: Ontology for liquids. In Hobbs and Moore (1985), pages 71–107.

P. J. Hayes. The second naive physics manifesto. In Hobbs and Moore (1985), pages 1–36.

F. Hayes-Roth. Using proofs and refutations to learn from experience. In R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, editors, *Machine Intelligence, an artificial intelligence approach*, Symbolic Computation, chapter 8, pages 221–240. Springer-Verlag, Berlin, 1984.

C. G. Hempel. *Philosophy of Natural Science.* Foundations of Philosophy Series. Prentice-Hall, Englewood Cliffs NJ, 1966.

C. G. Hempel. On the "standard conception" of scientific theories. In M. Radner and S. Winokur, editors, *Minnesota Studies in the Philosophy of Science*, volume IV, pages 142–163. University of Minnesota Press, Minneapolis, 1970.

M. Henle. On the relation between logic and thinking. *Psychological Review*, 69 (4):366–378, 1962.

J. Hintikka. Truth definitions, skolem functions and axiomatic set theory. *The Bulletin of Symbolic Logic*, 4:303–337, 1998.

J. R. Hobbs, T. Blenko, B. Croft, G. Hager, H. A. Kautz, P. Kube, and Y. Shoham. Commonsense summer: Final report. Technical Report CSLI-85-35, CSLI, Stanford University, 1985.

J. R. Hobbs and R. C. Moore, editors. *Formal Theories of the Commonsense World*. Ablex Publishing Corp. Norwood NJ, 1985.

W. Hodges. An editor recalls some hopeless papers. *The Bulletin of Symbolic Logic*, 4(1):1–16, 1998.

G. C. Homans. *The Human Group*. Hartcourt, Brace & World, New York, 1950.

G. C. Homans. *The Nature of Social Science*. Harcourt, Brace & World, New York, 1967.

T. K. Hopkins. *The Exercise of Influence in Small Groups*. The Bedminster Press, Totowa NJ, 1964.

Z. Huang, M. Masuch, and L. Pólos. ALX, an action logic for agents with bounded rationality. *Artificial Intelligence*, 82(1-2):75–127, 1996.

K. Inoue and C. Sakama. Abductive framework for nonmonotonic theory change. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 204–210. Morgan Kaufmann, 1995.

M. C. W. Janssen and Y.-H. Tan. Friedman's permanent income hypothesis is an example of diagnostic reasoning. *Economics and Philosophy*, 8:23–49, 1992.

J. Kamps. Computational support for logical formalization. In K. Carley, editor, *Proceedings of the Seventh Annual Workshop on Computational and Mathematical Organization Theory*, pages 14–15, May 3-4, San Diego CA, 1997a.

J. Kamps. Logic in organization theory. *Logic in the Netherlands*, 1(2):11–12, 1997b.

J. Kamps. Model generation and theory building: a case study of Hage's axiomatic theory of organizations. In U. Furbach, editor, *IJCAI-97 Workshop on Model-Based Automated Reasoning*, pages 63–72, Nagoya Congress Center, Nagoya, Japan, 1997c.

J. Kamps. Formal theory building using automated reasoning tools. In A. G. Cohn, L. K. Schubert, and S. C. Shapiro, editors, *Principles of Knowledge*

Representation and Reasoning: Proceedings of the Sixth International Conference (KR'98), pages 478–487. Morgan Kaufmann Publishers, San Francisco CA, 1998a.

J. Kamps. On formal theory building using automated reasoning tools. In H. La Poutré and J. van den Herik, editors, *NAIC'98: Proceedings of the Tenth Netherlands/Belgium Conference on Artificial Intelligence*, pages 285–286, 1998b.

J. Kamps. On criteria for formal theory building. In E. Postma and M. Gyssens, editors, *BNAIC 1999: Proceedings of the Eleventh Belgium-Netherlands Conference on Artificial Intelligence*, pages 229–230, 1999a.

J. Kamps. On criteria for formal theory building: Applying logic and automated reasoning tools to the social sciences. In J. Hendler and D. Subramanian, editors, *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*, pages 285–290. AAAI Press/The MIT Press, Menlo Park CA, 1999b.

J. Kamps. On the axiomatization of organization theory. In J. Cachro and K. Kijania-Placek, editors, *Volume of Abstracts of the Eleventh International Conference of Logic, Methodology and Philosophy of Science*, page 392. Faculty of Philosophy, Jagiellonian University, Cracow, Poland, 1999c.

J. Kamps. On the process of axiomatizing scientific theories: Using justification criteria in the context of discovery. In *Proceedings of the AISB'99 Symposium on Scientific Creativity*, pages 49–58. The Society for the Study of Artificial Intelligence and Simulation of Behaviour, 1999d.

J. Kamps and M. Masuch. Partial deductive closure: Logical simulation and management science. *Management Science*, 43(9):1229–1245, 1997.

J. Kamps, M. Masuch, and B. Ó Nualláin. Applying automated reasoning to theory building in the social sciences. In M. Abraham and T. Voss, editors, *Herbsttagung der DGS-Sektion Modellbildung und Simulation*, Institut für Soziologie, Universität Leipzig, D, 1997.

J. Kamps and G. Péli. Qualitative reasoning beyond the physics domain: The density dependence theory of organizational ecology. In B. Bredeweg, editor, *Ninth International Workshop on Qualitative Reasoning*, pages 114–122, University of Amsterdam, Amsterdam, The Netherlands, 1995.

J. Kamps and L. Pólos. Reducing uncertainty: A formal theory of *Organizations in Action*. *American Journal of Sociology*, 104:1776–1812, 1999.

A. Kaplan. *The Conduct of Inquiry: Methodology for Behavioral Science*. Chandler Publishing Company, Scranton Pennsylvania, 1964.

J. G. Kemeny and J. L. Snell. *Mathematical Models in the Social Sciences*. Blaisdell Publishing Company, New York, 1962.

S. O. Kimbrough, editor. *Special Issue on Logic Modeling*, volume 6(2) of *Decision Support Systems*. Elsevier Science Publishers, North-Holland, 1990.

