ASSESSING THEORIES

The Problem of a Quantitative Theory of Confirmation

A DISSERTATION

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY (DR. PHIL.)

OF THE
PHILOSOPHICAL FACULTY OF THE

UNIVERSITY OF ERFURT, GERMANY

By

Franz Huber

Erfurt 2002

First Reader: Prof. Dr. Gerhard Schurz
(Universities of Erfurt and Düsseldorf)

Second Reader: Prof. Dr. Alex Burri
(University of Erfurt)

Date: . . . .

urn:nbn:de:gbv:547-200400640
http://nbn-resolving.de/urn/resolver.pl?urn=nbn%3Ade%3Agbv%3A547-200400640

• • •

Acknowledgements

The list of people I would like to thank on this occasion is long, and I can name only a few of them.

First of all, I thank Prof. Dr. Gerhard Schurz under whose supervision this dissertation has been written, and to whom I have been assistant for two years at the University of Erfurt.

Most of what I have learned during my studies in Salzburg and Erfurt is the result of visiting his inspiring lectures and reading some of the numerous clearly written and profound articles covering most, if not all topics of today's analytic philosophy, in particular, the philosophy of science of the day.

From the close and fruitful cooperation with him I have profited very much, and his remarks showed me the direction in which to proceed. I am especially indebted to him for his great involvement in my work and the many discussions on the topic of confirmation and related issues which provided the basis for this dissertation. The latter would not have come about in its present form without his decisive suggestions, though he is, of course, not responsible for any of the mistakes I have made here or elsewehere.

Furthermore I would like to thank Prof. Dr. Alex Burri, who has been reader of this dissertation, and whose accomodating friendliness always ensured that things could proceed in the way it was best.

Special thanks go to Prof. Luc Bovens for discussing in length many of my ideas and relating them to probability theory, to Prof. Kenneth Gemes for his readiness to help and cooperate, and to Prof. Donald Gillies for his interest in and suggestions to some of my thoughts.

During my studies in Salzburg, I have also learned a lot from Prof.es Johannes Czermak, Georg J.W. Dorn, and, most notably, Paul Weingartner.

Since September 2002 I have been research fellow in the working group "Philosophy, Probability, and Modeling" led by Luc Bovens and Stephan Hartmann. The group is part of the Center for Junior Research Fellows at the University of Konstanz, and is supported by the Alexander von Humboldt Foundation, the Federal Ministry of Education and Research, and the Program for the Investment in the Future (ZIP) of the German Government through a Sofja Kovalevskaja Award. I am grateful to the latter for a grant enabling me to finish this dissertation – and to my colleagues in the group for providing such an inspiring atmosphere (an impression of which one may get at: www.uni-konstanz.de/ppm).

My personal thanks are first, and foremost, due to Anna Hülsmann, who knows more than anyone else how many hours it took me to write this dissertation, and whose constant love provided the kind of support and encouragement I needed for finishing it.

Among my friends and colleagues Karoline Krenn mostly deserves being mentioned for her persistent friendship – and many hours on the telephone.

Last, but not least, I may thank my parents and my grandmother, who have enabled my studies, and who have given me support throughout the years.

Abstract

This dissertation deals with the problem of a quantitative theory of confirmation. The latter can be sketched as follows: You are given a theory T, an evidence E, and a background knowledge B. The question is how much does E confirm T relative to B. A solution consists in the definition of a function C such that $C(T, E, B)$ measures the degree to which E confirms T relative to B.

In chapter 1 I make precise what is meant by a theory, an evidence, and a background knowledge. Next comes a chapter on *formal* conditions of adequacy for any formal theory (not only of confirmation): A formal theory has to be non-arbitrary, comprehensible, and computable in the limit. Chapter 2 closes with a critical remark on Bayesian confirmation theory.

In chapter 3 I list a set of *material* conditions of adequacy for any quantitative theory of confirmation: A measure of confirmation has to be sensitive to (and only to) the *confirmational virtues*.

These give rise to two strategies of solving the problem under consideration: The first is to argue that there is one distinguished property of theories in relation to evidences and background knowledges that takes into account all (and only) the confirmational virtues. The candidate here is coherence with respect to the evidence, which is discussed in chapter 4 on foundationalist coherentism. This approach is found to be unsuccessful.

The second strategy is first to define for every confirmational virtue V a function f_V such that $f_V(T, E, B)$ measures the extent to which V is exhibited by theory T, evidence E, and background knowledge B; and then to define the measure of confirmation C as a function of (some of) the functions f_V.

In chapter 5 it is argued that this strategy is successful. In a nutshell, it is observed that there are two conflicting concepts of confirmation, viz. loveliness and likeliness. I reason that it suffices to consider these two *primary* confirmational virtues. The two main approaches to confirmation are Hypothetico Deductivism and probabilistic theories of confirmation: The former is based on loveliness, whereas the focus of the latter is on likeliness. The idea is simple: Combine these two aspects, keep their merits, get rid of their drawbacks.

Chapter 6 is on evidential diversity, more generally: the goodness of the evidence. A goodness measure is defined which together with the loveliness-likeliness measure gives rise to the refined measure of confirmation C^*. C^* can

answer the question why scientists (should) gather evidence, and it provides a solution to the ravens paradox.

Contents

Chapter 1

Introduction

1.1 The Problem of a Quantitative Theory of Confirmation

The following monograph deals with the problem of a quantitative theory of confirmation of theory T by evidence E relative to background knowledge B. The latter may be sketched as follows:

You are given a theory T, an evidence E, and a background knowledge B, and you want to know how much E confirms T relative to B.

A solution to this problem consists in the definition of a (set of) function(s) C such that $C(T, E, B)$ measures the degree to which T is confirmed by E relative to B, for every theory T, every evidence E, and every background knowledge B.[1]

In order for this characterisation to be precise, one first has to make clear what is meant by a theory T, an evidence E, and a background knowledge B.

[1] If C is a set of functions, then this has to hold for every function $c \in C$. Corresponding to the problem of a quantitative theory of confirmation there is the problem of a qualitative theory of confirmation a solution to which consists in the definition a (set of) function(s) C such that

$$C(T, E, B) = \begin{cases} 1, & \text{if } E \text{ confirms } T \text{ relative to } B, \\ 0 & \text{otherwise,} \end{cases}$$

for every theory T, every evidence E, and every background knowledge B. In a similar way one may characterise the problem of a comparative theory of confirmation. I will only be concerned with the problem of a quantitative theory of confirmation a solution to which automatically is a solution to the problems of a comparative and of a qualitative theory of confirmation.

Before doing so let me stress that T, E, and B are all one is given. In particular, it is not assumed that one is equipped with some degree of belief function p (defined over some language containing T, E, and B) that could be used in determining the degree of confirmation $C(T, E, B)$. $C(T, E, B)$ has to be *squeezed out* of the logical structure of T, E, and B alone!

Squeezing is one of the most important features distinguishing the present approach from probabilistic theories of confirmation (not only Bayesian ones in the sense of Gillies 1998). In addition to T, E, and B, they all assume the existence of some probability function p defined over some language containing T, E, and B, which is then used to determine $C(T, E, B)$. Squeezing does not even hold of a *logical* probability function p, if the values of latter depend not only on T, E, and B, but on the *whole* language containing them.

In contrast to this, the values of the measure of confirmation C defined later on are the same for *any* language containing T, E, and B, provided the language is rich enough in order to express those features of T, E, and B that determine $C(T, E, B)$. This is the case for any predicate language with the identity sign, but for no propositional language. The latter are simply too poor in order to express the relevant information.

This phenomenon is not new, but is familiar from Quine's *On What There Is* (1948). His slogan "To be is to be the value of a (bound) variable"[2] does not make sense within the framework of propositional logic. That identity is also needed is known from another slogan – "No entity without identity" (cf. Quine 1958).

I note this, because the mentioned feature may be taken as an argument against the adequacy of C: After all, the propositional calculus PC is contained in $PL1 =$ in the sense that every statement which is logically true in the sense of PC is logically true in the sense of $PL1 =$; and if C cannot deal with the simple case, how should it be able to deal adequately with the more general case.

1.2 Theory

In general, there are two positions concerning the question of what a theory is.

The semantic position defines a theory T as the set of all models $\mathcal{M} = \langle Dom, \varphi \rangle$ such that A_T is true in $\mathcal{M}$, where A_T is some axiomatization (formulation) of T. $Dom = \langle D_1, \ldots, D_r \rangle$ is the *domain*, where each $D_i, 1 \leq i \leq r$, is a set of entities of some sort, and φ is an *interpretation function*.

[2] As noted by Prof. Paul Weingartner, one may want to add: first-order variable.

If A_T is a set of statements or wffs of the language $\mathcal{L}_{PL1=}$ of first-order predicate logic with identity (including function symbols), $PL1 =$, then for every set D_i there are denumerably many i-variables 'v_1^i', 'v_2^i', $\ldots$ and corresponding i-constants 'c_1^i', 'c_2^i', ... φ assigns each i-constant 'c_j^i' of $\mathcal{L}_{PL1=}$ an entity φ ('c_j^i') $\in D_i$, each n-ary ($k_1, \ldots, k_n$-) predicate 'P' = '$P\left(x^{k_1}, \ldots, x^{k_n}\right)$', $1 \leq k_l \leq r$, of $\mathcal{L}_{PL1=}$ a subset φ ('P') $\subseteq D_{k_1} \times \ldots \times D_{k_n}$, and each n-ary ($k_1, \ldots, k_n, k_{n+1}$-) function symbol '$f_j^{k+1}$' = '$f_j^{k_{n+1}}\left(x^{k_1}, \ldots, x^{k_n}\right)$', $1 \leq k_{n+1} \leq r$, of $\mathcal{L}_{PL1=}$ a function φ ('$f_j^{k_{n+1}}$'),

$$\varphi\left(\text{`}f_j^{k_{n+1}}\text{'} \right) : D_{k_1} \times \ldots \times D_{k_n} \rightarrow D_{k_{n+1}}.$$

According to the syntactic position a theory T is a set of statements A_T that formulates or axiomatizes T.

Personally I think that the semantic conception is more in accordance with our intuitive understanding of a theory, but for the purposes of a theory of confirmation the question of how a theory T has to be interpreted does not arise, if one takes it to be a *sine qua non* that an adequate measure of confirmation C is to be closed under equivalence transformations of T in the sense that

$$C\left(A_T, E, B\right) = C\left(A'_T, E, B\right), \quad \text{if} \quad A_T \dashv\vdash A'_T,$$

for any two axiomatizations A_T and A'_T of any theory T, every evidence E, and every background knowledge B. For then it must not matter how a theory T – syntactically construed as a set of statements – is formulated.[3] If, however, one takes the position that the way a theory T is formulated may matter, and that the values $C\left(A_T, E, B\right)$ of C for a given axiomatization A_T of T, a given evidence E, and a given background knowledge B may differ for different formulations A_T of T, one is forced to consider a theory T as a set of statements A_T.

Since none of these two positions should be ruled out right from the start, I will take a theory T to be a set of statements or wffs. If the measure of confirmation C turns out to be closed under equivalence transformations of T in the above

[3]Though one can, of course, make a function $f\left(\ldots, T, \ldots\right)$ invariant under equivalence transformations of T by recourse to some uniquely determined formulation A_T of T, and by defining

$$f\left(\ldots, T, \ldots\right) = f\left(\ldots, A_T, \ldots\right),$$

or by taking, say, the maximum function, and by defining

$$f\left(\ldots, T, \ldots\right) = \max\left\{f\left(\ldots, T', \ldots\right) : T' \dashv\vdash T, T' \subseteq \mathcal{L}_{PL1=}\right\}.$$

sense, then both the semantic and the syntactic interpretation are allowed for – and because of the adequacy of $PL1 =$ (with respect to its standard semantics) one can still interpret the set of statements T semantically as the set of models $mod\,(T)$. If, however, it turns out that $C\,(A_T, E, B)$ – in order to be an adequate measure of confirmation – is to be sensitive to the way T is formulated by A_T, then T must be interpreted syntactically as a set of statements A_T. So considering T as a set of statements does not rule out any of the above mentioned positions concerning the definition of theories, and thus does not put any restrictions on the behaviour of an adequate measure of confirmation C.

In speaking of theories I always mean *scientific* theories. I do not attempt to define these, but restrict myself to giving a necessary condition. Before doing so let us have again a look at the structure the models of a theory have according to the semantic position[4].

These consist of sequences of the form

$$\langle D_1, \ldots, D_r, R_1, \ldots, R_s \rangle \,,$$

where the $D_i, 1 \leq i \leq r$, are sets of entities which settle the ontology of theory, and the $R_j, 1 \leq j \leq s$, are relations among the objects in the sets D_i. The latter have been summarized as $Dom = \langle D_1, \ldots, D_r \rangle$, and the relations R_j have been subsumed under the interpretation function φ, because I am considering sets of statements A_T formulating theories T, and my interest is in the syntactical pendant of the sets D_i the domain Dom consists of.

Consider a theory of, say, physics. Here the domain Dom will consist of four sets: A set of material objects D, a set of space points S (usually $\Re^3$), a set of time points T (usually $\Re$), and a set of numbers R (usually $\Re$). The relations R_j among the objects in the sets D, S, T, and R need not concern us here.

The question of interest is: Which are the entities the theory is properly talking about? I think the natural answer is that it are the material objects in D about which the theory of physics is making claims. The space points in S, the time

[4]Though what I call the semantic position is usually called *structuralism*, I prefer the former term, because all I am concerned with here is the ontological skeleton of a theory, but not the many other questions normally associated with structuralism, as, for instance, T-theoreticity, the focus on constraints, links, and admissible blurs, or the distinction between the models, the potential models, and the partial potential models of a theory. For an introduction to the basic ideas of structuralism cf. Moulines (1996) or Balzer/Moulines (2000). For a critical discussion of T-theoreticity cf. Schurz (1990); a summary of recent developments including a defense against Schurz's criticism can be found in Balzer (1996). Another critical discussion of the structuralist position concerning theoretical terms is contained in Zoglauer (1993).

points in T, and the numbers in R are not the things physicists are investigating. No physicist will entertain an experiment in order to test a mathematical equation.

This gives rise to a distinction between the sets D_i containing the objects of proper investigation, those things about which claims are made by the theory – for the physical theory these are the material objects – and the remaining sets $D_{i'}$ – in physics, the sets of mathematical entities representing the space and time points, and the set of real numbers constituting the range of the functions among the relations R_j. Let us call the former *domains of proper investigation*.

Admittedly, there may cases where this classification of the sets D_i the domain Dom consists of is difficult to draw. The term I have chosen may also be misleading, for if one is concerned with historical claims such as Cesar won all wars he entertained or Cleopatra seduced all men she wanted to, then the domain of proper investigation consists of wars in the former case and of people (Cleopatra possibly could have wanted to seduce) in the latter – *ex*cluding Cleopatra herself! – though Cesar and Cleopatra may justifiedly be called the objects of proper investigation here.

Furthermore, one may question the epistemic significance of this distinction, and consider it a mere formal manipulation only making things more complicated. Nevertheless I propose that in considering a theory one should single out some set(s) of proper investigation. The reason for doing so will become more clear when the idea underlying the present proposal is presented. Roughly speaking, the latter consists in considering how many objects of the domains of proper investigation of the theory in question the evidence reports about, and how many of them confirm the theory.

As mentioned before, the syntactic pendant of the sets of entities D_i are the i-variables 'v_1^i', $\ldots$ and the i-constants 'c_1^i', $\ldots$. Since I am dealing with sets of statements A_T formulating the theories T under consideration, the assumption concerning scientific theories is expressed in terms of these.

Assumption 1.1 (Finite Axiomatizability Without Constants) If T is a scientific theory with domain $Dom_T = \langle D_1, \ldots, D_r \rangle$, and $D_{k_1}, \ldots, D_{k_n}$ as its domains of proper investigation, $1 \leq k_l \leq r$, for every $l, 1 \leq l \leq r$, then there is at least one finite axiomatization A_T of T without occurrences of k_l-constants, but with at least one essential occurrence of a k_l-variable, for every $l, 1 \leq l \leq r$.

Any such A_T is called a *wff-ication of* T.

If, for a given theory T, the domains of proper investigation cannot be specified in advance, then one may take recourse to the following definition.

Definition 1.1 (Domain of Proper Investigation) Let T be a scientific theory, and let D_i be a set of entities with corresponding i-variables and i-constants in $\mathcal{L}_{PL1=}$.

D_i is a *domain of proper investigation of* T iff there is at least one finite axiomatization A_T of T with at least one essential occurrence of an i-variable, and without occurrences of i-constants.

Definition 1.2 (Finite Axiomatization) Let T be a theory, and let A_T be a set of wffs, $A_T \subseteq \mathcal{L}_{PL1=}$[5]. A_T is a(n) (*finite*) *axiomatization of* T iff (A_T is finite, and)

1. $A_T \dashv\vdash T$, if T is a set of wffs, and
2. $T = mod\,(A_T)$, if T is a set of models.

Definition 1.3 (Essential Occurrence of a Variable) Let h be a wff[6], and let 'x_j^i', $j \geq 1$, be an i-variable. h contains *at least one essential occurrence of an i-variable* iff it holds for every wff h':

If $h \dashv\vdash h'$, then h' contains at least one occurrence of an i-variable.

The clause that A_T contains at least one essential occurrence of an i-variable should avoid that every set of entities D_i which is redundant or not among the sets $D_1, \ldots, D_r$ the domain Dom_T consists of is a domain of proper investigation of T.

Please note that in the example of before, the set of space points S and the set of time points T may come out as domains of proper investigation according to the above definition. This is as it should be, for confirmation is domain-relative, and with a suitable evidence one may perhaps confirm a theory of physics by investigating various space points.[7]

Theorem 1.1 (Domains of Proper Investigation) Let T be a scientific theory with domain $Dom_T = \langle D_1, \ldots, D_r \rangle$ and $D_{k_1}, \ldots, D_{k_n}$ as its domains of proper investigation, $1 \leq k_l \leq r$, for every $l, 1 \leq l \leq n$.

Then there is at least one finite axiomatization A_T of T with at least one essential occurrence of a k_l-variable, and without occurrences of k_l-constants, for every $k_l, 1 \leq l \leq n$.

[5]Here and in the following the language $\mathcal{L}_{PL1=}$ of first-order predicate logic with identity (including function symbols), $PL1 =$, is identified with the set of its well-formed formulas.

[6]If not specified otherwise, a wff is always meant to be a wff of the language $\mathcal{L}_{PL1=}$ of first-order predicate logic with identity (including function symbols), $PL1 =$.

[7]Confirmation or, more generally, assessment by means of the set of space points S may be appropriate for the hypothesis that the gravitational force is acting everywhere.

Assumption 1.1 is plausible with regard to the following two positions.

First, theories have to be finitely axiomatizable in order to be such that they can be put forth by some scientist and can be contemplated by us – and only theories of this kind are of interest in the sciences.

Second, scientific theories consist of lawlike statements, and these do not, among others, speak about particular entities of their domains of proper investigation, but express general regularities or patterns.

Please note that this assumption allows for that a theory contains constants for particular entities as e.g. constants of nature. For example, a theory of physics may well contain occurrences of constants for space points, time points, or (real) numbers, as is the case, for instance, with Galilei's law which contains occurrences of the gravitational constant 'g'.[8] In the same way a hypothesis about some particular historical person or event may contain occurrences of a constant denoting the person or event in question.[9]

Let me stress that I do not claim that every statement without occurrences of constant terms is lawlike. All I claim is that containing no (essential) occurrences of k_l-constants, $1 \leq l \leq n$, is a necessary condition for a statement to be lawlike in the sense of some theory T, where $D_{k_1}, \ldots, D_{k_n}$ are the domains of proper investigation of T.[10]

In the following a theory T will be identified with one of its by the above assumption existing finite axiomatizations A_T without occurrences of k_l-constants, where $D_{k_1}, \ldots, D_{k_n}$ are the domains of proper investigation of T. The set of all wff-ications A_T of any theory T is denoted by '$\mathcal{T}$'. Although assumption 1.1 is only a *necessary* condition for scientific theories, finite sets of statements without occurrences of i-constants, but with at least one essential occurrence of an i-variable, $1 \leq i \leq n$, are often called theories with domains of proper investigation $D_1, \ldots, D_n$.

---

[8]Here and on many other places I have profited very much from the discussions with my supervisor Prof. Dr. Gerhard Schurz. Though his influence on this dissertation is enormous, he is, of course, not responsible for any of the views expressed here.

[9]We do not claim that there are – or are not – laws of history. Their existence or non-existence has no impact on the questions discussed here.

[10]For more on the nature of a law of nature cf. the *locus classicus* Armstrong (1983). More recent monographs are Harré (1993) and Carroll (1994).

1.3 Evidence

I take the evidence E by which a given theory T is to be assessed relative to some background knowledge B to report our (uncontrolled) observations and the results of our (controlled) experiments. Because of the fact that

> we are damned *qua humans* to be able to examine only finitely many entities, and to describe these in only finitely many statements of finite length

the following *preliminary* assumption is plausible.

Preliminary Assumption 1.1 (Strong Finitism in the Evidence) If E is an evidence, then E is a *finite* set of wffs of *finite* length talking about *finitely* many entities.

A consequence of this is that quantifiers can be eliminated, for these are only necessary in order to speak about infinitely many entities.

With regard to the preceding section it seems appropriate to distinguish different "kinds" of evidences. Roughly speaking, these different kinds are determined by the sorts of entities an evidence reports about, i.e. the domains these entities are taken from. For instance, an evidence gathered by a physicist will report data about material objects, whereas an evidence gathered by an ornithologist will report data about the much narrower class of birds, and an evidence gathered by a psychologist may report data about such entities as neuroses. Finally, a historian's evidence perhaps reports about the wars Cesar entertained or the people Cleopatra wanted to seduce – though data about these entities cannot, of course, be directly observed but only inferred.

Yet an evidence may contain a statement to the effect that for all time points t after some given point of time t_0, some special event e, say the soccer championships in Japan and South Corea are over, $\forall t\,(t \geq t_0 \rightarrow O\,(e, t))$. Such a statement contains (essential) occurrences of quantifiers and time variables, and so possibly speaks about infinitely many time points. Strong Finitism in the Evidence does not allow for such statements to occur in an evidence.

It thus seems reasonable to relativize the above preliminary assumption to the sets D_i of entities data about which are reported by the evidence. I will therefore speak of an *evidence from the sets of entities* $D_1, \ldots, D_k, k \geq 1$, or of *data about the entities in* $D_1, \ldots, D_k$.

Assumption 1.2 (Finitism in the Evidence) If E is an evidence from $D_1, \ldots, D_k, k \geq 1$, then E is a finite set of wffs of finite length speaking about finitely many entities in $\bigcup_{1 \leq i \leq k} D_i$.

As noted before, this has the consequence that (quantifiers binding) variables which range over the sets of entities $D_1, \ldots, D_k$ can be eliminated. In order to avoid triviality it is furthermore assumed that E speaks about at least one entity (in the sense of containing at least one essential occurrence of an i-constant) from every set $D_i, 1 \leq i \leq k$ – otherwise every finite set of statements of finite length is an evidence for every set of entities except those which the variables occurring in E range over.

As stated above, the evidence E by which a given theory T is assessed relative to some background knowledge B is supposed to report our (uncontrolled) observations and the results of our (controlled) experiments. It may be that the language we use in describing these observations is not rich enough in order to express all the nuances of our observations – indeed, this is quite plausibly the case. Therefore I have to make an assumption possibly restricting the applicability of the present account: It is supposed that parts of our observations can be described in the language $\mathcal{L}_{PL1=}$ of standard standard first-order predicate logic with identity (including function symbols), $PL1 =$, and that these parts are large enough to contain all relevant aspects of our observations for the assessment of a given theory T relative to some background knowledge B. If they do not, this is, of course, a limitation; but note: this is a general problem and no specific one besetting only the approach presented here.

Assumption 1.3 (Expressability) The language $\mathcal{L}_{PL1=}$ of standard first-order predicate logic with identity (including function symbols), $PL1 =$, is rich enough in order to express all aspects of our observations that are relevant for the assessment of a given theory T relative to some background knowledge B.

Assumptions 1.2 and 1.3 give rise to the following definition of an evidence from the sets of entities $D_1, \ldots, D_k$.

Definition 1.4 (Evidence from $D_1, \ldots, D_k$) Let E be a set of wffs of $\mathcal{L}_{PL1=}$, let $D_1, \ldots, D_k, k \geq 1$, be sets of entities, let 'x^i_j' be the corresponding i-variables ranging over D_i, and let 'c^i_j' be the corresponding i-constants denoting entities of $D_i, j \geq 1, 1 \leq i \leq k$.

E is an *evidence from* $D_1, \ldots, D_k$ iff E is a finite set of wffs (of finite length) of $\mathcal{L}_{PL1=}$ such that it holds for every $i, 1 \leq i \leq k$: E contains at least one essential occurrence of an i-constant, but no occurrence of an i-variable.

If E is an evidence from $D_1, \ldots, D_k$, $D_1, \ldots, D_k$ are the (*evidential*) *domains of* E.

The set of all evidences from $D_1, \ldots, D_k$ is denoted by '$\mathcal{E}(D_1, \ldots, D_k)$'. The set of all evidences from any sets of entities $D_1, \ldots, D_k$ is denoted by '$\mathcal{E}$'.

In the following the reference to the evidential domains $D_1, \ldots, D_k$ of an evidence E is often suppressed. Note that the assumption $k \geq 1$ yields that an evidence E is contingent.

A difficulty is illustrated by the following example: Consider the statement '*a* is a white raven', $Ra \wedge Wa$, and the ravens-hypothesis 'All ravens are black', $\forall x\,(Rx \rightarrow Bx)$, and suppose that $Ra \wedge Wa$ is all the evidence E reports, i.e. $E = \{Ra, Wa\}$. Without recourse to some background knowledge B telling us that nothing white is black we cannot infer that – relative to E – the ravens-hypothesis is shown to be false. This illustrates that confirmation has to be construed as a ternary relation between a theory T, an evidence E, and a background knowledge B.

However, the notion of a background knowledge B is no precise one, at least if introduced by taking recourse to its "obvious" meaning. Care has to be taken what to put into the background knowledge B. Intuitively, B is conceived of as containing those and only those statements which are taken for granted and whose truth is out of question. A special sort of these statements are the definitions and meaning postulates and, more generally, those statements traditionally termed analytic. Yet, if Quine[11] is right, there is no sharp distinction between these analytic statements on the one hand and the remaining synthetic ones on the other. So it is neither clear what exactly the background knowledge B consists of, nor where the distinction between background knowledge B and theory T is to be drawn. Not only meaning, but also the assessment of theory T by evidence E relative to background knowledge B is holistic.

Nevertheless, it seems that in practice one can draw a distinction between the theory T – or the hypothesis h – that is to be assessed, and the background knowledge B that is taken for granted in this assessment. T is the set of those statements or propositions that are put to test and whose domains of proper investigation E is evidence from, whereas B is the set of those statements that are assumed to be true in this assessment of T by E.

Given this, there are at least two strategies for solving the problem just mentioned: Either to demand of the scientist to be *explicit* in the sense that she reports not only what she is or takes to be observing, but also everything she assumes to

[11] Cf. Quine (1951) and (1961).

be (logically) implied by her observations and her background knowledge B, in particular her knowledge of the language she is using – call this the *explicitness approach*; or else to expand the evidence E to a set E_B containing all statements of E, and all those statements in the background knowledge B which are related to these, where it is defined as follows:

Definition 1.5 (Related Wffs) Let h_1 and h_2 be two wffs. h_1 is *related to* h_2 iff $PR_{ess}(h_1) \cap PR_{ess}(h_2) \neq \emptyset$ or $C_{ess}(h_1) \cap C_{ess}(h_2) \neq \emptyset$.

Let h be any wff. *The set of essential predicates of* h, $PR_{ess}(h)$, is the set of all those predicates 'P' without which no wff h' with $h' \dashv\vdash h$ can be formulated, i.e.

$$PR_{ess}(h) = \bigcap_{h' \dashv\vdash h} PR(h'), \text{ for every } h' \in \mathcal{L}_{PL1=}.$$

The set of essential constant terms of h, $C_{ess}(h)$, is the set of all those constant terms 'c' without which no wff h' with $h' \dashv\vdash h$ can be formulated, i.e.

$$C_{ess}(h) = \bigcap_{h' \dashv\vdash h} C(h), \text{ for every } h' \in \mathcal{L}_{PL1=}.$$

Let us call this the *relatedness approach*.[12]

In the above example, the explicitness approach demands of the scientist to report not only that a is a white raven, but also that a is not black, if her background knowledge B contains the information that nothing white is black. The relatedness approach, on the other hand, demands to expand (proper) evidence $E = \{Ra, Wa\}$ to evidence $E_B = \{Ra, Ba, \ \forall x(Wx \rightarrow \neg Bx)\}$.

I prefer the relatedness approach to the explicitness approach, because the former is more sensitive to the fact that we are not *logically omniscient* in the sense that we know or believe all logical consequences of the statements (propositions) we know or believe, respectively. Furthermore, the relatedness approach enables to distinguish between those statements which are taken to report our (uncontrolled) observations and the results of our (controlled) experiments – call the set E of these statements the *proper* evidence – and those statements which (logically) follow from the proper evidence E in combination with our background knowledge B.

Finally, since I am considering a ternary relation of confirmation (of theory T by evidence E relative to background knowledge B), one can, after all,

[12]Note that logically determined statements do not have any essential predicates or essential constant terms.

forget about the expansion of E to E_B, for all the information contained in the background knowledge B will be available in the assessment of T by E.

Assumptions 1.2 and 1.3 put restrictions on the syntactical form of an evidence E. Besides these syntactical considerations there is another semantic and also pragmatic feature of an evidence E. Remember that the aim is a quantitative theory of confirmation telling one, for every theory T, every evidence E, and every background knowledge B, how much E confirms T relative to B. Suppose for the moment that we already have some adequate measure of confirmation C. What the value $C(E, T, B)$ for given T, E, and B tells us is how much T is confirmed by E relative to B. The assessment of T is therefore not absolute, but relative to E and B. $C(T, E, B)$ does not tell us how much T is confirmed absolutely, but how much T is confirmed relative to B and the assumption that E is true and contains all the data we can rely on in the assessment of T.

Furthermore, if – as I do – one wants such a measure of confirmation C to implicitely provide a rule of acceptance for rational theory choice[13], then the value $C(T, E, B)$ of C for given T, E, and B is of interest only if E is assumed to be true or accepted – otherwise this implicitely provided rule of acceptance for rational theory choice will misguide those adopting it.[14]

In my opinion this feature of the evidence E to be *epistemically distinguished* – in the sense that the assessment of a given theory T is not only relative to a background knowledge B, but also relative to E – fits well with the role our observations (respectively the statements describing or propositions representing them) play in our establishing a representation of the world. In contrast to our other assumptions about the world, they are assigned an epistemically special status: If inconsistencies (or incoherencies) arise in our representation of the world, and if we want to resolve them, then the statements describing what to take to have observed *usually* are the last we will drop. This finds its expression in

Assumption 1.4 (Epistemic Mark of Distinction) If E is an evidence from $D_1, \ldots, D_k$, then E is *assumed* to be *true in the actual world*, i.e.

$$\mathcal{A} \in mod(E), \quad \text{for every evidence } E \in \mathcal{E}.$$

[13] For instance, such a rule may tell one to accept that theory T_i in a given finite set of alternative theories $\{T_1, \ldots, T_n\}$ such that $C(T_i, E, B) \geq C(T_j, E, B)$, for every $j, 1 \leq j \leq n$. If there are several such theories T_i, then the rule may select one of them, or it may postpone the decision until new evidence comes in which settles the question.

[14] This does, of course, not mean that such a function is of interest, only if the evidence E is assumed to be true or accepted; nor does it mean that E is indeed true, or has to be true in order for such a function to make sense.

Note that this assumption does not put any restrictions on what to count as evidence. It simply expresses a feature of the epistemic status of an evidence E – a feature exhibited by us in relation to E; namely how we as epistemic subjects treat the statements in E in establishing and changing our representation of the world. In particular, this assumption does not mean that an evidence E actually contains only true statements. It is a commonplace that we are fallible, also in what we take to observe, and this assumption is not at all intended to call this commonplace into question.

Furthermore, in order for it to make sense we have to assume[15] that the actual world can be represented by some set-theoretical structure $\mathcal{A} = \langle Dom_A, \varphi_A \rangle$ (with the evidential domains of E among the sets of entities in the sequence Dom_A); for the standard Tarskian notion of truth in – which is adopted here – is defined between (sets of) statements and models $\mathcal{M} = \langle Dom, \varphi \rangle$, and the actual world can hardly be argued to be an ordered pair consisting of a sequence Dom of sets of entities $D_1, \ldots, D_r$ and an interpretation function φ.

It follows from assumption 1.4 that every evidence E is a description of the actual world.

Definition 1.6 (Description of a Model) Let D be a set of wffs, and let $\mathcal{M} = \langle Dom, \varphi \rangle$ be a model. D is a *description* of $\mathcal{M}$ (*in* $\mathcal{L}_{PL1=}$) iff $\mathcal{M} \models D$, and there is at least one model $\mathcal{M}' = \langle Dom', \varphi' \rangle$ such that $\mathcal{M}' \not\models D$.

So in order for a set of statements D to be a description of some model $\mathcal{M}$ (in $\mathcal{L}_{PL1=}$), D need not be complete in the sense that

$$\text{if} \quad \mathcal{M} \models h, \quad \text{then} \quad D \vdash h, \quad \text{for every wff } h \in \mathcal{L}_{PL1=};$$

it suffices (that D is not logically valid – otherwise D does not tell us anything about $\mathcal{M}$ – and) that D is correct in the sense that

$$\text{if} \quad D \vdash h, \quad \text{then} \quad \mathcal{M} \models h, \quad \text{for every wff } h \in \mathcal{L}_{PL1=}.$$

Observation 1.1 (E Is a Description of $\mathcal{A}$) Let $\mathcal{A} = \langle Dom_A, \varphi_A \rangle$ be a model representing the actual world. Then it holds for every evidence E from any sets of entities $D_1, \ldots, D_k$: E is a description of $\mathcal{A}$.

In the following it will be assumed that there is exactly one intended model $\mathcal{M}_E = \langle Dom_E, \varphi_E \rangle$ for every evidence E from any sets of entities $D_1, \ldots, D_k$. So the

[15] Apart from the supposition that the actual world exists.

interpretation of the constant i-terms[16], the predicates, and the function symbols occurring in an evidence E is always fixed, and it makes sense to speak of *the* individual t^i denoted by the constant i-term 't^i'. (It is the entity $\alpha^i \in D_i$ with $\varphi_E(\text{'}t^i\text{'}) = \alpha^i$, where $Dom_E = \langle D_1, \ldots, D_r \rangle, r \geq k$.[17]) The reference to the intended model $\mathcal{M}_E$ of evidence E will be suppressed henceforth, but it is to be kept in mind that the talk of the individual t^i denoted by the constant i-term 't^i' occurring in E is meaningful.

1.4 Background Knowledge

As already indicated, the questions what to count as background knowledge B, and where to draw the distinction between the theory T that is to be assessed (by some evidence E relative to B) on the one hand and the background knowledge B on the other, are difficult to answer. As in the case of theory, I will therefore only give a necessary condition for a background knowledge.

The intuitive understanding of a background knowledge B, which I assume to be construed as a set of statements expressing this alleged knowledge, is that it consists of that (and only that) information which we take for granted and assume to be true when we are concerned with the truth or some other epistemically distinguished property of other (sets of) statements; in particular, when we are concerned with the assessment of theories T by evidences E. Among others, B contains our linguistic knowlegde, and a formulation of the mathematical apparatus we use. This is a pragmatic feature of the background knowledge B in relation to us as epistemic subjects, which does not put any syntactical restrictions on what to count as a background knowledge B. The latter I will now turn to.

The only condition a set of statements B has to satisfy in order to be a background knowledge is that it is finitely axiomatizable. The reason for this is that

> the information (implicitely) assumed in the assessment of a given theory T by an evidence E has to be such that it can be made explicit, for otherwise it cannot be taken into into account by the measure of confirmation C.

[16] For a definition of constant i-terms see section 1.5.

[17] For ease of readability the evidential domains $D_1, \ldots, D_k$ of an evidence E are assumed to be the first k sets of entities in the sequence of r sets of entities constituting the domain Dom_E of the intendend model $\mathcal{M}_E$ of E.

Assumption 1.5 (Finite Axiomatizability of Background Knowledge) If B is a background knowledge, then there is at least one finite axiomatizaton A_B of B.

In the following a background knowledge B will be identified with one of its – by the above assumption existing – finite axiomatizations A_B. The set of all finite axiomatizations A_B of any background knowledge B is denoted by '$\mathcal{B}$'. Although assumption 1.5 is only a *necessary* condition for a background knowledge, finite sets of statements are often called background knowledges.

Let us now fix the basic terminology for the remainder of this monograph.

1.5 Terminology

For the following definition cf. Schurz (1998).

Definition 1.7 (Irreducible Representation) Let A and B be sets of wffs. B is an *irreducible representation of* A iff B is a non-redundant set of relevant elements of A such that $A \dashv\vdash B$. The set of all irreducible representations of A is denoted by '$\mathbf{I}(A)$'.

