# Formal Theory Building Using Automated Reasoning Tools

**Jaap Kamps**
Applied Logic Laboratory
Institute for Logic, Language and Computation
University of Amsterdam
Amsterdam, The Netherlands

## Abstract

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 paper 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, Hage's axiomatic theory of organizations.

## 1 INTRODUCTION

The theory building methodology outlined in this paper is by no means a new one. The use of formal logic to represent scientific theories dates, at least, back 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.

It is part of the research plan at the Applied Logic Laboratory (ALL) 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 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 3 we present a case study that shows how this can be applied in the social sciences, and we end in Section 4 by summarizing our experiences and drawing conclusions.

## 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,

[1] See for example Adorno et al. (1969) for interesting arguments in favor of, and against the use of formal logic in the social sciences.

it remains open to many interpretations. Provided they can interpret the formalism, a formal exposition of a scientific theory allows readers to understand the theory, to distinguish between alternative readings, to gauge its boundaries, and to compare it with alternatives (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, 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 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 argumentation of the theory valid?** *Can a given theorem 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. Many scientific theories have, when taken in a literal sense, fallacious argumentation. The severity of such a flaw depends on the following two questions:

**(a) What unstated background knowledge is necessary?** In any exposition of a theory, a certain amount of background must be 'taken for granted', that is to say, it is assumed to be indisputable as common knowledge. But such information must be explicitly added to a formalization. For example, the author of a theory which describes how properties of organizations vary over time will hardly find it necessary to explicitly state that the *before* relation between time-points is transitive and anti-symmetric, but this axiom may be necessary to formally derive some of the conclusions of the theory. In this case, counterexamples to such a theorem are 'non-intended' models. For example, a counterexample might be a model in which time-points $t_1 < t_2 < t_3 < t_1$ (a model with a circular notion of time, whereas time is conceived as linear). Finding this counterexample will immediate reveal the cause, in this case a missing axiom on the *before* relation.

**(b) What assumptions has the author neglected to make?** In contrast to the common knowledge of the previous point, occasionally the formalization of a theory reveals a genuine hiatus in the theory. This may have been due to an oversight by the author or perhaps a failure to consider all possible configurations of his/her assumptions. It can also occur that the author expresses him/herself somewhat infelicitously. In this case, counterexamples cannot be as easily dis-

[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).

carded: we have to deal with a genuine counterexample. This may necessitate either strengthening the premises, weakening the conjecture, or even discarding the conjecture altogether. Examining the counterexample(s) provides useful information for deciding between options for refining the theory.

**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.

**What is the domain that the theory describes?** *What do the models of the theory look like?* A model generator can be used to provide candidate models for exploration. This 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. At ALL we use the following:

**Automated theorem provers** are programs designed for finding proofs of conjectures. For the case study of this paper, we used OTTER (McCune 1994b), a resolution-style 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).

**Automated model generators** are programs that can enumerate the finite (small) models of a theory. For the case study of this paper, we used 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 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.

## 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 et al. 1990), Thompson's *Organizations in Action* (Kamps and Pólos 1998), 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 1997).[3] These and most other social science theories are 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 paper: Hage (1965) *An Axiomatic Theory of Organizations*. Although this is not a formal theory in the sense that it uses natural language exclusively, it is an axiomatic theory in which axioms and theorems are clearly outlined.[4] This greatly facilitates

[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 few social scientists state carefully formulated propositions, and even fewer authors attempt to make their underlying assumptions explicit. It is also possible to reconstruct less explicit

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.

### 3.1 HAGE'S AXIOMATIC THEORY OF ORGANIZATIONS

Hage's axiomatic theory of organizations postulates seven axioms based on the theoretical writing of Weber, Barnard, and Thompson and predicts 21 derived theorems. The theory concerns interrelations between eight organizational variables: *complexity*, *centralization*, *formalization*, *stratification*, *adaptiveness*, *production*, *efficiency*, and *job satisfaction*. Hage (1965, p.293) lists two 'indicators' for each of the eight variables:

**Complexity** Number of occupational specialties. Level of training required.

**Centralization** Proportion of jobs that participate in decision making. Number of areas in which decisions are made by decision makers.