S. O. Kimbrough and R. M. Lee. Logical modeling: A tool for management science. *Decision Support Systems*, 4(1):3–16, 1988a.

S. O. Kimbrough and R. M. Lee, editors. *Special Issue on Logic Modeling*, volume 4(1) of *Decision Support Systems*. Elsevier Science Publishers, North-Holland, 1988b.

E. Kiser. Comment: Evaluating qualitative methodologies. In Raftery (1997), pages 151–159.

M. S. Kraatz and E. J. Zajac. Exploring the limits of the new institutionalism: The causes and consequences of illegitimate organizational change. *American Sociological Review*, 61:812–836, 1996.

T. S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, 1962.

T. S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, third edition, 1996.

B. J. Kuipers. *Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge*. MIT Press, Cambridge MA, 1994.

T. A. F. Kuipers. A structuralist approach to truthlikeness. In T. A. F. Kuipers, editor, *What is closer-to-the-truth?*, volume 10 of *Poznań Studies in the Philosophy of the Sciences and the Humanities*, pages 79–99. Rodopi, Amsterdam, 1987.

T. A. F. Kuipers. Logic and philosophy of science: Current interfaces. In M. L. Dalla Chiara, K. Doets, D. Mundici, and J. van Benthem, editors, *Logic and Scientific Methods*, volume 259 of *Synthese Library*, pages 379–381. Kluwer Academic Publishers, Dordrecht, 1997.

H. E. Kyburg, Jr. *Philosophy of Science: A Formal Approach*. The Macmillan Company, New York, 1968.

I. Lakatos. Falsification and the methodology of scientific research programmes. In Lakatos and Musgrave (1970), pages 91–195.

I. Lakatos. *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge University Press, Cambridge, England, 1976.

I. Lakatos and A. Musgrave, editors. *Criticism and the Growth of Knowledge*. Cambridge University Press, Cambridge UK, 1970.

P. Langley, H. A. Simon, G. L. Bradshaw, and J. M. Żytkow. *Scientific Discovery: Computational Explorations of the Creative Processes*. The MIT Press, Cambridge MA, 1987.

P. R. Lawrence and J. W. Lorsch. *Organization and Environment.* Harvard University Press, Cambridge MA, 1967.

P. F. Lazarsfeld, editor. *Mathematical Thinking in the Social Sciences.* The Free Press, Glencoe IL, 1954.

P. F. Lazarsfeld. Philosophy of science and empirical social research. In Nagel, Suppes, and Tarski (1962), pages 463–473.

D. B. Lenat. AM*: An Artificial Intelligence Approach to Discovery in Mathematics.* PhD thesis, Stanford University, 1976.

D. A. Levinthal. Organizational adaptation and environmental selection—interrelated processes of change. *Organization Science*, 2:140–145, 1991.

K. Lewin. *Field Theory in Social Science: Selected Theoretical Papers.* Harper and Row, New York, 1951.

D. W. Loveland, editor. *Automated Deduction: Some Achievements and Future Directions*, 1996. Report to the National Science Foundation (NSF) on the Future Directions of Automated Deduction.

D. Makinson and P. Gärdenfors. Relations between the logic of theory change and nonmonotonic reasoning. In A. Fuhrmann and M. Morreau, editors, *The Logic of Theory Change*, volume 465 of *Lecture Notes in Artificial Intelligence*, pages 185–205. Springer-Verlag, Berlin, 1991.

K. Manhart. Strukturalistische Theorienkonzeption in den Sozialwissenschaften. *Zeitschrift für Soziologie*, 23:111–128, 1994.

J. G. March and H. A. Simon. *Organizations.* Wiley, New York, 1958.

M. Masuch, editor. *Organization, Management, and Expert Systems: models of automated reasoning.* De Gruyter, Berlin, New York, 1990.

M. Masuch. Introduction: Artificial intelligence in organization and management theory. In Masuch and Warglien (1992), pages 1–19.

M. Masuch, J. Bruggeman, J. Kamps, G. Péli, and L. Pólos. What theory is...: A reaction to Sutton and Staw's "what theory is not". *ASQ Research Listserver*, 2(1), 1996.

M. Masuch and Z. Huang. A case study in logical deconstruction: Formalizing J.D. Thompson's *organizations in action* in a multi-agent action logic. *Computational and Mathematical Organization Theory*, 2(2):71–114, 1996.

M. Masuch and M. Warglien, editors. *Artificial Intelligence in Organization and Management Theory: models of distributed activity.* North-Holland, Amsterdam, 1992.

J. McCarthy. Programs with common sense. In D. V. Blake and A. M. Uttley, editors, *Mechanisation of Thought Processes: Proceedings of a symposium held at* *the National Physics Laboratory*, pages 75–84. Her Majesty's Stationary Office, London, 1959.

W. McCune. A Davis-Putnam program and its application to finite first-order model search: Quasigroup existence problems. Technical report, Argonne National Laboratory, Argonne IL, 1994a. DRAFT.

W. McCune. OTTER: Reference manual and guide. Technical Report ANL-94/6, Argonne National Laboratory, Argonne IL, 1994b.

R. K. Merton. *Social Theory and Social Structure.* The Free Press, New York, 1968 enlarged edition, 1968.

H. Mintzberg. *The Structuring of Organizations.* Prentice-Hall, Englewood Cliffs NJ, 1979.

E. Nagel. The logic of social science. In *Logic Without Metaphysics*, chapter 12, pages 361–368. The Free Press, Glencoe IL, 1956.

E. Nagel. *The Structure of Science: Problems in the Logic of Scientific Explanation.* Routledge & Kegan Paul, London, 1961.

E. Nagel, P. Suppes, and A. Tarski, editors. *Logic, Methodology, and Philosophy of Science.* Stanford University Press, Stanford CA, 1962.