Let h be a wff. h is a *relevant element* of A iff h is an element, and $A \vdash_{crel} h$. The set of relevant elements of h is denoted by '$RE(h)$'; the set of relevant elements of A is denoted by '$RE(A)$'.

h *follows conclusion relevantly* from A, or h is a *relevant consequence of* A, $A \vdash_{crel} h$, iff (i) $A \vdash h$, and (ii) there are no (marked) occurrences of $n \geq 1$ predicates 'P_1', …, 'P_n' in h that can be replaced *salva validitate* of $A \vdash h$ by any n predicates 'P_1^*', …, 'P_n^*' of the same arity[18], i.e. such that $A \vdash h^*$, where h^* is the result of replacing these marked occurrences of 'P_i' in h by 'P_i^*', for every $i, 1 \leq i \leq n$.

h is an *element* iff h is an elementary normal form, and each quantifier scope in h is a conjunction of elementary wffs.

h is *elementary* iff there is no $n \geq 1$ such that $h \dashv\vdash h_1 \wedge \ldots \wedge h_n$, and each $h_i, 1 \leq i \leq n$, is shorter than h, where '$\rightarrow$' is eliminated and brackets are not counted.

Definition 1.8 (Redundancy) Let A be a set of wffs. A is (*formulated*) *redundant*(*ly*) iff there is at least one wff $h \in A$ such that $A \setminus \{h\} \vdash h$. Any such wff $h \in A$ with $A \setminus \{h\} \vdash h$ is called a *redundant part of* A.

A is (*formulated*) *non-redundant*(*ly*) iff A is not (formulated) redundant(ly).

[18]Equivalently, by n *new* predicates 'P_1^*', …, 'P_n^*' of the same arity.

For the following definition cf. Gemes (1994c) and Gemes (1997a).

Definition 1.9 (Content Part) Let A and B be wffs of the language $\mathcal{L}_{PC}$ of the classical propositional calculus. B is a *content part of* A iff

1. A and B are contingent, and $A \vdash B$, and

2. for some wff $C \in \mathcal{L}_{PC}$: $C \dashv\vdash B$, and there is no wff $D \in \mathcal{L}_{PC}$ such that $D \vdash C$, $C \nvdash D$, and $PV(D) \subseteq PV(C)$,

where, for any wff $A \in \mathcal{L}_{PC}$, $PV(A)$ is the set of all propositional variables occurring in A.

Let A and B be wffs of the language $\mathcal{L}_{PC'}$ of the classical propositional calculus enriched by countably infinite individual constants 'a_1', ..., 'a_n', ... and by finitely many predicates 'P_1', ..., 'P_m' of varying arity. B is a *content part of* A iff

1. A and B are contingent, and $A \vdash B$, and

2. for some wff $C \in \mathcal{L}_{PC'}$: $C \dashv\vdash B$, and there is no wff $D \in \mathcal{L}_{PC'}$ such that $D \vdash C$, $C \nvdash D$, and $AT(D) \subseteq AT(C)$,

where, for any wff $A \in \mathcal{L}_{PC'}$, $AT(A)$ is the set of all atomic wffs or propositional variables occurring in A.

Let A and B be wffs of the language $\mathcal{L}_{PL1}$ of first-order predicate logic without identity (excluding function symbols), $PL1$. B is a *content part of* A iff it holds for every non-empty set of individual constants IC with $IC(A) \cup IC(B) \subseteq IC$:

1. $Dev_{IC}(B)$ and $Dev_{IC}(A)$ are contingent, and $Dev_{IC}(A) \vdash Dev_{IC}(B)$, and

2. for some wff $C \in \mathcal{L}_{PC'}$: $C \dashv\vdash Dev_{IC}(B)$, and there is no wff $D \in \mathcal{L}_{PC'}$ such that $D \vdash C$, $C \nvdash D$, and $AT(D) \subseteq AT(C)$.

For the following definition cf. Gemes (1993).

Definition 1.10 (Natural Axiomatization) Let $\mathcal{L}$ be $\mathcal{L}_{PC}$, $\mathcal{L}_{PC'}$, or $\mathcal{L}_{PL1}$, and let T and T' be sets of wffs of $\mathcal{L}$. T' is a *natural axiomatization of* T iff

1. T' is finite, and $T \dashv\vdash T'$,

2. h is a content part of $\bigwedge_{h' \in T'}$, for every wff $h \in T'$,

3. there is no content part c_h of some wff $h \in T'$ such that $T' \setminus \{h\} \vdash c_h$, and

4. there is no set T'' of wffs of $\mathcal{L}$ satisfying (1)-(3) and such that $|T''| > |T'|$.

The set of all natural axiomatizations of T is denoted by '**NA** (T)'.

Definition 1.11 (Development) Let T be a set of wffs, and let $C = \{\text{'}c_1^i\text{'}, \ldots, \text{'}c_n^i\text{'}\}$ be a finite set of constant i-terms. The *development of* T *for* C, $Dev_C(T)$, is the development $Dev_C\left(\bigwedge_{h \in T} h\right)$ of the conjunction $\bigwedge_{h \in T} h$ of all wffs $h \in T$ for C.

Let h be a wff. The *development of* h *for* C, $Dev_C(h)$, is recursively defined as follows:

1. If h is atomic, i.e. if h is of the form '$P(t_1, \ldots, t_n)$', then $Dev_C(h) = h$.

2. If $h = \neg h_1$, then $Dev_C(h) = \neg Dev_C(h_1)$.

3. If $h = h_1 \wedge h_2$, then $Dev_C(h) = Dev_C(h_1) \wedge Dev_C(h_2)$.

4. If $h = h_1 \vee h_2$, then $Dev_C(h) = Dev_C(h_1) \vee Dev_C(h_2)$.

5. If $h = h_1 \rightarrow h_2$, then $Dev_C(h) = Dev_C(h_1) \rightarrow Dev_C(h_2)$.

6. If $h = \forall x^i A[x^i]$, then $Dev_C(h) = \bigwedge_{1 \leq j \leq n} A\left[c_j^i / x^i\right]$.

7. If $h = \forall x^k A\left[x^k\right], k \neq i$, then $Dev_C(h) = \forall x^k Dev\left(A\left[x^k\right]\right)$.

8. If $h = \exists x^i A[x^i]$, then $Dev_C(h) = \bigvee_{1 \leq j \leq n} A\left[c_j^i / x^i\right]$.

9. If $h = \forall x^k A\left[x^k\right]$, $k \neq i$, then $Dev_C(h) = \exists x^k Dev\left(A\left[x^k\right]\right)$.

Here, '$A\left[c_j^i / x^i\right]$' is the result of uniformly substituting the constant i-term 'c_j^i' for all free occurrences of the i-variable 'x^i' in A.

Definition 1.12 (Constant i-Term)

1. Every i-constant 'c_j^i,' $j \geq 1$, is a constant i-term, for every i.

2. If 'f^i' is an n-ary $k_1, \ldots, k_n, i$-function symbol, 't^{k_1}' is a constant k_1-term, ..., 't^{k_n}' is a constant n-term, then '$f^i\left(t^{k_1}, \ldots, t^{k_n}\right)$' is a constant i-term.

3. Nothing else is a constant i-term.

't' is a constant term iff there is an i such that 't' is a constant i-term.

Definition 1.13 (Description of an Individual) Let E be an evidence from $D_1, \ldots, D_k$, let 't' be a constant term occurring in E, and let 't^i', $1 \leq i \leq k$, be a constant i-term occurring in E, whence t^i is an individual of D_i.

The set of all constant i-terms 't^i' occurring in E is denoted by '$C_i(E)$'. The set of all constant terms 't' occurring in E is denoted by '$C(E)$'. The set of all constant i-terms 't^i' essentially occurring in E is denoted by '$C_{i,ess}(E)$'. The set of all constant terms 't' essentially occurring in E is denoted by '$C_{ess}(E)$'.

Let B be a set of wffs. The set of all constant i-terms 't_l^i' in $C_{ess}(E)$ for which there is no constant i-term 't_j^i' in $C_{ess}(E)$ such that $j < l$ and

$$B \cup E \vdash t_j^i = t_l^i$$

is called the *B-representative of* $C(E)$. It is denoted by '$C_{B-repr}(E)$'. If B is empty, I will speak of the representative of $C(E)$, $C_{repr}(E)$.

The *description of* 't' respectively t *in* E, $D_E(t)$, is defined as the set of relevant elements A of E with 't' $\in C(A)$, i.e.

$$D_E(t) = \{A \in RE(E) : \text{`}t\text{'} \in C(A)\}.$$

The set of entities $t^i \in D_i$ which are mentioned in E is denoted by '$I_i(E)$', i.e.

$$I_i(E) = \left\{\alpha^i : \varphi_E(\text{`}t^i\text{'}) = \alpha^i, \alpha^i \in D_i, \text{ for some `}t^i\text{'} \in C_i(E)\right\},$$

where $Dom_E = \langle D_1, \ldots, D_r\rangle, r \geq k$.

The set of entities t which are mentioned in E is denoted by '$I(E)$', i.e.

$$\begin{aligned} I(E) &= \{\alpha : \varphi_E(\text{`}t\text{'}) = \alpha, \alpha \in D_i, \text{ for some } i, 1 \leq i \leq r, \\ &\quad \text{and some `}t\text{'} \in C(E)\}, \end{aligned}$$

where $\mathcal{M}_E = \langle Dom_E, \varphi_E\rangle$ is the intended model of E.

Definition 1.14 (Unconditional Probability) A function $p(\cdot)$, $p(\cdot) : \mathcal{L}_{PC} \rightarrow \Re$, $\mathcal{L}_{PC}$ being the language of the classical propositional calculus PC, is a(n) (*unconditional*) *probability* iff it holds for any wffs $A, B \in \mathcal{L}_{PC}$:

1. $p(A) \geq 0$,

2. if $\vdash A$, then $p(A) = 1$, and

3. if $A \vdash \neg B$, then $p(A \vee B) = p(A) + p(B)$.

Definition 1.15 (Strict Unconditional Probability) A function $p(\cdot), p(\cdot) : \mathcal{L}_{PC} \rightarrow \Re$, is a *strict* (unconditional) probability iff $p(\cdot)$ is a(n) (unconditional) probability, and

$$p(A) = 1, \quad \text{only if} \quad \vdash A, \quad \text{for every wff } A \in \mathcal{L}_{PC}.$$

Definition 1.16 (Conditional Probability) Let $p(\cdot)$, $p(\cdot) : \mathcal{L}_{PC} \rightarrow \Re$, be a(n) (unconditional) probability. The partial function $p(\cdot \mid \cdot)$, $p(\cdot \mid \cdot) : \mathcal{L}_{PC} \times \mathcal{L}_{PC} \rightarrow \Re$, with

$$p(B \mid A) = \frac{p(B \wedge A)}{p(A)},$$

for any wffs $A, B \in \mathcal{L}_{PC}$ with $p(A) > 0$, is the *conditional probability based on* $p(\cdot)$.

If $p(\cdot)$ is a strict (unconditional) probability, then the conditional probability $p(\cdot \mid \cdot)$ based on $p(\cdot)$ is called the *strict* conditional probability based on $p(\cdot)$.

Theorem 1.2 (Strict Probabilities) Let $p(\cdot)$, $p(\cdot) : \mathcal{L}_{PC} \rightarrow \Re$, be a strict (unconditional) probability, and let $p(\cdot \mid \cdot)$ be the conditional probability based on $p(\cdot)$. Then it holds for any wffs $A, B \in \mathcal{L}_{PC}$ with $p(A) > 0$:

$$p(B \mid A) = 1, \quad \text{only if} \quad A \vdash B.$$

Chapter 2

The Problem of a Quantitative Theory of Confirmation

2.1 Criteria for a Solution

In my opinion, any solution to the problem of a quantitative theory of confirmation has to satisfy two sets of criteria: The first one is a set of high-level, meta-, or formal conditions of adequacy any formal theory has to satisfy. These criteria demand of a formal theory to be *formally handy* in the sense that it be non-arbitrary, comprehensible, and computable in the limit. They will be the topic of this chapter.

The second set of criteria is a set of low-level, object-, or material conditions of adequacy any quantitative theory of confirmation (whether or not it is intended to implicitely provide a rule of acceptance for rational theory choice) has to satisfy in my opinion. What these criteria amount to is that a quantitative theory of confirmation be *materially adequate* in the sense that all what matters in determining whether and to what degree a given evidence E confirms a given theory T relative to some background knowledge B are the so called *confirmational virtues* (of theory T in relation to evidence E and background knowledge B). These confirmational virtues are dealt with in the next chapter.

The challenge is the definition of a (set of) function(s) $C(\cdot,\cdot,\cdot)$,

$$C(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \to \Re,$$

such that $C(T, E, B)$ is a measure of confirmation of theory T by evidence E relative to background knowledge B which is formally handy and materially adequate.

2.2 High-Level, Meta-, or Formal Conditions of Adequacy

As mentioned I think that any formal theory, in particular, any quantitative theory of confirmation intended to implicitely provide a rule of acceptance for rational theory choice has to be formally handy in the sense that it be non-arbitrary, comprehensible, and computable in the limit. In this section these notions will be defined and applied, and I will try to motivate them.

Definition 2.1 (Arbitrariness) A (concept $\mathcal{C}_F$ defined by a) set F of functions f, $f : D \rightarrow R$, is *arbitrary* iff there are at least two functions $f_1, f_2 \in F$ and at least two arguments $x, y \in D$ such that

$$f_1(x) < f_1(y) \quad \text{and} \quad f_2(y) < f_2(x),$$

where $<$ is a strict order (i.e. an asymmetric and transitive relation) on R.[1]

A (concept $\mathcal{C}_F$ defined by a) set of functions F is *non-arbitrary* iff ($\mathcal{C}_F$ respectively) F is not arbitrary.

If (the *definiens* of a concept $\mathcal{C}_F$ defined by) a set F consists of a single function f, f is called non-arbitrary.

A theory T is *arbitrary* iff at least one of its central concepts $\mathcal{C}_F$ is defined by an arbitrary set of functions F.

A theory T is *non-arbitrary* iff at least one of its central concepts $\mathcal{C}_F$ is defined by a set of functions F, and none of its central concepts $\mathcal{C}_F$ is defined by an arbitrary set of functions F.[2]

In order for this definition to be comprehensible one has to make precise which concepts of a given theory are its central ones. For the cases dealt with here the matter is clear – or so I think – whence the notion of a central concept of a theory will not be defined. I will simply list which concepts of which theories I take be central ones.

Assumption 2.1 (Central Concepts)

[1] If F is a set of n-ary functions f, $f : D \rightarrow R$, $n \geq 1$, then the arguments $x, y \in D$ are n-tupels (and D is a set of n-tupels).

[2] Please note that a theory need not be non-arbitrary, if is not arbitrary. This is only the case, if the theory is *formal* in the sense that at least one of its central concepts is defined by a set of functions.

1. The concept of (degree of) confirmation of theory T by evidence E relative to background knowledge B is a central concept of any (quantitative) theory of confirmation of theory T by evidence E relative to background knowledge B.

2. The concept of (explanatory) coherence is a central concept of any theory of (explanatory) coherence.

3. The concept of probability is a central concept of any theory of probability, independently of whether probability is interpreted as logical probability, as (inter)subjective degree of belief, as objective chance or propensity, or as (limiting) relative frequency.[3]

Why should a formal theory, say, of (explanatory) coherence be non-arbitrary? Consider a typical situation at court, where two versions of the same event are presented – one by the accusing party, and another by the accused party. Suppose the judge who has to decide between these two versions reasons along coherentist lines. What she will do is to try to find out whether the version of the accusing party or that of the accused one is more coherent with the data, which are the materials admitted for proof.

Now assume the theory of (explanatory) coherence the judge is adopting is arbitrary for the case in question. This means that there are at least two coherence functions satisfying all constraints of this theory of (explanatory) coherence such that according to the first function the version of the accusing party is more coherent with the data, whereas according to the second function it is the version of the accused party which is more coherent with the data. So the theory of (explanatory) coherence is of no help for the judge.

Put differently, the arbitrary theory of (explanatory) coherence allows the judge to justify any decision, for instance, to acquit the accused party simply because the latter is supporter of the judge's favourite football team.

Application 2.1 (Arbitrariness)

1. Every set of Bayesian relevance measures[4] is arbitray; in particular, this

[3]For a clearly written discussion of all these views cf. Gillies (2000).

[4]A function $m_p(\cdot, \cdot \mid \cdot)$ is a *relevance measure* iff it holds for any wffs $H, E, K \in \mathcal{L}_{PC}$:

$$m_p(H, E \mid K) \begin{cases} > 0, & \text{if} \quad p(H \mid E \wedge K) > p(H \mid K), \\ < 0, & \text{if} \quad p(H \mid E \wedge K) < p(H \mid K), \\ = 0, & \text{if} \quad p(H \mid E \wedge K) = p(H \mid K), \end{cases}$$

holds of d, r, l, s, and c.[5]

2. The set P of all (unconditional) probabilities, and the set P_{cond} of all conditional probabilities are arbitrary.

3. The uncountable set F of functions $f_a(\cdot)$, $f_a(\cdot): \Re_0^+ \to \Re$, with

$$f_a(x) = x^a, \quad x \in \Re_0^+ = \{x : x \in \Re, x \geq 0\},$$
$$a \in \Re^+ = \{x : x \in \Re, x > 0\},$$

is not arbitrary.

---

where $p = p(\cdot \mid \cdot)$ is some conditional probability. The set of relevance measures m is defined as follows:

$$m = \{m_p : p \text{ is a conditional probability}\}.$$

Cf. Fitelson (2001). Fitelson's main thesis – '*the* central fact', Fitelson (2001), p. 6 – is that, for a given conditional probability p, the five Bayesian relevance measures d, r, l, s, and c are not ordinally equivalent, which means something slightly weaker than being arbitrary; namely that for any two sets of relevance measures $m, m' \in \{d, r, l, s, c\}$ there are conditional probabilities p and wffs $H, E, K, H', E', K' \in \mathcal{L}_{PC}$ such that

$$m_p(H, E \mid K) \geq m_p(H', E' \mid K') \quad \text{and} \quad m'_p(H, E \mid K) < m'_p(H', E' \mid K').$$

My point is not that, for a given conditional probability p, the set of relevance measures $\{d_p, r_p, l_p, s_p, c_p\}$ is arbitrary, but that every set of relevance measures is arbitrary. This means that for every set of relevance measures m there are conditional probabilities p_1 and p_2, and wffs $H, E, K, H', E', K' \in \mathcal{L}_{PC}$ such that

$$m_{p_1}(H, E \mid K) > m_{p_1}(H', E' \mid K') \quad \text{and} \quad m_{p_2}(H, E \mid K) < m_{p_2}(H', E' \mid K').$$

[5]The appendix to chapter 2 contains a proof of this arbitrariness claim. d, r, l, s, and c are defined as follows – cf. Fitelson (2001). Let $p(\cdot)$ be an (unconditional) pobability, let $p(\cdot \mid \cdot)$ be the conditional probability based on $p(\cdot)$, and let T, E, and B be single statements or propositions of $\mathcal{L}_{PC}$.

$$d_p(T, E \mid B) := p(T \mid E \wedge B) - p(T \mid B), \quad \text{cf. Earman (1992).}$$

$$r_p(T, E \mid B) := \log\left[\frac{p(T \mid E \wedge B)}{p(T \mid B)}\right], \quad \text{cf. Horwich (1982) and Milne (1996).}$$

$$l_p(T, E \mid B) := \log\left[\frac{p(E \mid T \wedge B)}{p(E \mid \neg T \wedge B)}\right], \quad \text{cf. Good (1983).}$$

$$s_p(T, E \mid B) := p(T \mid E \wedge B) - p(T \mid \neg E \wedge B), \quad \text{cf. Christensen (1999).}$$

$$c(T, E \mid B) := p(T \wedge E \wedge B) \cdot p(B) - p(T \wedge B) \cdot p(E \wedge B), \quad \text{Carnap (1962).}$$

4. No set of only one single function is arbitrary; in particular, the singletons containing *fuzzy*-negation $\neg_{fuzzy}$ as the fuzzy-logical interpretation of negation, and multiplication or *fuzzy*-conjunction $\wedge_{fuzzy}$ as the fuzzy-logical interpretation of conjunction are not arbitrary[6], where

$$\neg_{fuzzy} : [0,1] \to [0,1], \quad \neg_{fuzzy}(x) = 1 - x,$$
$$\wedge_{fuzzy} : [0,1] \to [0,1], \quad \wedge_{fuzzy}(x,y) = x \cdot y.$$

Let us turn to the second formal condition of adequacy which demands of a theory to be comprehensible.

Definition 2.2 (Comprehensibility) A theory T is *comprehensible* iff all its primitive concepts are comprehensible.

What this definition should capture is that a theory explicating one concept of interest – e.g. the concept of (explanatory) coherence – by means of another concept which is in need of explication itself – e.g. the concept of explanation – is no good theory, because the concept of interest is not explicated, but merely circumscribed in terms of another concept which is equally or even more unclear.

Though there may be an intuitive understanding of 'comprehensible', the question is which (primitive) concepts of a given theory are comprehensible. It may be argued that all concepts corresponding to observational terms are comprehensible – given that one has been able to clearly distinguish observational form theoretical terms – but already here doubts may be raised. I will therefore not argue for the comprehensibility of some distinguished set of concepts, but will restrict myself to the following pragmatically justified assumption.

Assumption 2.2 (Comprehensible Concepts)

- The primitive concepts of $PL1 =$ and ZF, i.e.

  $$\neg, \wedge, \vee, \rightarrow, =, \forall, \exists, x_1^1, x_2^1, \ldots, x_1^i, x_2^i, \ldots \in, \text{ and } \{\}$$

  are comprehensible.

- The concept of explanation is not comprehensible.

[6] That *fuzzy*-negation and *fuzzy*-conjunction are not arbitrary will be of importance in chapter 4.

The justification of this assumption is pragmatic since it consists in the fact that no formal theory – i.e. no theory at least one of whose central concepts is defined by a set of functions – can do without the primitive notions of logic ($PL1 =$) and set-theory (ZF).[7] In particular, this holds for every quantitative theory of confirmation aiming at the definition of an adequate measure of confirmation.

Application 2.2 (Comprehensibility)

- $PL1 =$ and ZF are comprehensible.
- No theory of (explanatory) coherence presupposing as primitive the concept of explanation is comprehensible.

Before defining the third formal condition of adequacy, remember the definiton of computability[8]: Roughly speaking, a function f, $f : D \to R$, is computable just in case there is an algorithm (a Turing machine) which yields for every argument $x \in D$ after finitely many steps the value $f(x) \in R$ of f for $x.$ That is, there is an algorithm which – when presented with input x – computes in finitely many steps the output $f(x)$, and then gives a sign that this is the value of f for $x.$ Such a *division* algorithm may be characterised as an assessment method which

> outputs the correct answer and then *halts*, thereby signaling that the answer is correct[9];

it

> is logically guaranteed to converge to the correct answer with *certainty*[10].

As stressed by Kelly (1996), a method – though not logically guaranteed to converge to the correct answer with certainty –

> may be logically guaranteed to stabilize to the truth without ever giving a sign that it has found the truth.[11]

[7] Apart from this, some concepts have to be assumed as primitive, for one cannot express anything without presupposing any concept at all.

[8] For the following cf. Kelly (1996).

[9] Kelly (1996), p. 4.

[10] Kelly (1996), p. 4.

[11] Kelly (1996), p. 4.

This weaker sense of convergence[12] to the correct answer gives rise to a weaker notion of computability: computability in the limit.

Definition 2.3 (Computability in the Limit[13]) A function f, $f : D \to R$, is *computable* iff there is an algorithm M such that it holds for every $x \in D$: If M is presented with x as input, then M outputs $f(x)$ after finitely many steps, and then halts; i.e. there is an $n \in \omega$ such that the output of M at n for x is $f(x)$, and M halts at n.

A function f, $f : D \to R$, is *computable in the limit* iff there is an algorithm M such that it holds for every $x \in D$: The output stream of M for x stabilizes to $f(x)$.

Let M be an algorithm. The output stream of M for x *stabilizes to* r iff there is an $n \in \omega$ such that it holds for every $m \geq n$, $m \in \omega$: The output of M at m for x is r.

A set of functions F is *computable in the limit* iff all functions $f \in F$ are computable in the limit.

A theory is *computable in the limit* iff at least one of its central concepts $\mathcal{C}_F$ is defined by a set of functions F, and none of its central concepts $\mathcal{C}_F$ is defined by a set of functions F which is not computable in the limit.

In short, the difference between a computable function and one which is only computable in the limit is that for the former there exists a method which outputs the correct answer after some finite time and additionally gives a sign that it has arrived at the correct answer, whereas for the latter there is a method which – though it also outputs the correct answer after some finite time – does not give a sign that it has arrived at the correct answer, but continues to output this correct answer forever. The method tells you the correct answer, but does not tell you that it is correct.

Two questions arise: First, why should a theory be computable in the limit? Second, why do I not demand that it be computable? The answers are that (1) a theory is useless for practical purposes if it is not computable in the limit, but that (2) demanding of it to be computable is demanding too much.

[12] Kelly (1996) introduces still further notions of convergence all of which give rise to corresponding notions of verification, refutation, and decision. I have adapted – and thereby changed – his definitions for the purposes I am concerned with here.

[13] The following definitions are rather informal. I have tried to define computability in the limit without introducing the notions of a Turing machine, a Turing-computable function, and several related concepts. For precise definitions of these notions and more about enumerability, decidability, and computability the reader is referred to Hermes (1961). An English introduction cited by Kelly (1996) is Cutland (1986).

Why should a formal theory be computable in the limit? Suppose you are concerned with the problem of determining how "good" a theory T is (relative to a given evidence E and some background knowledge B), where a theory T is good (relative to E and B), if it is true, can explain (together with B) many of the data in E, is simple, and so on. Suppose further that you can assume that there is exactly one "best" theory T (for given E and B). Finally, assume you are presented the following formal theory of the problem of "good" theories consisting in the measure $m(T, E, B)$ of the "goodness" of theory T (in relation to evidence E and background knowledge B):

$$m(T, E, B) = \begin{cases} 1, & \text{if } T \text{ is the "best" theory (relative to } E \text{ and } B\text{),} \\ 0 & \text{otherwise.} \end{cases}$$

Obviously, this measures is adequate in the sense that it always picks out the right theory. However, it is equally obvious that this measure is useless for practical purposes. The reason for this is that m is neither computable nor computable in the limit: In order to determine the value m takes on for given T, E, and B, one has to know whether an obviously not comprehensible concept applies, which, in this case, even coincides with the central concept to be explicated.

One may ask why I do not demand of a formal theory to be computable. In my opinion, this amounts to demanding too much, since it would rule out, among others, all theories that explicate a notion of interest by means of logical relations exhibited by various (sets of) statements. Doing so is not only common, but also reasonable practice in the philosophy of science. Much can be clarified, if one can determine whether a notion of interest applies (to a suitable argument), if all one has to assume are the logical relations between various (sets of) statements.

However, this practice does not satisfy the criterion of computability. In order to make this claim precise, note that a relation R on a set S is decidable just in case the characteristic function χ_R of R is computable.[14]

So decidability of relations (sets) is just computability of their characteristic functions. As $PL1 =$ is not decidable, the characteristic function $\chi_\vdash$ of theoremhood in $PL1 =$ is not computable, whereas the characteristic function of theoremhood in the classical propositional calculus PC is computable. As a consequence, if $PL1 =$ is the underlying logic, then no function the values of whose arguments

[14]If there is an algorithm that tells one for every $s \in S$ whether or not $s \in R$, then this algorithm tells one whether $\chi_R(s) = 1$ or $\chi_R(s) = 0$; for the former holds just in case $s \in R$, and the latter holds just in case $s \notin R$. On the other hand, if $\chi_R(s)$ is computable, then there is an algorithm that outputs for every $s \in S$ after finitely many steps the value $\chi_R(s)$ of χ_R for s, and thereby tells one whether or not $s \in R$.

depend on logical relations between various (sets of) statements is computable. If, however, PC is the underlying logic, then every function whose values depend only on logical relations between various (sets of) statements is computable.

Suppose you have achieved to explicate a notion of interest by means of logical relations between various sets of statements, and you have even achieved to define a function measuring the extent to which this notion is exhibited by given suitable arguments. If you have achieved this when PC is the underlying logic, then your function is computable. Now suppose you try to extend your results to $PL1 =$, and you succeed to do so after hard work. Again, the notion of interest is explicated by means of logical relations between various sets of statements. Yet in this case, the function measuring the extent by which the notion of interest is exhibited by suitable arguments is not computable, for $PL1 =$ is not decidable. Clearly, it is inappropriate to disallow this function simply because of the fact that it is not computable.

As an example consider the second axiom of the probability calculus:

$$p\left(A\right) = 1, \quad \text{if} \quad \vdash A.$$

Suppose $PL1 =$ is the underlying logic. For a given statement $A \in \mathcal{L}_{PL1=}$, one has to determine whether $\vdash A$ in order to know whether one has to assign A the value 1. But as $\vdash$ is not decidable, this cannot be done.

One might suggest that a theory need not be computable, but that it be computable under the assumption that the underlying logic is decidable. Call this kind of computability *near computability*.

It is easily seen that also near computability is too restrictive. For instance, define as follows: A set of statements T says something about the individual t iff there is a statement A such that t is mentioned in A and A logically follows from T. The corresponding characteristic function f,

$$f\left(T, t\right) = \begin{cases} 1, & \text{if there is a wff } A \in \mathcal{L}_{PL1=} \text{ such that `} t \text{' } \in C\left(A\right) \text{ and } T \vdash A, \\ 0 & \text{otherwise,} \end{cases}$$

is not computable even if it is assumed that the underlying logic is decidable. (There are infinitely many statements A that have to checked on their containing an occurrence of 't' and on their logically following from T.)

Nevertheless, the above definition and the corresponding function should not be disallowed simply for the *formal* reason that it is not nearly computable – though there may, of course, be other reasons for doing so. This would be very restrictive, and many concepts and corresponding characteristic functions would be ruled out thereby.

As shown by Kelly (1996) there are functions which are not even computable in the limit[15]:

> [W]ithout extra background knowledge, the hypothesis that matter is infinitely divisible is not decidable in the limit.[16]

The function f_{inf}, $f_{inf} : \mathcal{L}_{PL1=} \longrightarrow \Re$, is therefore not even nearly computable in the limit, where, for every statement $A \in \mathcal{L}_{PL1=}$,

$$f_{inf}(A) = \begin{cases} 1, & \text{if matter is infinitely divisible}, \\ 0 & \text{otherwise}. \end{cases}$$

Application 2.3 (Near Computability)

1. The characteristic function of theoremhood in PC, $\chi_{\vdash,PC}$, $\chi_{\vdash,PC} : \mathcal{L}_{PC} \longrightarrow \{0,1\}$, is not computable, but computable in the limit, where, for every wff $A \in \mathcal{L}_{PC}$,
$$\chi_{\vdash}(A) = \begin{cases} 1, & \text{if} \quad \vdash A, \\ 0 & \text{otherwise}. \end{cases}$$

2. The characteristic function of theoremhood in $PL1 =$, $\chi_{\vdash}$, $\chi_{\vdash} : \mathcal{L}_{PL1=} \longrightarrow \{0,1\}$, is not computable, but computable in the limit, where, for every wff $A \in \mathcal{L}_{PL1=}$,
$$\chi_{\vdash}(A) = \begin{cases} 1, & \text{if} \quad \vdash A, \\ 0 & \text{otherwise}. \end{cases}$$

3. If $p(\cdot)$, $p(\cdot) : \mathcal{L}_{PC} \longrightarrow \Re$, is a(n) (unconditional) probability, and the values
$$p\left(\bigwedge_{p_i \in D \subseteq PV} \pm p_i\right)$$

[15] Cf. Kelly (1996), p. 53f. What Kelly shows there is that the hypothesis that matter is infinitely divisible is refutable in the limit, but neither decidable in the limit nor verifiable in the limit. Decidability, verifiability, and refutability (in the limit) are defined for (sets of) hypotheses relative to the set of possible worlds, which he identifies with infinite data streams.

The example of the infinite divisibility of matter is construed in such a way that if particle p is divisible, then it will be divided after finitely many trials, but not necessarily after the first one. The latter is important, for otherwise a failure to divide p would show that p is not divisible, whence the hypothesis that matter is infinitely divisible were refutable with certainty, which it is not under the more realistic assumption – cf. Kelly (1996), p. 51.

[16] Kelly (1996), p. 53.

are not all given in advance (which amounts to cheating), then $p(\cdot)$ is not computable in the limit, where PV is the set of all propositional variables of $\mathcal{L}_{PC}$.

The same holds for every conditional probability.

4. No set of Bayesian relevance measures in the sense of Fitelson (2001) is computable in the limit.

2.3 Why a Quantitative Theory of Confirmation Is to Be Formally Handy

As already indicated, I think that a quantitative theory of confirmation of theory T by evidence E relative to background knowledge B should implicitely provide a rule of acceptance for rational theory choice. The typical problem situation of the latter consists in the question which theory T_i of a finite set of alternative theories $\{T_1, \ldots, T_n\}$ it is rational to accept with regard to a given evidence E (and some background knowledge B) belonging to the domain of application of each theory T_l, $1 \leq l \leq n$, where two theories T and T' are *alternative* relative to B just in case $B \cup T \cup T' \vdash \bot$. A rule of acceptance for rational theory choice implicitely provided by any quantitative theory of confirmation defining some measure of confirmation C is the following:

> $(\mathcal{R})$ Let E be an evidence, let B be a background knowledge, and let $\{T_1, \ldots, T_n\}$ be a finite set of alternative theories, where E belongs to the domain of application of each theory T_l, $1 \leq l \leq n$.
>
> If there is exactly one theory T_i, $1 \leq i \leq n$, among $T_1, \ldots, T_n$ such that $C(T_i, E, B) \geq C(T_j, E, B)$ for every $j, 1 \leq j \leq n$, then accept T_i.
>
> If not, continue gathering evidence.[17]

Suppose that a quantitative theory of confirmation defining the central concept of confirmation of theory T by evidence E relative to background knowledge B by a set of functions $\mathcal{C}$ is not formally handy. Then it is arbitrary, not comprehensible, or not computable in the limit.

[17] If one wants such a rule to choose one single theory also in case there are several such theories T_i, then one may arbitrarily choose the first one.

If it is arbitrary, then there are problem situations of the above kind and measures of confirmation C and C' in $\mathcal{C}$ such that according to C some theory T should be accepted with regard to a given evidence E and a given background knowledge B, whereas according to C' some *other* theory T' should be accepted with regard to the same evidence E and the same background knowledge B, where T and T' are alternative (relative to B), and evidence E belongs to the domain of application of both T and T'. As the quantitative theory of confirmation under consideration does not provide any criterion for choosing between the two measures of confirmation C and C', it does not decide between the two theories T and T', whence, after all, $(\mathcal{R})$ turns out to be no rule of acceptance (for rational theory choice) for the case in question.

If the quantitative theory of confirmation is not comprehensible, then the value $C(T, E, B)$ of the measure of confirmation C for given T, E, and B will depend on determining whether some primitive concept which is not comprehensible applies to T, E, and B. However, this determination is arbitrary to the extent to which the primitive concept in question is not comprehensible, whence $(\mathcal{R})$ is of no help again.[18]

Finally, if the quantitative theory of confirmation is not computable in the limit, then there is no method which stabilizes to the correct value $C(T, E, B)$ of C for all theories T, evidences E, and background knowledgees B. Once more $(\mathcal{R})$ is no guide in deciding which theory T to accept with regard to E and B, since for all methods M there are theories T, evidences E, and background knowledges B such that M does not stabilize to $C(T, E, B)$, i.e. M will continue to output false values forever.[19]

2.4 Down With Bayesianism?

The foregoing examples – showing that all Bayesian theories of confirmation are arbitrary (and, under realistic circumstances, not computable in the limit) – seem to urge the conclusion: Down with Bayesianism.

In the second subsection of this section I will briefly turn to the constructive

[18]One can see here – and in the example of the research project on "good" theories – that the three formal conditions of adequacy are intimately related in that violating the second (non-formal) criterion of comprehensibility often yields violations of the first criterion of non-arbitrariness and the third criterion of computability in the limit, both of which are criteria for formal theories.

[19]For all methods M there are T, E, and B such that for every point of time n there is a point of time $m \geq n$ such that the output of M at m on T, E, and B differs from $C(T, E, B)$).

part of my criticism – arguing that this conclusion is rush, and sketchily describing in very general terms three possible ways out. Before doing so an example is considered that should illustrate to what – in my opinion: unacceptable – consequences the subjective interpretation of probability as degree of belief leads when it is to provide the basis for a quantitative theory of confirmation.

I think the main problem of Bayesianism is its arbitrariness. Indeed, one may be inclined to call Bayesianism the paradigm[20] of arbitrariness. The reason for this lies in the fact that the three (four) axioms of the probability calculus are far too inclusive in the sense that any assignment of values in $[0, 1]$ to the propositional variables (atomic statements) is coherent with these axioms.

The subjective interpretation is very popular. I think this is – at least partly – due to the fact that in modeling various problems the subjective interpretation allows one to assume particular suitable values for the probabilities occurring in the model. These then yield the expected result. The justification for these values is very easy: After all, probabilities are just degrees of belief.

Though this liberality enables Bayesianism to explain lots of phenomena – not only in the philosophy of science, but also in economics, politics, and, more generally, the social sciences – it allows, so to speak, to explain too much: There follow things that should not. What I would like to illustrate here is that there are cases where, when taken seriously, the subjective interpretation leads to unacceptable consequences.

The rejection of the subjective interpretation of probability as degree of belief concerns the problem of a quantitative theory of confirmation. The reason for this is that I take confirmation to be a relation between theories, evidences, and background knowledges that holds (to a given extent) independently of anyone's subjective degrees of belief. Whether and to what degree some theory T is confirmed by an evidence E relative to a background knowledge B is not even dependent on the existence of someone's having certain beliefs – whether or not these are given as numerical degrees and, if so, whether or not they obey the probability calculus. I take confirmation to be *objective* in this sense.