**Formalization** Proportion of jobs that are codified. Range of variation allowed within jobs.

**Stratification** Differences in income and prestige among jobs. Rate of mobility between low- and high-ranking jobs.

**Adaptiveness** Number of new programs a year. Number of new techniques a year.

**Production** Number of units produced per year. Rate of increase in units produced per year.

**Efficiency** Cost per unit of output per year. Amount of idle resources per year.

**Job satisfaction** Satisfaction with working conditions. Rate of turnover in job occupants per year.

The first four variables are organizational means, and the second four variables are organizational ends.

Hage (1965) postulates seven axioms that interrelate the eight organizational variables. According to Hage, there are 21 theorems derivable from these seven natural language axioms. Table 1 reprints the 7 axioms and 21 corollaries used in (Hage 1965, p.300). We will reconstruct Hage's theory by formalizing it in first-order logic.

theories, especially if the original authors can be consulted. A case in point is (Péli and Masuch 1997).

### 3.2 A FIRST FORMALIZATION

We represent the eight organizational variables by unary functions, i.e., $\mathsf{comp}(x)$, $\mathsf{cent}(x)$, $\mathsf{form}(x)$, $\mathsf{stra}(x)$, $\mathsf{adap}(x)$, $\mathsf{prod}(x)$, $\mathsf{effi}(x)$, and $\mathsf{jobs}(x)$ respectively. For example, '$\mathsf{cent}(O_1)$' denotes the *centralization* of organization $O_1$. In this way, we can represent the fact that the *centralization* of $O_1$ is higher than that of $O_2$ by '$\mathsf{cent}(O_1) > \mathsf{cent}(O_2)$' (using a strict ordering '$>$'). Now, the fact that *the higher the centralization, the higher the production* (A.1) can be represented by:

$$\forall x, y\ [\ \mathsf{cent}(x) < \mathsf{cent}(y)\ \rightarrow\ \mathsf{prod}(x) < \mathsf{prod}(y)\ ]$$

Table 2 contains a first-order formalization of the axioms in Table 1. We can verify the consistency of the theory by attempting to generate a model of it. MACE can generate a model of F.1 through F.7 using domain size 2 (see Table 3). It is easy to verify that the assumptions hold in this model: F.1–F.3 hold vacuously, F.4–F.6 hold, and F.7 holds again vacuously. Consequently, theory F.1–F.7 is consistent.[5] In fact, MACE can derive 5398 models on a domain of size 2.

Table 2: The Formal Assumptions.

| | |
|---|---|
| **F.1** | $\forall x, y\ [\ \mathsf{cent}(x) < \mathsf{cent}(y)\ \rightarrow\ \mathsf{prod}(x) < \mathsf{prod}(y)\ ]$ |
| **F.2** | $\forall x, y\ [\ \mathsf{form}(x) < \mathsf{form}(y)\ \rightarrow\ \mathsf{effi}(x) < \mathsf{effi}(y)\ ]$ |
| **F.3** | $\forall x, y\ [\ \mathsf{cent}(x) < \mathsf{cent}(y)\ \rightarrow\ \mathsf{form}(x) < \mathsf{form}(y)\ ]$ |
| **F.4** | $\forall x, y\ [\ \mathsf{stra}(x) < \mathsf{stra}(y)\ \rightarrow\ \mathsf{jobs}(x) > \mathsf{jobs}(y)\ ]$ |
| **F.5** | $\forall x, y\ [\ \mathsf{stra}(x) < \mathsf{stra}(y)\ \rightarrow\ \mathsf{prod}(x) < \mathsf{prod}(y)\ ]$ |
| **F.6** | $\forall x, y\ [\ \mathsf{stra}(x) < \mathsf{stra}(y)\ \rightarrow\ \mathsf{adap}(x) > \mathsf{adap}(y)\ ]$ |
| **F.7** | $\forall x, y\ [\ \mathsf{comp}(x) < \mathsf{comp}(y)\ \rightarrow\ \mathsf{cent}(x) > \mathsf{cent}(y)\ ]$ |