O. Neurath. Foundations of the social sciences. In Neurath et al. (1970), chapter II.1, pages 1–51.

O. Neurath, R. Carnap, and C. Morris, editors. *Foundations of the Unity of Science: Towards an International Encyclopedia of Unified Science*, volume I(1–10) and II(1–9). The University of Chicago Press, Chicago IL, 1970.

A. Newell, J. C. Shaw, and H. A. Simon. Emperical explorations with the logic theory machine: a case study in heuristics. In *Proceedings Western Joint Computer Conference*, pages 218–239, 1957.

A. Newell and H. A. Simon. The logic theory machine. In *IRE Transactions on Information Theory*, volume IT-2(3), pages 61–79, 1956.

A. Newell and H. A. Simon. *Human Problem Solving.* Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1972.

N. Nicholson, editor. *Encyclopedic Dictionary of Organizational Behavior.* Blackwell Publishers, Oxford UK, 1995.

B. Ó Nualláin. Mixing `metafor`: a system description of the `metafor` theorem proving system. In A. Voronkov, editor, *Logic Programming and Automated Reasoning, St.Petersburg, Russia*, volume 698 of *Lecture Notes in Artificial Intelligence*, pages 357–359. Springer-Verlag, 1993.

T. Parsons. *Structure and Process in Modern Societies.* The Free Press, Glenncoe IL, 1960.

G. Péli. A logical formalization of population dynamics: the density dependence model of organizational ecology. Technical Report 93-106, Center for Computer Science in Organization and Management, University of Amsterdam, 1993.

G. Péli. The niche hiker's guide to population ecology: A logical reconstruction of organizational ecology's niche theory. In Raftery (1997), pages 1–46.

G. Péli, J. Bruggeman, M. Masuch, and B. Ó Nualláin. A logical approach to formalizing organizational ecology. *American Sociological Review*, 59:571–593, 1994.

G. Péli and M. Masuch. The logic of propagation strategies: Axiomatizing a fragment of organizational ecology in first-order logic. In D. P. Moore, editor, *Best Papers Proceedings of the* 54^{th} *Annual Academy of Management Convention, Dallas, Texas*, pages 218–222, 1994.

G. Péli and M. Masuch. The logic of propagation strategies: Axiomatizing a fragment of organizational ecology in first-order logic. *Organization Science*, 8: 310–331, 1997.

C. Perrow. *Complex Organizations: A Critical Essay.* McGraw-Hill, New York, third edition, 1986.

J. Pfeffer and G. R. Salancik. *The External Control of Organizations: A Resource Dependence Perspective.* Harper & Row Publishers, New York NY, 1978.

J. Piaget. *Structuralism.* Routledge and Kegan Paul, London, 1971.

M. Polanyi. *Personal Knowledge: Towards a Post-Critical Philosophy.* University of Chicago Press, Chicago IL, 1958.

L. Pólos. Organizations on holiday: Formalizing Thompson's organizations in action. Technical Report 91-35, Center for Computer Science in Organization and Management, University of Amsterdam, 1991.

L. Pólos and M. Masuch. Information states in situation theory. In L. Pólos and M. Masuch, editors, *Applied Logic: How, What, and Why*, volume 247 of *Synthese Library*, pages 195–236. Kluwer Academic Publishers, Dordrecht, 1995.

G. Pólya. *How to Solve It: a new aspect of mathematical method.* Princeton University Press, Princeton New Jersey, 1945.

G. Pólya. *Induction and Analogy in Mathematics*, volume I of *Mathematics and Plausible Reasoning.* Princeton University Press, Princeton New Jersey, 1954a.

G. Pólya. *Patterns of Plausible Inference*, volume II of *Mathematics and Plausible Reasoning.* Princeton University Press, Princeton New Jersey, 1954b.

G. Pólya. *How to Solve It: a new aspect of mathematical method.* Princeton University Press, Princeton New Jersey, second edition, 1957.

G. Pólya. *Mathematical Discovery: on understanding, learning, and teaching problem solving*, volume I. John Wiley & Sons, New York, 1962.

G. Pólya. *Mathematical Discovery: on understanding, learning, and teaching problem solving*, volume II. John Wiley & Sons, New York, 1965.

D. Poole. A logical framework for default reasoning. *Artificial Intelligence*, 36: 27–47, 1988.

K. R. Popper. *The Logic of Scientific Discovery.* Hutchinson, London, 1959.

K. R. Popper. *Conjectures and Refutations: The Growth of Scientific Knowledge.* Routledge and Kegan, London, 1963.

K. R. Popper. Die Logik der Sozialwissenschaften. In Adorno et al. (1969), pages 103–123.

K. R. Popper. Ayer on empiricism and against verisimilitude. In Schilpp (1974), pages 1100–1114.

W. W. Powell and P. J. DiMaggio, editors. *The New Institutionalism in Organizational Analysis.* University of Chicago Press, Chicago IL, 1991.

H. Putnam. What theories are not. In Nagel et al. (1962), pages 240–251.

W. V. O. Quine and J. S. Ulian. *The web of belief.* Random House, New York, 1970.

R. E. Quinn and K. S. Cameron, editors. *Paradox and Transformation: Towards a Theory of Change in Organization and Management.* Ballinger Publishing Company, Cambridge MA, 1988.

A. E. Raftery, editor. *Sociological Methodology 1997.* Blackwell, Oxford UK, 1997.

N. Rakić. *Common Sense Time and Special Relativity.* ILLC dissertation series 1997-02, University of Amsterdam, 1997a.

N. Rakić. Past, present, future, and special relativity. *British Journal for the Philosophy of Science*, 48:257–280, 1997b.

F. P. Ramsey. Theories. In *The Foundations of Mathematics*, chapter IX.A, pages 212–236. Routledge & Kegan Paul, London, 1931.

H. Reichenbach. *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge.* The University of Chicago Press, Chicago IL, 1938.

G. D. Ritchie and F. K. Hanna. AM: A case study in AI methodology. *Artificial Intelligence*, 23:249–268, 1984.

J. A. Robinson. A machine-oriented logic based on the resolution principle. *Journal of the Association of Computing Machinery*, 12(1):23–41, 1965.