This does not amount to a rejection of the subjective interpretation of probability as a whole. I do not attempt to argue against the position

> that the unconditional (and derivatively the conditional) probability axioms are a type of *consistency* constraint on partial beliefs[21],

where

[20] In the sense of Kuhn (1996).

[21] Howson (1997c), p. 518.

> [c]onsistency (also known as coherence) means that your evaluations of fair odds are consistent in the sense that they do not depend on the form in which the relevant gambles are presented, and hence are invulnerable to a Dutch Book; consistency is thus an extensional semantic criterion, like truth[22].

In Howson's opinion

> [t]he existence of a well-defined semantics and syntax, with a soundness and completeness theorem, supports the claim that in the Bayesian theory we have a genuine *logic* of consistent belief.[23]

Just as deductive logic is concerned with the question whether the truth of some set of statements necessitates the truth of another statement, the logic of partial belief is concerned with the question whether certain degrees of belief in some statements necessitate a certain degree of belief in another statement; and just as deductive logic is not concerned with the matter-of-fact question whether a given statement is true in some world, the logic of partial belief is not concerned with the matter-of-fact question which degree of belief a given statement is assigned by someone.

> Deductive logic, in other words, provides the conditions regulating what might be called coherent truth-value assignments. This objectivism is nicely paralleled in the interpretation of the probability axioms as the conditions regulating the assignment of coherent betting quotients.[24]

I do not object to considering the probability calculus as a logic of partial belief. But the relativisation to partial beliefs is an important one. I disagree with taking the probability calculus as

> a genuinely *inductive* logic[25],

if an inductive logic is a solution to the problem of a quantitative theory of confirmation.

It is unacceptable to me that the degree of confirmation depends on or is even determined by someone's subjective degrees of belief, as it is according to any

[22]Howson (1997b), pp. S185-S186.
[23]Howson (1997c), p. 521.
[24]Howson (1997c), pp. 521-522.
[25]Howson (1997a), p. 278.

Bayesian theory of confirmation, where these degrees of belief can be arbitrarily chosen (except for coherence with the probability caculus).

To continue Howson's analogy with deductive logic: Just as the logical consequence relation holds independently of the truth values of the statements, the relation of confirmation (to some degree) holds independently of the subjective degrees of belief someone assigns to the statements.

Howson compares coherence with truth, both of which are extensional semantic criteria. In deductive logic, an argument or inference from a set of statements $S = \{P_1, \ldots, P_n\}$ to a statement C may be characterised as deductively valid just in case it holds for all coherent truth value assignments φ to $P_1, \ldots, P_n, C$:

$$\text{If } \varphi(P_1) = \ldots = \varphi(P_n) = 1, \text{ then } \varphi(C) = 1,$$

i.e.

$$\varphi(C \wedge P_1 \wedge \ldots \wedge P_n) \geq \varphi(P_1 \wedge \ldots \wedge P_n) \cdot 1.$$

The analogon of this in inductive logic is to characterise an argument from S to C as inductively valid to degree r, $v(C, S) = r$, just in case (i) it holds for all coherent probability assignments p:

$$p(C \mid P_1 \wedge \ldots \wedge P_n) \geq r,$$

i.e.

$$p(C \wedge P_1 \wedge \ldots \wedge P_n) \geq p(P_1 \wedge \ldots \wedge P_n) \cdot r,$$

and (ii) there is no $s > r$ such that (i) holds for s. Here, a probability assignment is coherent just in case it satisfies the probability axioms – just as a truth value assignment is coherent iff it is a standard evaluation function.

The degree of probabilistic confirmation of T by E, $c(T, E)$, may be then be defined either as $v(T, E)$, in which case one would adopt a measure of confirmation as *firmness*; or else one may define it as, for instance, the greatest number $r \in \Re$ such that it holds for all coherent probability assignments p:

$$p(T \mid E) - p(T) \geq r,$$

i.e.

$$p(T \wedge E) - p(T) \cdot p(E) \geq p(E) \cdot r,$$ [26]

which would be a measure of confirmation as *increase in firmness*.

[26] Note that $c(T, E)$, if defined in this way, need not coincide with the difference between the degree of inductive validity of the inference from E to T, and the degree of inductive validity of the inference from $\top$ to T.

In this case the relation of confirmation to degree r holds independently of the probabilities assigned to the statements, just as the logical consequence relation holds independently of the truth values assigned to the statements.

2.4.1 The Less Reliable the Source of Information, the Higher the Degree of Bayesian Confirmation

Bayesians argue that they can take into account the fact that we are not always sure about our observations or, more generally, that the sources we take our data from are not always reliable.

Let $p_1(E)$ be my subjective degree of belief in (the atomic statement or proposition) E in case (I think) the source of information for E is not fully reliable, say, my subjective degree of belief in 'This chair in my room is red' when looking at my chair at time t_1 at night when the light is off. Let $p_2(E)$ be my subjective degree of belief in E in case (I think) the source of information for E is very reliable, say, my subjective degree of belief in 'This chair in my room is red' when looking at my chair at time t_2 at night when the light is on and I have checked that I am awake. As the source of information is less reliable in the first case than in the second, $p_1(E) < p_2(E)$, where it is assumed that $p_1(E) > 0$.

Consider the hypothesis $T =$ 'All furniture in my room is red', where T is taken to logically imply E (strictly speaking, it does not). According to one Bayesian theory of confirmation[27], the degree of confirmation of T by E[28] at time t_1, where the source of information is (thought to be) unreliable, is

$$d_{p_1}(T,E) = p_1(T \mid E) - p_1(T) = \frac{p_1(T)}{p_1(E)} - p_1(T) \quad T \vdash E.$$

At time t_2, where the source of information is (thought to be) reliable, the degree of confirmation is

$$d_{p_2}(T,E) = p_2(T \mid E) - p_2(T) = \frac{p_2(T)}{p_2(E)} - p_2(T) \quad T \vdash E.$$

There are two possibilities: Either $p_1(T) = p_2(T)$ or $p_1(T) \neq p_2(T)$.

In the first case T is the more confirmed by E, the less reliable the source of information for E, because

$$d_{p_1}(T,E) > d_{p_2}(T,E) \quad \text{iff} \quad p_2(E) > p_1(E).$$

[27] The distance measure d_p is considered by Earman (1992). Although – as shown by Fitelson (2001) – not all relevance measures are ordinally equivalent, I am only considering d_p.

[28] The relativisation to the background knowledge is dropped.

Assuming $p_1(T) = p_2(T)$ may be appropriate, if t_1 and t_2 are two possible scenarios: I consider hypothesis T, and because of knowing that T logically implies E I start to investigate whether E is true. In scenario t_1, the light is off and I have a low degree of belief in E. As a consequence, I get a low degree of confirmation. In scenario t_2, the light is on and I have a high degree of belief in E. Therefore I get a high degree of confirmation.

One may object that this is not very plausible, for a change in my subjective degree of belief in E will give rise to a change in my subjective degree of belief in T. This may be appropriate, if t_1 and t_2 are construed as two successive points of time.

Therefore consider the second case. In my view, the only Bayesian answer to the question whether $p_1(T)$ is smaller or greater than $p_2(T)$ is that the latter is given by Jeffrey conditionalisation (JC) on E, which yields strict conditionalisation in the limiting case of $p_2(E) = 1$:[29]

$$p_2(T) = p_1(T \mid E) \cdot p_2(E) + p_1(T \mid \neg E) \cdot p_2(\neg E) = p_1(T) \cdot \frac{p_2(E)}{p_1(E)} \quad T \vdash E.$$

As in the first case it follows that T is the more confirmed by E, the less reliable the source of information for E, because

$$d_{p_1}(T, E) > d_{p_2}(T, E) \quad \text{iff} \quad p_2(E) > p_1(E).^{30}$$

One[31] might reply that this is just a more general version of the problem of old evidence raised by Glymour (1980a): If E is old evidence so that $p(E) = 1$, then the degree of confirmation of any theory T by E is 0. For if $p(E) = 1$, then $p(T \mid E) = p(T)$, whence

$$d_p(T, E) = p(T \mid E) - p(T) = 0.$$

This is a problem for Bayesian confirmation theory, because there are many historical cases where old evidence provided confirmation to a theory.

[29] It is assumed that in going from t_1 to t_2 the only change is in E. Cf. Jeffrey (1967), esp. chapter 11.

[30] The same holds if $p_1(T)$ is obtained by Jeffrey conditionalisation as

$$p_1(T) = p_2(T \mid E) \cdot p_1(E) + p_2(T \mid \neg E) \cdot p_1(\neg E) = p_2(T) \cdot \frac{p_1(E)}{p_2(E)} \quad T \vdash E.$$

[31] Luc Bovens in personal correspondence.

In general, there are two solutions to the problem of old evidence.[32] The one is to condition on the entailment relation that holds between T and E; the other is to take recourse to counterfactual degrees of belief. I will argue that the first does not give a solution to the example of above, and that (a charitable reformulation of) the second gives no satisfying solution, but illustrates to what – in my opinion: unacceptable – consequences the subjective interpretation of probability as degree of belief leads if it is to provide the basis of a quantitative theory of confirmation.

2.4.1.1 Conditioning on the Entailment Relation

The first strategy is taken by Garber (1983)[33]. He distinguishes between a historical and an ahistorical problem of old evidence: The former

> concerns the scientist in the midst of his investigations who appears to be using a piece of old evidence to increase his confidence in a given theory.[34]

The second problem is that although

> [w]hen we are first learning a scientific theory, we are often in roughly the same epistemic position that the scientist was in when he first put the theory to test; the evidence that served to increase *his* degrees of belief will increase *ours* as well. But having absorbed the theory, our epistemic position changes. [...] Once we have learned the theories, the evidence has done its work on our beliefs, so to speak. But nevertheless, even though the old evidence no longer serves to increase our degrees of belief in the theories in question, there is still a sense in which the evidence in question remains good evidence, and there is still a sense in which it is proper to say that the old evidence confirms the theories in question.[35]

According to Garber[36], the ahistorical problem of old evidence may be solved by some version of the counterfactual strategy. His concern is the historical problem of old evidence. To him

---

[32] For a discussion of the problem of old evidence cf. Curd/Cover (1998), chap. 5. For the following cf. Earman (1992), chap. 5.

[33] A similar account is that of Jeffrey (1983).

[34] Garber (1983), p. 102.

[35] Garber (1983), pp. 102-103.

[36] Garber (1983), p. 103.

> it seems clear that in the cases at hand, what *increases* S's confidence in T is not E itself, but the *discovery* of some generally logical or mathematical relationship between T and E.[37]

Garber considers a language L consisting in the truth-functional closure of a countably infinite collection of atomic statements, to which he adds atomic statements of the form '$A \vdash_G B$', where '$\vdash_G$' stands for some relation of entailment that need not be further specified. The resulting language is denoted by 'L^*'. He then shows[38] that there is at least one probability function p on L^* that satisfies the following condition:

$$(G) \qquad p\left((A \vdash_G B) \wedge A\right) = p\left((A \vdash_G B) \wedge A \wedge B\right),$$

and which is such that

$$0 < p\left(A \vdash_G B\right) < 1,$$

if '$A \vdash_G B$' is an atomic statement of L^*, and A and $\neg B$ are not both tautologies of L. So,

> on [his] construction, it is *not* trivially the case that $p\left(T \mid T \vdash_G E\right) = p\left(T\right)$ when $p\left(E\right) = 1$, and the discovery that $T \vdash_G E$ *can* raise S's confidence in T.[39]

This approach does not solve the more general problem for two reasons. First, one may construe the example in such a way that just because of knowing that T logically implies E I start to investigate whether E is true. In this case $p\left(T \vdash_G E\right) = 1$. Second, by substituting '$T \vdash_G E$' for 'E' in the example one gets the same problem: T is more confirmed by $T \vdash_G E$ at t_1 than at t_2 just in case the source of information for $T \vdash_G E$ at t_1 is less reliable than at t_2. For instance, at t_1 my friend, a student of logic, tells me that T entails E, whereas at T it is her professor who tells me that this is the case, and also shows me how to deduce E from T.

---

[37] Garber (1983), p. 104. I have changed the notation.

[38] Garber (1983), p. 116.

[39] Garber (1983), p. 123. I have changed the notation. More precisely, Garber shows that

> [f]or L and L^* constructed as above, for any atomic sentence of L^* of the form '$A \vdash_G B$' where B is not a truth-functional contradiction in L and where A does not truth-functionally entail $\neg B$ in L and B does not truth-functionally entail A in L, for any r, s in $(0, 1)$, there exists an infinite number of probability functions on L^* that satisfy (G) and are such that $p\left(B\right) = 1$, $p\left(A \vdash_G B\right) = r$, $p\left(A\right) = s$, and $p\left(A \mid A \vdash_G B\right) > p\left(A\right)$.

Garber (1983), pp. 120-121. I have changed the notation.

2.4.1.2 The Counterfactual Strategy

Howson/Urbach (1993) write:

> the support of T by E is gauged according to the effect which one believes a knowledge of E *would now have* on one's degree of belief in T, on the (counter-factual) supposition that one does not yet know E.[40]

I take this to be a standard version of the counterfactual solution to the problem of old evidence.

Let us first see how the counterfactual strategy can solve the problem of old evidence. Let E be an evidential statement or proposition which is old in the sense that $p(E \mid B) = 1$, let T be a theory, let B be the background knowledge, and suppose $B - E$ (read: B without/minus E) is the (up to equivalence) uniquely determined weakest statement with

$$(B - E) \wedge E \dashv\vdash B.$$

According to one reading of the above quotation, the degree of confirmation of T by E relative to B is given by

$$p(T \mid B) - p(T \mid B - E),$$

which is positive if and only if

$$p(E \mid T \wedge (B - E)) > p(E \mid B - E),$$

provided $p(T \mid B - E) > 0$ and $p(B) > 0$. This result seems to be correct: T is confirmed by E relative to B just in case E is likelier given T and the restricted background knowledge $B - E$ – that without (the information bearing on) E – than without T being given.

Before continuing, note that although $p(E \mid B - E) = 1$, if $p(E) = 1$, the problem of old evidence can indeed be solved in this way, for the assumption that E is old evidence has to be expressed as $p(E \mid B) = 1$, from which it does not follow that $p(E) = 1$ or $p(E \mid B - E) = 1$.

More precisely, the probabilities here have to satisfy the following conditions: (1) $p(E \mid B) = 1$, because that constitutes the problem of old evidence; and (2) $p(E \mid B - E) < 1$, for otherwise the problem of old evidence cannot be

[40]Howson/Urbach (1993), pp. 404-405. I have changed the notation.

solved. If one additionally assumes (3) $p(B-E)=1$, because this is taken to follow from the meaning of a background *knowledge*, then one gets

$$p(E \mid B-E) = p(E).$$

Before applying the counterfactual strategy to the example of above, one has to generalize from the strict (i.e. $p(E \mid B)=1$) to the Jeffrey case, where $p(E \mid B)$ need not be 1. Remember that

$$p(T \mid B) = p(T \mid (B-E) \wedge E) \cdot p(E \mid B) + p(T \mid (B-E) \wedge \neg E) \cdot p(\neg E \mid B),$$

if $p(E \mid B)=1$, i.e. if the assumption constituting the problem of old evidence is given. Therefore it seems reasonable to consider

$$\begin{aligned} hu_p(T,E,B) \;:=\; & p(T \mid (B-E) \wedge E) \cdot p(E \mid B) + \\ & + p(T \mid (B-E) \wedge \neg E) \cdot p(\neg E \mid B) - p(T \mid B-E) \end{aligned}$$

as the degree to which E confirms T relative to B. This suggestion is strenghtened by noting that given some provisos, $hu_p(T,E,B)$ is positive if and only if[41]

$$\begin{array}{c} p(E \mid T \wedge (B-E)) > p(E \mid B-E) \quad \text{and} \quad p(E \mid B) > p(E \mid B-E) \\ \text{or} \\ p(E \mid T \wedge (B-E)) < p(E \mid B-E) \quad \text{and} \quad p(E \mid B) < p(E \mid B-E), \end{array}$$

which is the appropriate generalisation of the equivalence of before.

However, the problem is that in the Jeffrey case one does not know E, whence conditioning on $B-E$ – "the (counter-factual) supposition that one does not yet know E" – is of no help. What is needed is the degree of belief in T, on the counterfactual supposition that one does not yet belief in E with degree $p(E \mid B)$. I will consider two ways of arriving at this (counterfactual) degree of belief: A genuinely counterfactual one, and one sticking to actual degrees of belief.

2.4.1.2.1 Counterfactuals Degrees of Belief Let '$B \wr E$' denote the information that is left, when all information that bears on E is dropped from B – thereby neglecting the question what to do in case the degree of belief in E, $p(E \mid B)$, is

[41]The provisos and the calculations for the following claims are to be found in the appendix to chapter 2. The following equivalence is shown to hold by calculation 1.

not only due to information that can be expressed in terms of statements or propositions.[42] Suppose $B \wr E$ is well defined and such that it is the logically strongest consequence of B with $p(E) = p(E \mid B \wr E)$.

I take this independence to follow from the supposition that $B \wr E$ is what remains when all information bearing on E is dropped from B. Note that the independence follows from $p(B \wr E) = 1$, which is not assumed here, but may be assumed with regard to the meaning of a background *knowledge* – though, of course, one must not assume $p(B) = 1$ or $p(E \mid B \wr E) = p(E \mid B)$; otherwise one cannot the solve our problem.

Suppose therefore that $p(T \mid B \wr E)$ is my degree of belief in T, on the counterfactual supposition that I do not yet belief in E with degree $p(E \mid B)$. Consider the E and T of the example, where $T \vdash E$. What is the degree of confirmation of T by E relative to B at time t_2?

According to the above quotation, the calculation has to be based on my subjective degree of belief function at time t_2, p_2, because Howson/Urbach write

> *... would **now** have ...*[43]

So replacing '$B - E$' by '$B \wr E$' in the definition of $hu_p(T, E, B)$ yields[44] that the degree of confirmation of T by E at time t_2 is given by

$$\begin{aligned} hu_{p_2}(T, E, B_2) &= p_2(T \mid (B_2 \wr E) \wedge E) \cdot p_2(E \mid B_2) + \\ &\quad + p_2(T \mid (B_2 \wr E) \wedge \neg E) \cdot p_2(\neg E \mid B_2) - p_2(T \mid B_2 \wr E), \end{aligned}$$

which is positive if and only if

$$\begin{gathered} p_2(E \mid T \wedge (B_2 \wr E)) > p_2(E \mid B_2 \wr E) \quad \text{and} \quad p_2(E \mid B_2) > p_2(E \mid B_2 \wr E) \\ \text{or} \\ p_2(E \mid T \wedge (B_2 \wr E)) < p_2(E \mid B_2 \wr E) \quad \text{and} \quad p_2(E \mid B_2) < p_2(E \mid B_2 \wr E), \end{gathered}$$

where B_2 is the background knowledge at time t_2.

For our example, where T is assumed to logically imply E, this means that at time t_2, T is confirmed by E relative to B_2 just in case my actual degree of

[42] The latter is necessary in order to solve the puzzle, for in the example my degree of belief in E changes exogenously in going from t_1 to t_2.

[43] The bold letters are due to me, but the italics are from the original.

[44] Cf. calculation 1, which does not use $(B - E) \wedge E \dashv\vdash B$, whence the equivalence of before holds also with '$B \wr E$' instead of '$B - E$'. The provisos stated there with '$B \wr E$' substituted for '$B - E$' are assumed to hold here.

belief in E at t_2, $p_2(E \mid B_2)$, is greater than my degree of belief in E on the counterfactual supposition that I do not yet belief in E with degree $p_2(E \mid B_2)$, $p_2(E \mid B_2 \wr E)$. This seems to be reasonable.

Let us now compare the degree of confirmation of T by E at time t_2 with that at t_1. The only change in going from t_1 to t_2 is in E. Therefore it seems justified to assume that $B_1 \wr E \dashv\vdash B_2 \wr E$, although B_1, my background knowledge at t_1, will differ from my background knowledge at t_2, B_2 (for the sake of argument, it is currently assumed that the change in my degree of belief in E in going from t_1 to t_2 is *not* exogenous, but is due to some statement in B_2, which is not in B_1).

In order to solve our problem it has to be assumed that $p_1(B_1) < 1$ and $p_2(B_2) < 1$, though it may be the case that $p_1(B_1 \wr E) = 1$ and $p_2(B_2 \wr E) = 1$. Otherwise

$$p_1(E \mid B_1 \wr E) = p_1(E) = p_1(E \mid B_1), \quad \text{and so} \quad hu_{p_1}(T, E, B_1) = 0;$$

or

$$p_2(E \mid B_2 \wr E) = p_2(E) = p_2(E \mid B_2), \quad \text{and so} \quad hu_{p_2}(T, E, B_2) = 0,$$

provided $0 < p_i(E \mid B_i \wr E) < 1$.

Jeffrey conditionalisation then yields that

$$p_2(T \mid (B_2 \wr E) \wedge \pm E) = p_1(T \mid (B_1 \wr E) \wedge \pm E).$$ [45]

What about $p_2(T \mid B_2 \wr E)$, my degree of belief in T at t_2 on the counterfactual supposition that I do not yet belief in E with degree $p_2(E \mid B_2)$? Should this also be the result of conditioning on E? A little bit calculation yields that

$$\begin{aligned} p_2(T \mid B_2 \wr E) \quad = \quad & \frac{p_1(T \mid B_1 \wr E)}{p_1(E) \cdot (1 - p_1(E))} \cdot \\ & \cdot [p_1(E \mid T \wedge (B_1 \wr E)) \cdot (p_2(E \mid B_2) - p_1(E)) + \\ & + p_1(E) \cdot (1 - p_2(E \mid B_2)], \end{aligned}$$

which is equal to

$$p_2(T \mid B_2 \wr E) = p_1(T \mid B_1 \wr E) \cdot \frac{p_2(E \mid B_2)}{p_1(E \mid B_1 \wr E)},$$

[45] Cf. calculation 2 and the provisos stated there. This holds also for counterfactual Jeffrey conditionalisation, which results from Jeffrey conditionalisation by substituting '$p_2(\pm E \mid B_2 \wr E)$' for '$p_2(\pm E \mid B_2)$'.

if $T \vdash E$.[46] This means that my degree of belief in T at t_2 on the counterfactual supposition that I do not yet belief in E with degree $p_2(E \mid B_2)$ is greater than my degree of belief in T at t_1 on the counterfactual supposition that I do not yet belief in E with degree $p_1(E \mid B_1)$ just in case

$$\begin{array}{c} p_1(E \mid T \wedge (B_1 \wr E)) > p_1(E \mid B_1 \wr E) \quad \text{and} \quad p_2(E \mid B_2) > p_1(E \mid B_1 \wr E) \\ \text{or} \\ p_1(E \mid T \wedge (B_1 \wr E)) < p_1(E \mid B_1 \wr E) \quad \text{and} \quad p_2(E \mid B_2) < p_1(E \mid B_1 \wr E), \end{array}$$

provided $0 < p_1(E) < 1$ and $p_1(T \mid B_1 \wr E) > 0$, where $B_1 \wr E \dashv\vdash B_2 \wr E$ and $p_1(E) = p_1(E \mid B_1 \wr E)$.

However, these assumptions yield the following oddity.[47]

Observation 2.1 (Oddity) Suppose

$$B_1 \wr E \dashv\vdash B_2 \wr E \quad \text{and} \quad p_1(E \mid B_1 \wr E) = p_1(E).$$

If $p_2(T \mid B_2 \wr E)$ is the result of Jeffrey conditioning on E, then

$$\begin{array}{c} hu_{p_1}(T, E, B_1) \quad > \quad hu_{p_2}(T, E, B_2) \\ \text{iff} \\ p_1(E \mid T \wedge (B_1 \wr E)) > p_1(E \mid B_1 \wr E) \quad \text{and} \quad p_1(E \mid B_1) > p_1(E \mid B_1 \wr E) \\ \text{or} \\ p_1(E \mid T \wedge (B_1 \wr E)) < p_1(E \mid B_1 \wr E) \quad \text{and} \quad p_1(E \mid B_1) < p_1(E \mid B_1 \wr E). \end{array}$$

In case $T \vdash E$, this means that

$$hu_{p_1}(T, E, B_1) > hu_{p_2}(T, E, B_2) \quad \text{iff} \quad p_1(E \mid B_1) > p_1(E \mid B_1 \wr E).$$

It seems to be rather clear that this oddity arises from obtaining $p_2(T \mid B_2 \wr E)$ by Jeffrey conditionalisation on E. This is not allowed, because $p_2(T \mid B_2 \wr E)$ should express my degree of belief in T at t_2 on the counterfactual supposition that I do not yet believe in E with degree $p_2(E \mid B_2)$.

$p_2(T \mid B_2)$ should not be obtained by Jeffrey conditionalisation on E, but by *counterfactual* Jeffrey conditionalisation on E, which is just JC but with *my degree of belief in E at t_2 on the counterfactual supposition that I do not yet belief*

[46] Cf. calculation 3.

[47] Cf. calculation 4, which also gives the provisos under which the following holds.

in E with degree $p_2(E \mid B_2)$, $p_2(E \mid B_2 \wr E)$, instead of my actual degree of belief in E at t_2, $p_2(E \mid B_2)$, i.e.

$$\begin{aligned} p_2(T \mid B_2 \wr E) &= p_1(T \mid (B_2 \wr E) \wedge E) \cdot p_2(E \mid B_2 \wr E) + \\ &\quad + p_1(T \mid (B_2 \wr E) \wedge \neg E) \cdot p_2(\neg E \mid B_2 \wr E), \end{aligned}$$

which reduces to

$$p_1(T \mid B_1 \wr E) \cdot \frac{p_2(E \mid B_2 \wr E)}{p_1(E \mid B_1 \wr E)},$$

if $T \vdash E$, provided $p_1((B_1 \wr E) \wedge E) > 0$.

This means that my degree of belief in T at t_2 on the counterfactual supposition that I do not yet belief in E with degree $p_2(E \mid B_2)$, $p_2(T \mid B_2 \wr E)$, equals my degree of belief in T at t_1 on the counterfactual supposition that I do not yet belief in E with degree $p_1(E \mid B_1)$, $p_1(T \mid B_1 \wr E)$, just in case[48]

$$p_1(E \mid T \wedge (B_1 \wr E)) = p_1(E \mid B_1 \wr E) \quad \text{or} \quad p_1(E \mid B_1 \wr E) = p_2(E \mid B_2 \wr E).$$

In case $T \vdash E$ this means that, given $0 < p_1(E \mid B_1 \wr E) < 1$,

$$p_2(E \mid B_2 \wr E) = p_1(E \mid B_1 \wr E)$$

is necessary and sufficient for

$$p_2(T \mid B_2 \wr E) = p_1(T \mid B_1 \wr E).$$

Assuming the latter seems to be natural, for, after all, $p_i(E \mid B_i \wr E)$ is my degree of belief in E at t_i on the counterfactual supposition that I do not yet believe in E with degree $p_i(E \mid B_i)$, and the only change in going from t_1 to t_2 is in E.

And indeed – as shown by the theorem below – with these assumptions one gets the desired result that T is more confirmed by E, which is assumed to be positively relevant for T, at t_2 than at t_1 if and only if the source of information for E is more reliable at t_2 than at t_1. More generally (T is not assumed to logically imply E):

Theorem 2.1 (NecSuff) Given

$$B_1 \wr E \dashv\vdash B_2 \wr E, \quad p_1(E \mid B_1 \wr E) = p_1(E), \quad \text{and} \quad p_1(T \mid B_1 \wr E) > 0,$$

[48] Provided $p_1(T \mid B_1 \wr E) > 0$. Cf. the proof of theorem 2.1.

the equality

$$p_1 (T \mid B_1 \wr E) = p_2 (T \mid B_2 \wr E)$$

is necessary and sufficient for the equivalence

$$\begin{array}{rcl} hu_{p_2} (T, E, B_2) & > & hu_{p_1} (T, E, B_1) \\ & \text{iff} & \\ p_1 (E \mid T \wedge (B_1 \wr E)) > p_1 (E \mid B_1 \wr E) & \text{and} & p_2 (E \mid B_2) > p_1 (E \mid B_1) \\ & \text{or} & \\ p_1 (E \mid T \wedge (B_1 \wr E)) < p_1 (E \mid B_1 \wr E) & \text{and} & p_2 (E \mid B_2) < p_1 (E \mid B_1), \end{array}$$

provided

$$\begin{array}{l} p_2 (B_2) > 0, \quad p_2 ((B_2 \wr E) \wedge E) > 0, \\ p_2 ((B_2 \wr E) \wedge \neg E) > 0, \quad \text{and} \quad 1 > p_1 (E) > 0. \end{array}$$

With counterfactual Jeffrey condition this means that

$$p_1 (E \mid T \wedge (B_1 \wr E)) = p_1 (E \mid B_1 \wr E) \quad \text{or} \quad p_1 (E \mid B_1 \wr E) = p_2 (E \mid B_2 \wr E)$$

is necessary and sufficient for this equivalence.

The result obtained seems to be the intuitively correct answer. Yet, is it in accordance with what Howson/Urbach say on the problem of old evidence? According to them, the source of the latter lies

> in relativising all the probabilities to the *totality* of current knowledge. They should, of course, have been relativised to current knowledge *minus* E. The reason for the restriction is, of course, that *your current assessment of the support of* T *by* E *measures the extent to which, in your opinion, the addition of* E *to your current stock of knowledge would cause a change in your degree of belief in* T.[49]

As noted, $B - E$ cannot be taken in the Jeffrey case where I do not know E, for in this case E is not part of B, whence $B = B - E$. Howson/Urbach say that I have to measure the extent to which, in my opinion, the addition of E to my current stock of knowledge (minus E) would cause a change in my degree of belief in T. But

[49] Howson/Urbach (1993), p. 404. I have changed the notation.

to consider that change in case E is not part of B, i.e. $p(T \mid B \wedge E) - p(T \mid B)$, where $B \nvdash E$, takes us back to where we have started off, for

$$\begin{array}{rcl} p_1(T \mid B \wedge E) - p_1(T \mid B) & > & p_2(T \mid B \wedge E) - p_2(T \mid B) \\ & \text{iff} & \\ p_2(T \mid B) & > & p_1(T \mid B) \quad JC \\ & \text{iff} & \\ p_1(E \mid T \wedge B) > p_1(E \mid B) & \text{and} & p_2(E \mid B) > p_1(E \mid B) \\ & \text{or} & \\ p_1(E \mid T \wedge B) < p_1(E \mid B) & \text{and} & p_2(E \mid B) < p_1(E \mid B), \end{array}$$

provided $0 < p_1(E) < 1$ and $p_1(T \mid B) > 0$.[50]

So, what is wrong with the Howson/Urbach-prescription? In my opinion the trouble is caused by their relativisation to my *current* stock of knowledge. The latter may contain information highly relevant for E, although it does not contain E itself. In this case I may already be quite sure of E and assign it a

[50] For a Bayesian, it is also no help to consider

$$p_1(T \mid (B_1 \wr E) \wedge E) - p_1(T \mid B_1 \wr E) \quad \geq \quad p_2(T \mid (B_2 \wr E) \wedge E) - p_2(T \mid B_2 \wr E),$$

since (counterfactual) Jeffrey conditionalisation yields

$$p_2(T \mid (B_2 \wr E) \wedge E) = p_1(T \mid (B_1 \wr E) \wedge E),$$

and $p_2(T \mid B_2 \wr E)$ is definitely not smaller than $p_1(T \mid B_1 \wr E)$. If $p_2(T \mid B_2 \wr E)$ is obtained by counterfactual Jeffrey conditionalisation, then

$$\begin{array}{rcl} p_2(T \mid B_2 \wr E) & > & p_1(T \mid B_1 \wr E) \\ & \text{iff} & \\ p_1(E \mid T \wedge (B_1 \wr E)) > p_1(E \mid B_1 \wr E) & \text{and} & p_2(E \mid B_2 \wr E) > p_1(E \mid B_1 \wr E) \\ & \text{or} & \\ p_1(E \mid T \wedge (B_1 \wr E)) < p_1(E \mid B_1 \wr E) & \text{and} & p_2(E \mid B_2 \wr E) < p_1(E \mid B_1 \wr E), \end{array}$$

and

$$\begin{array}{rcl} p_2(T \mid B_2 \wr E) & = & p_1(T \mid B_1 \wr E) \\ & \text{iff} & \\ p_1(E \mid T \wedge (B_1 \wr E)) = p_1(E \mid B_1 \wr E) & \text{or} & p_2(E \mid B_2 \wr E) = p_1(E \mid B_1 \wr E). \end{array}$$

Cf. calculation 3 and the provisos stated there.

very high degree of belief. As a consequence, the extent to which, in my opinion, the additon of E to my *current* stock of knowledge would cause a change in my degree of belief in T is usually only very small. In the limiting case, where E is known (in the sense of being assigned a degree of belief of 1), there is no increase at all. Here Howson/Urbach tell one to drop E; yet in case E is not known, but only believed (in the sense that $p(E) < 1$), E cannot be dropped. So their solution to the problem of old evidence is no genuine solution, because it is no solution to the more general problem.

I think a Bayesian has to make two corrections. First she should consider

> the extent to which, in her opinion, the addition of E to some part of her stock of knowledge which contains no information bearing on E, e.g. $B \wr E$, would cause a change in her degree of belief in T.

Second, she should additionally take into account her actual degree of belief in E (cf. the preceding footnote).

The result I arrived at with the above prescription seemed to be correct according to Bayesian intuitions: T (which logically implies E) is more confirmed by E at t_2 than at t_1 if and only if the source of information for E is more reliable at t_2 than at t_1, where it is assumed that $p_1(T \mid B_1 \wr E) = p_2(T \mid B_2 \wr E)$, which I derived with counterfactual Jeffrey conditionalisation and by assuming $p_1(E \mid B_1 \wr E) = p_2(E \mid B_2)$.

But what are these degrees of belief on counterfactual suppositions; and how are they related to my actual degrees of belief? After all, Bayesian confirmation theory aims at determining the degree of confirmation by means of someone's actual degrees of belief. Finding the strongest consequence (or subset) $B \wr E$ of B with $p(E \mid B \wr E) = p(E)$ is not only a difficult task; it may even be an impossible one, for there may be several $B \wr E$ which are probabilistically independent of E (in the sense of p), but which cannot be compared with respect to their logical strength.

It also remains questionable what to do in case my degree of belief in E changes exogenously in going from t_1 to t_2, for here $B_1 \dashv\vdash B_2$.

Furthermore, in order to obtain the desired result it was – and had to be – assumed that $p_1(T \mid B_1 \wr E) = p_2(T \mid B_2 \wr E)$. Given counterfactual Jeffrey conditionalisation, this reduces to assuming $p_1(E \mid B_1 \wr E) = p_2(E \mid B_2 \wr E)$, provided E is positively relevant for T (under p_1). With the independence assumption given by the intended meaning of $B \wr E$, it follows that

$$p_1(E) = p_1(E \mid B_1 \wr E) = p_2(E \mid B_2 \wr E) = p_2(E).$$

This has to hold not only for t_1 and t_2, a particular E, and a particular T, but for *all* time points t_i and t_j, for *every* piece of evidence E, and for *every* theory T.

That is, it has to hold for every theory T, every evidence E, and all points of time t_i and t_j: $p_i\left(T \mid B_i \wr E\right) = p_j\left(T \mid B_j \wr E\right)$, and, given the independence assumption and counterfactual Jeffrey conditionalisation,

$$p_i\left(E\right) = p_i\left(E \mid B_i \wr E\right) = p_j\left(E \mid B_j \wr E\right) = p_j\left(E\right).$$

This means that in order to avoid our problem, and to get confirmation right, the counterfactual degrees of belief in T and E have to be stable over time. In particular, $p_0\left(T \mid B_0 \wr E\right) = p_i\left(T \mid B_i \wr E\right)$, and

$$p_0\left(E\right) = p_0\left(E \mid B_0 \wr E\right) = p_i\left(E \mid B_i \wr E\right) = p_i\left(E\right),$$

for every theory T, every evidence E, and every point of time t_i.

So the degree of confirmation of T by E at time t_i is given by

$$\begin{aligned} hu_{p_i}\left(T, E, B_i\right) \;&:=\; p_i\left(T \mid \left(B_i \wr E\right) \wedge E\right) \cdot p_i\left(E \mid B_i\right) + \\ &\quad + p_i\left(T \mid \left(B_i \wr E\right) \wedge \neg E\right) \cdot p_i\left(\neg E \mid B_i\right) - p_0\left(T \mid B_0 \wr E\right), \end{aligned}$$

where $p_0\left(T\right) = p_0\left(T \mid B_0 \wr E\right)$, if, as seems to be justified in view of the meaning of a background *knowledge*, $p_0\left(B_0 \wr E\right) = 1$, or, more generally, $p_i\left(B_i \wr E\right) = 1$.

Here, B_i is the background knowledge at time t_i, $B_i \wr E$ is what remains of B_i if all information bearing on E is dropped from B_i, and t_0 is the first point of time *in* the beginning when I first built up my probability space and made my absolutely first guess in terms of p_0.

Since Jeffrey conditionalisation and $B_i \wr E \dashv\vdash B_j \wr E$ – both of which are justified by assuming that the only change in going from t_i to t_j is in E – yield that $p_i\left(T \mid B_i \wr E\right) = p_j\left(T \mid B_j \wr E\right)$, it follows that

$$\begin{aligned} hu_{p_i}\left(T, E, B_i\right) \;&=\; p_0\left(T \mid \left(B_0 \wr E\right) \wedge E\right) \cdot p_i\left(E \mid B_i\right) + \\ &\quad + p_0\left(T \mid \left(B_0 \wr E\right) \wedge \neg E\right) \cdot p_i\left(\neg E \mid B_i\right) - p_0\left(T \mid B_0 \wr E\right). \end{aligned}$$

In other words, the degree of confirmation of T by E crucially depends on my absolutely first guess in terms of p_0!