Table 3: A Model Of Assumptions **F.1** Through **F.7**.

| | cent | prod | form | effi | stra | jobs | adap | comp |
|---|---|---|---|---|---|---|---|---|
| $O_1$ | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 |
| $O_2$ | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 |

But what about theorems? Hage (1965) derives 21 theorems "...by applying the simple rules of the syllogism" (p.299) to the seven axioms. We use OTTER to test whether the theorems can be soundly derived from the axioms F.1 through F.7. For example, C.2

[5] One can argue that theories consisting solely of universally quantified conditional statements are by definition consistent, since one can construct numerous models in which none of the conditions is satisfied, making all axioms vacuously true. Nevertheless, it is comforting that we can construct a less trivial model of the theory.

Table 1: The Assumptions (**A.1**–**A.7**) And Derived Theorems (**C.1**–**C.21**).

| | |
|---|---|
| **A.1** | The higher the centralization, the higher the production. |
| **A.2** | The higher the formalization, the higher the efficiency. |
| **A.3** | The higher the centralization, the higher the formalization. |
| **A.4** | The higher the stratification, the lower job satisfaction. |
| **A.5** | The higher the stratification, the higher the production. |
| **A.6** | The higher the stratification, the lower the adaptiveness. |
| **A.7** | The higher the complexity, the lower the centralization. |
| **C.1** | The higher the formalization, the higher the production. |
| **C.2** | The higher the centralization, the higher the efficiency. |
| **C.3** | The lower the job satisfaction, the higher the production. |
| **C.4** | The lower the job satisfaction, the lower the adaptiveness. |
| **C.5** | The higher the production, the lower the adaptiveness. |
| **C.6** | The higher the complexity, the lower the production. |
| **C.7** | The higher the complexity, the lower the formalization. |
| **C.8** | The higher the production, the higher the efficiency. |
| **C.9** | The higher the stratification, the higher the formalization. |
| **C.10** | The higher the efficiency, the lower the complexity. |
| **C.11** | The higher the centralization, the lower job satisfaction. |
| **C.12** | The higher the centralization, the lower the adaptiveness. |
| **C.13** | The higher the stratification, the lower the complexity. |
| **C.14** | The higher the complexity, the higher job satisfaction. |
| **C.15** | The lower the complexity, the lower the adaptiveness. |
| **C.16** | The higher the stratification, the higher the efficiency. |
| **C.17** | The higher the efficiency, the lower job satisfaction. |
| **C.18** | The higher the efficiency, the lower the adaptiveness. |
| **C.19** | The higher the centralization, the higher the stratification. |
| **C.20** | The higher the formalization, the lower job satisfaction. |
| **C.21** | The higher the formalization, the lower the adaptiveness. |

states that *the higher the centralization, the higher the efficiency*. A formal version of the second theorem would read:

$$\forall x, y \; [\; \mathsf{cent}(x) < \mathsf{cent}(y) \;\rightarrow\; \mathsf{effi}(x) < \mathsf{effi}(y) \;]$$

This theorem, T.2, is derivable from axiom F.3 and F.2 using the input-file:

```
set(auto).
formula_list(usable).
% F.1
all x y (form(x)<form(y) -> effi(x)<effi(y)).
% F.3
all x y (cent(x)<cent(y) -> form(x)<form(y)).
% negation of T.2
-( all x y (cent(x)<cent(y) -> effi(x)<effi(y)) ).
end_of_list.
```

OTTER's resolution-style proofs require the theorem to be negated in the input-file. If now the theorem (in its original form) is indeed derivable, we will have a contradiction (by *reductio ad absurdum*).