H. Rott. *Reduktion und Revision: Aspekte des nichtmonotonen Theorienwandels.* Number 290 in Europäische Hochschulschriften: Reihe XX, Philosophie. Verlag Peter Lang GmbH, Frankfurt am Main, 1991.

R. S. Rudner. *Philosophy of Social Science.* Foundations of Philosophy Series. Prentice-Hall, Englewood Cliffs NJ, 1966.

W. A. Rushing and M. N. Zald, editors. *Organizations and Beyond: Selected Essays of James D. Thompson.* D.C. Heath, Lexington MA, 1976.

P. A. Schilpp, editor. *The Philosophy of Karl Popper*, number XIV–1,2 in The Library of Living Philosophers. Open Court, La Salle IL, 1974.

W. R. Scott. *Organizations: Rational, Natural, and Open Systems.* Prentice-Hall, Englewood Cliffs NJ, fourth edition, 1998.

E. Shapiro. Inductive inference of theories from facts. In J.-L. Lassez and G. Plotkin, editors, *Computational Logic*, chapter 7, pages 199–254. The MIT Press, Cambridge MA, 1991.

J. Shrager and P. Langley, editors. *Computational Models of Scientific Discovery and Theory Formation*, The Morgan Kaufmann Series in Machine Learning. Morgan Kaufmann Publishers, San Mateo CA, 1990.

H. A. Simon. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization.* The Macmillan Company, New York, 1947.

H. A. Simon. A formal model of interaction in social groups. *American Sociological Review*, 17:202–211, 1952.

H. A. Simon. *Models of Man: Social and Rational.* John Wiley & Sons, New York NY, 1957.

H. A. Simon. *Models of Discovery*, volume 114 of *Synthese Library.* D. Reidel Publishing Company, Dordrecht, The Netherlands, 1977.

H. A. Simon. Fitness requirements for scientific theories. *British Journal for the Philosophy of Science*, 34:355–365, 1983.

H. A. Simon. Machine discovery. *Foundations of Science*, 2:171–200, 1995.

H. A. Simon and G. J. Groen. Ramsey eliminability and the testability of scientific theories. *British Journal for the Philosophy of Science*, 24:357–408, 1973.

R. A. Singleton, Jr. and B. C. Straits. *Approaches to Social Research.* Oxford University Press, New York, Oxford, third edition, 1999.

J. Slaney. The crisis in finite mathematics: Automated reasoning as cause and cure. In Bundy (1994), pages 1–13.

J. Slaney. FINDER: Finite domain enumerator. In Bundy (1994), pages 789–801.

J. D. Sneed. *The Logical Structure of Mathematical Physics.* D. Reidel Publishing Company, Dordrecht, 1971.

Social Sciences Citation Index. *Social Sciences Citation Index: an international interdisciplinary index to the literature of the social sciences.* Institute for Scientific Information, Philadelphia PA, 1995.

Social Sciences Citation Index. *Social Sciences Citation Index: an international interdisciplinary index to the literature of the social sciences.* Institute for Scientific Information, Philadelphia PA, 1996.

Social Sciences Citation Index. *Social Sciences Citation Index: an international interdisciplinary index to the literature of the social sciences.* Institute for Scientific Information, Philadelphia PA, 1997.

A. Sørensen. Mathematical models in sociology. *Annual Review of Sociology*, 4: 354–371, 1978.

W. Stegmüller. *Logische Analyse der Struktur ausgereifter physikalischer Theorien, 'Non-statement view' von Theorien.* Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie: Band II Theorie und Erfahrung, Teil D. Springer-Verlag, Berlin, 1973.

W. Stegmüller. Anwendungsbeispiele außerhalb der Physik. In *Theorie und Erfahrung: Band II/3*, Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, chapter 14, pages 360–449. Springer-Verlag, Berlin, 1986.

A. L. Stinchcombe. Social structure and organizations. In J. G. March, editor, *Handbook of Organizations*, pages 152–193. Rand McNally, Chicago IL, 1965.

A. L. Stinchcombe. *Constructing Social Theories.* Harcourt, Brace & World, New York, 1968.

F. Suppe, editor. *The Structure of Scientific Theories.* University of Illinois Press, Urbana IL, 1974.

P. Suppes. The desirability of formalization in science. *Journal of Philosophy*, LXV(20):651–664, 1968.

P. Suppes. Heuristics and the axiomatic method. In *Models and Methods in the Philosophy of Science*, volume 226 of *Synthese Library*, chapter 5, pages 55–65. Kluwer Academic Publishers, Dordrecht, Boston, 1993.

G. Sutcliffe, C. Suttner, and T. Yemenis. The TPTP problem library. In Bundy (1994), pages 252–266.

J. R. Sutton and F. Dobbin. The two faces of governance: Responses to legal uncertainty in U.S. firms, 1955 to 1985. *American Sociological Review*, 61: 794–811, 1996.

J. R. Sutton, F. Dobbin, J. W. Meyer, and W. R. Scott. The legalization of the workplace. *American Journal of Sociology*, 99:944–971, 1994.

R. I. Sutton and B. M. Staw. What theory is *not*. *Administrative Science Quarterly*, 40:371–384, 1995.

Y.-H. Tan. *Non-Monotonic Reasoning: Logical Architecture and Philosophical Applications.* PhD thesis, Free University, Amsterdam, The Netherlands, 1992.

A. Tarski. *Introduction to Logic and to the Methodology of Deductive Sciences.* Oxford University Press, New York, second, revised edition, 1946.

A. Tarski. *Logic, Semantics, Metamathematics: Papers from 1923 to 1928.* Oxford University Press, New York, 1956.

P. Thagard. *Computational Philosophy of Science.* The MIT Press, Cambridge MA, 1988.

J. D. Thompson. *Organizations in Action: Social Science Bases of Administrative Theory.* McGraw-Hill, New York, 1967.

M. L. Tushman and E. Romanelli. Organizational evolution: A metamorphosis model of convergence and reorientation. *Research in Organizational Behavior*, 7:171–222, 1985.