Before trying to relate $p_0\left(T \mid \left(B_0 \wr E\right) \wedge E\right)$ and $p_0\left(T \mid B_0 \wr E\right)$ to my actual degrees of belief[51], and discussing the consequences of all this, let us see

---

[51] It seems reasonable to set

$$p_0\left(T \mid B_0 \wr E\right) = p_0\left(T\right) \quad \text{and} \quad p_0\left(T \mid \left(B_0 \wr E\right) \wedge E\right) = p_0\left(T \mid E\right),$$

whether these difficulties can be overcome by keeping more in touch with reality, and by sticking to the Bayesian aim of determining the degree of confirmation in terms of someone's *actual* degrees of belief. In particular, this seems to be a good advice with regard to the fact that we still do not have a solution for the case where my degree of belief in E changes exogenously in going from t_1 and t_2.

2.4.1.2.2 Actual Degrees of Belief In the preceding paragraph I tried to use the counterfactual approach to the problem of old evidence to solve the more general puzzle presented at the beginning of this section. The problems with it are (1) how to obtain the counterfactual degrees of belief from the actual ones, (2) what to do with exogenous belief changes, and (3) that the degree of confirmation at any time t_i crucially depends on my first guess in terms of p_0 – that (3) is the most serious of these problems is argued for later on. In this paragraph I will therefore try to determine my degree of belief in T on the counterfactual supposition that I do not yet believe in E to some degree by keeping more in touch with reality in the sense of using only actual degrees of belief.

Remember: In case of known evidence E Howson/Urbach tell one to consider "the extent to which, in your opinion, the addition of E to your current stock of knowledge would cause a change in your degree of belief in T." In case E is not known but only believed, it may therefore be appropriate to consider the extent to which, in my opinion, coming to believe E with degree $p(E)$ would cause a change in my degree of belief in T, where background knowledge B is suppressed.

In terms of actual degrees of belief, this extent, which should yield the degree of confirmation of T by E at t_2, may be measured in one of the following two ways: Either by the difference between my (actual) degree of belief in T at t_1 (where I do not yet believe in E with degree $p_2(E)$) conditional on the evidence E, and my degree of belief in T at t_1 *before* I came to believe in E with degree $p_2(E)$, i.e.

$$a_{p_2}(T, E) = p_1(T \mid E) - p_1(T),$$

which is positive if and only if

$$p_1(E \mid T) > p_1(E),$$

both of which are consequences, if – as is natural for a (restricted) background knowledge – $B_0 \wr E$ is assigned a degree of belief of 1. This is even more so, if it is assumed that *in* the beginning there is *no* background knowledge at all, so that $B_0 \wr E \dashv\vdash B_0 \dashv\vdash \top$. Let me stress that whether or not this is the case does not affect the discussion here.

provided $p_1(T) > 0$ and $p_1(E) > 0$, where it is assumed that my degree of belief in E changes exogenously in going from t_1 to t_2 so that the background knowledge, which is suppressed, is the same at t_1 and at t_2.[52]

Or else, the degree of confirmation is measured by the difference between my actual degree of belief in T at t_2, and my degree of belief in T at t_1 *before* I came to believe in E with degree $p_2(E)$, i.e.

$$b_{p_2}(T, E) = p_2(T) - p_1(T),$$

which is positive just in case

$$\begin{gathered} p_1(E \mid T) > p_1(E) \quad \text{and} \quad p_2(E) > p_1(E) \\ \text{or} \\ p_1(E \mid T) < p_1(E) \quad \text{and} \quad p_2(E) < p_1(E), \end{gathered}$$

provided $p_1(T) > 0$ and $1 > p_1(E) > 0$, where Jeffrey conditionalisation has been used.

This means that T is confirmed by E at time t_2 either iff T is positively relavant for E in the sense of p_1; or iff *in addition* to this, my degree of belief in E increased in passing from t_1 to t_2.

For a Bayesian, at least the second option seems to be reasonable – or so I think. Note that in order to get the degree of confirmation for the example, where T logically implies E, it must be assumed that I am logically omniscient in the first sense that all logical truths are transparant to me.[53]

So far, so good. Now consider the degree of confirmation of T by E at time t_1. Here, I have to consider my subjective degree of belief function p_0 at time t_0, where t_0 is the point of time just before t_1. In order to arrive at

$$p_0(T \mid E) - p_0(T) \quad \text{and} \quad p_1(T) - p_0(T),$$

I have to assume that I am logically omniscient in the second sense that I am aware of all statements or propositions in my probability space (otherwise it is not guaranteed that $p_0(E), \ldots$ are defined).

[52] Note that I cannot consider

$$p_1(T \mid E) \cdot p_1(E) + p_1(T \mid \neg E) \cdot p_1(\neg E) - p_1(T),$$

because this is always 0.

[53] Cf. Earman (1992), p. 122.

Suppose again that the only change in my degree of belief in passing from t_0 to t_1 is in E, where $p_0(E)$ is my subjective degree of belief in 'This chair in my room is red' at time t_0 at night when I wake up because of some noice, but before I am looking at my chair at time t_1 when the light is off. The source of information for E at t_0 is less reliable than that at t_1, because at t_0 I am not even looking at my chair, whence $p_0(E) < p_1(E)$, where $p_0(E)$ is assumed to be positive.

Calculating the degree of confirmation yields that in both cases T is more confirmed by E at t_1 than at t_2. More generally, it holds that

$$\begin{array}{ccc} a_{p_1}(T,E) & > & a_{p_2}(T,E) \\ & \text{iff} & \\ p_0(E \mid T) > p_0(E) & \text{and} & p_1(E) > p_0(E) \\ & \text{or} & \\ p_0(E \mid T) < p_0(E) & \text{and} & p_1(E) < p_0(E), \end{array}$$

and

$$\begin{array}{ccc} b_{p_1}(T,E) & > & b_{p_2}(T,E) \\ & \text{iff} & \\ p_0(E \mid T) > p_0(E) & \text{and} & p_1(E) - p_0(E) > p_2(E) - p_1(E) \\ & \text{or} & \\ p_0(E \mid T) < p_0(E) & \text{and} & p_1(E) - p_0(E) < p_2(E) - p_1(E), \end{array}$$

provided $p_0(T) > 0$ and $1 > p_0(E) > 0$.

What went wrong? I think it is obvious that at t_2 I must not consider my subjective degree of belief in T at t_1, $p_1(T)$, but my subjective degree of belief in T at time t_0, $p_0(T)$, and that therefore the degree of confirmation of T by E at t_2 is given by

$$a'_{p_2}(T,E) = p_0(T \mid E) - p_0(T)$$

or

$$b'_{p_2}(T,E) = p_2(T) - p_0(T),$$

where the latter is positive if and only if

$$\begin{array}{ccc} p_0(E \mid T) > p_0(E) & \text{and} & p_2(E) > p_0(E) \\ & \text{or} & \\ p_0(E \mid T) < p_0(E) & \text{and} & p_2(E) < p_0(E), \end{array}$$

provided $p_0(T) > 0$ and $1 > p_0(E) > 0$. In this case the desired result follows indeed for the second measure, since

$$\begin{array}{rcl} b'_{p_2}(T,E) & > & b'_{p_1}(T,E) \\ & \text{iff} & \\ p_2(T) - p_0(T) & > & p_1(T) - p_0(T) \\ & \text{iff} & \\ p_1(E \mid T) > p_1(T) & \text{and} & p_2(E) > p_1(E) \\ & \text{or} & \\ p_1(E \mid T) < p_1(T) & \text{and} & p_2(E) < p_1(E), \end{array}$$

provided $p_1(T) > 0$ and $1 > p_1(E) > 0$. In the first case, the degree of confirmation of T by E at t_1 does not differ from that at t_2.

As before, this has to hold not only for t_1 and t_2, but for any time points t_i and t_j, for every piece of evidence E, and every theory T. So is

$$\begin{array}{rcl} c_{p_i}(T,E) & := & p_i(T) - p_0(T) \\ & = & p_i(T \mid E) \cdot p_i(E) + p_i(T \mid \neg E) \cdot p_i(\neg E) - p_0(T) \\ & = & p_0(T \mid E) \cdot p_i(E) + p_0(T \mid \neg E) \cdot p_i(\neg E) - p_0(T) \end{array}$$

the solution to the puzzle[54]; the one which gives the degree of confirmation of T by E at time t_i without recourse to counterfactual degrees of belief, and which can also deal with exogenous belief changes? Note that c_{p_i} is very similar to hu_{p_i}; indeed, they coincide if

$$p_0(T \mid B_0 \wr E) = p_0(T) \quad \text{and} \quad p_0(T \mid (B \wr E) \wedge E) = p_0(T \mid E),$$

both of which are consequences of setting $p_0(B_0 \wr E) = 1$, which, as already mentioned several times, seems to be natural for a background knowledge, even more so, if it is restricted.

I think c_p – or its counterfactual relative hu_p – are the best response a Bayesian can give to the puzzle under consideration. Yet, they do not provide an adequate measure of confirmation in terms of degrees of belief, but show what is at the heart of confirmation theory. As there may be times before t_0, one has

[54]The last equality holds, because the only change is in E. The important point – namely that the degree of confirmation at any time t_i crucially depends on my first guess in terms of p_0 – is also true without this assumption.

to consider the earliest time when E first appeared in the probability space. This amounts to consider the point of time in my history, say t^*, when I built up my probability space and made my absolutely first assignment p^*.

In order for $p_j(T)$ to be defined, where t_j is any point of time after t^*, one first has to assume that $p_i(E) > 0$, for every $i < j$, for otherwise one cannot condition on E. In particular, this holds of $p^*(E)$.

If T logically implies E as in the example, then the degree of confirmation of T by E at any time t_i is uniquely determined by my actual degree of belief in E at t_i, $p_i(E)$, and my first guesses in E and T, $p^*(E)$ and $p^*(T)$. That is, my absolutely first assignment p^* uniquely determines the degree of confirmation of T by E at any time t_i in case E is known and logically implied by T!

Why the exclamation mark? The reason is that this shows that the idea behind any Bayesian theory of confirmation – namely to determine the degree of confirmation by means of someone's degrees of belief – fails. For what is this absolutely first assignment p^*? *Any arbitrary assignment of values in* $[0,1]$ to the atomic statements – among which I take to be at least E – is consistent/coherent with the axioms of the probability calculus, whence any possible value for $c_p(T,E)$ can be obtained as degree of confirmation of T by E – at least, if $T \vdash E$.[55] For let r be any possible value for $c_{p_i}(T,E)$, i.e. let

$$r \in [p_i(T) - p_i(T \mid E), p_i(T)).^{56}$$

Then the function p^*,

$$p^*(E) \quad := \quad \frac{p_i(T \mid E) \cdot p_i(E) - r}{p_i(T \mid E)} = \frac{p_i(T) - r}{p_i(T \mid E)},$$

[55]For reasons of time the following can only be conjectured at the moment:

Conjecture 2.1 (Anything Goes) For any Boolean algebra of propositions $\mathcal{M}$, for any probability function $p_{i,\mathcal{M}}$ defined on $\mathcal{M}$, for any two propositions T and E of $\mathcal{M}$, and for any possible value r for $c_{p_{i,\mathcal{M}}}(T,E)$ there exists a probability function $p^*_{\mathcal{M}}$ on $\mathcal{M}$ such that (i) $p_{i,\mathcal{M}}$ results from $p^*_{\mathcal{M}}$ by i times Jeffrey conditioning on E, and (ii)

$$c_{p_{i,\mathcal{M}}}(T,E) = p_{i,\mathcal{M}}(T) - p^*_{\mathcal{M}}(T) = r.$$

As I got to know only shortly before finishing this dissertation, there is a similar point in Albert (2001), which I can only refer to.

[56]r cannot be smaller than $p_i(T) - p_i(T \mid E)$, because $T \vdash E$, whence

$$p^*(T) \leq p^*(T \mid E) = p_i(T \mid E).$$

$$\begin{aligned} p^*(\cdot \mid \pm E) &:= p_i(\cdot \mid \pm E), \\ p^*(\cdot) &:= p_i(\cdot \mid E) \cdot p^*(E) + p_i(\cdot \mid \neg E) \cdot p^*(\neg E), \end{aligned}$$

is a probability function (defined on the same language as p_i) which yields that the degree of confirmation of T by E at time t_i equals r, and where p_i results from p^* by Jeffrey conditioning on E.[57]

It seems that we are back at the problem of assigning prior probabilities: According to Earman (1992), there are three answers to this problem.

> The first is that the assignment of priors is not a critical matter, because as the evidence accumulates, the differences in priors "wash out." [...] it is fair to say that the formal results apply only to the long run and leave unanswered the challenge as it applies to the short and medium runs. [...] The second response is to provide rules to fix the supposedly reasonable initial degrees of belief. [...] We saw that, although ingenious, Bayes's attempt is problematic. Other rules for fixing priors suffer from similar difficulties. And generally, none of the rules cooked up so far are capable of coping with the wealth of information that typically bears on the assignment of priors. [...] The third response is that while it may be hopeless to state and justify precise rules for assigning numerically exact priors, still there are plausibility considerations that can be used to guide the assignments. [...] This response [...] opens the Bayesians to a new challenge[.] [...] That is, Bayesians must hold that the appeal to plausibility arguments does not commit them to the existence of a logically prior sort

[57] It suffices to show that $0 \leq p^*(E) \leq 1$, and that $c_{p_i}(T, E) = r$. The former holds, because $r < p_i(T)$ and $p_i(T \mid E) \cdot p_i(E) - p_i(T \mid E) \leq r$; the latter holds, because

$$\begin{aligned} c_{p_i}(T, E) &= p_i(T) - p^*(T) \\ &= p_i(T \mid E) \cdot p_i(E) - p^*(T) \quad T \vdash E \\ &= p^*(T \mid E) \cdot p_i(E) - p^*(T) \quad JC \\ &= \frac{p^*(T) \cdot p_i(E)}{p^*(E)} - p^*(T) \quad T \vdash E \\ &= \frac{(p_i(T) - r) \cdot p_i(E) \cdot p_i(T \mid E)}{p_i(T) - r} - (p_i(T) - r) \\ & \quad p^*(T) = p_i(T \mid E) \cdot p^*(E) = p_i(T) - r \\ &= r. \end{aligned}$$

> of reasoning: plausibility assessment. Plausibility arguments serve to marshall the relevant considerations in a perspiciuous form, yet the assessment of these considerations comes with the assignment of priors. But, of course, this escape succeeds only by reactivating the original challenge. The upshot seems to be that some form of the washout solution had better work not just for the long run but also for the short and medium runs as well.[58]

I take the standard Bayesian answer to be that differences in the priors do not matter, because they are "washed out" in the long run.

The point of the above example is that the limiting theorems of convergence to certainty and merger of opinion are of no help, and would not even be of help, if they worked for the medium and short runs: It shows that differences in the priors do matter. For in case T logically implies E my first guess in E, $p^{*}(E)$, can be used to obtain any possible value for $c_p(T, E)$ as degree of confirmation of T by E (in the sense of c_p) – provided E is among the atomic statements.

I do not see how this difficulty can be overcome – and how one can intersubjectify (objectify) Bayesian confirmation theory – without recourse to some objective (logical) prior probability function p^{*}.

However – and that is the pinpointing upshot of all this – the difficulty of determining such an objectively reasonable or logical probability function p^{*} was just the reason for turning to the subjective interpretation.

2.4.2 Steps Towards a Constructive Probabilism

As mentioned, I think the main problem of Bayesianism is its arbitrariness, which is caused by the fact that the three (four) axioms of the probability calculus are far too inclusive in the sense that any assignment of values in $[0, 1]$ to the atomic statements of the underlying language is consistent with these axioms – and the subjective interpretation of probability as degree of belief does not put any restrictions on these assignments.

This suggests the following way out: To add new axioms to the three (four) axioms of the probability calculus so that the set of all (unconditional) probabilities is narrowed down more and more. A first step along these lines is Abner Shimony's strengthening of the second axiom to

$$p(A) = 1 \quad \text{iff} \quad \vdash A, \quad \text{for every wff } A \in \mathcal{L}_{PC},^{59}$$

[58] Earman (1992), p. 57-59.

which narrows down the set of all (unconditional) probabilities P to the set of all strict (unconditional) probabilities P_{strict}.

For a given problem, e.g. the problem of a quantitative theory of confirmation, the aim is to restrict P in such a way that the *order* induced by the probabilistic measure of confirmation C_p among all theories T, evidences E, and background knowledges B, is the same for every (unconditional) probability p. For then it does not matter which (unconditional) probability p the measure of confirmation C_p is based on in order to determine whether T_1 is more confirmed by E_1 relative to B_1 than T_2 by E_2 relative to B_2, for any theories T_1, T_2, evidences E_1, E_2, and background knowledges B_1, B_2.

Of course, the *values* $C_p(T, E, B)$ of C_p for given T, E, and B will vary with the (unconditional) probability p. But there will be no theories T_1, T_2, evidences E_1, E_2, background knowledges B_1, B_2, and (unconditional) probabilities p, p' such that

$$C_p(T_1, E_1, B_1) > C_p(T_2, E_2, B_2) \quad \text{and} \quad C_{p'}(T_1, E_1, B_1) < C_{p'}(T_2, E_2, B_2)\,;$$

and this is enough in order for the probabilistic measure of confirmation

$$C = \{C_p(\cdot, \cdot, \cdot) : p \text{ is a(n) (unconditional) probability}\}$$

to implicitely provide a rule of acceptance for rational theory choice.

Another, perhaps more promising strategy may be sketched as follows: In general, a Bayesian approach to some problem is a probabilistic modeling of the problem under consideration. If the model thus established is dependent on particular values of the (unconditional) probabilities p used in it, and if it varies with varying p – as is the case for any Bayesian relevance measure – then, other things being equal, this model will be arbitrary.

[59] Cf. Shimony (1955). Shimony considers conditional probabilities as primitive, but the reformulation in terms of (unconditional) probabilities roughly amounts to the same, since a(n) (unconditional) probability $p(\cdot)$ can be defined in terms of a conditional probability $p(\cdot \mid \cdot)$ as

$$p(A) := p(A \mid \top)\,, \quad \text{for every wff } A \in \mathcal{L}_{PC},$$

and – if (unconditional) probabilities are defined in this – it can be shown that $p(\cdot)$ is a strict (unconditional) probability, if $p(\cdot \mid \cdot)$ is a strict conditional probability in the sense that

$$p(B \mid A) = 1 \quad \text{iff} \quad A \vdash B, \quad \text{for any wffs } A, B \in \mathcal{L}_{PC} \text{ with } A \nvdash \bot.$$

If, however, the problem in question can be modeled not by functions f_p depending on particular values of the (unconditional) probabilities p, but by a set of (un)equations between the functions f_p, then the model is not exposed to the argument of arbitrariness, if – as is to be expected – this set of (un)equations determines a non-arbitrary set F of functions f_p. Probabilistic modeling along these lines is illustrated by Bovens/Olsson (2000), whereas the ordinal measure of coherence defined in Hartmann/Bovens (2000) is arbitrary[60].

A third approach towards solving the problem of arbitrariness for Bayesianism is to define (for as many statements as possible) a set of uniquely determined – in some[61] sense: logical – (conditional or unconditional) probabilities. In the limiting (and most desirable, but hardly imaginable) case, this set consists of a single probability p^* which is defined for all (sets of) statements (of $\mathcal{L}_{PL1=}$).

[60] For a proof see the appendix to chapter 4.

[61] That of Keynes and (the early) Carnap.

Chapter 3

The Two Approaches

3.1 Preliminaries

The problem of a (quantitative) theory of confirmation has been – and still is – a "hot topic" in the philosophy of science for over a half century, starting with such classics as Carl Gustav Hempel's *Studies in the Logic of Confirmation* (1945)[1], Rudolf Carnap's work on inductive logic and probability[2], and various contributions by Nelson Goodman, Olaf Helmer, Janina Hosiasson-Lindenbaum, John G. Kemeny, R. Sherman Lehman, Paul Oppenheim, Abner Shimony, and others.[3] Despite these efforts there is still no generally accepted definition of (degree of) confirmation.

In my opinion one reason for this is that there are at least two conflicting concepts of confirmation[4]: On the one hand there is the *likeliness* concept of confirmation expressing our acknowledging theories T that are true or *likely* (probable, truthlike) relative to evidence E and background knowledge B. On the other hand there is the *loveliness*[5] concept of confirmation expressing our acknowledging theories T that are informative and which imply (predict, explain, account for) together with the background knowledge B many of the data in the evidence E.

---

[1]Cf. also Hempel (1943) and (1965).

[2]Cf. Carnap (1945), (1946), (1950), (1952) and Carnap/Stegmüller (1959).

[3]Cf. Goodman (1946), Helmer/Oppenheim (1945), Hosiasson-Lindenbaum (1940), Kemeny (1953) and (1955), Lehman (1955), Shimony (1955). If one understands the problem of a theory of confirmation in a broad sense so that it includes the issue(s) of induction (and probability), it can be traced back even to the ancients.

[4]Cf. Smokler (1968).

[5]The term is borrowed from Lipton (1993); cf. p. 114ff.

Proponents of the likeliness concept of confirmation are all Bayesian and, more generally, all probabilistic theories of confirmation. Their measures of confirmation C_p either measure the probability of theory T given evidence E (and background knowledge B); or else they measure the boost in the probability of (subjective degree of belief in) T (given B) which is caused by the addition of E to B, i.e. the difference between $p(T \mid B \wedge E)$ and $p(T \mid B)$. In the former case, the focus is on confirmaton as *firmness*; in the latter it is on confirmation as *increase in firmness*.[6]

All what matters for the likely-ist is whether T is probable or more probable given B and E than without E being given; questions as the informativeness of T (and B for E), its simplicity, and its coherence with respect to, henceforth w.r.t., E are neglected (except if they bear on T's boost in probability by the extension of B by E). This exclusive focus on truth (probability) will be referred to as *theory enmity* of *theory hostility*.

Among the approaches based on the loveliness concept of confirmation one can cite the various versions of Hypothetico-Deductivism (HD)[7]; but also Bootstrap-Theory[8] may be argued to be a case in point. According to (HD)[9], evidence E confirms theory T (relative to background knowledge B) if(f) E is logically implied by T (and B) in some suitable (relevant) way – the way depending on the version of (HD) under consideration.

[6]Gillies' (1998) distinction between *confirmation* on the one hand and *support* on the other amounts to the same.

[7]For a discussion of (HD) and its (alleged) hopelessness see Glymour (1980c). An (unsuccessful – cf. Glymour 1980c) attempt to rescue (HD) can be found in Merrill (1979). Approaches to (HD) which *replace* the underlying (classical) logic by an *alternative* logic are provided by Waters (1987) and Sylvan/Nola (1991). Along similar lines Grimes (1990) tries to solve the problems of (HD), most notably the *tacking by conjunction* problem, by considering the relation of *narrow consequence* instead of the classical logical consequence relation.

Far more promising are the accounts by Gerhard Schurz and Kenneth Gemes, which do not replace, but *restrict* the classical consequence relation. Schurz demands of a (classically) valid inference $A \vdash B$ to be in addition *relevant*; Gemes demands that B be not only a (classical) consequence of A, but that it be a *content part* of A. Cf. Schurz (1991a), (1991b), (1998), Schurz/Weingartner (1987), and Weingartner/Schurz (1986) for his (their) theory of relevance, and Schurz (1994) for an application to (HD); cf. Gemes (1994c) and (1997a) for his "New Theory of Content", and Gemes (1990), (1993), (1994a), (1998), and (1999) for a discussion of (HD).

[8]Cf. especially Glymour (1980a), but also Glymour (1975), (1977), (1980b), and (1983). For discussions see Christensen (1990), Culler (1995), Edidin (1983), and Mitchell (1995).

[9]In contrast to probabilistic theories of confirmation, (HD) does not attempt do define a *quantitative* (or *comparative*) concept of confirmation, but confines itself to the definition of a *qualitative* one.

A third class of theories of confirmation are the accounts of (explanatory) coherence, which may be argued to take into account both the likeliness and the loveliness concept of confirmation.

Loveliness and likeliness are called *primary confirmational virtues*. (The term 'virtue' should make clear that I consider confirmation theory to be a *normative* theory of theory assessment – just as logic is a normative theory of truth-preserving reasoning.) They are conflicting in the following sense: Other things being equal, a theory T implies (together with background knowledge B) the more data of the evidence E, the logically stronger it is, whereas T is the likelier relative to E and B, the logically weaker it is.

Apart from the two primary confirmational virtues of loveliness and likeliness, there are the derived confirmational virtues of *simplicity* and *natural formulation*: The theories we aim at should be *simple*, and – if they are interpreted as sets of statements, and the measure of confirmation C need not be closed under equivalence transformations of T – they should be *formulated naturally*.

A further property often cited as being of relevance for the assessment of theory by evidence is the (*explanatory*) *coherence* of T, B, and E, or the (explanatory) coherence of T and B w.r.t. E.

In addition, evidence E is argued to be preferable, if it is "*big*" and *varied* or *diverse* in the sense that it reports about "different classes of facts". Size and variety together determine the "goodness" of the evidence which is the topic of chapter 6. There I argue that the variety of evidence E depends on the theory T and the background knowledge B under consideration. A non-arbitrary and comprehensible function $G(\cdot,\cdot,\cdot)$ is defined which is computabe in the limit, and such that $G(T,E,B)$ measures the goodness of E in relation to T and B by measuring (i) how many classes of facts E consists of, (ii) how much these differ from each other, and (iii) how "big" they are. A class of facts is construed as a set of individuals, because I take the latter to be ontologically basic.

In this chapter I will try to make precise what I mean by the confirmational virtues. The next chapters deal with attempts to solve the problem of a quantitative theory of confirmation by defining a measure of confirmation C which is formally handy and materially adequate, i.e. non-arbitrary, comprehensible, computable in the limit, and sensitive to all (and only) the confirmational virtues.

Prima facie, there are two possible approaches for a solution to the problem of a quantitative theory of confirmation. The one is (1) to argue that there is one single property of theory T in relation to evidence E and background knowledge B which takes into account all (and only the) confirmational virtues of T in relation to E and B; (2) to define a function $Coh(\cdot,\cdot,\cdot)$ such that $Coh(T,E,B)$

measures the degree to which T, E, and B exhibit this property – and thereby the confirmational virtues; and (3) to identify the measure of confirmation C with the function Coh. As indicated, the candidate here is the coherence of theory T and background knowledge B w.r.t. evidence E, measured by the formally handy function Coh presented in the next chapter.

The second approach is (1) to define, for each confirmational virtue V, a function $f_V(\cdot, \cdot, \cdot)$ such that $f_V(T, E, B)$ measures the degree to which V is exhibited by T, E, and B, for every theory T, every evidence E, and every background knowledge B; and (2) to define the measure of confirmation C as a function of (some of) these functions f_V.

It will turn out that it suffices to consider the two primary confirmational virtues of loveliness and likeliness (and the goodness of the evidence). So in a sense, the idea here is to preserve and combine those elements of HD on the one hand and Bayesian confirmation theory on the other which are worth being preserved, and, at the same time, to get rid of their respective drawbacks.

It remains to be argued for one of these two approaches. In principle, there is a simple criterion that would do the job, namely the answer to the question whether the measure of confirmation C is to be closed under equivalence transformations of T, i.e. whether it should matter how T – construed as a set of statements – is formulated. Though this invariance does not hold of the coherence measure Coh, it can be made to hold of Coh by referring to some canonical formulation of T. So the answer to the question of how a theory T is to be defined – as a set of models, or as a set of statements – does not automatically give an answer to the question which approach to take.

Nevertheless, the quantitative theory of confirmation resulting from the first approach, which identifies degree of confirmation with degree of coherence w.r.t. the evidence, can hardly be argued to be materially adequate in the sense that coherence w.r.t. the evidence is sensitive to all (and only) the confirmational virtues. This is definitely not the case for the definitions of these virtues given in the next sections.

In contrast to this, the result of the second approach, which defines the measure of confirmation C as a function of (some of) the functions f_V (measuring the confirmational virtues V), can be shown to be sensitive to all (and only) the (primary) confirmational virtues (of loveliness and likeliness). I will therefore conclude that the second approach is a more promising way towards a solution to the problem of a quantitative theory of confirmation.

In the last chapter, the measure of confirmation C is combined with the measure of the "goodness" of evidence G to yield the refined measure of confirmation

$C^* (\cdot, \cdot, \cdot)$. C^* additionally takes into account how much evidence E is worth for the assessment of theory T.

3.2 The Confirmational Virtues

The measure of confirmation C should be materially adequate in the sense that it is sensitive to all (and only) the confirmational virtues. Two questions arise: Which are these confirmational virtues, and what does it mean to be *sensitive* to the confirmational virtues?

First, there are the conflicting primary confirmational virtues of *loveliness* and *likeliness*. Loveliness expresses our acknowledging theories T that are informative or lovely in the sense that they imply (together with background knowledge B) many of the data in evidence E. Likeliness expresses our acknowledging theories T that are likely in the sense that E – when combined with B – speaks in favour of T.

Then there are the *secondary* or *derived* confirmational virtues of *simplicity* and *natural formulation*, where demanding the latter makes sense only if T, E, and B are construed as sets of statements, and the measure of confirmation C need not be closed under equivalence transformations of T.

Apart from these confirmational virtues, there is the (explanatory) coherence of T and B w.r.t. E, which seems to combine the former. Intuitively, T, B, and E cohere not only, if T is likely relative to E and B; they cohere also, if T and B imply (account for) many of the data in E. Furthermore, our pretheoretical understanding of coherence tells us that this concept is implicitely sensitive to the simplicity of T, and perhaps also to the way T is formulated. In sum, coherence w.r.t. the evidence seems to take into account both the primary and the derived confirmational virtues, which gives rise to the hypothesis that the problem of a quantitative theory of confirmation is subsidiary to the problem of a quantitative theory of coherence w.r.t. the evidence. Therefore, coherence w.r.t. the evidence is *not* called *a* confirmational virtue, for if this talk is proper one should better speak of *the* confirmation virtue.

A final property of importance is the "goodness" (size plus variety) of the evidence E, which, at first sight, seems to differ from the confirmational virtues in that it is a property of E which is independent of theory T and background knowledge B. That this is not the case, but that variety and goodness of evidence differ from the confirmational virtues in another respect, will be argued for in the last chapter.

3.3 The Primary Confirmational Virtues

In the foregoing sections I have appealed to an intuitive understanding of the loveliness or power a theory T in relation to an evidence E and a background knowledge B. I think that any adequate measure $\mathcal{LO}(T, E, B)$ of the loveliness of T and B for E should be *searching power* in the following sense:

Definition 3.1 (Searching Power) Let E be an evidence. A function $f(\cdot, E, \cdot)$, $f(\cdot, E, \cdot) : \mathcal{T} \times E \times \mathcal{B} \to \Re$, is *searching power for* $mod(E)$ iff it holds for any theories T and T', and every background knowledge B: If $T \cup B \not\vdash \perp$ and $T' \cup B \not\vdash \perp$, then

1. $f(T, E, B) \geq 0$,
2. if $T \cup B \vdash E$, then $f(T, E, B) = 1$, and
3. if $T' \vdash T$, then $f(T', E, B) \geq f(T, E, B)$.

A function $f(\cdot, \cdot, \cdot)$, $f(\cdot, \cdot, \cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \to \Re$, is a *power searcher* iff $f(\cdot, E, \cdot)$ is searching power for $mod(E)$, for every evidence E.

The notion of searching power can, of course, be generalised to any sets of statements T, E, and B respectively functions f with domains $\wp(\mathcal{L}_{PL1=})$. However, it will turn out that the restriction to theories T, evidences E, and background knowledges B is necessary in order for several theorems to hold.

The first and second condition set lower and upper bounds, respectively, for the values a power searcher can take on, where the second condition in addition tells one that the power of T for E relative to B is maximal, if T and B guarantee (in the sense of logical implication) that E is true. The third condition is a condition of monotonicity saying that the power of T' for E relative to B is greater than or equal to the power of T for E relative to B, if T is logically implied by T'. That is, power or loveliness increases with logical strength.

A consequence of the third condition is that every power searcher $\mathcal{LO}$ is closed under equivalence transformations of T. More precisely:

$$\text{If } T \cup B \not\vdash \perp \text{ and } T' \dashv\vdash T, \text{ then } f(T, E, B) = f(T', E, B),$$

for any theories T, T', every evidence E, every background knowledge B, and every power searcher $\mathcal{LO}(\cdot, E, \cdot)$ for $mod(E)$.

The intuitive understanding of likeliness I have appealed to in the last section is made precise by demanding of any measure $\mathcal{LI}(T, E, B)$ of the likeliness of theory T relative to evidence E and background knowledge B to be *indicating truth* in the following sense:

Definition 3.2 (Indicating Truth) Let E be an evidence. A function $f(\cdot, E, \cdot)$, $f(\cdot, E, \cdot) : \mathcal{T} \times E \times \mathcal{B} \to \Re$, is *indicating truth in* $mod(E)$ iff it holds for any theories T and T', and every background knowledge B: If $E \cup B \nvdash \bot$, then

1. $f(T, E, B) \geq 0$,
2. if $E \cup B \vdash T$, then $f(T, E, B) = 1$, and
3. if $T' \vdash T$, then $f(T', E, B) \leq f(T, E, B)$.

A function $f(\cdot, \cdot, \cdot)$, $f(\cdot, \cdot, \cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \to \Re$, is a *truth indicator* iff $f(\cdot, E, \cdot)$ is indicating truth in $mod(E)$, for every evidence E.

This definition can, of course, also be generalised to any sets of statements. As mentioned before, the restriction to theories T, evidences E, and background knowledges B is necessary in order for several theorems to hold. In particular, this is the case for the truth indicativeness of the likeliness function LI presented in chapter 5.

The first and second conditions set again lower and upper bounds, respectively, for the values a truth indicator can take on, where the second condition in addition tells one that the likeliness of T relative to E and B is maximal, if E and B guarantee the truth of T. As in the previous case, the third condition is a condition of monotonicity saying that the likeliness of T relative to E and B is greater than or equal to the likeliness of T' relative to E and B, if T' logically implies T. In other words, likeliness decreases with logical strength.[10]

A consequence of the third condition is that every truth indicator $\mathcal{LI}$ is closed under equivalence transformations of T.

There are many power searchers and truth indicators.

Theorem 3.1 (Power Searcher and Truth Indicator) Let T, E, and B range over wffs of $\mathcal{L}_{prop}$ (instead of theories, evidences, and background knowlegdes, respectively, which are sets of wffs of $\mathcal{L}_{PL1=}$) in the definitions of searching power and

[10]One might want to add the condition that likeliness increases with the logical strength of the background knowledge B, i.e.

$$\text{if} \quad B' \vdash B, \quad \text{then} \quad \mathcal{LI}(T, E, B') \geq \mathcal{LI}(T, E, B).$$

In my opinion this is inadequate, because new background information may even lead to the refutation of a theory. A similar remark applies to the definition of searching power.

indicating truth. Then it holds for every contingent wff E and every strict (unconditional) probability $p(\cdot)$:

1. $p(\cdot \mid E \wedge \cdot)$ is indicating truth in $mod(E)$.
2. $i(\cdot, E, \cdot) := 1 - p(\cdot \wedge \cdot \mid \neg E)$ is searching power for $mod(E)$.
3. $i'(\cdot, E, \cdot) := 1 - p(\cdot \mid \neg E \wedge \cdot)$ is searching power for $mod(E)$, if it is defined, i.e. if $\neg E \wedge B \nvdash \bot$.

What is needed are not only two functions $\mathcal{LO}$ and $\mathcal{LI}$ which are searching power and indicating truth, respectively. In addition these functions have to be formally handy, i.e. non-arbitrary, comprehensible, and computable in the limit. Arbitrariness will be avoided by defining two single functions; comprehensibility will be achieved by purely syntactical definitions in the terms of $PL1 =$ and ZF; and computability in the limit will be a consequence of these definitions.

Let me stress that the measure of confirmation C should not be both searching power and indicating truth, for such functions are constant.

Theorem 3.2 (Truth Indicating Power Searchers Are Constant) Let E be an evidence, and let $f(\cdot, E, \cdot)$, $f(\cdot, E, \cdot) : \mathcal{T} \times E \times \mathcal{B} \to \Re$, be searching power for $mod(E)$.

If $f(\cdot, E, B)$ is indicating truth in $mod(E)$, then it holds for every theory T and every background knowledge B with $E \cup B \nvdash \bot$: $f(T, E, B) = 1$.

The measure of confirmation C should be *sensitive* to loveliness and likeliness; it should balance between these two conflicting concepts of confirmation.

If the likeliness of T relative to E and B equals the likeliness of T' relative to E' and B', then the degree of confirmation $C(T, E, B)$ of T by E relative to B should be greater than the degree of confirmation $C(T', E', B')$ of T' by E' relative to B' just in case the loveliness or power of T and B for E is greater than the loveliness or power of T' and B' for E'. Similarly, if the loveliness or power in the first case is equal to the loveliness or power in the second case, then the degree of confirmation should be greater in the first case if and only if the likeliness is. Furthermore, confirmation should be minimal just in case loveliness or likeliness is minimal; and it should be maximal if and only if both are maximal. This is expressed in the following definition.