Surprisingly, we can only derive a small fraction of the claimed theorems, namely the three theorems T.2, T.6, and T.7 in Table 4.[6]

Table 4: Theorems Derivable From **F.1** Through **F.7**.

| | |
|---|---|
| **T.2** | $\forall x, y \; [\; \mathsf{cent}(x) < \mathsf{cent}(y) \;\rightarrow\; \mathsf{effi}(x) < \mathsf{effi}(y) \;]$ |
| **T.6** | $\forall x, y \; [\; \mathsf{comp}(x) < \mathsf{comp}(y) \;\rightarrow\; \mathsf{prod}(x) > \mathsf{prod}(y) \;]$ |
| **T.7** | $\forall x, y \; [\; \mathsf{comp}(x) < \mathsf{comp}(y) \;\rightarrow\; \mathsf{form}(x) > \mathsf{form}(y) \;]$ |

Why are we unable to derive the other theorems? Let us examine in detail one of Hage's theorems that we cannot derive. The first theorem discussed in the text is *the higher the centralization, the higher the stratification* (C.19), supposedly derivable using A.1 and A.5 (p.299/300). In our formal set-up this theorem, T.19, would read:

$$\forall x, y \; [\; \mathsf{cent}(x) < \mathsf{cent}(y) \;\rightarrow\; \mathsf{stra}(x) < \mathsf{stra}(y) \;]$$

OTTER cannot derive this theorem (neither from F.1 and F.5, nor from the total set of axioms). Since we fail to prove this conjecture, we can attempt to disprove it.

We use the automated model generator to look for counterexamples, that is, models in which the axioms hold, but the conjecture is falsified. Using MACE we can construct counterexamples to T.19 using input-file (w.l.o.g., we use only F.1 and F.5):

```
set(auto).
formula_list(usable).
% F.1
all x y (cent(x)<cent(y) -> prod(x)<prod(y)).
% F.5
all x y (stra(x)<stra(y) -> prod(x)<prod(y)).
% negation of T.19
-( all x y (cent(x)<cent(y) -> stra(x)<stra(y)) ).
end_of_list.
```

Note that both OTTER (when trying to prove a conjecture) and MACE (when trying to disprove it) use exactly the same input-file!

After running MACE using a domain size 2 (for example `-n2 -p -m10`) we find the four counterexamples of Table 5. In the counterexamples, F.1 is satisfied, F.5 holds vacuously, but the claimed theorem, T.19, is falsified. This proves that the claimed theorem is not derivable. Is the claimed theorem a false conjecture? Or has something gone wrong when we translated the natural language axioms into first-order logic?

Table 5: Counterexamples To **T.19**.

| | cent | prod | stra |
|---|---|---|---|
| $O_1$ | 0 | 0 | 0 |
| $O_2$ | 1 | 1 | 0 |
| $O_1$ | 0 | 0 | 1 |
| $O_2$ | 1 | 1 | 1 |
| $O_1$ | 1 | 1 | 0 |
| $O_2$ | 0 | 0 | 0 |
| $O_1$ | 1 | 1 | 1 |
| $O_2$ | 0 | 0 | 1 |

## 3.3 A SECOND FORMALIZATION

From a logical perspective, three options present themselves:

1. Discard the intended theorem as a false conjecture.
2. Rescue the intended theorem by weakening it sufficiently such that it becomes derivable.
3. Rescue the intended theorem by qualifying these models as an unintended one, and strengthening the premises such that these models are excluded.

Option 1 basically means that we stick to the formalization F.1 through F.7 and limit the explanatory power of the theory from 21 'theorems' to just the three theorems T.2, T.6, and T.7. This seems like an outcome we would like to avoid.

For option 2 we need to transform the counterexamples to T.19 into examples, i.e., we have to weaken

[6] The converse of T.10, $\forall x, y \; [\; \mathsf{comp}(x) < \mathsf{comp}(y) \;\rightarrow\; \mathsf{effi}(x) > \mathsf{effi}(y) \;]$, is also derivable.

the intended theorem such that it does hold for the counterexamples. Let us analyze the counterexamples more precisely. They have the following form (let $O_a$ denote $O_1$ in the first two models and $O_2$ in the second two):

$$\begin{aligned} \mathsf{cent}(O_a) < \mathsf{cent}(O_b) \quad &\wedge \\ \mathsf{prod}(O_a) < \mathsf{prod}(O_b) \quad &\wedge \\ \mathsf{stra}(O_a) = \mathsf{stra}(O_b) \quad & \end{aligned}$$