U.S. Small Business Administration. *The State of Small Business: A Report of the President.* United States Government Printing Office, Washington DC, 1985.

R. E. Valdés-Pérez. Research note: Principles of human-computer collaboration for knowledge discovery in science. *Artificial Intelligence*, 107:335–346, 1999.

J. van Benthem. The logical study of science. *Synthese*, 51:431–472, 1982.

A. H. Van de Ven. Nothing is quite so practical as a good theory. *Academy of Management Review*, 14:486–489, 1989a. Special forum on theory building.

A. H. Van de Ven. Special forum on theory building. *Academy of Management Review*, 14:486–594, 1989b.

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, Princeton, 1944.

M. Weber. *Methodologische Schriften.* Fisher, Frankfurt am Main D, 1968.

K. E. Weick. Theory construction as disciplined imagination. *Academy of Management Review*, 14:516–531, 1989. Special forum on theory building.

K. E. Weick. What theory is *not*, theorizing *is*. *Administrative Science Quarterly*, 40:385–390, 1995.

C. Weidenbach. Spass: Version 0.49. *Journal of Automated Reasoning*, 18:247–252, 1997.

D. S. Weld and J. de Kleer, editors. *Readings in Qualitative Reasoning about Physical Systems.* Morgan-Kaufman CA, 1990.

A. N. Whitehead. *The Organisation of Thought: Educational and Scientific.* Williams and Norgate, London, 1917.

A. N. Whitehead and B. Russell. *Principia Mathematica to *56.* Cambridge University Press, Cambridge, 1962.

J. H. Woodger. *Axiomatic Method in Biology.* Cambridge University Press, 1937.

J. H. Woodger. *The Technique of Theory Construction*, volume II.5 of *Foundations of the Unity of Science*. University of Chicago Press, Chicago, 1939.

L. Wos. *Automated Reasoning, 33 Basic Research Problems*. Prentice Hall, Englewood Cliffs, New Jersey, 1988.

L. Wos. *The Automation of Reasoning: An Experimenter's Notebook with OTTER Tutorial*. Academic Press, San Diego CA, 1996.

L. Wos. Programs that offer fast, flawless, logical reasoning. *Communications of the ACM*, 41(6):87–95, 1998.

L. Wos, R. Overbeek, E. Lusk, and J. Boyle. *Automated Reasoning: Introduction and Applications*. McGraw-Hill, Inc., second edition, 1992.

M. N. Zald. More fragmentation? unfinished business in linking the social sciences and the humanities. *Administrative Science Quarterly*, 41:251–261, 1996.

H. L. Zetterberg. On axiomatic theories in sociology. In P. F. Lazarsfeld and M. Rosenberg, editors, *The Language of Social Research*, pages 533–540. The Free Press, Glencoe IL, 1955.

H. L. Zetterberg. *On Theory and Verification in Sociology*. The Bedminster Press, Totowa NJ, third enlarged edition, 1965.

J. Zhang and H. Zhang. SEM: a system for enumerating models. In *Proceedings of the 14th International Joint Conference on AI*, pages 298–303, 1995.

Index

Samenvatting

De algemene beschikbaarheid van krachtige computers kan de manier waarop wetenschap bedreven wordt radicaal veranderen. Om dit potentieel te verwezenlijken moeten we onze wetenschappelijke kennis voor onze computers toegankelijk maken. Tijdens de laatste decennia heeft de vooruitgang van de informatica het mogelijk gemaakt om op grote schaal wetenschappelijke gegevens geautomatiseerd vast te leggen. Dezelfde vooruitgang is vooralsnog niet bereikt voor de meest gerenommeerde vorm van wetenschappelijke kennis – onze theorieën. Dit proefschrift presenteert *een logische benadering van computer-ondersteunde theorie-ontwikkeling*, hetgeen wil zeggen dat wij de representatie van wetenschappelijke theorieën met behulp van symbolische logica zullen onderzoeken.

Wetenschappelijk theorieën worden doorgaans gezien als systemen van beweringen: de theorie doet bepaalde voorspellingen die volgen uit een aantal onderliggende aannamen. De argumentatie voor een dergelijke voorspelling omvat doorgaans zowel aannames over het empirische domein, alsmede algemene afleidingsregels (waarvan de geldigheid niet afhankelijk is van de gegeven toepassing). De logica is het wetenschapsgebied dat van oudsher onderzoek doet naar deze algemene afleidingsregels. De formele logica heeft logisch geldige formalismen opgeleverd die, op zijn minst, een normatieve rol kunnen spelen voor de juistheid van argumenten. De voordelen van het logisch representeren van theorieën zijn algemeen bekend en vaak uitgebreid besproken, en tevens is gebleken dat interessante methodologische onderwerpen kunnen worden besproken onder de aanname dat theorieën in een logische taal zijn geaxiomatiseerd. Het is echter belangrijk om op te merken dat, tot nu toe, logische axiomatiseringen van significante theorieën uit de empirische wetenschappen zelden zijn verwezenlijkt. Met andere woorden, over het in de praktijk toepassen van formele logica voor de representatie van theorieën uit de empirische wetenschappen is veel minder bekend. Deze praktische aspecten vormen een belangrijk onderdeel van dit proefschrift. In het bijzonder zullen we in ieder hoofdstuk formele versies van theorieën of fragmenten van theorieën construeren of bespreken.

Als toepassingsgebied voor de representatie van theorieën in formele logica hebben we gekozen voor de sociale wetenschappen, in het bijzonder theorieën uit het vakgebied van de sociologie en de organisatiesociologie. Dit is misschien een verrassende keuze aangezien de paradigmatische voorbeelden van theorieën in de empirische wetenschappen uit andere vakgebieden komen, met name uit de natuurkunde. De sociologie staat zelfs bekend om haar gebrek aan systematische theorieën. Niettemin, voor zover de sociologie een wetenschap is, moet zij voldoen aan dezelfde methodologische principes als de andere wetenschappen. Dat betekent, bijvoorbeeld, dat de vraag of een theoretisch argument in de sociologie valide is, moet worden beoordeeld volgens dezelfde criteria als een argument in de natuurkunde. In het kort, de "logica" van de sociale wetenschappen is dezelfde als die van de andere wetenschappen.