Definition 3.3 (Sensitivity to Loveliness and Likeliness) Let $\mathcal{LO}(\cdot, \cdot, \cdot)$, $\mathcal{LO}(\cdot, \cdot, \cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \to \Re$, be a power searcher, and let $\mathcal{LI}(\cdot, \cdot, \cdot)$, $\mathcal{LI}(\cdot, \cdot, \cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \to \Re$, be a truth indicator.

A function $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, is *sensitive to loveliness and likeliness in the sense of* $\mathcal{LO}$ *and* $\mathcal{LI}$ iff it holds for any theories T and T', any evidences E and E', and any background knowledges B and B', where $X = \langle T, E, B\rangle$ and $X' = \langle T', E', B'\rangle$:

1. If $\mathcal{LI}(X) = \mathcal{LI}(X') \neq 0$, then $f(X) \geq f(X')$ iff $\mathcal{LO}(X) \geq \mathcal{LO}(X')$,
2. if $\mathcal{LO}(X) = \mathcal{LO}(X') \neq 0$, then $f(X) \geq f(X')$ iff $\mathcal{LI}(X) \geq \mathcal{LI}(X')$,[11]
3. $f(X) = 0$ iff $\mathcal{LO}(X) = 0$ or $\mathcal{LI}(X) = 0$, and
4. $f(X) = 1$ iff $\mathcal{LO}(X) = 1$ and $\mathcal{LI}(X) = 1$.

It is straightforward that sensitivity to loveliness and likeliness in the sense of some power searcher $\mathcal{LO}$ and some truth indicator $\mathcal{LI}$ is sufficient for invariance under equivalence transformations of T.

3.4 The Derived Confirmational Virtues

Let us turn to the *secondary* or *derived confirmational virtues* of simplicity and natural formulation. I will not define when a theory T is simple (w.r.t. to some evidence E and some background knowledge B), or when it is formulated naturally, but will restrict myself to giving necessary conditions. As it stands, the necessary condition for being formulated naturally is a consequence of that for being simple (w.r.t. some evidence E and some background knowledge B).

Though I think that the concept of simplicity applies to theories T in relation to evidences E, background knowledges B, and power searchers $\mathcal{LO}$, this four-place concept of simplicity can also be construed as a one-place concept applying to theories. Intuitively, if a theory T is simple w.r.t. some evidence E, some background knowledge B, and some power searcher $\mathcal{LO}$, then T contains no statement h that is superfluous for E and B w.r.t. $\mathcal{LO}$ in the sense that the power of T *without* h and B for E equals the power of T and B for E; that is, T must not contain a statement h such that

$$\mathcal{LO}(T \setminus \{h\}, E, B) = \mathcal{LO}(T, E, B).$$

[11] Note that conditions (1) and (2) are equivalent with '$>$' instead of '$\geq$'.

Necessary Condition 3.1 ($\mathcal{LO}$-Simplicity) Let E be an evidence, let B be a background knowledge, and let $\mathcal{LO}$ be a power searcher.

If a theory T is *$\mathcal{LO}$-simple w.r.t. E and B*, then there is no wff $h \in T$ such that

$$\mathcal{LO}(T \setminus \{h\}, E, B) = \mathcal{LO}(T, E, B).$$

Any such wff h is called a *$\mathcal{LO}$-superfluous part of T for E and B*.

This four-place concept of simplicity gives rise to a one-place concept of simplicity *per se*. The necessary condition for the latter is the following.

Necessary Condition 3.2 (Simplicity) If a theory T is *simple*, then there is at least one power searcher $\mathcal{LO}$ for which there is no wff $h \in T$ such that it holds for every evidence E, and every background knowledge B:

$$\mathcal{LO}(T \setminus \{h\}, E, B) = \mathcal{LO}(T, E, B).$$ [12]

Any such wff h is called a *$\mathcal{LO}$-superfluous part of T*; i.e. h is a *$\mathcal{LO}$-superfluous part of T* iff it holds for every evidence E, and every background knowledge B: h is a $\mathcal{LO}$-superfluous part of T for E and B.

h is a *superfluous part of T* iff there is at least one power searcher $\mathcal{LO}$ such that h is a $\mathcal{LO}$-superfluous part of T.

Let us now briefly turn to the derived confirmational virtue of being formulated naturally. It is rather doubtless that a theory T which is formulated naturally should or does not contain any redundant part that is already logically implied by the rest of T.

Necessary Condition 3.3 (Natural Formulation) If a theory T is *formulated naturally*, then T is formulated non-redundantly.

Clearly, every simple theory T is formulated non-redundantly.

[12] Demanding of a simple theory T to be such that there is at least one power searcher $\mathcal{LO}$ for which there is no statement $h \in T$ such that it holds for at least one evidence E, and at least one background knowledge B:

$$\mathcal{LO}(T \setminus \{h\}, E, B) = \mathcal{LO}(T, E, B),$$

would yield that no theory T is simple. The reason is that for every theory T, every wff $h \in T$, every power searcher $\mathcal{LO}$, and every evidence E with $E \cup T \nvdash \bot$ there is at least one background knowledge B – e.g. $B = E \cup T$ or $B = E$ – such that: $\mathcal{LO}(T \setminus \{h\}, E, B) = \mathcal{LO}(T, E, B) = 1$.

Observation 3.1 (Non-Redundancy) If there is an evidence E, a power searcher $\mathcal{LO}$, and a background knowledge B such that T is $\mathcal{LO}$-simple w.r.t. E and B, then T is formulated non-redundantly.

A measure of confirmation C should be sensitive to simplicity considerations[13], and it should not be impressable by redundancies. Before presenting these notions let me make a point concerning their definition: In the section on theories in chapter 1 I pointed out that a theory T is taken to be a set of statements in order to allow for both the semantic and the syntactic definition of theories, and in order to put no restrictions on the behaviour of an adequate measure of confirmation C.

Now I am concerned with putting restrictions on the behaviour of an adequate measure of confirmation. The question is whether these may be so strong as to rule out the semantic interpretation of theories; i.e. whether they may be such that a measure of confirmation satisfying them cannot be closed under equivalence transformations of T. It turns out that if the following definitions are formulated with '$\geq$' (and not '$>$'), then this does not follow. However, if these definitions are fomulated with '$>$' instead of '$\geq$', it follows that no function satisfying any of these conditions can be closed under equivalence transformations of T. Therefore the following definitions are formulated with '$\geq$' instead of '$>$'.

Definition 3.4 (Sensitivity to Simplicity Considerations i.w.s.) A function $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot):\mathcal{T}\times\mathcal{E}\times\mathcal{B}\rightarrow\Re$, is *sensitve to simplicity considerations in the weak sense* iff there is at least one power searcher $\mathcal{LO}$ such that it holds for every theory T, every evidence E, every background knowledge B, and every wff $h\in T$:

> If h is a $\mathcal{LO}$-superfluous part of T, then $f(T\setminus\{h\},E,B)\geq f(T,E,B)$;

i.e. which is such that it holds for every theory T, and every wff $h\in T$:

> If $\mathcal{LO}(T\setminus\{h\},E,B)=f(T,E,B)$, for every evidence E, and every background knowledge B, then $f(T\setminus\{h\},E,B)\geq f(T,E,B)$, for every evidence E, and every background knowledge B.

Definition 3.5 (Sensitivity to Simplicity Considerations i.s.s.) A function $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot):\mathcal{T}\times\mathcal{E}\times\mathcal{B}\rightarrow\Re$, is *sensitive to simplicity considerations in the strong sense* iff there is at least one power searcher $\mathcal{LO}$ such that it holds for every theory T, every evidence E, every background knowledge B, and every wff $h\in T$:

[13] The measure of confirmation presented in the chapter on loveliness and likeliness is sensitive to simplicity considerations in the very strong sense.

> If h is a $\mathcal{LO}$-superfluous part of T for E and B, then $f(T \setminus \{h\}, E, B) \geq f(T, E, B)$;

i.e. which is such that it holds for every theory T, every evidence E, every background knowledge B, and every wff $h \in T$:

> If $\mathcal{LO}(T \setminus \{h\}, E, B) = \mathcal{LO}(T, E, B)$, then $f(T \setminus \{h\}, E, B) \geq f(T, E, B)$.

A generalisation of the last definition (in the sense that sensitivity to simplicity considerations i.s.s. is a consequence of sensitivity to simplicity considerations i.v.s.s.) is the following.

Definition 3.6 (Sensitivity to Simplicity Considerations i.v.s.s.) A function $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, is *sensitive to simplicity considerations in the very strong sense* iff there is at least one power searcher $\mathcal{LO}$ such that it holds for any theories T and T', every evidence E, and every background knowledge B:

> If $T' \vdash T$ and $\mathcal{LO}(T, E, B) = \mathcal{LO}(T', E, B)$, then $f(T, E, B) \geq f(T', E, B)$.

Definition 3.7 (Unimpressability by Redundancies) A function $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, *cannot be impressed by redundancies* iff it holds for every theory T, every evidence E, every background knowledge B, and every wff $h \in T$:

> If h is a redundant part of T, then $f(T \setminus \{h\}, E, B) \geq f(T, E, B)$.[14]

Before turning to coherence w.r.t. the evidence respectively the first approach to a solution of the problem of a quantitative theory of confirmation in the next chapter, let me note some relations between sensitivity to loveliness and likeliness (in the sense of some power searcher $\mathcal{LO}$ and some truth indicator $\mathcal{LI}$), sensitivity to simplicity considerations (in some sense), and invariance under equivalence transformations.

Theorem 3.3 (SensSimplCons and Unimpressability) Let $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, be a function.

1. If f is sensitive to simplicity considerations in the very strong sense, then f is sensitive to simplicity considerations in the strong sense.

[14] Note that no function which is closed under equivalence transformations of T can be impressed by redundancies.

2. If f is sensitive to simplicity considerations in the strong sense, then f is sensitive to simplicity considerations in the weak sense.

3. If f is sensitive to simplicity considerations in the weak sense, then f cannot be impressed by redundancies.

The last theorem holds also in case '$\geq$' is relaced by '$>$' in the definitions of sensitivity to simplicity considerations in any sense and unimpressability by redundancies. As no function which is closed under equivalence transformations of T satisfies *strict* unimpressability by redundancies, i.e. unimpressability with '$>$' instead of '$\geq$', no such function can be *strictly* sensitive to simplicity considerations in any sense.

It is obvious that sensitivity to simplicity considerations in the very strong sense implies invariance under equivalence transformations of T. This does not hold of sensitivity to simplicity considerations in the strong sense.

Theorem 3.4 (SensSimplCons i.s.s. Does Not Imply InvEquTrans) Let $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \to \Re$, be a function. If f is sensitive to simplicity considerations in the strong sense, then f need not be closed under equivalence transformations of T in the sense that

$$f(T, E, B) = f(T', E, B), \quad \text{if} \quad T \dashv\vdash T',$$

for any theories T and T', every evidence E, and every background knowledge B.

Theorem 3.5 (InvEquTrans Implies SensSimplCons i.w.s.) If f is closed under equivalence transformations of T, then f is sensitive to simplicity considerations in the weak sense.

Theorem 3.6 (InvEquTrans Does Not Imply SensSimplCons i.s.s.) If f is closed under equivalence transformations of T, then f need not be sensitive to simplicity considerations in the strong sense.

A consequence of these theorems is that if there is a property which implies sensitivity to simplicity considerations in the very strong sense, then a function having this property is sensitive to all derived confirmational virtues; i.e. such a function is sensitive to simplicity considerations in any sense, (it is invariant under equivalence transformations of T, and) it cannot be impressed by redundancies.

The following theorem states that sensitivity to loveliness and likeliness in the sense of some power searcher $\mathcal{LO}$ and some truth indicator $\mathcal{LI}$ is such a property, whence every function which is sensitive to the primary confirmational virtues is automatically sensitive to all derived confirmational virtues.

Theorem 3.7 (SensLoveLike Implies SensSimplCons i.v.s.s.) Let $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, be a function. If f is sensitive to loveliness and likeliness in the sense of some power searcher $\mathcal{LO}$ and some truth indicator $\mathcal{LI}$, then f is sensitive to simplicity considerations in the very strong sense.

Chapter 4

Coherence with Respect to the Evidence

4.1 Coherence as Truth-Indicator

Coherence plays a prominent role in the philosophy of science – in the theory of confirmation – and, more generally, in epistemology – in the theory of justification – as indicator of truth.

There is an enduring discussion[1] whether the coherence of a set of statemets or propositions S is indicative of the truth of (the statements or propositions in) S, or as it is often put: whether coherence is *truth conducive*. I think the answer to this question is straightforward: Either one adopts a coherence theory of truth according to which a statement s is true just in case s is an element of at least one *coherent* set of statements C (and a set of statements S is true if and only if all statements in S are elements of at least one such set C). Then the coherence of S is not only indicative of the truth of the statements in S; it is *guaranteeing* their truth.

Or else one adopts a correspondence theory of truth according to which truth is a binary relation between a statement s on the one hand and a world or model $\mathcal{M}$ on the other. Then the coherence of a set of statements S cannot be indicative of the truth of S in some world or model $\mathcal{M}$, if the coherence of S is independent of $\mathcal{M}$.[2] More formally:

---

[1]Cf. Akiba (2000), Barker (1994), Cross (1999), Klein/Warfield (1994) and (1996), Merricks (1995), Millgram (2000), Olsson (2001) and (2002), and Shogenji (1999), (2001a), and (2001b).

[2]The claim that coherence is truth indicative may also be read as the claim that coherence is

Necessary Condition 4.1 (Coherence as Truth Indicator) Let $\mathcal{M} = \langle Dom, \varphi \rangle$ be a model, and suppose $Coh\,(\cdot, \mathcal{M})$, $Coh\,(\cdot, \mathcal{M}) : D \times \mathcal{M} \to \Re$, $D \subseteq \wp\,(\mathcal{L}_{PL1=})$, is a function such that $Coh\,(S, \mathcal{M})$ measures the coherence of S w.r.t. $\mathcal{M}$, for every set of wffs $S \in D$.

If $Coh\,(\cdot, \mathcal{M})$ is indicative of truth in $\mathcal{M}$, then it does not hold for every set of wffs $S \in D$, and every model $\mathcal{M}'$: $Coh\,(S, \mathcal{M}) = Coh\,(S, \mathcal{M}')$.[3]

Obviously, this condition is not satisfied by any function $Coh\,(\cdot)$, $Coh\,(\cdot) : D \to \Re$, $D \subseteq \wp\,(\mathcal{L}_{PL1=})$, which is independent of the world whose truth in one is interested in.[4]

I adopt a theory of truth along the lines of Tarski[5], whence coherence – if construed in the usual way as a one-place concept applying to sets of statements or propositions – is not indicative of truth (in any model). The reason being that the coherence of a set of statements S is independent of the model whose truth in one is interested in.

If this one-place property of coherence *per se* is not indicative of truth in some model, because it is independent of every model, then the concept of coherence has to be relativised to the model whose truth in one is interested in. Let me

---

indicative of truth in the actual world. The reason may be seen in a principle of the coherence of the world – in a way similar to the justification of induction (as a valid inference for the actual world) by reference to a principle of the uniformity of nature. Before accepting such a principle of the coherence of the world I would rather accept the claim of the truth conduciveness of coherence itself.

[3]The restriction to a subset D of $\wp\,(\mathcal{L}_{PL1=})$ should avoid that this condition does not make sense, if there is no complete or total coherence measure $m\,(\cdot, \mathcal{M})$, where $m\,(\cdot, \mathcal{M})$ is *complete* or *total* iff $m\,(S, \mathcal{M})$ is defined for *every* set of wffs $S \subseteq \mathcal{L}_{PL1=}$. The existence of such a complete coherence measure may be questioned, but nothing really substantial hinges on this for the necessary condition for coherence as truth indicator.

Note that it is not even unplausible that there is no connected *ordinal* coherence measure $\succeq (\cdot, \cdot, \mathcal{M})$, where $\succeq (S, S', \mathcal{M})$ says that S is as coherent w.r.t. $\mathcal{M}$ as or more coherent w.r.t. $\mathcal{M}$ than S'. $\succeq (\cdot, \cdot, \mathcal{M})$ is *connected* iff it holds for *any* two sets of wffs $S, S' \subseteq \mathcal{L}_{PL1=}$: $\langle S, S', \mathcal{M} \rangle \in \succeq (\cdot, \cdot, \mathcal{M})$ or $\langle S', S, \mathcal{M} \rangle \in \succeq (\cdot, \cdot, \mathcal{M})$. Cf. Hartmann/Bovens (2000).

[4]Note that the necessary condition for coherence as truth indicator allows for cases, where $Coh\,(S, \mathcal{M}_1) = Coh\,(S, \mathcal{M}_2)$, for *some* set of statements $S, S \in D$, and *some* models $\mathcal{M}_1, \mathcal{M}_2$. There may even be some set of statements S such that this holds for any models $\mathcal{M}, \mathcal{M}'$.

[5]It does not matter whether it is adequate to call Tarski's theory of truth a correspondence theory of truth. Although Tarski himself does so, the adequacy of this may be questioned on the grounds that the actual world is no model consisting of a domain Dom and an interpretation function φ, and that a correspondence theory of truth seeks a correspondence with the actual world. However, this does no harm, if the actual world can be adequately represented by some model $\mathcal{A}$ of the mentioned form.

stress that I do not claim that coherence is indicative of truth in some model $\mathcal{M}$, if it is relativised to $\mathcal{M}$; all I claim is that *if* coherence is to be indicative of truth in the model $\mathcal{M}$, then it has to be relativised to $\mathcal{M}$.

Which is the world whose truth in we are interested in? The answer to this question may depend on the set of statements S under consideration, but in general we are interested in truth in the *actual* world. So for most cases it will be appropriate to relativise the coherence of S to the actual world – or a model $\mathcal{A} = \langle A, \varphi_A \rangle$ adequately representing the actual world as a set-theoretical structure consisting of a domain A and an interpretation function φ_A. Thus the question is not whether coherence *per se* is truth indicative – it is not – but whether coherence w.r.t. to the actual world respectively a model $\mathcal{A}$ adequately representing the latter is indicative of truth in the actual world respectively in $\mathcal{A}$.

Assume $\mathcal{A} = \langle A, \varphi_A \rangle$ is a model adequately representing the actual world, which will be identified with $\mathcal{A}$ in the following. How can such a relativisation of the coherence of S to the actual world $\mathcal{A}$ look like? After all, our pretheoretical and intuitive understanding of coherence tells us that this concept applies to sets of statements (or propositions). Furthermore, under the assumption that the actual world exists at all, we hardly have access to it – and, for sure, the aim is a theory of coherence that not only explicates the notion of coherence (w.r.t. some model), but that also enables one to determine whether (and to what degree) a given set of statements S is coherent w.r.t. the actual world. In order to achieve this one needs a (true) description of the actual world which allows for this determination.

However, there seems to be no fully reliable method – no algorithm – that tells one, for a given set of statements D, whether D is a description of the actual world $\mathcal{A}$. So what to do? Well, simply *assume* of some set of statements $D_{\mathcal{A}}$ that it is a description of $\mathcal{A}$. Then one can determine whether a set of statements S is coherent w.r.t. the actual world $\mathcal{A}$, if one can determine whether S is coherent w.r.t. $D_{\mathcal{A}}$.

Of course, the reliability of the determination of the degree of coherence w.r.t. the actual world $\mathcal{A}$ by means of the degree of coherence w.r.t. one of its descriptions $D_{\mathcal{A}}$ depends on the detailedness or accuracy of $D_{\mathcal{A}}$. The latter is maximal, only if $D_{\mathcal{A}}$ is complete in the sense that it holds for every statement h: If $\mathcal{A} \models h$, then $D_{\mathcal{A}} \vdash h$, which is not neccessary in order for a set of statements $D_{\mathcal{M}}$ to be a description of some model $\mathcal{M}$.

The chosen set of statements $D_{\mathcal{A}}$ should be such that assuming of it to be a description of $\mathcal{A}$ is as weak an assumption as possible. In my opinion there is one special candidate that is epistemically distinguished in just this respect: the set E of those statements that we take to express what we take to be the case because of perceiving it; in other words: the *evidence* E which is available at some given point of time. In the following it is *assumed* that an evidence E is true in the actual world $\mathcal{A}$; i.e. I will make assumption 1.4, which is restated here as assumption 4.1.

Assumption 4.1 (Epistemic Mark of Distinction) If E is an evidence from $D_1, \ldots, D_k$, then E is *assumed* to be *true in the actual world*, i.e.

$$\mathcal{A} \in mod\,(E)\,, \quad \text{for every evidence } E \in \mathcal{E}.$$

Let me stress that this assumption should only enable me to make sense of the claim that coherence (w.r.t. the actual world $\mathcal{A}$) is indicative of truth (in $\mathcal{A}$). I do not claim that an evidence E is true in the actual world $\mathcal{A}$, nor do I claim that coherence is indicative of truth in the actual world $\mathcal{A}$, if it is relativised to an evidence E. On the contrary, it will turn out that coherence w.r.t. evidence E is not even indicative of truth in $mod\,(E)$, and thus (under the above assumption that $\mathcal{A} \in mod\,(E)$) not indicative of truth in $\mathcal{A}$ – given that the measure of coherence w.r.t. the evidence defined below properly models our pretheoretical and intuitive concept of coherence (w.r.t. the evidence).

This is one reason for preferring the second approach of a solution to the problem a quantitative theory of confirmation: the definition of the measure of confirmation C by means of a function of the functions $\mathcal{LO}$ and $\mathcal{LI}$ measuring the primary confirmational virtues of loveliness and likeliness. Another reason for not adopting the first approach, which argues that coherence w.r.t. the evidence is *the* confirmation value, and takes account of all (and only) the primary and derived confirmational virtues, is the following: Coherence w.r.t. the evidence E – in its formalisation $Coh\,(\cdot, E, \cdot)$ of below – is neither indicating truth in $mod\,(E)$, nor is it sensitive to loveliness and likeliness in the sense of any power searcher $\mathcal{LO}$ and any truth indicator $\mathcal{LI}$.

Before continuing remember the definition of a description $D_{\mathcal{M}}$ of some model $\mathcal{M}$, and the fact that every evidence E is a description every model $\mathcal{A}$ adequately representing the actual world.

As already indicated, one has to assume that there is at least one model $\mathcal{A} = \langle A, \varphi_A \rangle$ which adequately represents the actual world in order for Tarski's theory of truth to be able to define a notion of truth in the actual world – the reason being that the actual world can hardly be argued to be an ordered pair $\mathcal{M} = \langle Dom, \varphi \rangle$ consisting of a domain Dom and an interpretation function φ.

Assumption 4.2 (Existence of a Model of the Actual World) There is at least one model of the actual world for $\mathcal{L}_{PL1=}$, i.e. there is at least one model $\mathcal{A} = \langle A, \varphi_A \rangle$ such that it holds for every wff $h \in \mathcal{L}_{PL1=}$:

h is true in the actual world if and only if $\mathcal{A} \models h$,

where the concept of being true in the actual world is a primitive concept which is assumed to be meaningful.

So in order to construe coherence w.r.t. the actual world as coherence w.r.t. the evidence, one has to assume as primitive a meaningful concept of truth in the actual world, and the existence of at least one model $\mathcal{A} = \langle A, \varphi_A \rangle$ of the actual world for $\mathcal{L}_{PL1=}$. Otherwise the claim that coherence w.r.t. evidence E is indicative of truth in the actual world cannot be based on the claim that $Coh\,(\cdot, E, \cdot)$ is indicating truth in $mod\,(E)$.

As mentioned, the indication of truth in the actual world by means of coherence w.r.t. the actual world is not fully reliable, if the evidence is no complete description of the actual world, which, in general, it is not. A measure of the reliability of the indication of truth in the actual world by means of a function $f\,(\cdot, \ldots, \cdot, E)$ which is indicating truth in $mod\,(E)$, for some evidence E, may be seen in the measure of the "goodness" of evidence E, $G\,(\cdot, E, \cdot)$, presented in the last chapter.[6]

4.2 Arbitrary Theories of (Explanatory) Coherence

4.2.1 Introductory Remarks

Against promoting an own account of the coherence of a set of statements T w.r.t. an evidence E (and a background knowledge B) it may be objected that there have already been proposed several theories of (explanatory) coherence. Why not adopt one of these? The answer to this is twofold: First, I am aiming at a formal theory of coherence that enables me to *measure* the coherence of a set of statements T w.r.t. an evidence E (and a background knowledge B) or, at least, to *compare* triples $\langle T, E, B \rangle$ and $\langle T', E', B' \rangle$ with regard to their coherence w.r.t. the evidence; i.e. the aim is the definition of a *quantitative*, at least *comparative* concept of coherence w.r.t. the evidence. Second, the theory of coherence w.r.t. the evidence should be formally handy, in particular non-arbitrary.

[6] As the measure of coherence w.r.t. the evidence E, $Coh\,(\cdot, E, \cdot)$, defined below is not indicating truth in $mod\,(E)$, for any evidence E, the measure $G\,(\cdot, E, \cdot)$ is of no help for a coherentist adopting $Coh\,(\cdot, E, \cdot)$.

To demand of a theory to be non-arbitrary makes sense, only if this theory is formal in the sense that it defines a quantitative or comparative concept of (explanatory) coherence by means of some (set of) function(s). Apart from the fact that non-formal theories cannot fulfill this *desideratum*, there is still another reason that justifies an own approach as concerns the non-formal coherence theories of BonJour (1985), Lehrer (1990), and Bartelborth (1996).

As to the formal theories of (explanatory) coherence, I rely on the formal condition of adequacy that any such formal theory be non-arbitrary, comprehensible, and computable in the limit. The most popular of these theories is the theory of explanatory coherence TEC of Thagard (1989) respectively its formalisation $ECHO$. Apart from TEC, there are the fuzzy measure of explanatory coherence of Schoch (2000), and the probabilistic theory of the coherence of an information set of Hartmann/Bovens (2000). As a matter of fact, these three theories are arbitrary.

Since the account of Schoch (2000) is, according to his own words, a formalisation of the theory of coherence of Bartelborth (1996), and as I consider the notion of coherence of Bartelborth (1996) as an improvement of the notions of coherence of both BonJour (1985) and Lehrer (1990)[7], I take the introduction of Coh to be independently justified as concerns these three theories of (explanatory) coherence.

The following two subsections deal with the theory of explanatory coherence TEC of Thagard (1989) (and its formal model $ECHO$), and the fuzzy measure for explanatory coherence of Schoch (2000). This should make familiar with the concept of explanatory coherence, which is similar to the concept of coherence w.r.t. the evidence. Despite this similarity, the function Coh is definitely not a measure of explanatory coherence.[8]

In the last section I have argued that the coherence of a set of statements S has to be relativised to the model $\mathcal{M}$ whose truth in one is interested in; otherwise

[7] I do not have the place to argue for this here. Let me only note that the conception of the coherence of a system of beliefs S of Lehrer (1990) is not very elaborated in that hardly anything else is demanded of S except that it must not contain alternative or concurring beliefs. Though BonJour (1985) adds some conditions – in particular, he demands that S must not consist of several unrelated subsystems – it is not precisely determined when these conditions are fulfilled. For a discussion cf. Bartelborth (1996), who argues that his coherence theory of justification is not exposed to objections that may be raised against the accounts of BonJour and Lehrer.

[8] In particular, the concept of accounting for, which the notion of coherence w.r.t. the evidence is based on, is not at all intended to be an explication or even definition of the concept of explanation. For more on this see the section on foundationalist coherentism, and the section on accounting for in the chapter on loveliness and likeliness.

the coherence of S cannot be indicative of the truth of S in $\mathcal{M}$. I have indicated to do this by relativising the coherence of S to an evidence E. Although the accounts of Thagard (1989) and Schoch (2000) do not explicitely relativise the coherence of S to an evidence E or some other epistemically distinguished set of statements[9], there is a similar element in their accounts: Those statements in the set of statements S, whose explanatory coherence is to be assessed,

> that describe the results of observation

are epistemically distinguished in that they

> have a degree of acceptability on their own.[10]

In contrast to this, the probabilistic theory of the coherence of an information set of Hartmann/Bovens (2000) does not have such an epistemically distinguished element which would enable their account to explicate a concept of coherence which is indicative of truth in some model. Furthermore, their account is based on a somewhat different concept of coherence than that of coherence w.r.t. the evidence[11], which differs also from the concept of explanatory coherence. Therefore their theory will not be discussed. Let me only note the following.

Theorem 4.1 ($\succeq$ Is Arbitrary) The ordinal measure of coherence $\succeq$ of Hartmann/Bovens (2000) is arbitrary.

Finally, it is to be noted that, for reasons of space and time, the related topic of (explanatory) unification is not dealt with.[12] This shortcoming is in particular serious for the account of Schurz/Lambert (1994) and Schurz (1999) according to which

> *coherence minus circularity = unification*[13],

which is in accordance with the claim that coherence has to be relativised to the evidence.

---

[9]Like Hartmann/Bovens (2000), Thagard (1989) and Schoch (2000) consider propositions instead of statements.

[10]Thagard (1989), p. 437. In case of Schoch this finds its expression in the principle of data evidence which says that the singleton $\{E\}$ containing the proposition E is coherent, if there is positive evidence for E (if there is negative evidence for E, then there is positive evidence for $\neg E$). Cf. Schoch (2000), p. 298.

[11]This is – at least partly – due to the fact that their concept of coherence is one of coherence *per se*.

[12]Cf. Friedman (1974), (1979), and (1990), Kitcher (1981) and (1990), Morrison (1990), Schurz/Lambert (1994), and Schurz (1999). For a recent comment on Schurz (1999) cf. Weber/van Dyck (2002).

[13]Schurz/Lambert (1994), p. 72. Cf. also Schurz (1999), p. 98.

4.2.2 The Theory of Explanatory Coherence of Thagard (1989)

The theory of explanatory coherence TEC of Thagard (1989), which is applied to case studies from the history of the sciences in Eliasmith/Thagard (1997) and Nowak/Thagard (1992)[14], is the most popular non-Bayesian theory of (explanatory) coherence. TEC is modeled by the computer program $ECHO$ (Explanatory Coherence by Harmany Optimization), which generates connectionist networks. The various exhibitory and inhibitory links between the units – standing for hypotheses – in such a network are assigned numbers representing the strength of the links. A link between two units i and j is excitatory, if the two hypotheses represented by i and j cohere; it is inhibitory, if they incohere, which is something stronger than not to cohere:

> The term 'incohere' is used to mean more than just that two propositions do not cohere: to incohere is to resist holding together.[15]

The arbitrariness of $ECHO$ can already be seen here: There are no restrictions on the numbers which are assigned to the links between two units i and j – representing the strength of the (in)coherence relation between the two hypotheses represented by i and j.

TEC consists of the following series of principles:

> *Principle 1. Symmetry.*
> (a) If P and Q cohere, then Q and P cohere.
> (b) If P and Q incohere, then Q and P incohere.
> *Principle 2. Explanation.*
> If $P_1, \ldots, P_m$ explain Q, then:
> (a) For each P_i in $P_1, \ldots, P_m, P_i$ and Q cohere.
> (b) For each P_i and P_j in $P_1, \ldots, P_m, P_i$ and P_j cohere.
> (c) In (a) and (b), the degree of coherence is inversely proportional to the number of propositions $P_1, \ldots, P_m$.
> *Principle 3. Analogy.*
> (a) If P_1 explains Q_1, P_2 explains Q_2, P_1 is analogous to P_2, and Q_1 is analogous to Q_2, then P_1 and P_2 cohere, and Q_1 and Q_2 cohere.

[14] Cf. also Eliasmith/Thagard (2001), Holyoak/Thagard (1997), O'Laughlin/Thagard (2000), Thagard (1997), (1999), and (2000), Thagard/Kunda (1998), Thagard/Millgram (1995), Thagard/Shelley (1997) and (2001), and Thagard/Verbeurgt (1998), all of which can be found on http://cogsci.uwaterloo.ca/Articles/Pages/Coherence.html. In addition, the latter contains articles that are forthcoming or in progress.

[15] Cf. Nowak/Thagard (1992), p. 274.

> (b) If P_1 explains Q_1, P_2 explains Q_2, Q_1 is analogous to Q_2, but P_1 is disanalogous to P_2, then P_1 and P_2 incohere.
> *Principle 4. Data Priority.*
> Propositions that describe the results of observation have a degree of acceptability on their own.
> *Principle 5. Contradiction.*
> If P contradicts Q, then P and Q incohere.
> *Principle 6. Acceptability.*
> (a) The acceptability of a proposition P in a system S depends on its coherence with the proposition[s] in S.
> (b) If many results of relevant experimental observations are unexplained, then the acceptability of a proposition P that explains only a few of them is reduced.
> *Principle 7. System Coherence.*
> The global explanatory coherence of a system S of propositions is a function of the pairwise local coherence of those propositions.[16]

An additional principle is introduced in Nowak/Thagard (1992):

> *Principle C. Competition.*
> If P and Q both explain evidence E, and if P and Q are not explanatorily connected, then P and Q incohere. Here P and Q are explanatorily connected if any of the following conditions holds:
> (a) P is part of the explanation of Q.
> (b) Q is part of the explanation of P.
> (c) P and Q are together part of the explanation of some proposition R.
> (d) P and Q are both explained by some higher-level proposition R.[17]

The global coherence of a system S of propositions is thus traced back to the local coherence between pairs of propositions. This is just the critic of Schoch (2000):

> [Thagard's] measure of coherence is shown to be incapable of dealing adequately with explanatorily relations between more than two sentences.[18]

[16]Thagard (1989), pp. 436-437.
[17]Nowak/Thagard (1992), p. 277.
[18]Schoch (2000), p. 292; cf. also Schoch (2000), pp. 295-296.

Thagard presupposes – as does Schoch (2000) – as primitive the notion of explanation (and that of analogy). This is problematic, not because explanatorily relations have no impact on the coherence of a set of statements S – on the contrary – but because the notion of explanation is itself in need of explication; in particular, the concept of explanation is not comprehensible. One may be of a different opinion – as is, for instance, Thagard, who even praises his theory for not depending on a particular notion of explanation:

> Our account of theory acceptance and our input to $ECHO$ [...] do not presuppose any special theory of explanation. [...] Explanation, however, has many aspects and construing theory choice in terms of explanatory coherence is compatible with various ways of understanding causality and explanation.[19]

But what, if coherence is itself an indispensable ingredient of explanation, so that any adequate definition of explanation presupposes the concept of coherence?[20]

Furthermore, the central principle 2 of explanation makes the questionable assumption (as does the principle 6 of acceptance) that propositions can be counted. Though this point will be discussed later on once more, let me note that without any restrictions on the way a given set of propositions S has to be represented (or formulated, if S is a set of statements), there seems to be no way of uniquely determining how many propositions S consists of.[21]

A consequence of principle 2 is that, if several propositions $P_1, \ldots, P_n$ together explain a proposition Q, and each proposition P_i is necessary for this explanation of Q, then the relation of coherence holds between Q and any single proposition P_i, though, intuitively, Q coheres only with the set (conjunction) of all

[19] Eliasmith/Thagard (1997), p. 11.

[20] This is not even bizarre. For instance, one may define as follows, where sets of statements are considered instead of (sets of) propositions: T explains E (relative to B) just in case T and E cohere (relative to B), where such an explanation is the better, the greater the degree of coherence.

In this manner one may define an inference from E (and B) to T_i as abductively valid or as an inference to the best explanation just in case $Coh\,(T_i, E, B) \geq Coh\,(T_j, E, B)$, for every $j, 1 \leq j \leq n$, where $T_1, \ldots, T_n$ are the finitely many available alternative theories (whose domain of application E belongs to), and B is the background knowledge. If one prefers a quantitative concept of abductive validity, then one may adopt $Coh\,(T_i, E, B)$ or

$$\min\{Coh\,(T_i, E, B) - Coh\,(T_j, E, B) : 1 \leq j \leq n\}$$

as the degree of abductive validity of the inference from E (and B) to T_i.

[21] An elegant way of representing knowledge by relevant elements – and thereby solving this problem – can be found in Schurz/Lambert (1994), p. 88ff.

propositions $P_1, \ldots, P_n$. For instance, if $\forall x\,(Fx \to Gx)$ and Fa together explain Ga, it follows from principle 2 that Ga coheres with $\forall x\,(Fx \to Gx)$, and that Ga coheres with Fa, though, intuitively, Ga coheres only with $\{\forall x\,(Fx \to Gx)\,, Fa\}$. That is, we would say that the set $\{\forall x\,(Fx \to Gx)\,, Fa, Ga\}$ is coherent, but neither would we say that the set $\{\forall x\,(Fx \to Gx)\,, Ga\}$ is coherent, nor would we say that the set $\{Fa, Ga\}$ is coherent.

Eventually, in order to escape the reproach of arbitrariness, Thagard (1989) would have to presuppose as primitive a *quantitative* notion of explanation – something which, to the best of my knowledge, no theory of explanation discussed in the literature even aims at.

Apart from all this, Thagard's TEC and $ECHO$ are not adopted for the following reasons: First, both TEC and $ECHO$ are not comprehensible.[22] Though this gives no ground for rejecting them, if one considers the goal of a theory of explanatory coherence the explication of the concept of explanatory coherence in terms of the concept of explanation, the following theorem is a case in point even if the concept of explanation is assumed to be comprehensible.

Theorem 4.2 ($ECHO$ Is Arbitrary) The computer program $ECHO$, which models the theory of explanatory coherence TEC of Thagard (1989), is arbitrary.[23]

4.2.3 The Fuzzy Measure for Explanatory Coherence of Schoch (2000)

Let me now turn to the fuzzy measure for explanatory coherence of Schoch (2000)[24], which may be considered as a formalisation of the theory of coherence of Bartelborth (1996). Schoch himself notes that

> [his] approach satisfies all these requirements except the last without further restrictions,[25]

[22]That TEC is not comprehensible follows from assumption 2.2 (Comprehensible Concepts), and the fact that TEC presupposes as primitive, apart from the concept of analogy, the concept of explanation. That $ECHO$ is not comprehensible follows from the same assumption, and the fact that $ECHO$ presupposes as primitive a quantitative concept of explanation. The latter finds its expression in the weights w_{ij} representing the strength of the explanatory relation between the propositions represented by the units i and j. Cf. the proof of the next theorem in the appendix to this chapter.