An obvious way to implement option 2 is to formalize the intended theorem as the weaker *the higher the centralization, the higher **or equal** the stratification*:

$$\forall x, y \; [\; \mathsf{cent}(x) < \mathsf{cent}(y) \;\rightarrow\; \mathsf{stra}(x) \leq \mathsf{stra}(y) \;]$$

This weaker version of the theorem holds in the models of Table 5, turning the former counterexamples into examples. Now, we make a second attempt at proving this (weaker version of the) theorem using OTTER. Note that, although we have dealt with the (type of) counterexamples in Table 5, this gives no guarantee that there are no other counterexamples. There turns out to be none, because OTTER can prove the weaker version of T.19.[7]

This same strategy also works for T.10 and T.13, but not for the remaining other 15 claimed theorems. Consider for example the first conjecture, T.1:

$$\forall x, y \; [\; \mathsf{form}(x) < \mathsf{form}(y) \;\rightarrow\; \mathsf{prod}(x) < \mathsf{prod}(y) \;]$$

MACE generates 12 counterexamples (cardinality 2, w.l.o.g. using only F.1 and F.3), two of which are listed in Table 6.

Table 6: Counterexamples To **T.1**.

| | cent | form | prod |
|---|---|---|---|
| $O_1$ | 0 | 0 | 1 |
| $O_2$ | 0 | 1 | 0 |
| $O_1$ | 0 | 0 | 0 |
| $O_2$ | 0 | 1 | 0 |

These counterexamples have the following forms:

- $\mathsf{form}(O_1) < \mathsf{form}(O_2) \;\wedge\; \mathsf{prod}(O_1) > \mathsf{prod}(O_2)$
- $\mathsf{form}(O_1) < \mathsf{form}(O_2) \;\wedge\; \mathsf{prod}(O_1) = \mathsf{prod}(O_2)$

Moreover, there are also models (of F.1 and F.3) in which the theorem does hold (note that these are no counterexamples but examples). These have the form:

- $\mathsf{form}(O_1) < \mathsf{form}(O_2) \;\wedge\; \mathsf{prod}(O_1) < \mathsf{prod}(O_2)$

There is no relation between the variables of formalization and production: a weaker version of T.1 that holds in all these models, will be a tautology. Pursuing option 2 gives us three additional theorems, i.e., the weaker versions of T.10, T.13 and T.19. Although this doubles the explanatory power of the theory, it remains somewhat doubtful that Hage did overlook counterexamples to the remaining 15 of his 21 theorems.

It may be more reasonable to assume that these counterexamples were not among the models that Hage intended for his theory (option 3). Based on our analysis above, a way to exclude the models that are counterexamples to T.19 is to add the axiom that a higher *production* will imply a higher *stratification* (the converse of F.5):

$$\forall x, y \; [\; \mathsf{prod}(x) < \mathsf{prod}(y) \;\rightarrow\; \mathsf{stra}(x) < \mathsf{stra}(y) \;]$$

After adding this axiom, the models of Table 5 are no longer models of the (modified) theory, making these counterexamples disappear. This is confirmed by OTTER which can now prove the theorem T.19.

One way to view this revision is as adding an axiom to the premise set, but there's another way to view it. We can combine both F.5 and its converse to form F.5′, a revised formalization of axiom A.5:

$$\forall x, y \; [\; \mathsf{stra}(x) < \mathsf{stra}(y) \;\leftrightarrow\; \mathsf{prod}(x) < \mathsf{prod}(y) \;]$$

As a result, we have reformalized the natural language axiom *the higher the stratification, the higher the production* by interpreting it as a bi-implication. This interpretation can be justified considering the ambiguity of the natural language axioms.

This strategy works for all the theorems in (Hage 1965) that were not derivable from F.1 through F.7, causing similar revisions to the other axioms.[8] Table 7 contains the revised first-order formalization of the axioms. Now that we have interpreted all of Hage's axioms as bi-implications we can, using OTTER, derive the all the corollaries that are mentioned in Table 1. Not only can we derive the conditional version of the corollaries, but we can also derive the 21 corresponding bi-implications. For example, OTTER can derive

[7] OTTER requires an axiom expressing that the ordering is strict, for example $\forall x, y \, \neg[\; x < y \;\wedge\; y < x \;]$. In MACE the order "<" is build-in.