Hoofdstuk 2 treft de voorbereidingen voor een computationele methodologie voor het axiomatiseren van wetenschappelijke theorieën. Het eerste doel is te verkennen hoe logische criteria voor het evalueren van theorieën in de praktijk kunnen worden getoetst, daarbij gebruik makend van generieke computerprogramma's uit de kunstmatige intelligentie. Het tweede doel is om na te gaan welke logische criteria overeen komen met vragen die van nature over sociaalwetenschappelijke theorieën te stellen zijn. Het grootste gedeelte van dit hoofdstuk bestaat uit een case studie van de formalisering van een sociologische theorie, Hopkins' "The Exercise of Influence in Small Groups", in eerste-orde logica.

Er zijn overtuigende filosofische argumenten dat sociaal-wetenschappelijke theorieën moeten voldoen aan dezelfde logische criteria als theorieën uit andere disciplines. Dit impliceert echter niet dat deze criteria direct toepasbaar zijn op theoretische beschrijvingen in de sociologie. Het is zelfs mogelijk dat deze criteria (nog) niet relevant lijken voor theoretische onderzoekers in de sociologie. Uit de case studie blijkt dat diverse van de standaard criteria voor het evalueren van theorieën relevant zijn voor deze specifieke theorie. Tevens slagen we erin om ieder van deze criteria computationeel te toetsen met behulp van automatisch-redeneerprogramma's. De functionaliteit van deze programma's is beperkt door een aantal restricties, zoals de onbeslisbaarheid van eerste-orde logica en praktische beperkingen in CPU-kracht, geheugen en tijd. Desalniettemin kan, gebruikmakend van slechts de standaardinstellingen van de programma's, ieder van de criteria binnen vijf seconden worden getoetst door het vinden van een bepaald bewijs of (tegen)voorbeeld. De evaluatie van sociologische theorieën lijkt dus binnen het bereik van de huidige automatisch-redeneerprogramma's te liggen.

Hoofdstuk 3 presenteert een logische axiomatisering van een klassieke organisatietheorie, Thompsons "Organizations in Action". Thompsons boek is een van de klassieke bijdragen aan de organisatietheorie: het geeft een raamwerk dat het gezichtspunt dat organisaties als gesloten systemen behandelt, en het gezichtspunt dat uitgaat van de afhankelijkheden tussen organisaties en hun omgeving, met elkaar verenigt. Dit raamwerk heeft veel van het daaropvolgende onderzoek beïnvloed. "Organizations in Action" is een theorie in natuurlijke taal, waarin alleen de belangrijkste beweringen duidelijk zijn aangegeven. Dit hoofdstuk geeft een formele vertaling van de eerste hoofdstukken van Thompsons boek, door de tekstuele argumentatie te reconstrueren. Alhoewel de tekst geen expliciete definities geeft, suggereert het gebruik van terminologie strikte afhankelijkheden tussen een aantal belangrijke concepten. Dit gaf aanleiding tot het definiëren van deze concepten in termen van een klein aantal primitieve begrippen uit de organisatietheorie. In de formele theorie kunnen de belangrijkste beweringen worden afgeleid als theorema's. De bewijzen van deze theorema's zijn gebaseerd op een reconstructie van de argumentatie in de tekst. Verder verklaart de formele theorie waarom de beweringen slechts geldig zijn voor een bepaald type van organisaties. Tevens is er een tot nu toe onbekende implicatie van de theorie afleidbaar, die Thompsons theorie relateert aan recent empirisch onderzoek en huidige ontwikkelingen in de organisatietheorie.

De sociale wetenschappen staan bekend om het gebruik van een ruim vocabulaire – een van de meest in het oog springende verschillen met theorieën in andere wetenschappen. Sociaal-wetenschappelijke theorieën gebruiken vaak veel gerelateerde concepten met slechts subtiele verschillen in betekenis. Het resultaat hiervan is dat een logische representatie van sociaal-wetenschappelijke theorie ook een groot vocabulaire gebruikt. In de formalisering van "Organizations in Action" wordt geëxperimenteerd met het gebruik van definities als middel om de voordelen van een groot vocabulaire te combineren met de voordelen van een klein aantal primitieve termen. Indien een theorie definities bevat, kunnen de gedefinieerde termen worden geëlimineerd door de definities uit te schrijven. Hoofdstuk 4 vervolgt onze discussie van logische criteria voor het evalueren van theorieën. Onze eerdere discussie maakte geen onderscheid tussen verschillende soorten van premissen. Hier onderzoeken we hoe deze criteria moeten worden aangepast voor theorieën die definities bevatten. We zullen praktische operationaliseringen geven van criteria voor het evalueren van theorieën, zoals de consistentie (bestaan van een model) of inconsistentie (afleidbaarheid van een tegenstelling) van de theorie; de betrouwbaarheid (bestaan van een bewijs) of onbetrouwbaarheid (bestaan van een tegenvoorbeeld) van theorema's; de vervulbaarheid en falsificeerbaarheid van theorema's; en de onafhankelijkheid van de axioma's. De tests voor deze criteria, in de praktijk neerkomend op het vinden van een bepaald bewijs of model, kunnen direct worden uitgevoerd door bestaande automatisch-redeneerprogramma's. We gebruiken deze criteria voor het evalueren van de formele theorie van "Organizations in Action" uit hoofdstuk 3. Het bepalen van deze criteria leidt tot een precieze evaluatie van de theorie, hetgeen in sommige gevallen belangrijke gegevens over de theorie kan onthullen. We laten bijvoorbeeld zien dat een van de afgeleide beweringen onfalsificeerbaar is – empirische toetsing kan, op zijn best, de triviale geldigheid van deze bewering bevestigen.