[23]That TEC is not arbitrary has its reason in the fact that it does not define explanatory coherence by a (set of) function(s) to which the concept of arbitrariness could apply.

[24]Cf. also Schoch (2001).

[25]Schoch (2000), p. 302.

where he refers to the principles of systematic coherence and of incoherence, which form the theory of coherence within the coherence theory of justification of Bartelborth (1996).[26]

The basic structure of Schoch's fuzzy measure is the same as that of Thagard's theory. It consists of (1) a set of propositions $\mathcal{E}$; (2) a set $\mathcal{R}$ of rules of the form '$\mathcal{P}$ explains Q', $\mathcal{P} \subseteq \mathcal{E}$, $Q \in \mathcal{E}$, '$\mathcal{P}$ is contradictory', or 'E is a fact', $E \in \mathcal{E}$; (3) a closed interval $I \subseteq \Re$ representing truth values ($I = [0, 1]$ in case of Schoch, and $I = [-1, 1]$ in case of Thagard); (4) a set of real-valued variables $x_1, \ldots, x_n$ with domain I (for Schoch x_i is the fuzzy truth value of the i-th proposition in $\mathcal{E}$; for Thagard $x_i = a_i(t)$ is the degree of acceptance of the i-th proposition in $\mathcal{E}$ at some given point of time t); (5) a first-degree polynomial[27]; and (6) an algorithm translating the rules in $\mathcal{R}$ to the weights $a_{r_1,\ldots,r_n}$ respectively w_{ij}.

Schoch considers his account as a generalisation of Thagard's TEC, which

> [...] does not adequately represent explanatory relations between more than two propositions.[28]

Schoch's measure is defined for pairs of sets of constituents, where a constituent is a subset $\mathcal{P}$ of a set of signed propositions $\mathcal{E}$ over a set of propositions $\mathcal{PR}$, $\mathcal{E} = \mathcal{PR} \cup \{\neg P : P \in \mathcal{PR}\}$, such that there is no proposition $P \in \mathcal{E}$ with $P \in \mathcal{P}$ and $\neg P \in \mathcal{P}$. So the measure is defined for pairs of sets of sets of propositions not containing both a proposition and its negation.

The coherence value $V_{\langle \mathbf{C}, \mathbf{I} \rangle}$ of the pair of sets **C** and **I** of coherent respectively incoherent constituents is recursively defined as follows:

$$V_{P_i}(x_1, \ldots, x_n) = x_i,$$

[26] Cf. Bartelborth (1996), p. 193. The requirement not satisfied by Schoch's account is that

> the degree of (systematical) coherence of a belief system [...] decreases with the number of unconnected subsystems.

Schoch (2000), p. 302.

[27]
$$V(x_1, \ldots, x_n) = \sum_{0 \leq r_1, \ldots, r_n \leq 1} a_{r_1,\ldots,r_n} \cdot x_1^{r_1} \cdots x_n^{r_n}$$

in case of Schoch. In Thagard's model it is

$$H(\mathcal{E}, t) = V_T(a_1(t), \ldots, a_n(t)) = \sum_{0 \leq i \leq n} \sum_{0 \leq j \leq n} w_{ij} \cdot a_i(t) \cdot a_j(t).$$

Cf. the appendix to this chapter, which includes a presentation of the basic structure of $ECHO$.

[28] Schoch (2000), p. 291.

$$V_{\neg P_i}(x_1,\ldots,x_n) = 1 - x_i,$$

$$V_{\mathcal{P}}(x_1,\ldots,x_n) = c_{\mathcal{P}} \cdot \prod_{P\in\mathcal{P}} V_P,$$

$$V_{\langle \mathbf{C},\mathbf{I}\rangle}(x_1,\ldots,x_n) = \sum_{\mathcal{P}\in\mathbf{C}} V_{\mathcal{P}} - \sum_{\mathcal{P}\in\mathbf{I}} V_{\mathcal{P}}.$$

The constants $c_{\mathcal{P}}$, called the weight factor of coherence, can be considered as the strength of explanation or competition respectively.[29]

The function $V_{\langle \mathbf{C},\mathbf{I}\rangle}$ satisfies the following principles:

(1) *Principle of Explanation*.
If $\mathcal{P} = \{P_1,\ldots,P_m\}$ explains Q and both $\mathcal{P}\cup\{Q\}$ and $\mathcal{P}\cup\{\neg Q\}$ are constituents, then $\mathcal{P}\cup\{Q\}$ coheres and $\mathcal{P}\cup\{\neg Q\}$ incoheres with the same weight factor $c_{\mathcal{P}}$.
(2) *Principle of Competition*.
If $\mathcal{P}$ is contradictory or competing and $\mathcal{P}$ is a constituent, then $\mathcal{P}$ incoheres.
(3) *Principle of Data Evidence*.
If there is positive evidence for E, then $\{E\}$ is coherent. If there is negative evidence for E, then there is positive evidence for $\neg E$.
(4) *Principle of Fuzzy Confirmation*.
The measure of coherence only depends on the coherent and incoherent constituents. If $\mathcal{P}$ coheres ($\mathcal{P} \in \mathbf{C}$), the degree of coherence is proportional to the fuzzy truth value of the conjunction of its elements. If $\mathcal{P}$ incoheres ($\mathcal{P} \in \mathbf{I}$), the degree of coherence is proportional to the negative fuzzy truth value of the conjunction of its elements.
(5) *Principle of Language Independence*.
Let P be a proposition which does not occur in any rule in $\mathcal{R}$. Then the rule system $\mathcal{R}'$ obtained from $\mathcal{R}$ by replacing each rule of the form '$\mathcal{Q}$ explains R' by the two rules '$\mathcal{Q}\cup\{P\}$ explains R', '$\mathcal{Q}\cup\{\neg P\}$ explains R' and each rule of the form '$\mathcal{Q}$ incoheres' by '$\mathcal{Q}\cup\{P\}$ incoheres', '$\mathcal{Q}\cup\{\neg P\}$ incoheres' induces the same order of coherence over $\mathcal{E}\cup\{P\}$ irrespective of the value of P.[30]

[29] Schoch (2000), p. 299.
[30] Cf. Schoch (2000), pp. 297-299.

Besides that it will not always be clear whether a set of propositions is competing[31], Schoch's fuzzy theory of explanatory coherence[32] is not comprehensible.[33]

However, let me stress that it would be unfair if I took this as a point against Schoch's account: He explicitely[34] distinguishes between a micro- and a macro-level on which theories of explanation can be formulated, where

> [...] the macro-level view takes the concept of explanation as an undefined primitive. It either inquires into the general properties of explanations, or uses explanatory relations in certain contextual frameworks.[35]

The general question Schoch is

> interested in is the problem of choice between concurrent hypotheses[36],

whence he takes on the macro-level view.

Apart from the above mentioned minor points, I consider Schoch's theory as a refinement of Thagard's TEC. In particular, Schoch is aware of the fact that the weight factors $c_{\mathcal{P}}$ have to be specified (in order to avoid arbitrariness).

> We introduce the [...] concept of an irreducible 'proper' explanation and define the weight factors only for them. [Footnote:] This must

[31] Schoch (2000), p. 298, mentions as example – which he ascribes to Thagard – two competing theories of dinosaur extinction, which could be caused by meteorite impact or a drop in sea-level; though

> these events are not mutually exclusive, scientists are interested in establishing the best explanation and therefore regard the two theories as competing.

I agree; but according to principles (2a) and (4b) of Bartelborth (1996) p. 193, the existence of several explanations of one and the same event may also lead to an increase in the coherence of a system of propositions – and Schoch does not give a criterion deciding whether the fact that a proposition R is explained by two distinct sets of propositions $\mathcal{P}$ and $\mathcal{Q}$ leads to an increase in the coherence of some set of propositions $\mathcal{T}$ containing $\mathcal{P}$, $\mathcal{Q}$, and R, or whether this yields $\mathcal{T}$ competing. A more modest principle of *contradiction* may be easier to handle.

[32] I take the fuzzy theory of explanatory coherence of Schoch (2000) to be given by the five principles mentioned above.

[33] This follows from assumption 2.2 (Comprehensible Concepts), and the fact that the fuzzy theory of explanatory coherence of Schoch (2000) presupposes as primitive, apart from the concept of competition, the concept of explanation.

[34] Cf. Schoch (2000), p. 291.

[35] Schoch (2000), p. 291.

[36] Schoch (2000), p. 291.

> also be done in order to avoid ambiguities in the weight factors, otherwise redundant parts of the explanations will effectively enlarge the weight factor if they are added.

The concept of a proper explanation is defined for rules: The rule '$\mathcal{P}$ explains Q' in the rule system $\mathcal{R}$ is a *proper explanation* if and only if it holds for every rule '$\mathcal{S}$ explains Q' with $\mathcal{S} \subseteq \mathcal{P}$ that $\mathcal{S} = \mathcal{P}$. The weight factor $c_{\mathcal{P}}$ for the set of propositions $\mathcal{P}$ is then defined as $c_{\mathcal{P}} = 2^{N_{\mathcal{R}}(P)}$, where $N_{\mathcal{R}}(P)$ is the number of propositions which are properly explained by $\mathcal{P}$ in the rule system $\mathcal{R}$.[37]

By doing so Schoch seems to escape the reproach of arbitrariness, for the function $V(x_1, \ldots, x_n)$ is uniquely determined for given values of the variables $x_1, \ldots, x_n$, if the weight factors $c_{\mathcal{P}}$ of all coherent *and* incoherent constituents $\mathcal{P} \subseteq \mathcal{E}$ are fixed.

However, despite the fact that in order for the concept of a proper explanation to be meaningful one has to assume that the data and the hypotheses can be partitioned into distinct atomic propositions so that counting propositions makes sense; and apart from the strange consequence that the weight factors of the constituents $\{E\}$ containing the data E are all equal to $2^{N_{\mathcal{R}}(\emptyset)}$, and thus increase exponentially with the number $N_{\mathcal{R}}(\emptyset)$ of data in $\mathcal{R}$[38]; there is no corresponding function which uniquely determines the weight factors $c_{\mathcal{Q}}$ of the *in*coherent constituents $\mathcal{Q} \in \mathbf{I}$, whence the weight factors are uniquely determined only for rule systems $\mathcal{R}$ without competing constituents, and the fuzzy measure for explanatory coherence turns out to be arbitrary, after all – and this it does in two respects.

Obviously, the function $V(x_1, \ldots, x_n)$ takes on different values for different values of the variables $x_1, \ldots, x_n$.

> The problem is to find a truth value assignment which maximizes explanatory coherence[39],

so that the set $\mathcal{E}$ of signed propositions can be partitioned into two disjoint sets of accepted and rejected propositions. It turns out that there are examples of rule systems $\mathcal{R}$ such that the sets of accepted and rejected propositions, into which

[37] Presumably $\mathcal{S}$ is supposed to be non-empty. Otherwise the weight factor $c_{\mathcal{P}}$ of any constituent $\mathcal{P} \neq \emptyset$ explaining at least one datum E is equal to 1, because

> [d]ata evidence is handled as a special instance of the explanation rule with an empty explanans.

Schoch (2000), p. 298.

[38] Cf. the preceding footnote.

[39] Schoch (2000), p. 292.

the underlying set $\mathcal{E}$ of signed propositions is partitioned by $\mathcal{R}$, differ with the weights assigned to some single incoherent constituent $\mathcal{Q} \in \mathbf{I}$.

This problem may be solved by fixing the weight factors for the incoherent constituents. The more important point is that there are examples of rule systems $\mathcal{R}_1, \mathcal{R}_2$ on a common set of signed propositions $\mathcal{E}$ (over some set of propositions $\mathcal{PR}$) such that (i) the explanatory coherence of rule system $\mathcal{R}_1$ is strictly greater than the explanatory coherence of the rule system $\mathcal{R}_2$, if the truth value assignment φ_1 which maximizes the explanatory coherence of $\mathcal{R}_1$ is adopted; (ii) the explanatory coherence of rule system $\mathcal{R}_2$ is strictly greater than the explanatory coherence of the rule system $\mathcal{R}_1$, if the truth value assignment φ_2 which maximizes the explanatory coherence of $\mathcal{R}_2$ is adopted; and (iii) the explanatory coherence of both the rule system $\mathcal{R}_1$ and the rule system $\mathcal{R}_2$ is 0, if any other truth value assignment φ is adopted, where the truth values are restricted to 0 and 1.[40] If, however, one considers the combined rule system $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$, then the explanatory coherence of $\mathcal{R}$ is 0 for every truth value assignment φ (including φ_1 and φ_2). This is shown in the proof of the next theorem.

Theorem 4.3 (Fuzzy Measure V Is Arbitrary) The fuzzy measure V for explanatory coherence of Schoch (2000) is arbitrary.

Let me note that the arbitrariness of $V(x_1, \ldots, x_n)$ is not caused by its being stated in the framework of fuzzy-logic.[41]

As already noted at the beginning of this chapter, coherence plays an important role as indicator of truth. What is characteristic of the problem situations where coherence enters is that one is given a set of statements S whose truth values are not known, and one wants to know whether believing or accepting (the statements in) S is justified. The coherence of S is then taken to provide the justification for believing S or accepting S as true (in some world or model $\mathcal{M}$) – the reason being that coherence is indicative of truth (in $\mathcal{M}$)[42].

If, however, the truth values (in $\mathcal{M}$) of all statements in S are known in advance, then there is no need of an indicator of truth (in $\mathcal{M}$) which justifies believing S or accepting S as true (in $\mathcal{M}$), and coherence can be dispensed with.

In short, coherence is of interest[43], only if the truth values – fuzzy or not – in some $\mathcal{M}$ of the statements in S are not all known in advance, and if coherence

[40] This can be done because of lemma 1.1 of Schoch (2000). Cf. Schoch (2000), p. 293.

[41] Cf. application 2.1 (Arbitrariness).

[42] Which it is only if it is not wholly independent of the model $\mathcal{M}$.

[43] Namely as justifier for believing S or accepting S as true (in $\mathcal{M}$) *via* being indicative of truth (in $\mathcal{M}$).

is adopted as truth indicator; otherwise – and under the assumption that the aim is to accept true statements, and to reject false ones – it is clear, or so I think, that one accepts the true and rejects the false ones.[44] So if one has to adopt in advance a truth value assignment to the statements in the sets of statements S_1 and S_2 in order to determine which of these two sets is more coherent, this presupposes too much: The whole idea of coherence as truth indicator is lead *ad absurdum*, if the truth values of those statements whose truth (in some model $\mathcal{M}$) should be indicated have to be assumed.

Before concluding let me note that I do not claim that no truth value (in some model $\mathcal{M}$) of any statement in a set S of such can be assumed to be given. On the contrary, in order for coherence to be indicative of truth in $\mathcal{M}$ this is even necessarily so (as argued above). However, supposing that the truth values (in $\mathcal{M}$) of all statements in S are given, trivially yields a truth-guaranteer (for $\mathcal{M}$), i.e. an algorithm for truth in $\mathcal{M}$.

The challenge is to define a function $Coh\,(\cdot)$ which determines the degree of coherence $Coh\,(S)$ of any set of statements S, and which does not presuppose the truth values of all statements in S, where the assumption of being true (in $\mathcal{M}$) or, more generally, of being epistemically distinguished (w.r.t. $\mathcal{M}$), should be as light as possible.

This will be done below by partitioning S into two disjoint subsets T and E, where the latter has to be an evidence.[45] This has the consequence that the measure of coherence is defined only for sets of statements with an evidence as a subset. However, as my interest is in the role of coherence in the context of assessing theories by evidences (relative to background knowledges), this is no serious restriction. Anyway, it will turn out that coherence w.r.t. evidence E, if modeled by the function $Coh(\cdot, E, \cdot)$ of below, is neither indicative of truth in any model $\mathcal{M} \in mod\,(E)$; nor sensitive to loveliness and likeliness in the sense of any power searcher $\mathcal{LO}$ and any truth indicator $\mathcal{LI}$; nor closed under equivalence transformations of T.

---

[44] Lemma 1.1 of Schoch (2000) states that one can restrict oneself to the classical truth values 'true' and 'false', if these are taken to be represented by the fuzzy truth values 1 and 0, respectively. This means that if the function $V\,(x_1, \ldots, x_n)$ has a global maximum for a distribution of values for the n real-valued variables $x_1, \ldots, x_n$ which are not all in the set $\{0, 1\}$, then there is always another distribution of values over $x_1, \ldots, x_n$ such that all these values are in the set $\{0, 1\}$, and such that this "classical" distribution yields the same coherence judgement (according to $V\,(x_1, \ldots, x_n)$), i.e. such that $V\,(x_1, \ldots, x_n)$ takes on the same global maximum for the "classical" distribution as it does for the first distribution.

[45] Besides this, the concept of coherence (w.r.t. the evidence) is relativised to a background knowledge B.

4.3 Foundationalist Coherentism

In this section I will present an own proposal for a measure of coherence. Before doing so, let me note why no probabilistic account is given.

4.3.1 Why No Probabilistic Measure of Coherence?

Intuitively, a set of statements H is coherent, if the statements in H hang together or fit each other. A probabilistic characterisation of this is the following: A (finite) set of statements $H = \{h_1, \ldots, h_n\}$ is coherent, if every statement h_i in H is likely to be true, given that the remaining statements h_j in H are true.

Consider a story: If the statements the story is composed of hang together or fit each other, and if each of them is likely to be true, given that the rest of them is – and if there is evidence for at least some of these statements – then, so the coherentist line of argument, one will be inclined to believe the story. If, however, the statements of the story do not hang together, or if not all of them are likely to be true, given that the remaining ones are true[46] – or if there is no evidence for any of them – then, according to the coherentist, one will not be inclined to believe the story.

How can this notion of coherence be made precise? The preceding paragraph suggests a probabilistic modeling that runs as follows: For a given conditional probability $p(\cdot \mid \cdot)$, the probabilistic degree of coherence of a finite set of statements $H = \{h_1, \ldots, h_n\}$, $ProbCoh_p(H)$, is given as

$$ProbCoh_p(H) = \sum_{1 \leq i \leq n} p\left(h_i \mid \bigwedge_{1 \leq j \neq i \leq n} h_j\right).$$

A measure with range $[0, 1]$ could then be defined as

$$ProbCoh'_p(H) = \frac{ProbCoh_p(H)}{n}.$$

Apart from the question whether coherence has anything to do with subjective degrees of belief (and, again, how propositions are counted), this – and any similarly defined – probabilistic measure of coherence is exposed to the reproach of arbitrariness.

[46] In particular, if there are some statements which are unlikely to be true or likely to be false, given that the remaining ones are true.

Bayesians[47] often argue that the prior distribution to the statements in a given language $\mathcal{L}$ does not really matter much, for the limit theorems of convergence to certainty and merger of opinion yield that the differences in the prior probabilities are "washed out"[48] in the long run. I disagree. In my opinion, the Bayesian faces the following dilemma:

(1) Either she admits that the distribution of the prior probabilities is not justified; and that one can explain nearly everything by choosing the "right" prior distribution – where, in principle, nothing is really explained, because this distribution is not justified.

(2) Or else she tries to justify the prior distribution by recourse to the limit theorems; and argues that different prior distributions do not matter, because the differences in the conditional probabilities based on them go to zero in the long run, i.e. if the number of statements which the conditional probabilities are conditional on goes to infinity.[49] But then the Bayesian has to assume the (fourth) axiom of countable additivity. The latter is problematic, because it forces one to play favourites in the sense that one has to assign different degrees of belief to the statements of a countably infinite set of mutually exclusive statements[50]. Furthermore, countable additivity yields the dogmatism of Bayesianism: Every agent that is rational in the sense of Bayesianism, i.e. coherent (consistent with the axioms of the probability calculus), has to be sure (in the sense of having a subjective degree of belief of 1) that every statement – of whatever complexity and quantifier structure – is equivalent to a verifutable statement; where a statement is *verifutable* just in case it is a truth-functional combination of verifiable and/or refutable statements.[51]

Moreover, the limit theorems do not really help much, for they do not tell one anything about the convergence rate.[52] Consider convergence to certainty. What this theorem says is that for every statement h, every subjective degree of belief function $p(\cdot)$, nearly[53] every possible world w, and every real number $\varepsilon > 0$

[47] In the tradition of Horwich (1982), Earman (1992), Howson/Urbach (1993). Cf. also Skyrms (2000).

[48] Cf. Earman (1992), p. 141.

[49] That there are cases where such a justification is of no help, even if the limit theorems would work for the short and medium runs, is argued for in the last section of chapter 2.

[50] Otherwise the disjunction of all these mutually exclusive statements would be assigned a degree of belief greater than 1.

[51] Cf. Earman (1992), p. 192.

[52] For the following cf. Earman (1992), p. 144ff; cf. also Gaifman/Snir (1982).

[53] All possible worlds w except those whose probability measure $Pr(w)$ is 0. The probability measure $Pr(\cdot)$, which is defined on the set of all possible worlds, is uniquely determined by the

there is at least one (time-) point m such that it holds for every (later time-) point $n > m$: the difference between the subjective degree of belief in h, given evidence $\bigwedge_{1\leq i\leq n} e_i(w)$ of world w at time n, $p\left(h \mid \bigwedge_{1\leq i\leq m} e_i(w)\right)$, and the truth value $\varphi(h, w)$ of h in w is at most ε, i.e.

$$\forall h \in \mathcal{L} \forall p(\cdot) \forall w \forall \varepsilon > 0 \exists m \forall n > m :$$

$$Pr(w) > 0 \rightarrow \left| p\left(h \mid \bigwedge_{1\leq i\leq n} e_i(w)\right) - \varphi(h, w)\right| < \varepsilon,$$

where $p(\cdot \mid \cdot)$ is the conditional probability based on the (unconditional) probability $p(\cdot)$, and it is assumed that for any two distinct possible worlds w_1 and w_2 there is at least one statement e_i in the set of statements $E = \{e_1, \ldots, e_n, \ldots\}$ such that e_i is true in w_1, but false in w_2, i.e. $\varphi(e_i, w_1) = 1$ and $\varphi(e_i, w_2) = 0$. What convergence to certainty does not say is when this (time-) point n is reached, whence one never knows that one believes in a true statement, if one believes in a true statement – under the assumption that the actual world is at all among nearly all possible worlds.

Something similar holds of merger of opinion, except that it is additionally assumed that the merging subjective degree of belief or (unconditional) probability functions $p(\cdot)$ and $p'(\cdot)$ are *equally dogmatic*. This means that the probability measures $Pr(\cdot)$ and $Pr'(\cdot)$, which are uniquely determined by $p(\cdot)$ and $p'(\cdot)$, respectively, assign the measure 0 to the same possible worlds w, so that the set of nearly all possible worlds is the same for p and p' respectively Pr and Pr'.

Eventually (and as already noted), Bayesianism is *theory enemy* or *theory hostile* in the sense that all what matters for the assessment of a hypothesis or theory T by some evidence E relative to some background knowledge B is the probability of T given E and B, or the boost in the probability of T that is caused by the addition of E to B, i.e. the difference (in whichever manner it is measured) of $p(T \mid E \wedge B)$ and $p(T \mid B)$. That is, the focus of Bayesianism is exclusively on the likeliness concept of confirmation. Other aspects[54], in particular those

(unconditional) probability $p(\cdot)$, which is defined on the underlying formal language $\mathcal{L}$. It thus depends on the (unconditional) probability $p(\cdot)$ (and the language $\mathcal{L}$), which possible worlds w are among nearly all possible worlds.

[54]Bayesian arguments to the effect that relevance measures as the distance measure

$$d_p(T, E \mid B) = p(T \mid E \wedge B) - p(T \mid B)$$

are sensitive to the variety of E depend on the right choice of the prior probability of E. These arguments run as follows (background knowledge B is suppressed): The prior probability $p(E)$ of corresponding to the loveliness concept of confirmation are neglected, except if they bear on (the boost in) the probability of T that results by adding E to B.

4.3.2 No Evidence Without Relevance

At the beginning of this chapter it has been argued that in order for coherence to be indicative of truth in the actual (or some other) world, coherence has to be relativised to this world. This has been done by relativising the coherence of a set of statements T to an evidence E, which is assumed to be true in the actual world. In a certain sense, this is *foundationalist coherentism*.

As an evidence is in general no complete description of the actual world, coherence w.r.t. the evidence is, properly speaking, not indicative of truth in the actual world – if it is truth indicative at all – but indicative of truth in $mod(E)$. The idea behind the concept of coherence w.r.t. the evidence can be sketched as follows:

Idea 1 (Informal Characterisation of Coherence w.r.t. E) Two statements h_1 and h_2 cohere with the world or the data, if their conjuntion $h_1 \wedge h_2$ says something about the world or the data which is not already said by one of h_1, h_2 alone.

Two statements h_1 and h_2 cohere the more with the world or the data, the more their conjuntion $h_1 \wedge h_2$ says about the world or the data which is not already said by one of h_1, h_2 alone.

This relation of coherence is symmetric in the sense that h_2 and h_1 cohere with the world or the data, if h_1 and h_2 do. It is stipulated that two statements logically contradicting each other do not cohere with the world; their degree of coherence w.r.t. the data is minimal. Furthermore, the evidential statements describing data about the world have a special status: They are epistemically distinguished in the sense of assumption 1.4 respectively 4.1. A more difficult question is whether it makes sense to call a single statement coherent with the data.

evidence E is the smaller, the greater the variety or diversity of E. As the conditional probability $p(T \mid E)$ of T given E is the greater, the smaller $p(E)$, it follows that, other things being equal, the degree of confirmation of T by E is the greater, the greater the variety of E – the other things being $p(T)$ and $p(E \mid T)$.

This is clearly seen in case T logically implies E, for here $p(T \mid E) = \frac{p(T)}{p(E)}$, which is the greater, the smaller $p(E)$, provided $p(T)$ is held constant.

As already noted, by choosing the "right" prior distribution one can explain nearly everything; for instance, that T is more confirmed by E, if the weather is nice than if it is not, for on sunny days one is inclined to assign high priors to T and low priors to E, whereas on rainy days it is the other way round.

The idea of coherence w.r.t. the evidence as informally characterised above is similar to the concept of relevance of Sperber/Wilson (1995) according to which

> [a]n assumption is relevant in a context if and only if it has some contextual effect in that context.[55]

Here,

> [...] the various types of possible contextual effects [include]: contextual implications, strengthenings, and contradictions resulting in the erasure of premises from the context.[56]

The important concept of a contextual implication is defined as follows:

> A set of assumptions **P** *contextually implies* an assumption *Q* in the context **C** if and only if (i) the Union of **P** and **C** non-trivially implies *Q*, (ii) **P** does not non-trivially imply *Q*, and (iii) **C** does not non-trivially imply *Q*.[57]

Without restrictions, the idea of above results in triviality in the sense that any two statements h_1 and h_2 (none of which logically implies the other) cohere, because there is always something the conjunction $h_1 \wedge h_2$ says which is not already said by one of h_1, h_2 alone – namely the conjunction $h_1 \wedge h_2$. In order to avoid this, Sperber/Wilson (1995) restrict the consequences of the union $\mathbf{P} \cup \mathbf{C}$ – in our case: the consequences of the conjunction $h_1 \wedge h_2$ – to non-trivial logical implications involving only elimination rules:

> A set of assumptions **P** *logically and non-trivially implies* an assumption *Q* if and only if, when **P** is the set of initial theses in a derivation involving only elimination rules, *Q* belongs to the set of final theses.

Another possibility[58] is to restrict the consequences of the conjunction $h_1 \wedge h_2$ to relevant (consequence-) elements in the sense of Schurz (1998) respectively Schurz/Weingartner (1987), and to consider

$$RE\left(h_1 \wedge h_2\right) \setminus \left(RE\left(h_1\right) \cup RE\left(h_2\right)\right).$$

[55] Sperber/Wilson (1995), p. 122.

[56] Sperber/Wilson (1995), p. 115.

[57] Sperber/Wilson (1995), p. 107-108.

[58] Note that restricting the consequences of $h_1 \wedge h_2$ to content parts in the sense of Gemes (1994c) and (1997a) is not sufficient, for $h_1 \wedge h_2$ is a content part of $h_1 \wedge h_2$, for any two statements h_1 and h_2. For more on the notion of a content part see below.

A third way of solving the problem that any two statements h_1 and h_2 none of which logically implies the other cohere is not to let it arise at all: This is the case if, for a given statement h, to say something about the world or the data means to account for some entity t mentioned in some evidence E.[59]

Definition 4.1 (Account) Let T, B, and S be (not necessarily finite) sets of wffs, let E be an evidence, and let 't' be a constant term occurring in E. T *accounts for* t respectively 't' *in* E *relative to* B iff there is a finite and non-redundant[60] $D \subseteq D_E(t)$ and a wff $A \in D$ such that

$$T \cup B \cup (D \setminus \{A\}) \vdash A.$$

The set of all constant terms 't' accounted for by T in E relative to B is called *the account of* T *in* E *relative to* B; it is denoted by '$A(T, E, B)$'.

The set of all constant i-terms 't_l^i' in $A(T, E, B) \cap C_{ess}(E)$ for which there is no $j < l$ such that

1. T accounts for t_j^i in E relative to B, and
2. $S \cup E \vdash t_j^i = t_l^i$,

is called the S*-representative of* $A(T, E, B)$. It is denoted by '$A_{S-repr}(T, E, B)$'.[61]

If T consists of a single wff h, '$A(h, E, B)$' and '$A_{S-repr}(h, E, B)$' are written instead of '$A(\{h\}, E, B)$' and '$A_{S-repr}(\{h\}, E, B)$', respectively.

In order for the problem of above to arise it would have to hold that for any statements h_1, h_2 (not logically implying each other), every evidence E, and every background knowledge B there is at least one constant term 't' $\in C(E)$ such that $h_1 \wedge h_2$ acounts for 't' in E relative to B, but h_1 does not, and h_2 does not either. Clearly, this is not the case – it suffices to give an example of two statements h_1, h_2 (not logically implying each other), an evidence E, and a background knowledge B such that it holds for every constant term 't' $\in C(E)$: If $h_1 \wedge h_2$ accounts for 't' in E relative to B, then so does one of h_1, h_2 alone.[62] In this sense there is no evidence without conclusion-relevance.

[59] Insofar as the notion of accounting for is defined in terms of relevant elements, this way is subsidiary to the second one of restricting the consequences to relevant elements.

[60] Non-redundancy should avoid triviality. As $t = t$ is a relevant element of E, for any 't' and any E, there is always a finite (but redundant) set of relevant elements of E – namely $D = \{t = t\}$ – and a wff $A \in D$ (namely $t = t$) such that $T \cup B \cup (D \setminus \{A\}) \vdash A$.

[61] The representative should avoid that an entity t with more than one name is counted more than once.

[62] $h_1 = \forall x Fx$, $h_2 = \forall x Gx$, $E = \{Fa, Gb\}$, and $B = \emptyset$ do the job.

In the following the distinction between the constant terms and the entities denoted by them is handled loosely, if no confusion can arise. Before turning to the measure of coherence w.r.t. the evidence, let me introduce a notion which will provide useful below: Power.

Definition 4.2 (Power) Let T and B be (not necessarily finite) sets of wffs, and let E be an evidence. The *power of T for E relative to B, $\mathcal{P}(T, E, B)$,* is given by the following equation:

$$\mathcal{P}(T, E, B) = \frac{|A_{B-repr}(T, E, B)|}{|C_{B-repr}(E)|}.$$ [63]

If T consists of a single wff h, '$\mathcal{P}(h, E, B)$' is written instead of '$\mathcal{P}(\{h\}, E, B)$'.

The power function $\mathcal{P}$ is discussed to a greater extent in the chapter on loveliness and likeliness. For the moment it suffices to note that $\mathcal{P}$ is a power searcher which is formally handy for finite sets of statements T and B.

4.3.3 The Measure of Coherence w.r.t. the Evidence

The informal characterisation of coherence w.r.t. the evidence is generalized by the surplus of a set of statements T w.r.t. an evidence E and a set of statements B.

Definition 4.3 (Surplus) Let T, B, and V be (not necessarily finite) sets of wffs, and let E be an evidence. The *surplus of T in E relative to B, $S(T, E, B)$*, is the set of constant terms 't' which are accounted for by T in E relative to B, but by none of its proper subsets T', i.e.

$$S(T, E, B) = A(T, E, B) \setminus \bigcup_{T' \subset T} A(T', E, B).$$

The set of constant i-terms 't_l^i' in $S(T, E, B) \cap C_{ess}(E)$ for which there is no $j < l$ such that

$$\text{`}t_j^i\text{'} \in S(T, E, B) \quad \text{and} \quad E \cup V \vdash t_j^i = t_l^i$$

is called the *V-representative of $S(T, E, B)$*. It is denoted by '$S_{V-repr}(T, E, B)$'.

If T consists of a single wff h, '$S(h, E, B)$' and '$S_{V-repr}(h, E, B)$' are written instead of '$S(\{h\}, E, B)$' and '$S_{V-repr}(\{h\}, E, B)$', respectively.

[63] $C_{B-repr}(E)$ is always non-empty.

Some immediate consequences of this definition are the following.

Theorem 4.4 (Surplus) Let T and B be (not necessarily finite) sets of wffs, and let E be an evidence.

1. $S(T, E, B) = \emptyset$, if T is infinite,
2. $S(\emptyset, E, B) = A(\emptyset, E, B) = A(B, E, B) = A(T, E, B)$, if $B \vdash T$,
3. $S(B, E, B) = \emptyset$, if $B \neq \emptyset$,
4. $S(T, E, B) = \emptyset$, if $T \neq \emptyset$ and $B \vdash T$, and
5. $S(h_T, E, B) = A(h_T, E, B) = A(T, E, B)$, for every single wff h_T with $h_T \dashv\vdash T$, if $A(\emptyset, E, B) = \emptyset$.

The measure of coherence w.r.t. the evidence is defined as follows.

Definition 4.4 (Coherence w.r.t. the Evidence) Let T be a finite set of wffs, let E be an evidence, and let B be a (not necessarily finite) set of wffs. The *degree of coherence of T w.r.t. E relative to B*, $Coh(T, E, B)$, is defined as follows:

If $T \neq \emptyset$ and $T \cup B \cup E \nvdash \bot$, then

$$Coh(T, E, B) = \sum_{\emptyset \neq T' \subseteq T} \frac{|S_{B-repr}(T', E, B)|}{|C_{B-repr}(E)| \cdot (2^{|T|} - 1)};^{64}$$

otherwise, $Coh(T, E, B) = 0$.

Let $T = \{h_1, \ldots, h_n, \ldots\}$ be a countably infinite set of wffs, and let $T_i := \{h_1, \ldots, h_i\}$ for some enumaration $h_1, \ldots, h_n, \ldots$ of the wffs in T. The *degree of coherence of T w.r.t. E relative to B*, $Coh(T, E, B)$, is defined as follows:

If $\lim_{i\to\infty} Coh(T_i, E, B)$ exists, and is the same for every enumeration $h_1, \ldots, h_n, \ldots$ of the wffs in T, then

$$Coh(T, E, B) = \lim_{i\to\infty} Coh(T_i, E, B);$$

otherwise $Coh(T, E, B) = 0$.

If T consists of a single wff h, '$Coh(h, E, B)$' is written instead of '$Coh(\{h\}, E, B)$'.

[64] $C_{B-repr}(E)$ is always non-empty.

The notion of the surplus and consequently also the measure of coherence w.r.t. evidence Coh are implicitely sensitive to aspects of premise-relevance: If a constant term 't' is in the surplus of some T w.r.t. some evidence E relative to some B, then every statement $h \in T$ is necessary in order for T to account for 't' in E relative to B.

Theorem 4.5 (Coh Is Formally Handy) $Coh(\cdot,\cdot,\cdot)$,

$$Coh(\cdot,\cdot,\cdot) : \wp_{fin}(\mathcal{L}_{PL1=}) \times \mathcal{E} \times \wp_{fin}(\mathcal{L}_{PL1=}) \to \Re,$$

is non-arbitrary, comprehensible, and computable in the limit, where $\wp_{fin}(\mathcal{L}_{PL1=})$ is the set of all finite sets of wffs of $\mathcal{L}_{PL1=}$.

This holds in particular, if T is a theory and B is a background knowledge. Let us turn to some examples.

4.3.4 Examples

In the following T is a finite set of statements and B is empty. '$Coh(T,E)$' stands for '$Coh(T,E,\emptyset)$'; similarly for '$\mathcal{P}(T,E)$'.