[8] The converse of F.1 (because of T.1, 8-9, 13, 16), F.2 (T.10, 17, 19), F.3 (T.10, 17-18, 20-21), F.4 (T.3-4), F.5 (T.5, 11-12, 14-15, 17-21), and F.7 (T.10, 13) are necessary for the derivation of theorems. The converse of F.6 is not! We decide to include the converse of F.6 in order to give similar interpretations of all natural language axioms.

Table 7: The Revised Formal Assumptions.

| | |
|---|---|
| **F.1′** | $\forall x, y$ [ $\mathsf{cent}(x) < \mathsf{cent}(y) \leftrightarrow \mathsf{prod}(x) < \mathsf{prod}(y)$ ] |
| **F.2′** | $\forall x, y$ [ $\mathsf{form}(x) < \mathsf{form}(y) \leftrightarrow \mathsf{effi}(x) < \mathsf{effi}(y)$ ] |
| **F.3′** | $\forall x, y$ [ $\mathsf{cent}(x) < \mathsf{cent}(y) \leftrightarrow \mathsf{form}(x) < \mathsf{form}(y)$ ] |
| **F.4′** | $\forall x, y$ [ $\mathsf{stra}(x) < \mathsf{stra}(y) \leftrightarrow \mathsf{jobs}(x) > \mathsf{jobs}(y)$ ] |
| **F.5′** | $\forall x, y$ [ $\mathsf{stra}(x) < \mathsf{stra}(y) \leftrightarrow \mathsf{prod}(x) < \mathsf{prod}(y)$ ] |
| **F.6′** | $\forall x, y$ [ $\mathsf{stra}(x) < \mathsf{stra}(y) \leftrightarrow \mathsf{adap}(x) > \mathsf{adap}(y)$ ] |
| **F.7′** | $\forall x, y$ [ $\mathsf{comp}(x) < \mathsf{comp}(y) \leftrightarrow \mathsf{cent}(x) > \mathsf{cent}(y)$ ] |

T.19′:

$$\forall x, y \; [\; \mathsf{cent}(x) < \mathsf{cent}(y) \;\leftrightarrow\; \mathsf{stra}(x) < \mathsf{stra}(y) \;]$$

Also the theorems, formulated in the same way as the axioms, can be interpreted as bi-implications. This result provides some confidence for the interpretation in Table 7 (and for the choice to regard the counterexamples as non-intended models of the theory).[9]

Using F.1′ through F.5′, Mace generates 258 models on a domain of size 2 (one of which is shown in Table 8). Consequently, the revised formalization is

Table 8: A Model Of Assumptions **F.1′** Through **F.7′**.

| | cent | prod | form | effi | stra | jobs | adap | comp |
|---|---|---|---|---|---|---|---|---|
| $O_1$ | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| $O_2$ | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |

still a consistent theory. Inspecting this model, we can confirm that the theorems are satisfiable; for example, it is a model of theorem T.19′. We can also check whether theorems are falsifiable by constructing a model in which the theorem does not hold (disregarding the premises). For example, the models in Table 5 still prove that theorem T.19′ is falsifiable. Note that these models are necessarily no longer models of the revised axioms, otherwise they would still be counterexamples to the theorem. Since T.19′ is both satisfiable and falsifiable, the theorem is contingent—it is neither a tautology nor a contradiction.

We can explore the theory's domain by examining its models. Interestingly, in 256 of the 258 models of F.1′ through F.5′ with cardinality 2, *all* axioms hold vacuously (because each function is equal for both elements of the domain). For each of the eight functions there are 2 options to be equal, both zero or both one, giving 256 ($= 2^8$) models. As soon as one of the functions is unequal, all functions are unequal. This results in a model as depicted in Table 8 (and the isomorphic copy with $O_1$ and $O_2$ interchanged being the remaining last model).