Wetenschappelijke activiteiten worden traditioneel onderverdeeld in de context van rechtvaardiging en de context van ontdekking. Binnen deze verdeling wordt de axiomatisering van theorieën doorgaans gezien als de ultieme stap in de rechtvaardiging van een theorie. Hoofdstuk 5 stelt vraagtekens bij deze zienswijze door het proces van axiomatiseren te onderzoeken. Een theorie in eerste-orde logica geeft een expliciete, ondubbelzinnige uiteenzetting van de theorie. Door toetsing van de criteria uit hoofdstuk 4 kunnen we een nauwgezette evaluatie maken van een dergelijke theorie. Echter, wij zien deze criteria niet als een laatste, statische evaluatie van een theorie. Integendeel, in onze ervaring zijn deze criteria juist het meest bruikbaar tijdens het axiomatiseringsproces. De gebruikte operationaliseringen van de criteria bewijzen niet alleen het criterium, maar geven ook een specifiek bewijs of model dat beschikbaar is voor verdere inspectie.

Iedere uiteenzetting van een theorie maakt diverse onderliggende theoretische vooronderstellingen. De formalisering van een wetenschappelijke theorie vereist dat ook deze vooronderstellingen expliciet worden gemaakt. Het identificeren van deze impliciete onderliggende aannames is een van de moeilijkste problemen tijdens de formalisering van een theorie, waarvoor een diep begrip van het desbetreffende vakgebied nodig is. Indien de van toepassing zijnde onderliggende aannames niet aan de formele theorie worden toegevoegd, zijn bepaalde beweringen niet afleidbaar. Echter, indien een automatisch-redeneerprogramma een klein tegenvoorbeeld tegen een dergelijke niet-afleidbare bewering kan vinden, is het meteen duidelijk waarom een poging het bewijs te vinden zal falen. In het bijzonder kunnen deze tegenvoorbeelden onthullen welke impliciete aannames aan de theorie moeten worden toegevoegd.

Het product van een axiomatisering, een theorie gerepresenteerd in eerste-orde logica, is een deductieve theorie. Het proces van axiomatisering is echter noodzakelijkerwijs niet-deductief. We zullen dit in detail illustreren door het formaliseren van een theorie-fragment uit Zetterbergs “On Theory and Verification in Sociology”. Zoals blijkt moeten we herhaaldelijk de formele representatie van de theorie aanpassen door het toevoegen van impliciete achtergrond aannames, door het herformuleren van niet-afleidbare beweringen, en door het aanpassen van de axioma’s van de theorie. Deze revisies van de formele theorie kunnen ook gevolgen hebben voor de originele theorie. Dit kan van groot belang zijn, aangezien zelfs een kleine verduidelijking van een theorie kan voorkomen dat er onnodige kosten worden gemaakt bij de empirische toetsing van onjuiste of irrelevante hypothesen. De gebruikte criteria faciliteren een stapsgewijze revisie van de theorie, resulterend in een cyclisch proces van theorie-ontwikkeling.

Hoofdstuk 6 presenteert eerder onderzoek naar een geïmplementeerd programma om een bepaalde klasse van theorema’s te enumereren. In de formele logica wordt de deductieve geslotenheid van een theorie als vanzelfsprekend beschouwd – de formele definitie van een theorie is een verzameling zinnen gesloten onder de deductieve afleidingsregels. In de praktijk is het onmogelijk om een complete, deductief gesloten verzameling zinnen te construeren, aangezien deze noodzakelijkerwijs oneindig groot is (bijvoorbeeld omdat die alle tautologieën bevat, i.e., zinnen die altijd waar zijn en dus volgen uit iedere verzameling zinnen). Een gevolg hiervan is dat een complete deductieve sluiting van een verzameling premissen noch realiseerbaar, noch gewenst is – alleen gedeeltelijk gesloten verzamelingen kunnen bruikbare resultaten opleveren. In dit hoofdstuk presenteren we een algoritme dat een efficiënte, gedeeltelijke deductieve sluiting voor een belangrijke klasse van theorema's realiseert, i.e., voor conditionele formules die twee relevante domein concepten relateren. Beweringen van deze klasse vormen de ruggengraat van veel empirische sociale wetenschappen; het is, wellicht, de belangrijkste klasse van beweringen in de sociale wetenschappen. Het geïmplementeerde programma wordt toegepast on een eerste-orde logica formalisering van Hannan en Freemans theorie van "Structural Inertia and Organizational Change". Het algoritme blijkt meer theorema's te genereren dan de oorspronkelijke theorie expliciet maakt, waarvan sommige van theoretisch belang zijn.

Het vermogen om kwalitatief te redeneren over fysieke systemen is belangrijk voor het begrip en interactie met de fysieke wereld, zowel voor mensen als voor machines. Dientengevolge is de studie hiervan een belangrijk onderwerp binnen de kunstmatige intelligentie. De meeste sociaal-wetenschappelijke theorieën zijn van nature kwalitatieve theorieën. Hoofdstuk 7 gaat niet over de representatie van theorieën in logica, maar over het modelleren van theorieën met kwalitatieve redeneertechnieken. Kwalitatieve redeneertechnieken worden bijna uitsluitend toegepast binnen het domein van de natuurkunde. Dit hoofdstuk onderzoekt de toepasbaarheid van deze technieken buiten de natuurkunde. Als case studie wordt in dit hoofdstuk een kwalitatief simulatiemodel van de Hannan en Carrolls "Density Dependence Theory" geconstrueerd. De case studie laat zien dat uit het model en de simulaties inzichten van theoretisch belang kunnen volgen. Tevens bespreekt dit hoofdstuk de verschillen tussen de toepassing van kwalitatieve redeneertechnieken in de natuurkunde en in de sociale wetenschappen.

Hoofdstuk 8 bevat, ten slotte, enige discussie en verdere verwijzingen naar gerelateerd onderzoek.