(1) The first example illustrates that coherence coincides with power if single hypotheses h are considered. Let

$$E = \{Fa_1, Ga_1, \ldots, Fa_n, Ga_n\} \quad \text{and} \quad T = \{\forall x\,(Fx \to Gx)\}.$$

Then

$$Coh(T,E) = \frac{n}{n \cdot (2^1 - 1)} = 1 = \frac{n}{n} = \mathcal{P}(T,E).$$

(2) The second example is one where a unified theory T_G is more coherent w.r.t. evidence E than the union T_H of two theories T_{H_1} and T_{H_2}. Let

$$\begin{aligned} E &= \{Fa_1, Ga_1, H_1a_1, Fa_2, Ga_2, H_2a_2\}, \\ T_{H_1} &= \{\forall x\,(Fx \to H_1x)\}, \\ T_{H_2} &= \{\forall x\,(Fx \to H_2x)\}, \\ T_H &= T_{H_1} \cup T_{H_2} = \{\forall x\,(Fx \to H_1x), \forall x\,(Fx \to H_2x)\}, \\ T_G &= \{\forall x\,(Fx \to Gx)\}. \end{aligned}$$

Then

$$Coh\left(T_{H_1}, E\right) = Coh\left(T_{H_2}, E\right) = \frac{1}{2 \cdot (2^1 - 1)} = 1/2,$$

and

$$Coh(T_H, E) = \frac{1+1+0}{2 \cdot (2^2 - 1)} = 1/3 < 1 = \frac{2}{2 \cdot (2^1 - 1)} = Coh\left(T_G, E\right).$$

(3) The third and fourth example show that $Coh\left(T, E\right)$ is not closed under equivalence transformations of T. Let

$$\begin{aligned}
E &= \{Fa, Ga\}, \\
T_1 &= \{\forall x\,(Fx \rightarrow Gx), \forall x\,(Gx \rightarrow Fx), \forall x\,(Fx \rightarrow Hx), \forall x\,(Hx \rightarrow Fx)\}, \\
T_2 &= \{\forall x\,(Fx \rightarrow Gx), \forall x\,(Gx \rightarrow Fx), \forall x\,(Fx \rightarrow Hx), \forall x\,(Hx \rightarrow Gx)\}, \\
T_3 &= \{\forall x\,(Fx \rightarrow Gx), \forall x\,(Gx \rightarrow Hx), \forall x\,(Hx \rightarrow Fx)\}, \\
T_4 &= \{\forall x\,(Fx \rightarrow Hx), \forall x\,(Hx \rightarrow Gx), \forall x\,(Gx \rightarrow Fx)\}, \\
T_5 &= \{\forall x\,(Fx \leftrightarrow Gx), \forall x\,(Gx \leftrightarrow Hx)\}, \\
T_6 &= \{\forall x\,(Fx \leftrightarrow Hx), \forall x\,(Hx \leftrightarrow Gx)\}.
\end{aligned}$$

$T_i \dashv\vdash T_j$, for every i and $j, 1 \leq i, j \leq 6$, but

$$\begin{aligned}
Coh\left(T_1, E\right) = \frac{2}{1 \cdot (2^4 - 1)} = 2/15 \quad &< \quad 3/15 = \frac{3}{1 \cdot (2^4 - 1)} = Coh\left(T_2, E\right) \\
&< \quad 2/7 = \frac{2}{1 \cdot (2^3 - 1)} = Coh\left(T_3, E\right) \\
&= \quad 2/7 = \frac{2}{1 \cdot (2^3 - 1)} = Coh\left(T_4, E\right) \\
&< \quad 1/3 = \frac{1}{1 \cdot (2^2 - 1)} = Coh\left(T_5, E\right) \\
&= \quad 1/3 = \frac{1}{1 \cdot (2^2 - 1)} = Coh\left(T_6, E\right).
\end{aligned}$$

(4) Let

$$\begin{aligned}
E &= \{Fa, Ga, Ra\}, \\
T_1 &= \{\forall x\,(Fx \rightarrow Gx), \forall x\,(Gx \rightarrow Hx), \forall x\,(Hx \rightarrow Fx), \forall x\,(Rx \rightarrow Hx)\}, \\
T_2 &= \{\forall x\,(Fx \leftrightarrow Gx), \forall x\,(Gx \leftrightarrow Hx), \forall x\,(Rx \rightarrow Hx)\}, \\
T_3 &= \{\forall x\,(Fx \leftrightarrow Hx), \forall x\,(Hx \leftrightarrow Gx), \forall x\,(Rx \rightarrow Hx)\}.
\end{aligned}$$

$T_i \dashv\vdash T_j$, for every i and $j, 1 \leq i, j \leq 3$, but

$$\begin{aligned}
Coh\,(T_1, E) &= \frac{1+0+0+0+0+0+0+1+0+1+0+0+0+0+0}{1 \cdot (2^4 - 1)} \\
&= 3/15 \\
&< 2/7 = \frac{1+0+0+0+0+1+0}{1 \cdot (2^3 - 1)} = Coh\,(T_2, E) \\
&< 3/7 = \frac{0+0+0+1+1+1+0}{1 \cdot (2^3 - 1)} = Coh\,(T_3, E)\,.
\end{aligned}$$

(5) The fifth example shows that coherence decreases with the number of statements in T that are $\mathcal{P}$-superfluous for E and $\emptyset$. Let

$$E = \{Fa_1, \ldots, Fa_n\} \quad \text{and} \quad T = \{\forall x Fx, \forall x G_1 x, \ldots, \forall x G_m x\}\,.$$

Then

$$Coh\,(T, E) = \frac{n}{n \cdot (2^{m+1} - 1)} \to 0, \quad \text{if} \quad m \to \infty.$$

(6) The sixth example shows that theories T which are, in an intuitive sense, “internally” coherent may have a higher degree of coherence w.r.t. an evidence E than theories consisting of isolated subtheories, *if* this “internal” coherence of T yields that more entities (constant terms) are accounted for by T, or that some of them are accounted for in different ways. Let

$$\begin{aligned}
E &= \{Ha_1, \ldots, Ha_n\}\,, \\
T_1 &= \{\forall x Fx, \forall x Gx, \forall x Hx\}\,, \\
T_2 &= \{\forall x Fx, \forall x Gx, \forall x\,(Fx \vee Gx \to Hx)\}\,.
\end{aligned}$$

Although $T_1 \dashv\vdash T_2$,

$$\begin{aligned}
Coh\,(T_1, E) &= \frac{0+0+n+0+0+0+0}{n \cdot (2^3 - 1)} \\
&= 1/7 \\
&< 2/7 \\
&= \frac{0+0+0+0+n+n+0}{n \cdot (2^3 - 1)} = Coh\,(T_2, E)\,.
\end{aligned}$$

(7) The seventh example shows that theories T which are in the above intuitive sense “internally” coherent need not have a higher degree of coherence w.r.t. an

evidence E than theories consisting of isolated subtheories. They may even have a lower degree of coherence w.r.t. an evidence E, *if* this "internal" coherence of T does *not* yield that more entities (constant terms) are accounted for by T, or that some of them are accounted for in different ways. Let

$$\begin{aligned} E &= \{Ha_1, \ldots, Ha_l\}, \\ T_1 &= \{\forall x F_1 x, \ldots, \forall x F_m x, \forall x Hx, \forall x (F_1 x \to G_1 x), \ldots \\ &\quad \ldots, \forall x (F_m x \to G_m x), \forall x (Hx \to G_{m+1} x), \\ &\quad \forall x (G_{m+1} x \to G_{m+2} x), \ldots, \forall x (G_{n-1} x \to G_n x)\}, \\ T_2 &= \{\forall x F_1 x, \ldots, \forall x F_m x, \forall x (F_1 x \vee \ldots \vee F_m x \to Hx), \\ &\quad \forall x G_1 x, \ldots, \forall x G_n x\}, \end{aligned}$$

where $l \geq 1$ and $1 < m < n$. Although $T_1 \dashv\vdash T_2$,

$$\begin{aligned} Coh(T_1, E) &= \frac{l}{l \cdot (2^{m+n+1} - 1)} \\ &< \frac{l \cdot m}{l \cdot (2^{m+n+1} - 1)} = Coh(T_2, E). \end{aligned}$$

(8) As mentioned, the necessary condition 4.1 (Coherence as Truth Indicator) allows for the existence of sets of statements T and models $\mathcal{M}_1$ and $\mathcal{M}_2$ with

$$Coh(T, \mathcal{M}_1) = Coh(T, \mathcal{M}_2).$$

For $Coh(\cdot, \cdot)$, this is illustrated by the last example. Let

$$\begin{aligned} E_1 &= \{Pa, Qa, Fb_1, \ldots, Fb_n\}, \\ E_2 &= \{Pa, Qa, Gc_1, \ldots, Gc_n\}, \\ T &= \{\forall x (Px \to Qx)\}. \end{aligned}$$

Then

$$Coh(T, E_1) = \frac{1}{(n+1) \cdot (2^1 - 1)} = Coh(T, E_2).$$

4.3.5 Properties of Coh

As is obvious from the examples of the last subsection, the values of Coh depend heavily on the formulation of T. In particular, Coh is not closed under equivalence transformations of T.

Theorem 4.6 (No InvEquTrans of T for Coh) For every evidence E, and every set of wffs B there are theories T and T' such that

$$T \dashv\vdash T' \quad \text{and} \quad Coh\,(T, E, B) \neq Coh\,(T', E, B)\,,$$

provided there is at least one theory T with $Coh\,(T, E, B) \neq 0$.

Theorem 4.7 (Coherence Versus Power) Let T be a finite set of wffs, let E be an evidence, and let B be a set of wffs. If $T \cup B \cup E \not\vdash \bot$ and $A_{B-repr}\,(\emptyset, E, B) = \emptyset$, then

$$Coh\,(T, E, B) \leq Coh\left(\bigwedge_{h \in T}, E, B\right) = \mathcal{P}\,(T, E, B)\,,$$

where $\mathcal{P}$ is closed under equivalence transformations of T and B.

Remember, $\mathcal{P}$ is not only closed under equivalence transformations of T and B; it is also searching power and formally handy for finite sets of statements T and B.

What do the above theorems tell us? There are at least the following four interpretations.

1. The definition of coherence (w.r.t. the evidence) respectively its measure Coh is not adequate – e.g. because the relation of accounting for is monotone w.r.t. T and B.[65]

2. The concept of coherence (w.r.t. the evidence) can be dispensed with. The concept of power (for the evidence) is sufficient and has the advantage that its measure $\mathcal{P}$ is closed under equivalence transformations of T, whence theories T may be defined as sets of models $mod\,(T)$ without restricting oneself to some "canonical" formulation of T.

3. It does not make sense to call a single statement – as the conjunction $\bigwedge_{h \in T} h$ – coherent w.r.t. an evidence E, because the concept of coherence (w.r.t. an evidence) makes only sense, if several statements (propositions) are considered.

4. The set of statements T whose coherence w.r.t. some evidence E is to be assessed, has to be formulated in some special, perhaps uniquely determined way.

[65]That is, if T accounts for 't' in E relative B, then so does every T' logically implying T; and T does so relative to every B' logically implying B. This is not the case for E, because $\vdash_{crel}$ is not monotone.

Obviously, (3) and (4) are intimately related. If one considers these four alternatives as the only serious interpretations of the above theorems; if one adopts Coh as measure of coherence w.r.t. the evidence; and if one does not already give up the concept of coherence at this point of the discussion, then (3) and (4) are the only possible alternatives.

How, then, has the set of statements T be formulated? Intuitively, T should be formulated naturally in the sense of being split up into its smallest (content) parts. There are at least two approaches to this end: The first is based on Schurz' notion of a relevant (consequence-) element.[66]

1. The formulation of T has to be an *irreducible representation of* T, i.e. a non-redundant set I of relevant elements of T such that $I \dashv\vdash T$.

The second is based on Gemes' notion of a content part.[67]

2. The formulation of T has to be a *natural axiomatization of* T, i.e. a finite set of wffs A such that

   2.1 $A \dashv\vdash T$,

   2.2 every wff $h \in A$ is a content part of (the conjunction of all wffs in) A,

   2.3 there is no content part c_h of some wff $h \in A$ such that $A \setminus \{h\} \vdash c_h$, and

   2.4 there is no finite set of wffs A' satisfying (2.1)-(2.3) with $|A'| > |A|$.[68]

However, if Coh has to be closed under equivalence transformations of T, then neither (1) nor (2) is viable, for there are theories T_1 and T_2, evidences E, and background knowledges B such that both T_1 and T_2 are irreducible representations and natural axiomatizations of T_1, and such that $Coh\,(T_1, E, B) \neq Coh\,(T_2, E, B)$ – this is shown by T_1 and T_2 of example (3) of the preceding subsection.[69]

[66] Cf. Schurz (1991a), (1998), and Schurz/Weingartner (1987).

[67] Cf. Gemes (1993), (1994c), and (1997a).

[68] The fourth clause is added by Gemes in a footnote – cf. Gemes (1993), p. 483. Without it the concept of a natural axiomatization is of no help here, for the conjunction $\bigwedge_{h \in A} h$ of all wffs of a finite set of wffs A satisfying clauses (2.1)-(2.3) for a given set of wffs T also satisfies (2.1)-(2.3) for T.

[69] If '$\leftrightarrow$' is not eliminated, and the notion of content part is formulated as follows:

> For any wffs A and B: B is a content part of A iff (i) A and B are contingent, (ii) $A \vdash B$, and (iii) there is no wff C such that $A \vdash C$, $C \vdash B$, $B \nvdash C$, and C is formulated in the vocabulary of B,

Nevertheless, suppose these cases are only very rare, so that for the most part $Coh\,(T', E, B)$ is the same for all irreducible representations or natural axiomatizations T' of T. Then the measure of coherence w.r.t. the evidence Coh can be made invariant under equivalence transformations of T by defining it in one of the following two ways:

$$\begin{aligned} Coh_{irr}\,(T, E, B) &= \max\left\{Coh\,(T', E, B) : T' \in \mathbf{I}\,(T)\right\},\\ Coh_{na}\,(T, E, B) &= \max\left\{Coh\,(T', E, B) : T' \in \mathbf{NA}\,(T)\right\}. \end{aligned}$$

One could, of course, take some other function instead of the maximum function. Note that it would not be of help to consider all sets of statements T' with $T' \dashv\vdash T$, for then one would nearly always – whenever T is a theory and $A_{B-repr}\,(\emptyset, E, B) = \emptyset$ – consider the singleton $\{\bigwedge_{h' \in T'} h'\}$ containing the conjunction of all statements of some finite axiomatization T' of T. Though this does not hold of the minimum function, a similar problem arises in this case, for one would have to deal with the set of all statements logically following from T.

A question not yet answered is whether for every set of statements T of $\mathcal{L}_{PL1=}$ there is at least one irreducible representation or natural axiomatization of T. For the propositional calculus, the answer is affirmative for irreducible representations: In Schurz/Weingartner (1987) it is shown that for every statement A of $\mathcal{L}_{PC}$ there is a statement A' such that $A \dashv\vdash A'$ and $A' \vdash_{crel} A'$.[70]

Suppose, however, all these problems can be dealt with in a satisfying way. Is coherence w.r.t. the evidence under these assumptions indicating truth in the actual world? The answer is no, for it is not even indicating truth in $mod\,(E)$: For a given evidence E, there are always theories T_E and background knowledges B_E such that

$$E \cup B_E \vdash T_E \quad \text{and} \quad Coh\,(T_E, E, B_E) = 0$$

– which violates the second clause of the definition of indicating truth in $mod\,(E)$.

Still, one may argue that although the feature of interest usually ascribed to coherence is that of being truth indicative, this is not what coherence should do in case of the assessment of theory by evidence relative to background knowledge.

then only T_5 and T_6, but neither T_1 nor T_2 of example (3) are natural axiomatizations of T_1. However, for this notion – which is the one Gemes (in personal correspondence) favours – T_2 and T_3 of example (4) are sufficient to show that Coh is not closed under equivalence transformations of T, even if T has to be a natural axiomatization of itself (based on the notion of content part just stated). T_1 of this example is an irreducible representation, but no natural axiomatization of itself (in second sense of this footnote).

[70]Cf. Schurz/Weingartner (1987), p. 58.

Here the job of coherence w.r.t. the evidence is not to indicate truth in $mod\,(E)$, for some evidence E, but to be sensitive to loveliness and likeliness in the sense of some power searcher $\mathcal{LO}$ and some truth indicator $\mathcal{LI}$. However, this does not either hold of Coh – even if it is assumed that there is exactly one canonical formulation F_T for every set of statements T.

Theorem 4.8 (No SensLoveLike of Coh) For every power searcher $\mathcal{LO}$, every truth indicator $\mathcal{LI}$, and every evidence E there is a theory T_E and a background knowledge B_E such that it holds for any sets of wffs T and B, and every evidence E': If $T \dashv\vdash T_E$, $E' \dashv\vdash E$, and $B \dashv\vdash B_E$, then

1. $T \cup B \vdash E'$, and thus $\mathcal{LO}\,(T, E', B) = 1$,
2. $E' \cup B \vdash T$, and thus $\mathcal{LI}\,(T, E', B) = 1$, and
3. $Coh\,(T, E', B) = 0$.

I conclude that if Coh captures to some extent the concept of coherence (w.r.t. the evidence), then the latter has to be given up as indicator of truth in the actual world, provided the second clause of the definition of indicating truth in $mod\,(E)$ is adopted as minimal requirement for any truth indictor f: If evidence E together with background knowledge B guarantees (in the sense of logical implication) the truth of some theory T, then the degree to which f indicates the truth of T in $mod\,(E)$ relative to B is maximal.

I conclude further that Coh is no adequate measure of confirmation, because it is not sensitive to loveliness and likeliness in the sense of any power searcher $\mathcal{LO}$ and any truth indicator $\mathcal{LI}$ – even if it is assumed that there is exactly one canonical formulation F_T for every set of statements T.

In the next chapter I will therefore pursue the second approach to a solution of the problem of a quantitative theory of confirmation: First, to define for every (primary) confirmational virtue V a function $f_V\,(\cdot,\cdot,\cdot)$ such that $f_V\,(T, E, B)$ measures the degree to which (primary) confirmational virtue V is exhibited by T, E, and B, for every theory T, every evidence E, and every background knowledge B; and then to define the measure of confirmation C as a function of (some of) these functions f_V.

Chapter 5

Loveliness and Likeliness

5.1 Recapitulation

This chapter contains the definition of a power searcher $\mathcal{P}(\cdot,\cdot,\cdot)$, $\mathcal{P}(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, and a truth indicator $\mathcal{LI}(\cdot,\cdot,\cdot)$, $\mathcal{LI}(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, which together determine the measure of confirmation $C(\cdot,\cdot,\cdot)$.

We know from theorem 3.1 that there are lots of power searchers and truth indicators. As has been noted there, what is needed are a power searcher and a truth indicator which are formally handy, i.e. non-arbitrary, comprehensible, and computable in the limit. Arbitrariness will be avoided by defining two single functions (without parameters that can be chosen arbitrarily); comprehensibility will be achieved by purely syntactical definitions in the terms of $PL1 =$ and ZF; computability in the limit will be a consequence of these definitions.

Before defining these functions remember that non-arbitrariness, comprehensibility, and computability in the limit are *formal* conditions of adequacy for any formal theory, in particular, any quantitative theory of confirmation intended to implicitely provide a rule of acceptance for rational theory choice. Sensitivity to (and only to) the confirmational virtues is a *material* condition of adequacy for any quantitative theory of confirmation. The second approach to a solution of the problem of a quantitative theory of confirmation defines the measure of confirmation C as a function of (some of) the functions $f_V(\cdot,\cdot,\cdot)$ measuring the confirmational virtues V. The formal conditions of adequacy for these functions are inherited from those for C. The material conditions of adequacy are those of chapter 3: The function f_{LO} which measures loveliness has to be a power searcher; the function f_{LI} which measures likeliness has to be a truth indicator.

There will not be any functions for the derived confirmational virtues. These enter only when the resulting measure of confirmation C is considered. C has to be sensitive to the primary and derived confirmational virtues, i.e. sensitive to loveliness and likeliness in the sense of some power searcher $\mathcal{LO}$ and some truth indicator $\mathcal{LI}$ – sensitivity to simplicity considerations and unimpressability by redundancies (and invariance under equivalence transformations of T) being consequences of this.

In the next chapter C is combined with a function $G(\cdot,\cdot,\cdot)$, where $G(T,E,B)$ measures the "goodness" of evidence E in relation to theory T and background knowledge B. The reason for this is that C does not and is not intended to take into account that evidence which is varied or diverse is better than evidence which is uniform or homogenous; and that E is the better, the more information it contains. $C(T,E,B)$ only tells you how much T is confirmed by E relative to B, if E is all the evidence available. The refined measure of confirmation $C^*(\cdot,\cdot,\cdot)$ which is the result of combining C and G can be shown to be sensitive to diversity considerations in the sense of C and G.[1] Before continuing, let me note that G is independent of C, and may be combined with any measure of confirmation – or coherence (w.r.t. the evidence).

5.2 A Power Searcher and a Truth Indicator

The basic ideas behind the definitions of the functions $\mathcal{P}$ and $\mathcal{LI}$ are due to Carl Gustav Hempel, and can be found in his *Studies in the Logic of Confirmation* (1945) under the headings of the prediction criterion and the satisfaction criterion, respectively.

It is crucial that these functions are only defined, if the evidential domains and the domains of proper investigation overlap. Any domain which is among both is called a *confirmational domain* (of T and E). Though the definitions are stated in semantic terms, they are purely syntactic, because the domains are only distinguished by means of the different sorts of variables and constants occurring in T, E, and B.

The evidential domains and the domains of proper investigation overlap whenever there is an essential occurrence of an i-variable, but no occurrence of a constant i-term in T, and no occurrence of an i-variable, but a constant i-term essentially occurring in E, for some sort i of variables and constants. The domains

[1] For a definition of the concept of sensitivity to diversity considerations in the sense of some functions $\mathcal{C}_{\mathcal{LO},\mathcal{LI}}$ and $\mathcal{G}$ see chapter 6.

of T, E, and B are also only distinguished by the sorts of variables and constants occurring in T, E, and B. Strictly speaking, these "domains" are *domain variables* taking domains as their values.

So referring to the domains of T, E, and B, $D_1, \ldots, D_n$, is just another way of referring to n different sorts of variables and constants occurring in T, E, and B. The functions $\mathcal{P}$ and $\mathcal{LI}$ therefore have three argument places; the confirmational domains are uniquely determined by T and E. Technically, the value $\mathcal{P}$ takes on for given T, E, and B, is a *vector* whose length equals the number of confirmational domains of T and E, and is of the form: $\langle \mathcal{P}(T, E, B; D_1), \ldots, \mathcal{P}(T, E, B; D_c) \rangle$, where $D_1, \ldots, D_c$ are the confirmational domains of T and E. The claim that $\mathcal{P}$ is a power searcher means that conditions (1)-(3) in the definition of searching power for $mod(E)$ are satisfied by $\mathcal{P}(T, E, B; D_i)$ for *every* confirmational domain D_i of T and E, for all T, E, and B. Similar remarks apply to $\mathcal{LI}$ and its being a truth indicator.

The function $\mathcal{P}$ is already familiar from the chapter on coherence w.r.t. the evidence.

Definition 5.1 (Confirmational Domain) Let T be a theory with domains of proper investigation $D_1^T, \ldots, D_m^T$, let E be an evidence from $D_1^E, \ldots, D_n^E$, and let D_i be a domain (with corresponding i-variables and constant i-terms).

D_i is a *confirmational domain of T and E* iff D_i is among both the evidential domains of E, $D_1^E, \ldots, D_n^E$, and the domains of proper investigation of T, $D_1^T, \ldots, D_m^T$; i.e. iff T contains an essential occurrence of an i-variable, but no occurrence of a constant i-term, and E contains an essential occurrence of a constant i-term, but no occurrence of an i-variable.

Definition 5.2 (Power) Let T be a theory, let E be an evidence, let B be a background knowledge, and let D_i be a confirmational domain of T and E (with corresponding i-variables and constant i-terms).

The *power of T for E relative to B in D_i*, $\mathcal{P}(T, E, B; D_i)$, is given by the following equation:

$$\mathcal{P}(T, E, B; D_i) = \frac{|A_{B-repr}(T, E, B) \cap C_i|}{|C_{B-repr}(E) \cap C_i|},$$

where C_i is the set of constant i-terms.[2]

[2] $C_{B-repr}(E) \cap C_i \neq \emptyset$, because D_i is a confirmational domain of T and E, and hence among the evidential domains of E.

The function $\mathcal{LI}$ has not been dealt with so far.

Definition 5.3 (Likeliness) Let T be a theory, let E be an evidence, let B be a background knowledge, and let D_i be a confirmational domain of T and E (with corresponding i-variables and constant i-terms).

The *likeliness of* T *w.r.t.* E *and* B *in* D_i, $\mathcal{LI}(T, E, B; D_i)$, is given by the following equation:

$$\mathcal{LI}(T, E, B; D_i) = \frac{\max_{\mathcal{LI}}(T, E, B; D_i)}{|C_{B-repr}(E) \cap C_i|},^{3}$$

provided $E \cup B \nvdash \bot$, where

$$\begin{aligned}\max_{\mathcal{LI}}(T, E, B; D_i) \quad := \quad & \max\{|C \cap C_{B-repr}(E)| : C \subseteq C_{E,B,i}, \\ & E \vdash Dev_{C_{E,B,i}}(B) \rightarrow Dev_C(T)\Big\},\end{aligned}$$

$C_{E,B,i} := C(E \cup B) \cap C_i = C_i(E \cup B)$, $C_i(X)$ is the set of constant i-terms occurring in X, and C_i is the set of constant i-terms.

Concerning likeliness in domain D_i, it is important to note that only the i-variables in T are replaced by the constant i-terms of C in the development of T for C, $Dev_C(T)$; the k-variables, $k \neq i$, occurring in T and the quantifiers binding them remain unchanged (cf. definition 1.11).

The following theorems yield that $\mathcal{P}$ and $\mathcal{LI}$ satisfy the formal and material conditions of adequacy.

Theorem 5.1 ($\mathcal{P}$ Is a Formally Handy Power Searcher) $\mathcal{P}(\cdot,\cdot,\cdot), \mathcal{P}(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, is a power searcher which is non-arbitrary, comprehensible, and computable in the limit, provided for every $E \in \mathcal{E}$ and every 't' $\in C_{ess}(E)$ there is a contingent[4] $A \in RE(E)$ with 't' $\in C(A)$.

More precisely, $\mathcal{P}$ is formally handy, and for any theories T and T', every evidence E, every background knowledge B, and every confirmational domain D_i of T and E, and of T' and E:

1. $\mathcal{P}(T, E, B; D_i) \geq 0$,

2. if $T \cup B \vdash E$, then $\mathcal{P}(T, E, B; D_i) = 1$, and

[3]Cf. the preceding footnote.

[4]Contingency should rule out $t = t$, which is a relevant consequence of any E.

3. if $T' \vdash T$, then $\mathcal{P}(T', E, B; D_i) \geq \mathcal{P}(T, E, B; D_i)$,

provided for every $E \in \mathcal{E}$ and every 't' $\in C_{ess}(E)$ there is a contingent $A \in RE(E)$ with 't' $\in C(A)$.

Theorem 5.2 ($\mathcal{LI}$ Is a Formally Handy Truth Indicator) $\mathcal{LI}(\cdot,\cdot,\cdot)$, $\mathcal{LI}(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \to \Re$, is a truth indicator which is non-arbitrary, comprehensible, and computable in the limit.

More precisely, $\mathcal{LI}$ is formally handy, and for any theories T and T', every evidence E, every background knowledge B, and every confirmational domain D_i of T and E, and of T' and E: If $E \cup B \nvdash \perp$, then

1. $\mathcal{LI}(T, E, B; D_i) \geq 0$,

2. if $E \cup B \vdash T$, then $\mathcal{LI}(T, E, B; D_i) = 1$, and

3. if $T' \vdash T$, then $\mathcal{LI}(T', E, B; D_i) \leq \mathcal{LI}(T, E, B; D_i)$.

If the proviso in theorem 5.1 does not hold for some constant term 't' $\in C_{B-repr}(E)$, for some E and B, then *no* T can account for 't' in E relative to B. The proviso is satisfied, if (1) $RE(E) \vdash E$; or if (2) E is *minimally observational* in the sense that for every 't' essentially occurring in E (and thus for every 't' $\in C_{B-repr}(E)$) there is at least one contingent statement A containing only one predicate occurrence such that 't' $\in C(A)$ and $E \vdash A$. (Any such statement A is a relevant element of any evidence E logically implying A. There is just one predicate occurrence that can be replaced, whence substituting a logically determined predicate for it would yield E inconsistent.)

The term 'minimally observational' arises from the following consideration: One may define an evidence to be *observational* just in case it consists only of (possibly negated) atomic statements, because – so it may be argued – we do not observe (negative or) disjunctive properties, but only whether some entity has a property (whether some entities stand in some relation); disjunctive (and negative) properties are not observed, but *inferred*.

Any (possibly negated) atomic statement has only one predicate occurrence, and thus is of the required form. But other statements – e.g. Popperian *Basissätze* of the form 'At space-time point k *there are* $x_1, \ldots, x_n$ such that $A[x_1, \ldots, x_n]$'[5] – do not have the form of (possibly negated) atomic statements; nor do they imply

[5]Cf. Popper (1994), p. 66ff.

such statements. However, they usually entail a statement with just one predicate occurrence.

The proviso is superfluous, if $C_{B-repr}(E)$ is restricted to those constant terms 't' for which there is at least one contingent relevant element A of E with 't' $\in C(A)$. The reason for not doing so is that I conjecture that the proviso is satisfied anyway – for lack of mathematical skill I just cannot prove it.

5.3 The Measure of Confirmation

This section contains the definition of the degree of confirmation $C(T, E, B)$ of theory T by evidence E relative to background knowledge B, which is the result of pursuing the second approach to a solution of the problem of a quantitative theory of confirmation. C is defined as the product of the functions $\mathcal{P}$ and $\mathcal{LI}$ measuring the primary confirmational virtues of loveliness and likeliness, respectively. An immediate consequence of this definition is that C is sensitive to loveliness and likeliness in the sense of $\mathcal{P}$ and $\mathcal{LI}$. The formal handiness of C is straightforward, because $\mathcal{P}$ and $\mathcal{LI}$ are both non-arbitrary, comprehensible, and computable in the limit, and the multiplication function $\cdot$ preserves these properties, as it is itself a single and thus non-arbitrary computable function that can be defined in the terms of $PL1 =$ and ZF. Sensitivity to simplicity considerations i.v.s.s., unimpressability by redundancies, and invariance under equivalence transformations of T result from C being sensitive to loveliness and likeliness in the sense of $\mathcal{P}$ and $\mathcal{LI}$.

As in the case of power and likeliness, confirmation of T by E relative to B is only defined for the confirmational domains of T and E. Strictly speaking, the value $C(T, E, B)$ of C for T, E, and B is a vector $\langle C(T, E, B; D_i), \ldots, C(T, E, B; D_c)\rangle$, whose length c equals the number of confirmational domains of T and E, $D_1, \ldots, D_c$.

Definition 5.4 (Degree of Confirmation) Let T be a theory, let E be an evidence, let B be a background knowledge, and let D_i be a confirmational domain of T and E.

The *degree of confirmation of T by E relative to B in D_i*, $C(T, E, B; D_i)$, is given by the following equation:

$$C(T, E, B; D_i) = \mathcal{P}(T, E, B; D_i) \cdot \mathcal{LI}(T, E, B; D_i),$$

provided both $\mathcal{P}(T, E, B; D_i)$ and $\mathcal{LI}(T, E, B; D_i)$ are defined.

The claim that C is sensitive to loveliness and likeliness in the sense of $\mathcal{P}$ and $\mathcal{LI}$ means that $C(T, E, B; D_i)$ satisfies the four conditions in the definition of sensitivity to loveliness and likeliness, for *every* confirmational domain D_i.

Main Theorem 1 (C Is Formally Handy and Materially Adequate) $C(\cdot,\cdot,\cdot)$, $C(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, is non-arbitrary, comprehensible, computable in the limit, and sensitive to loveliness and likeliness in the sense of $\mathcal{P}$ and $\mathcal{LI}$, provided for every $E \in \mathcal{E}$ and every 't' $\in C_{ess}(E)$ there is an $A \in RE(E)$ with 't' $\in C(A)$.

More precisely, C is formally handy, and for any theories T and T', any evidences E and E', any background knowledges B and B', every confirmational domain D of T and E, and every confirmational domain D' of T' and E', where $X = \langle T, E, B; D\rangle$ and $X' = \langle T', E', B'; D'\rangle$:

1. If $\mathcal{LI}(X) = \mathcal{LI}(X') \neq 0$, then $C(X) \geq C(X')$ iff $\mathcal{P}(X) \geq \mathcal{P}(X')$,
2. if $\mathcal{P}(X) = \mathcal{P}(X') \neq 0$, then $C(X) \geq C(X')$ iff $\mathcal{LI}(X) \geq \mathcal{LI}(X')$,
3. $C(X) = 0$ iff $\mathcal{P}(X) = 0$ or $\mathcal{LI}(X) = 0$, and
4. $C(X) = 1$ iff $\mathcal{P}(X) = 1$ and $\mathcal{LI}(X) = 1$,

provided $\mathcal{P}(X)$, $\mathcal{P}(X')$, $\mathcal{LI}(X)$, and $\mathcal{LI}(X')$ are defined, and for every $E \in \mathcal{E}$ and every 't' $\in C_{ess}(E)$ there is an $A \in RE(E)$ with 't' $\in C(A)$.[6]

As a corollary we get that C is sensitive to simplicity considerations i.v.s.s.; it cannot be impressed by redundancies; and it is closed under equivalence transformations of T.

Observation 5.1 (Derived ConfVirtues and InvEquTransf) For any theories T and T', every evidence E, every background knowledge B, every confirmational domain D_i of T and E, and of T' and E, and every wff $h \in T$:

1. If $T' \vdash T$ and $\mathcal{P}(T, E, B; D_i) = \mathcal{P}(T', E, B; D_i)$, then $C(T, E, B; D_i) \geq C(T', E, B; D_i)$,
2. if h is a redundant part of T, then $C(T \setminus \{h\}, E, B; D_i) \geq C(T, E, B; D_i)$, and
3. if $T \dashv\vdash T'$, then $C(T, E, B; D_i) = C(T', E, B; D_i)$,

provided $C(T, E, B; D_i)$ and $C(T', E, B; D_i)$ are defined, and for every $E \in \mathcal{E}$ and every 't' $\in C_{ess}(E)$ there is an $A \in RE(E)$ with 't' $\in C(A)$.

[6]Proof omitted.

5.4 On Accounting

The adequacy of the definition of $\mathcal{P}$, and therefore that of C, depend on the notion of accounting for, which is a generalisation of an idea due to C. G. Hempel[7], and is defined between sets of statements T and B on the one hand and constant terms 't' occurring in some evidence E on the other. Which concept is this definition intended to capture?

Most importantly, it is *not* supposed to grasp the notion of explanation – this being the reason why I have avoided to speak of the *explanatory* power (of T for E relative to B).

Note first that only constant terms 't' occurring in some evidence E can be accounted for by a some set of statements T relative to another such set B, whereas in case of explanation, if it is defined for pairs of sets of statements T_1 and T_2, T_2 need not be an evidence. Also laws of nature, at least empirical generalisations or observational law hypotheses of the form $\forall \boldsymbol{x} \left(A \left[\boldsymbol{x}\right] \rightarrow C \left[\boldsymbol{x}\right]\right)$, '$\boldsymbol{x}$' being a vector of individual variables, and '$A\left[\boldsymbol{x}\right]$' and '$C\left[\boldsymbol{x}\right]$' being conjunctions or disjunctions of observational predicates, can be explained by being *subsumed* under more general laws. This is not the case for the relation of accounting for.

Furthermore, in contrast to the relation of explanation, the relation of accounting for is monotone with respect to T and B: If T accounts for 't' in E relative to B, then so does every T' logically implying T; and if T accounts for 't' in E relative to B, then T does so relative to every B' logically implying B. This holds in particular, if T' or B' is inconsistent (though one may, of course, exclude this in the definition of accounting for). All this need not be the case for explanation.

Enough has been said to show that accounting for and explanation are two different things. Which concept, then, is to be captured? The definition of accounting for is the formal characterisation of the relation that holds between (i) an individual t and the properties we observe t to have (respectively a set of statements describing this), (ii) a theory T, in particular, a set of empirical generalisations or observational law hypotheses, whose intended domain of application t belongs to, and (iii) a set of statements B expressing the available background knowledge, if there are properties t is expected to have with regard to T and B on the basis of the remaining properties already observed on t. In other words, if T together with B could have *predicted* some of t's properties on the basis of its remaining ones. More precisely, if a statement describing part of what has been

[7] Cf. the prediction criterion in Hempel (1945).

observed of t is logically implied by T and B together with statements describing the rest of what has been observed of t.

5.5 An Objection

An objection that may be raised against this approach is that the degree of confirmation is – in contrast to Bayesian theories of confirmation[8] and the various versions of (HD)[9] – not defined between any (sets of) statement(s) T, E, and B, but only for Ts satisfying assumption 1.1 (Finite Axiomatizability Without Constants), evidences E, and finite sets of statements B.

My response is that it is, of course, always better to do with as few and as weak assumptions as possible (cf. the remark after the proof of theorem 5.2), but that $C(T, E, B)$ is intended to measure how much a *scientific theory* T is confirmed by an (observational) *evidence* E relative to a *background knowledge* B – and for these the restrictions on T, E, and B are appropriate, or so I have argued.

There may be some relation of "making plausible with regard to" that is defined between any three (sets of) statements. However, this concept is different from the concept of confirmation which is the topic discussed here: Confirmation of theory T by an evidence E relative to background knowledge B is more than E merely making it plausible with regard to B that T (is true). If E confirms T relative to B, then E also makes it plausible with regard to B that T (is true); but there are many cases where E makes it plausible with regard to B that T is true, which are no cases of confirmation: That Peter is happy may make it plausible – with regard to the information that Peter and Mary love each other – that Mary is happy, too; and that there are demons not liking George may make it plausible

[8]Cf., however, Gillies (1998), p. 150ff., who argues that Bayesian confirmation theory has to be restricted to singular statements (!) – apart from the further condition of the fixity of the theoretical framework. For the latter cf. Gillies (2001).