If one interprets the 0 as 'low' and the 1 as 'high' then this model represents to the 'two ideal types' of organization that are discussed on (Hage 1965, p.304-307). One 'ideal type' is an 'organic model' that has high complexity, low centralization, low formalization, low stratification, high adaptiveness, low production, low efficiency, and high job satisfaction (corresponding to $O_1$ in the model of Table 8). This organization type emphasizes adaptiveness. The other 'ideal type' is a 'mechanistic model'. This opposite of the 'organic model' has low complexity, high centralization, high formalization, high stratification, low adaptiveness, high production, high efficiency, and low job satisfaction (corresponding to $O_2$ in the model of Table 8). This organization type emphasizes production.

### 3.4 SUMMARIZING

In order to derive the claimed theorems, we had to interpret the natural language assumptions as logical bi-implications. On the one hand, this rescues the theory: the intended theorems are soundly derivable, the theorems are contingent (both satisfiable and falsifiable), and the theory is consistent. On the other hand, this trivializes the theory: now all eight functions become indistinguishable.[10] In short, one has to conclude that the axioms of (Hage 1965) are too strong. Nevertheless, the attempt that Hage undertook, i.e., to construct a general, axiomatic theory of organizations, remains an important enterprise. Any effort to formulate some of the basic axioms of organization theory, should take Hage's attempt into account.

A critical analysis of Hage's theory is unfair without noting the incomparability of his and our positions. Science has progressed significantly since the sixties: much more is known about formal logics and about their application; automated reasoning tools are operational and are valuable research assistants; and also the social sciences have advanced, e.g., empirical data have become available. This has led to renewed interest in building axiomatic, formal theories in the social sciences.

[9] Jerry Hage later confirmed that the axioms and theorems should indeed be interpreted as biconditionals, and mentioned that he explicitly included the words "and *vice versa*" in later references to the theory. Note that without adding the converse of F.6 the theorems T.4-5, 12, 15, 18, 21 can only be derived as conditionals.

[10] From a strictly formal point of view, one would need a trichotomy axiom for each of the eight functions to ensure that values for each function are comparable. For example, $\forall x, y$ [ $\mathsf{cent}(x) < \mathsf{cent}(y) \lor \mathsf{cent}(x) = \mathsf{cent}(y) \lor \mathsf{cent}(x) > \mathsf{cent}(y)$ ].

Table 9: Theoretical Criteria And Automated Reasoning Tools.

| CRITERION | THEOREM PROVER | MODEL GENERATOR |
|---|---|---|
| Consistency | | × |
| Inconsistency | × | |
| Soundness | × | |
| Unsoundness | | × |
| Falsifiability | | × |
| Unfalsifiability | × | |
| Contingency | | × |
| Noncontingency | × | |
| Domain | | × |

# 4 CONCLUSIONS AND DISCUSSION

In this paper we outlined a theory building methodology that is 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, such as consistency, soundness, falsifiability, and contingency (see Table 9). 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. However, none of the proofs and models searches for the case study in Section 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 cardinality). Still, current automated model generators are powerful enough to have solved several open problems in (finite) mathematics (Slaney 1994).

We started this paper by referring to the positivist heritage that we share—the logical analysis for evaluating and justifying scientific theories. However, the use of the tools goes beyond a rigid, final justification of theories. We use them extensively during the process of formalization. As a result, we also enter the context of discovery: during the formalization process, we will repeatedly refine the (formal) theory. This makes our methodology more in line with more recent philosophy of science, in particular with (Lakatos 1976). Lakatos gives a logical analysis of the development of theories over time, with which the case study of section 3 shows remarkable resemblance (especially his classroom dialogue attempting to prove the *polyhedron* conjecture).

In our experience, the tools are especially useful during the process of formalizing a theory—intermediate versions of a theory under construction are more likely to contain logical flaws. When formalizing a larger theory, a short lapse of attention may result in an inconsistent theory. Proving consistency by generating a model can be a fast and easy safeguard against such an unfortunate event. Moreover, on many occasions (especially in the social sciences), theorems cannot be derived because some background knowledge is missing. The need for such an assumption can become clear immediately by examining the counterexamples. It is often difficult to find such non-intended models by hand, because they conflict with our common sense.