Titles in the ILLC Dissertation Series:

ILLC DS-1993-01: **Paul Dekker**
Transsentential Meditations; Ups and downs in dynamic semantics

ILLC DS-1993-02: **Harry Buhrman**
Resource Bounded Reductions

ILLC DS-1993-03: **Rineke Verbrugge**
Efficient Metamathematics

ILLC DS-1993-04: **Maarten de Rijke**
Extending Modal Logic

ILLC DS-1993-05: **Herman Hendriks**
Studied Flexibility

ILLC DS-1993-06: **John Tromp**
Aspects of Algorithms and Complexity

ILLC DS-1994-01: **Harold Schellinx**
The Noble Art of Linear Decorating

ILLC DS-1994-02: **Jan Willem Cornelis Koorn**
Generating Uniform User-Interfaces for Interactive Programming Environments

ILLC DS-1994-03: **Nicoline Johanna Drost**
Process Theory and Equation Solving

ILLC DS-1994-04: **Jan Jaspars**
Calculi for Constructive Communication, a Study of the Dynamics of Partial States

ILLC DS-1994-05: **Arie van Deursen**
Executable Language Definitions, Case Studies and Origin Tracking Techniques

ILLC DS-1994-06: **Domenico Zambella**
Chapters on Bounded Arithmetic & on Provability Logic

ILLC DS-1994-07: **V. Yu. Shavrukov**
Adventures in Diagonalizable Algebras

ILLC DS-1994-08: **Makoto Kanazawa**
Learnable Classes of Categorial Grammars

ILLC DS-1994-09: **Wan Fokkink**
Clocks, Trees and Stars in Process Theory

ILLC DS-1994-10: **Zhisheng Huang**
Logics for Agents with Bounded Rationality

ILLC DS-1995-01: **Jacob Brunekreef**
On Modular Algebraic Protocol Specification

ILLC DS-1995-02: **Andreja Prijatelj**
Investigating Bounded Contraction

ILLC DS-1995-03: **Maarten Marx**
Algebraic Relativization and Arrow Logic

ILLC DS-1995-04: **Dejuan Wang**
Study on the Formal Semantics of Pictures

ILLC DS-1995-05: **Frank Tip**
Generation of Program Analysis Tools

ILLC DS-1995-06: **Jos van Wamel**
Verification Techniques for Elementary Data Types and Retransmission Protocols

ILLC DS-1995-07: **Sandro Etalle**
Transformation and Analysis of (Constraint) Logic Programs

ILLC DS-1995-08: **Natasha Kurtonina**
Frames and Labels. A Modal Analysis of Categorial Inference

ILLC DS-1995-09: **G.J. Veltink**
Tools for PSF

ILLC DS-1995-10: **Giovanna Cepparello**
Studies in Dynamic Logic

ILLC DS-1995-11: **W.P.M. Meyer Viol**
Instantial Logic. An Investigation into Reasoning with Instances

ILLC DS-1995-12: **Szabolcs Mikulás**
Taming Logics

ILLC DS-1995-13: **Marianne Kalsbeek**
Meta-Logics for Logic Programming

ILLC DS-1995-14: **Rens Bod**
Enriching Linguistics with Statistics: Performance Models of Natural Language

ILLC DS-1995-15: **Marten Trautwein**
Computational Pitfalls in Tractable Grammatical Formalisms

ILLC DS-1995-16: **Sophie Fischer**
The Solution Sets of Local Search Problems

ILLC DS-1995-17: **Michiel Leezenberg**
Contexts of Metaphor

ILLC DS-1995-18: **Willem Groeneveld**
Logical Investigations into Dynamic Semantics

ILLC DS-1995-19: **Erik Aarts**
Investigations in Logic, Language and Computation

ILLC DS-1995-20: **Natasha Alechina**
Modal Quantifiers

ILLC DS-1996-01: **Lex Hendriks**
Computations in Propositional Logic

ILLC DS-1996-02: **Angelo Montanari**
Metric and Layered Temporal Logic for Time Granularity

ILLC DS-1996-03: **Martin H. van den Berg**
Some Aspects of the Internal Structure of Discourse: the Dynamics of Nominal Anaphora

ILLC DS-1996-04: **Jeroen Bruggeman**
Formalizing Organizational Ecology

ILLC DS-1997-01: **Ronald Cramer**
Modular Design of Secure yet Practical Cryptographic Protocols

ILLC DS-1997-02: **Nataša Rakić**
Common Sense Time and Special Relativity

ILLC DS-1997-03: **Arthur Nieuwendijk**
On Logic. Inquiries into the Justification of Deduction

ILLC DS-1997-04: **Atocha Aliseda-LLera**
Seeking Explanations: Abduction in Logic, Philosophy of Science and Artificial Intelligence

ILLC DS-1997-05: **Harry Stein**
The Fiber and the Fabric: An Inquiry into Wittgenstein's Views on Rule-Following and Linguistic Normativity

ILLC DS-1997-06: **Leonie Bosveld - de Smet**
On Mass and Plural Quantification. The Case of French 'des'/'du'-NP's.

ILLC DS-1998-01: **Sebastiaan A. Terwijn**
Computability and Measure

ILLC DS-1998-02: **Sjoerd D. Zwart**
Approach to the Truth: Verisimilitude and Truthlikeness

ILLC DS-1998-03: **Peter Grunwald**
The Minimum Description Length Principle and Reasoning under Uncertainty

ILLC DS-1998-04: **Giovanna d'Agostino**
Modal Logic and Non-Well-Founded Set Theory: Translation, Bisimulation, Interpolation

ILLC DS-1998-05: **Mehdi Dastani**
Languages of Perception

ILLC DS-1999-01: **Jelle Gerbrandy**
Bisimulations on Planet Kripke

ILLC DS-1999-02: **Khalil Sima'an**
Learning efficient disambiguation

ILLC DS-1999-03: **Jaap Maat**
Philosophical Languages in the Seventeenth Century: Dalgarno, Wilkins, Leibniz

ILLC DS-1999-04: **Barbara Terhal**
Quantum Algorithms and Quantum Entanglement

ILLC DS-2000-01: **Renata Wassermann**
Resource-Bounded Belief Revision