**Acknowledgements**

Thank you to Jeroen Bruggeman, Jerry Hage, Michael Masuch, Breanndán Ó Nualláin, and László Pólos. This research was in part supported by the Netherlands Organization for Scientific Research (NWO grant # PGS 50-334, # 030-50-334, and # SIR 12-2918). Part of the material in section 3 was presented at the IJCAI'97 Workshop on Model-Based Automated Reasoning. Address correspondence to: Jaap Kamps, Applied Logic Laboratory, Universiteit van Amsterdam, Sarphatistraat 143, 1018 GD Amsterdam, The Netherlands, Tel. +31-20 5252538, Fax. +31-20 5252800, Email: ⟨`kamps@ccsom.uva.nl`⟩.

## References

Adorno, T. W., R. Dahrendorf, H. Pilot, H. Albert, J. Habermas, and K. R. Popper (1969). *Der Positivismusstreit in der deutchen Soziologie*. Hermann Luchterhand Verlag, Darmstadt und Neuwied.

Ayer, A. J. (Ed.) (1959). *Logical Positivism*, The library of philosophical movements. The Free Press, New York.

Bruggeman, J. (1997). Niche width theory reappraised. *Journal of Mathematical Sociology 22*(2), 201–220.

Bundy, A. (Ed.) (1994). *Automated Deduction — CADE-12*, Volume 814 of *Lecture Notes in Artificial Intelligence*. Springer-Verlag, Berlin.

Glorie, J. C., M. Masuch, and M. Marx (1990). Formalizing organizational theory: A knowledge-based approach. In M. Masuch (Ed.), *Organization, Management, and Expert Systems: models of automated reasoning*, Chapter 4, pp. 79–81. De Gruyter, Berlin, New York.

Hage, J. (1965). An axiomatic theory of organizations. *Administrative Science Quarterly 10*(3), 289–320.

Hannan, M. T. (1997). Rethinking age dependence in organizational mortality: Logical formalizations. Technical report, Stanford University.

Kamps, J. and L. Pólos (1998). Reducing uncertainty: A formal theory of *organizations in action*. CCSOM Working Paper 98-167.

Lakatos, I. (1976). *Proofs and Refutations. The logic of mathematical discovery*. Cambridge University Press, Cambridge, England.

McCune, W. (1994a). A Davis-Putnam program and its application to finite first-order model search: Quasigroup existence problems. Technical report, Argonne National Laboratory, Argonne IL. DRAFT.

McCune, W. (1994b). OTTER: Reference manual and guide. Technical Report ANL-94/6, Argonne National Laboratory, Argonne IL.

Péli, G. (1997). The niche hiker's guide to population ecology: A logical reconstruction of organizational ecology's niche theory. In A. E. Raftery (Ed.), *Sociological Methodology 1997*, pp. 1–46. Blackwell, Oxford UK.

Péli, G., J. Bruggeman, M. Masuch, and B. Ó Nualláin (1994). A logical approach to formalizing organizational ecology. *American Sociological Review 59*(4), 571–593.

Péli, G. and M. Masuch (1997). The logic of propagation strategies: Axiomatizing a fragment of organizational ecology in first-order logic. *Organization Science 8*(3), 310–331.

Popper, K. R. (1959). *The Logic of Scientific Discovery*. Hutchinson, London.

Popper, K. R. (1969). Die Logik der Sozialwissenschaften. See Adorno, Dahrendorf, Pilot, Albert, Habermas, and Popper (1969), pp. 103–123.

Rudner, R. S. (1966). *Philosophy of Social Science*. Foundations of Philosophy Series. Prentice-Hall, Englewood Cliffs NJ.

Slaney, J. (1994). The crisis in finite mathematics: Automated reasoning as cause and cure. See Bundy (1994), pp. 1–13.

Suppes, P. (1968). The desirability of formalization in science. *Journal of Philosophy LXV*(20), 651–664.

Sutcliffe, G., C. Suttner, and T. Yemenis (1994). The TPTP problem library. See Bundy (1994), pp. 252–266.

Tarski, A. (1956). *Logic, Semantics, Metamathematics*. Oxford University Press, New York.