[9]In case of (HD) the conditions specifying the suitable way E has to follow logically from T and B in order for E to confirm T relative to B may restrict the class of statements for which the qualitative concept of (HD)-confirmation is defined.

The concept of confirmaton of Glymour's Bootstrap-Theory is not defined for any statements T and E, because here evidence E has to be, roughly speaking, a particular instance of theory T.

Of course, if one wants to define the concept of confirmation between any three sets of statements T, E, and B, then the restrictions may be circumvented by setting the degree of confirmation to 0, if T does not satisfy assumption 1.1, E is no evidence, or B is not finite. Obviously, this amounts to cheating.

(with regard to some suitable information about demons) that George is hindered in daily life; but in neither case would we say that some evidence confirms a scientific theory relative to some background knowledge.

5.6 Properties of C

C is sensitive to loveliness and likeliness in the sense of $\mathcal{P}$ and $\mathcal{LI}$, and thus to the derived confirmational virtues. What is this good for?

If you have got a body of evidence E, a theory T, and some background knowledge B, and you are to assess T relative to E and B in the sense of confirmation combining the likeliness and the loveliness concept, then – other things being equal[10] – if you make T logically stronger, for instance, by adding new hypotheses, T will not become likelier relative to E and B, and may become less likely, but T may get more power for E relative to B, and its power will not decrease.

On the other hand, if you make T logically weaker, say, by deleting some of the hypotheses in T, then T will – other things being equal[11] – not get more power for E relative to B, and may become less powerful, but T may become likelier relative to E and B, and its likeliness will not decrease. This is the contribution of the third conditions in the definitions of searching power and indicating truth, respectively, to the definition of being sensitive to loveliness and likeliness.

However, whether or not the power and the likeliness of T relative to E and B are increased by making T logically stronger respectively weaker, depends on the *way* this is done. It depends on the added or deleted hypotheses whether a change in logical content results in an increase or a decrease of one of the primary confirmational virtues.

What is the point of changing the logical content of T (by adding and/or deleting hypotheses)? By adding new hypotheses we hope to make T more powerful for E relative to B, but thereby not to make it less likely; by deleting hypotheses we aim at an increase in the likeliness of T relative to E and B, which should not also result in a decrease of its power.

Furthermore, the power of T for E relative to B should be increased, only if the addition of new hypotheses makes T not only logically stronger, but enables it to account for data that have not been accounted for so far by; thereby the cost in likeliness should be as small as possible. In the same way the likeliness of

[10] I.e. if evidence E and background knowledge B are held constant.
[11] The other things being E, B, and the remaining hypotheses in T.

T relative to E and B should be increased, only if the deletion of hypotheses results in T becoming more likely relative to E and B, which is the case, if those hypotheses are dropped which are not related to or unlikely relative to E and B; also here, the cost in power should be as small as possible.

What the measure of confirmation C is expected to do with regard to this is that it *balances* between the two primary confirmational virtues: The job of C is to take into account both of these aspects, and to weigh between them in such a way that there results an *equilibrium* between loveliness and likeliness.

This balance is optimal, if T can account for all the data in E relative B, and is also maximally likely relative to E and B. In particular, if the theory T to be assessed coincides with the evidence E in the sense that T is just a reformulation of E as a set of statements satisfying assumption 1.1, then the balance between these two aspects is optimal, whence $C(T, E, B)$ should be maximal in this case. This is exactly what the following theorem states.

Observation 5.2 (Maximal Confirmation for $T \dashv\vdash E$) Let T be a theory, and let E be an evidence. Then it holds for every background knowledge B, and every confirmational domain D_i of T and E:

$$\text{If } T \dashv\vdash E \text{ or } E \cup B \dashv\vdash T \text{ or } T \cup B \dashv\vdash E, \text{ then } C(T, E, B; D_i) = 1,$$

provided $C(T, E, B; D_i)$ is defined.

Now this result may seem to be quite odd, for after all, what it tells us is that no theory T can be better confirmed than some odd reformulation of the evidence. Moreover, in case of the rule of acceptance for rational theory choice $(\mathcal{R})$, the above theorem tells us for every typical problem situation with a finite set of alternative theories $\{T_1, \ldots, T_n\}$, and E belonging to the domain of application of each T_i that we should accept some reformulation T_E of E.

So, does our measure of confirmation C force us to draw the conclusion that we can do without all the theories proposed in the history of the sciences by various ingenious scientists, and that we better stick to a theory-like formulation of all data gathered so far? It would, if there were no reasons for restricting the class of sets of statements T that may be considered as serious candidates for scientific theories to those which are finitely axiomatizable without constant i-terms, where D_i is among the domains of proper investigation of T.

Given this restriction, the consequences of the preceding observation need not be drawn.

Observation 5.3 (There Is No T with $T \dashv\vdash E$) Let E be an evidence. Then it holds for every theory T for which there is at least one confirmational domain of T and E:

$$T \nvdash E \quad \text{or} \quad E \nvdash T.$$

(By assumption, there is at least one confirmational domain D_i of T and E. This D_i is among the evidential domains of E, whence there is a constant i-term *essentially* occurring in E; but D_i is also among the domains of proper investigation of T, whence T contains no occurrence of a constant i-term.)

This does, of course, not mean that no theory T can be maximally confirmed by some evidence E relative to some background knowledge B. Let $E = \{Fa_1, \ldots, Fa_n\}, n \geq 1$, $B = \emptyset$, and $T = \{\forall x Fx\}$. Then

$$C(T, E, B) = \mathcal{P}(T, E, B) \cdot \mathcal{LI}(T, E, B) = 1 \cdot 1 = 1.$$

5.7 A Shortcoming?

The last example does not only show that there are theories T, evidences E, and background knowledges B such that $C(T, E, B) = 1$. It also illustrates that the degree of confirmation is determined by the *proportion* of those constant terms 't' which – in case of $\mathcal{P}$ – are in the account of T in E relative to B to all constant terms 't' $\in C_{B-repr}(E)$; and similarly for $\mathcal{LI}$. The size of the evidence E in the sense of the cardinality of $C_{B-repr}(E)$ and its variety or diversity do not matter for power, likeliness, and confirmation.

Is this a point against the approach presented here? I think it is not. The measure of confirmation C does not – and *is not intended to* – measure the overall support there is for a given theory. $C(T, E, B)$ tells us how much T is confirmed *given* that E is *all* the evidence available and B is the *whole* background knowledge. Whether there is a lot of overall support for T does not only depend on its degree of confirmation by E relative to B; it additionally depends on whether E is *good* evidence.

What follows? Do we have to rely on some principle of *total evidence* telling us that in assessing a given theory T we always have to consider the total available evidence E, and the total available background knowledge B – at least if we expect the measure of confirmation C to implicitely provide a rule of acceptance for rational theory choice?

I do not think so. If C is not to provide a rule of acceptance for rational theory choice, then all $C(T, E, B)$ is expected to tell us is how much T is confirmed by E relative to B – and this it does. If, however, C is to provide such a rule of acceptance for rational theory choice, then there are two possibilities: Either the problem situations this rule is to handle are of the type described in chapter 2, in which case C does its job, for these problem situations are relative to some evidence E (and some background knowledge B). Or else the problem situations this rule is to handle are not relative to some E (and B), but ask which theory to accept independently of the evidence under consideration (and the background knowledge taken for granted).

If, among others, the theory to be chosen should be true in the actual world – and I take this to be one of the features we aim at – then every such problem situation has to be understood as asking which theory to accept with regard to a complete description of the actual world (or at least the total available evidence). For, after all, (sorry, I am repeating myself), truth is a binary relation between a (set of) statement(s) on the one hand and a world or model on the other, and thus cannot be taken into account without recourse to the world or model whose truth in one is interested in. However, establishing this link is just the purpose of the evidence, and the reason why it is assumed to be true in the actual world. So every problem situation of the latter kind is relative to a complete description of the actual world, whence the second type of problem situation is only a special kind of the first one.

There is a peculiarity of demanding to consider a complete description of the actual world or the total available evidence. The idea behind a rule of accpetance for rational theory choice is to be a guide in deciding which theory to accept with regard to a given evidence and a given background knowledge. If the answer to this question demands of us to collect all the data there are, or even to consider a complete description of the actual world, then we will never be in the position to apply this rule, for we will never have collected all the data there are – nor will we ever possess a complete description of the actual world. So if a rule of acceptance for rational theory choice is to be meaningfully combined with a principle of total evidence, then all this principle can demand of us is to consider all the evidence that is *practically* available (at a given point of time). The question is whether there is not a better strategy for dealing with all this.

I think there is: In assessing a given theory T relative to some evidence E and some background knowledge B, one has to consider not only the degree of confirmation of T by E relative to B, but must also take into account the "goodness" of the evidence E. What the latter consists of, and how it can be measured, is the topic of the last chapter.

Chapter 6

Variety and Goodness of the Evidence

6.1 Introductory Remarks

As already mentioned the measure of confirmation C does not tell us anything about the overall degree of confirmation of a theory, which additionally depends on the "goodness" of the evidence. Similarly, the reliability of the rule $(\mathcal{R})$ of acceptance for rational theory choice of chapter 2 depends not only on the degree of confirmation, but also on the goodness of the evidence, which I take to consist in its size and its variety or diversity.

In this chapter a function $G(\cdot, \cdot, \cdot)$ is defined on the set of all evidences $\mathcal{E}$, the set of all theories $\mathcal{T}$, and the set of all background knowledges $\mathcal{B}$, and it is argued that, for a given evidence E, $G(T, E, B)$ measures the goodness of E relative to theory T and background knowledge B in the sense of the formers size and variety (diversity). I will reason that the refined measure of confirmation C^*, which is based on C and G, gives an answer to the question why scientists (should) gather evidence, and that it resolves the ravens paradox. The chapter ends with some comments on the reliability of truth indicators.

Intuitively, an evidence is the better, the more data it reports about, the more different classes of facts it consists of, the greater these classes of facts are, the more detailed or accurate they are described, and the more they differ from each other. The concept of evidential diversity or variety of evidence thus clearly depends on the notion of a class of facts, in particular, on when two classes of facts count as different ones, on when they (are big and) described in detail, and on when two classes of facts differ more from each other than two other ones. A class of facts is construed as a set of individuals mentioned in some evidence E – respectively a set of constant terms occurring in E – because I take individuals to be ontologically fundamental.

Whether two classes of facts count as different ones depends on the hypothesis or theory one is concerned with. Therefore determining the goodness of an evidence E – by determining the number, size, accuracy, and difference of the classes of facts E consists of – involves considering the theory in question. The background knowledge B has to say something, too, whence the measure of the goodness of evidence G is construed as a function with three argument places. For convenience, a fourth argument place is added for the confirmational domains the individuals in the various classes of facts are taken from; but again, strictly speaking the value of G for given T, E, and B is a vector whose length equals the number of confirmational domains of T and E.

Let us consider why the notion of a class of facts has to be relativised to the hypothesis or theory under consideration. Relative to a theory that claims to account for the colour of people's hair a black haired man and a black haired woman belong to the same class of facts, whereas a black haired woman and a red haired woman belong to two different classes of facts. On the other hand, relative to a theory about the sexual behaviour of humans the black haired man and the black haired woman belong to two different classes of facts, whereas the black haired woman and the red haired woman belong to the same class of facts – the reason being that the colour of humans' hair is irrelevant for their sexual behaviour, but relevant for the colour of their hair, whereas the sex of humans is irrelevant for the colour of their hair, but relevant for their sexual behaviour.

Furthermore, enlarging the data may yield that two individuals which belong to the same class of facts relative to a given theory in the old evidence belong to two different classes of facts relative to the same theory in the enlarged evidence, because the new data may be relevant for this theory. For instance, by taking into account the age of humans the black haired and the red haired woman of before, which belong to the same class of facts relative to the theory about the sexual behaviour in the old evidence, will no longer belong to the same class of facts relative to this theory, because their age, which is assumed to be very different, is relevant for and will make a difference in their sexual behaviour.

6.2 (Maximal) Classes of Facts

A class of facts is construed as a set of individuals mentioned in some evidence E. The information we have about these individuals, and on which we can rely in classifying them, is contained in E and the background knowledge B. Since I want to classify single individuals, and not whole n-tupels, I have to consider one-place predicates instead of n-ary ones. Finally, as is familiar by now, individuals enter the scence *via* their names, the constant terms in $C_{B-repr}(E)$.

Let us first consider an n-ary predicate '$P(x_1, \ldots, x_n)$', where all variables 'x_i' are of the same sort. Such a predicate gives rise to $2^{n-1} \cdot n$ one-place predicates

$$Q_1 x_1 \ldots Q_{i-1} x_{i-1} Q_{i+1} x_{i+1} \ldots Q_n x_n P(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n),$$

$1 \leq i \leq n$, where Q_j is an existential quantifier $\exists$ or a univeral quantifier $\forall$, $1 \leq j \neq i \leq n$. By rearranging these quantifiers (changing their order) one gets $2^{n-1} \cdot (n-1)! \cdot n = 2^{n-1} \cdot n!$ one place predicates from an n-ary predicate (some of them denote the same property, because the order of the quantifiers does not always matter).

Binding argument places with quantifiers is not the only way to get one-place predicates from n-ary ones. In combination with a set of constant terms 'c_1', …, 'c_m' of the appropriate sort, '$P(x_1, \ldots, x_n)$' gives rise to $m^{n-1} \cdot n$ predicates

$$P\left(c_{k_1}, \ldots, c_{k_{i-1}}, x, c_{k_{i+1}}, \ldots, c_{k_n}\right),$$

$1 \leq i \leq n$, $1 \leq k_j \leq m$, for every $j, 1 \leq j \neq i \leq n$. Together these two methods yield

$$\sum_{0 \leq r \leq n-1} \binom{r}{n-1} \cdot m^r \cdot 2^{n-1-r}$$

one-place predicates

$$Q_1 x_1 \ldots Q_{i-1} x_{i-1} Q_{i+1} x_{i+1} \ldots Q_n x_n P(t_1, \ldots, t_{i-1}, x_i, t_{i+1}, \ldots, t_n),$$

$1 \leq i \leq n$, out of one n-ary predicate, where Q_j is either $\exists$ or $\forall$, and 't_j' = 'x_j' or 't_j' $\in \{$'c_1'$, \ldots,$ 'c_m'$\}$, $1 \leq j \neq i \leq n$, whence some (namely r) of the quantifiers occur vacuously. By rearranging the $n - 1 - r$ quantifiers occurring non-vacuously one thus gets

$$\sum_{0 \leq r \leq n-1} \binom{r}{n-1} \cdot m^r \cdot 2^{n-1-r} \cdot (n-1-r)!$$

one-place predicates out of one single n-ary predicate.

A set PR of p n_i-ary predicates then gives rise to

$$\sum_{1\leq i\leq p}\sum_{0\leq r\leq n_i-1}\binom{r}{n_i-1}\cdot m^r\cdot 2^{n_i-1-r}\cdot(n_i-r-1)!$$

one-place predicates, where n_i is the arity of the i-th predicate in PR, $1\leq i\leq p$.

Things get more complicated, when one considers different sorts. I will not show how to get one-place i-predicates 'Px^i' out of n-ary $k_1,\ldots,k_n$-predicates '$P\left(x^{k_1},\ldots,x^{k_n}\right)$' and various sets of constant i-terms. I hope the above is sufficient to show that this *can* be done, and that the result is a *finite* set of one-place i-predicates.

I have argued that whether two individuals belong to the same class of facts depends on the theory under consideration. This appears in the definition of a class of facts by taking as the set of predicates PR the set of predicates $PR_{ess}(T)$ essentially occurring in theory T which is to be assessed by E relative to B.[1]

The restriction to the predicates *essentially* occurring in T is necessary, because otherwise the set of predicates PR can be chosen arbitrarily (by adding hypotheses which are logically valid and contain occurrences of the predicates one wants to have added).

The *ratio* behind taking $PR_{ess}(T)$ is that if two individuals t and t' (should) belong to two different classes of facts as far as some theory T is concerned, then this must be due to some property of t that t' does not have. If every property that can be expressed in terms of the predicates essentially occurring in T is either possessed by both t and t' or by none of them, then all properties distinguishing between t and t' are irrelevant for T, whence t and t' cannot belong to two different classes of facts as far as T is concerned. So the predicates essentially occurring in T settle the relevant conceptual space for the classification of the individuals the evidence is talking about.

In the example of before, the predicates 'male' and 'female' are among the predicates essentially occurring in the theory about the sexual behaviour of humans, whereas the predicates 'black haired' and 'red haired' do not belong to the essential vocabulary of this theory. Therefore the black haired man and the black haired woman can be distinguished by means of the conceptual framework of this theory, but not by the conceptual framework of the hair colour theory.

All we can rely on in determining the size and the variety – the *goodness* – of evidence E, is contained in E or the background knowledge B. In particular, the

[1] So PR is empty, if T is logically determined.

information that there is a property P theory T is talking about which is possessed by individual t, but not by individual t', must be obtained from E and B. In the definition of a class of facts, this finds its expression in considering whether E and B logically imply Pt, where 'Px' $\in PR_1$, 't' $\in C_{B-repr}(E)$, and PR_1 is the set of one-place predicates the set of predicates $PR_{ess}(T)$ gives rise to in combination with the constant terms in $C_{B-repr}(E)$.

Based on these considerations we can now define the notion of a (maximal) class of facts.

Definition 6.1 ((Maximal) Class of Facts) Let T be a theory, let E be an evidence, let B be a background knowledge, and let D_i be a confirmational domain of T and E (so $C_{B-repr}(E) \cap C_i$ is not empty). Let

$$PR = PR_{ess}(T) = \bigcap_{T' \dashv\vdash T} PR(T'),$$

and let 'P_1' = '$P_1^{n_1}\left(x^{k_1}, \ldots, x^{k_{n_1}}\right)$', ..., '$P_p$' = '$P_p^{n_p}\left(x^{k_1}, \ldots, x^{k_{n_p}}\right)$' be an enumeration of the predicates in PR, where $p = |PR|$. Let PR_1^i be the set of all one-place i-predicates which result from any of the following one-place i-predicates by rearranging the quantifiers:

$$Q_1 x^{k_1} \ldots Q_{l-1} x^{k_{l-1}} Q_{l+1} x^{k_{l+1}} \ldots Q_{n_q} x^{k_{n_q}} P_q\left(t^{k_1}, \ldots, t^{k_{l-1}}, x^{k_l}, t^{k_{l+1}}, \ldots, t^{k_{n_q}}\right),$$

$1 \leq l \leq n_q$, where 't^{k_j}' = 'x^{k_j}' or 't^{k_j}' $\in C_{B-repr}(E) \cap C_{k_j}$ (C_{k_j} is the set of constant k_j-terms), $1 \leq j \neq l \leq n$, 'P_q' $\in PR$, $1 \leq q \leq p$, and '$x^{k_l} = x^i$' (otherwise one does not get i-predicates).

Let PR_1^i be partitioned into $N := 2^{\left|PR_1^i\right|}$ sets $C_1^i, \ldots, C_N^i$ of negated or unnegated one-place i-predicates such that it holds for every such C_j^i, $1 \leq j \leq N$, and every one-place i-predicate 'P' $\in PR_1^i$:

$$\text{`}P\text{'} \in C_j^i \quad \text{iff} \quad \text{`}\neg P\text{'} \notin C_j^i.$$

For each of these N sets C_j^i, let $C_{j_k}^i$ be the k-th subset of C_j^i in some enumeration $C_{j_1}^i, \ldots, C_{j_N}^i$ of its N subsets. Let 't' $\in C_{B-repr}(E) \cap C_i$.

't' respectively t *belongs to* $C_{j_k}^i$ iff it holds for every negated or unnegated one-place i-predicate '$\pm P$' $\in C_{j_k}^i$, $E \cup B \vdash \pm Pt$.

't' respectively t belongs *maximally* to $C_{j_k}^i$ iff

1. 't' belongs to $C_{j_k}^i$, and

2. there is no C_l^i, $1 \leq l \leq N$, for which there is at least one $C_{l_p}^i \subseteq C_l^i$, $1 \leq p \leq N$, satisfying (1) and such that $C_{j_k}^i \subset C_{l_p}^i$.[2]

The (*maximal*) *class of i-facts induced by $C_{j_k}^i$ relative to T, E, and B*, $CF_{j_k}^i$, is given as follows:

$$CF_{j_k}^i = \left\{ \text{`}t\text{'} \in C_{B-repr}(E) \cap C_i : \text{`}t\text{' belongs (maximally) to } C_{j_k}^i \right\}.$$

The set of all (maximal) classes of i-facts $CF_{j_k}^i$ induced by $C_{j_k}^i$ relative to T, E, and B, for any j and $k, 1 \leq j, k \leq N$, is the set of (maximal) classes of i-facts T, E, and B *give rise to*.

Let $CF_{j_k}^i$ be the (maximal) class of i-facts induced by $C_{j_k}^i$ relative to T, E, and B, for some set of negated or unnegated one-place i-predicates $C_{j_k}^i$, $1 \leq j, k \leq N$.

$CF_{j_k}^i$ is a *non-empty* (maximal) class of i-facts relative to T, E, and B iff $CF_{j_k}^i \neq \emptyset$. Otherwise $CF_{j_k}^i$ is an *empty* (maximal) class of i-facts relative to T, E, and B.

---

[2] It is important to demand that there be no such proper superset of $C_{j_k}^i$. Demanding that there be no $C_{l_p}^i$ with $C_{l_p}^i \vdash C_{j_k}^i$ and $C_{j_k}^i \nvdash C_{l_p}$ has the consequence that there may be constant i-terms 't' belonging maximally to more than one set of negated or unnegated one-place i-predicates.

For the definition given, maximal classes of i-facts are disjoint. For suppose there is a constant i-term 't' $\in C_{B-repr}(E) \cap C_i$ that belongs maximally to at least two different sets of negated or unnegated one-place i-predicates C_1 and C_2. Then

$$E \cup B \vdash \pm Pt, \quad \text{for every `}\pm P\text{' } \in C_1 \cup C_2,$$

and there is no $C_1' \supset C_1$ or $C_2' \supset C_2$ such that the above holds of 't' and C_1' respectively C_2'. As $C_1 \neq C_2$ there is at least one negated or unnegated one-place i-predicate '$\pm P^*$' with '$\pm P^*$' $\in C_1$ and '$\pm P^*$' $\notin C_2$ (or the other way round). By the above, $E \cup B \vdash \pm P^* t$. But then there is a proper superset C_2^* of C_2 – namely $C_2 \cup \{\pm P^*\}$ – such that

$$E \cup B \vdash \pm Pt, \quad \text{for every `}\pm P\text{' } \in C_2^*$$

– a contradiction.

With $\vdash$ instead of $\subseteq$ this need not be the case, because 'P^*' may be of the form $\exists x \exists y P(x, y, a)$, and both C_1 and C_2 may contain '$\exists y \exists x P(x, y, a)$'.

Here the logical consequence relation $\vdash$ between such sets $C_{j_k}^i$ and $C_{l_p}^i$ of negated or unnegated one-place i-predicates holds iff this relation holds between the sets $\{\pm Pt : \text{`}\pm P\text{'} \in C_{j_k}^i\}$ and $\{\pm Pt : \text{`}\pm P\text{'} \in C_{l_p}^i\}$, where '$t$' is a constant i-term.

6.3 Proper Classes of Facts

Consider the maximal classes of facts

$$\begin{aligned} CF &= \{\text{‘}t\text{’} \in C_{B-repr}(E) : E \cup B \vdash Ft \wedge Gt\} = \{a_1, \ldots, a_m\}, ^3 \\ CF_1 &= \{\text{‘}t\text{’} \in C_{B-repr}(E) : E \cup B \vdash Ft \wedge Gt \wedge Q_1 t\} = \{b_1\}, \\ &\ldots, \\ CF_k &= \{\text{‘}t\text{’} \in C_{B-repr}(E) : E \cup B \vdash Ft \wedge Gt \wedge Q_k t\} = \{b_k\}, \quad k \geq 1, \end{aligned}$$

which are induced by the sets of negated or unnegated predicates

$$C = \{\text{‘}Fx\text{’}, \text{‘}Gx\text{’}\}, C_1 = \{\text{‘}Fx\text{’}, \text{‘}Gx\text{’}, \text{‘}Q_1x\text{’}\}, \ldots, C_k = \{\text{‘}Fx\text{’}, \text{‘}Gx\text{’}, \text{‘}Q_kx\text{’}\},$$

respectively, relative to

$$\begin{aligned} T &= \{\forall x (Fx \to Gx), \forall x (P_1 x \to Q_1 x), \ldots, \forall x (P_k x \to Q_k x)\}, \\ E &= \{Fa_1, Ga_1, \ldots, Fa_m, Ga_m, Fb_1, Gb_1, Q_1 b_1, \ldots, Fb_k, Gb_k, Q_k b_k\}, \quad \text{and} \\ B &= \emptyset, \end{aligned}$$

because it holds for every ‘t’ $\in CF$:

$$E \cup B \vdash Ft, \quad E \cup B \vdash Gt, \quad \text{and} \quad E \cup B \nvdash Q_j t, \quad \text{for every } j, 1 \leq j \leq k;$$

because it holds for every ‘b_i’ $\in CF_i$, $1 \leq i \leq k$:

$$\begin{aligned} &E \cup B \vdash Fb_i, \quad E \cup B \vdash Gb_i, \quad E \cup B \vdash Q_i b_i, \quad \text{and} \\ &E \cup B \nvdash Q_j b_i, \quad \text{for every } j, 1 \leq j \neq i \leq k; \end{aligned}$$

and because

$$PR_1 = PR = PR_{ess}(T) = \{\text{‘}Fx\text{’}, \text{‘}Gx\text{’}, \text{‘}P_1x\text{’}, \text{‘}Q_1x\text{’}, \ldots, \text{‘}P_kx\text{’}, \text{‘}Q_kx\text{’}\}.$$

T accounts for (every individual respectively constant term of) the maximal class of facts CF in E relative to B; and T accounts for (every individual of) every maximal class of facts CF_i in E relative to B, $1 \leq i \leq k$. However, the information about the individulas b_i which goes beyond that of their having the properties F and G is not necessary in order for T to account for b_i in E relative to B. It suffices to know that b_i has properties F and G. Let

$$E' = \{Fa_1, Ga_1, \ldots, Fa_m, Ga_m, Fb_1, Gb_1, \ldots, Fb_k, Gb_k\}.$$

[3]The quotation marks are dropped.

The only maximal class of facts E' gives rise to in combination with T and B is

$$\begin{aligned} CF' &= \{\text{`}t\text{'} \in C_{B-repr}(E') : E' \cup B \vdash Ft \wedge Gt\} \\ &= \{\text{`}a_1\text{'}, \ldots, \text{`}a_m\text{'}, \text{`}b_1\text{'}, \ldots, \text{`}b_k\text{'}\} = C_{B-repr}(E'), \end{aligned}$$

which is the maximal class of facts induced by the set of negated or unnegated predicates

$$C' = \{\text{`}Fx\text{'}, \text{`}Gx\text{'}\} \subseteq PR_1 = PR = PR_{ess}(T)$$

relative to T, E', and B. As before, T accounts for (every individual of) the maximal class of facts CF' in E' relative to B.

In both examples, T can account for all individuals mentioned in the evidence. The only difference between the evidences E and E' is that the redundant[4] information of individual b_i having property Q_i is missing in E'. Now suppose one claims that T is better confirmed by E than by E' (each time relative to B), because evidence E is varied, whereas evidence E' is not. The question is: Does this strike us as counterintuitive? If it does *not*, then the concept of a maximal class of facts as defined in the preceding section is sufficient.

If, however, it does, then it seems that in determining the variety of evidence E relative to theory T and background knowledge B we have to rely only on that information about the individuals t in the class of facts CF which is necessary in order for T to account for t in E relative to B. This is exactly what the notion of a proper class of facts is intended to capture.

Definition 6.2 (Proper Class of Facts) Let T be a theory, let E be an evidence, let B be a background knowledge, and let D_i be a confirmational domain of T and E. Let $CF_1^i, \ldots, CF_n^i$ be the classes of i-facts T, E, and B give rise to, and let $C_1^i, \ldots, C_n^i$ be the corresponding sets of negated or unnegated one-place i-predicates which induce $CF_1^i, \ldots, CF_n^i$, respectively, relative to T, E, and B.

T accounts for CF_j^i in E relative to B, $1 \leq j \leq n$, iff there is a non-redundant $C \subseteq C_j^i$ and a (contingent) negated or unnegated one-place i-predicate '$\pm P^*$' $\in C$ such that

$$T \cup B \cup \{\pm Pt : \text{`}\pm P\text{'} \in C \setminus \{\text{`}P^*\text{'}\}\} \vdash \pm P^* t,$$

where such a set C of negated or unnegated one-place i-predicates is non-redundant iff the set $\{\pm Pt : \text{`}\pm P\text{'} \in C\}$ is non-redundant, and 't' is a constant i-term.

Let CF_j^i be a class of i-facts relative to T, E, and B. CF_i is a *proper* class of i-facts relative to T, E, and B iff

[4] This information is redundant for the question whether T accounts for b_i in E relative to B, but, of course, not in general.

1. T accounts for CF_j^i in E relative to B, and

2. there is no $C_k^i \subset C_j^i$, $1 \leq k \leq n$, such that T accounts for CF_k^i in E relative to B.

In speaking of a (proper or maximal) class of i-facts I will always mean a *non-empty* (proper or maximal) class of i-facts.

6.4 The Measure of the Goodness of the Evidence

I take the goodness measure to be defined in terms of proper classes of i-facts. Nevertheless, if one considers the preceding example as one of two evidences E and E' with the same diversity relative to the theory T and the background knowledge B of the example, the definition may be based on maximal classes of i-facts.

Definition 6.3 (Goodness of Evidence) Let T be a theory, let E be an evidence, let B be a background knowledge, and let D_i be a confirmational domain of T and E. Let $CF_1^i, \ldots, CF_n^i$ be the non-empty proper (or maximal) classes of i-facts T, E, and B give rise to, and let $C_1^i, \ldots, C_n^i$ be the corresponding sets of negated or unnegated one-place i-predicates which induce the non-empty proper (or maximal) classes of facts $CF_1^i, \ldots, CF_n^i$, respectively, relative to T, E, and B.

The *goodness of* E *relative to* T *and* B *in* D_i, $G(T, E, B; D_i)$, is given by the following equation:

$$G(T, E, B; D_i) = 1 - \frac{1}{\log(g(T, E, B; D_i) + 1) + 1},$$

where

$$\begin{aligned} g(T, E, B; D_i) \quad = & \sum_{1 \leq j \neq k \leq n} \left| C_j^i \triangle C_k^i \right| \cdot \\ & \cdot \left[1 - \frac{1}{\log\left(\left|CF_j^i\right| + 1\right) + \log\left(\left|CF_k^i\right| + 1\right) + 1} \right]. \end{aligned}$$

G is not called a measure of evidential diversity, because it additionally takes into account the *size* of the evidence.

Both G and g increase with

1. the number n of non-empty proper (or maximal) classes of i-facts CF_j^i, $1 \leq j \leq n$;

2. the size of the proper (or maximal) classes of i-facts CF_j^i, i.e. the number $\left|CF_j^i\right|$ of constant i-terms in CF_j^i;

3. the detailedness or accuracy of the descriptions of the non-empty proper (or maximal) classes of i-facts CF_j^i, which I take to be proportional to $\left|C_j^i\right|$; and

4. the degree to which the non-empty proper (or maximal) classes of i-facts CF_j^i and CF_k^i differ from each other – the latter being proportional to $\left|C_j^i \triangle C_k^i\right|$, $1 \leq j \neq k \leq n$.

The formal conditions of adequacy for G are familiar by now.

Theorem 6.1 (G Is Formally Handy) $G(\cdot,\cdot,\cdot)$, $G(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, is non-arbitrary, comprehensible, computable in the limit, and closed under equivalence transformations of T.

6.5 The Refined Measure of Confirmation

The measure of confirmation C is not sensitive to the size or the diversity of evidence E (relative to some T and B), both of which are taken into account by the goodness measure G. A measure of confirmation which is additionally sensitive to diversity considerations in this sense, is the refined measure of confirmation C^*.

Definition 6.4 (Refined Degree of Confirmation) Let T be a theory, let E be an evidence, let B be a background knowledge, and let D_i be a confirmational domain of T and E.

The *refined degree of confirmation of T by E relative to B in D_i*, $C^*(T, E, B; D_i)$, is given by the following equation:

$$C^*(T, E, B; D_i) = C(T, E, B; D_i) \cdot G(T, E, B; D_i),$$

provided $C(T, E, B; D_i)$ and $G(T, E, B; D_i)$ are defined.

It is straightforward that C^* is formally handy and closed under equivalence transformations of T.

Observation 6.1 (Formal Handiness and InvEquTransf of C^*) $C^*(\cdot,\cdot,\cdot)$, $C^*(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \to \Re$, is non-arbitrary, comprehensible, computable in the limit, and closed under equivalence transformations of T.

Definition 6.5 (Sensitivity to Diversity Considerations) Let $\mathcal{LO}$ be a power searcher, let $\mathcal{LI}$ be a truth indicator, and let $\mathcal{C}_{\mathcal{LO},\mathcal{LI}}$ be sensitive to loveliness and likeliness.

A function $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \to \Re$, is *sensitive to diversity considerations in the sense of* $\mathcal{C}_{\mathcal{LO},\mathcal{LI}}$ *and* G iff it holds for any theories T and T', any evidences E and E', and any background knowledges B and B', where $X = \langle T, E, B\rangle$ and $X' = \langle T', E', B'\rangle$:

1. If $\mathcal{C}_{\mathcal{LO},\mathcal{LI}}(X) = \mathcal{C}_{\mathcal{LO},\mathcal{LI}}(X') \neq 0$, then $f(X) \geq f(X')$ iff $G(X) \geq G(X')$,
2. if $G(X) = G(X') \neq 0$, then $f(X) \geq f(X')$ iff $\mathcal{C}_{\mathcal{LO},\mathcal{LI}}(X) \geq \mathcal{C}_{\mathcal{LO},\mathcal{LI}}(X')$,
3. $f(X) = 0$ iff $\mathcal{C}_{\mathcal{LO},\mathcal{LI}}(X) = 0$ or $G(X) = 0$, and
4. $f(X) = 1$ iff $\mathcal{C}_{\mathcal{LO},\mathcal{LI}}(X) = 1$ and $G(X) = 1$.

C^* is sensitive to diversity considerations in the sense of $C = C_{\mathcal{P},\mathcal{LI}}$ and G, which means that the above holds for *every* confirmational domain D of T and E, and *every* confirmational domain D' of T' and E'.

Observation 6.2 (SensDivCons of C^*) For any theories T and T', any evidences E and E', any background knowledges B and B', every confirmational domain D of T and E, and every confirmational domain D' of T' and E', where $X = \langle T, E, B; D\rangle$ and $X' = \langle T', E', B'; D'\rangle$:

1. If $C(X) = C(X') \neq 0$, then $C^*(X) \geq C^*(X')$ iff $G(X) \geq G(X')$,
2. if $G(X) = G(X') \neq 0$, then $C^*(X) \geq C^*(X')$ iff $C(X) \geq C(X')$,
3. $C^*(X) = 0$ iff $C(X) = 0$ or $G(X) = 0$, and
4. $C^*(X) = 1$ iff $C(X) = 1$ and $G(X) = 1$,

provided $C(X)$ and $C(X')$ are defined.

6.6 Why Scientists Gather Evidence

This section deals with the question – posed by Maher (1990) – why scientists (should) gather evidence. Interestingly, it is of importance that G is based on *proper* classes of i-facts.

One answer to this question is that scientists (should) gather evidence, because the "bigger" the evidence, the "better" it is, and, as a consequence, the more reliable the inferences based on it. In particular, the determination of the degree of confirmation of some theory T by E relative to background knowledge B says the more about the overall degree of confirmation of T, the better the evidence E.

This goodness of the evidence E (in relation to theory T and background knowledge B) is exactly what G is intended to measure. Furthermore, the refined measure of confirmation C^* differs from C just in the respect that C^* additionally takes into account the goodness of the evidence.

So if it can be shown that, other things being equal[5], the refined degree of confirmation $C^*(T, E, B)$ of T by E relative to B is the greater, the bigger the evidence E, then C^* can explain why scientists (should) gather evidence.

Apart from this it may be argued that it is a material condition of adequacy anyway that a measure of the goodness of the evidence $\mathcal{G}(\cdot, \cdot, \cdot)$ increases with E getting bigger, i.e. that $\mathcal{G}(T, E, B)$ is the greater, the bigger E, for all theories T, evidences E, and background knowledges B.

What does it mean for an evidence E_1 to be bigger than some evidence E_2? There are at least the following two answers:

1. Evidence E_1 is bigger than or equally big as evidence E_2 iff $E_1 \vdash E_2$.

2. Evidence E_1 is bigger than or equally big as evidence E_2 iff $E_2 \subseteq E_1$.

I will stick to (1), because it is the more general claim (so the claims below hold also for 2). By the definition of C^*, it suffices to show that G satisfies the mentioned material condition of adequacy in order to show that $C^*(T, E, B)$ is the greater, the bigger the evidence E, provided the degree of confirmation $C(T, E, B)$ is held constant.

The goodness measure G satisfies this material condition of adequacy, if it is based on proper classes of i-facts, but not, if it is based on maximal classes of i-facts.

[5]If $C(T, E, B)$ is held constant.

Definition 6.6 (To Support Gathering Evidence) A function $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T}\times\mathcal{E}\times\mathcal{B}\rightarrow\Re$, *supports gathering evidence* iff it holds for every theory T, any evidences E and E', and every background knowledge B:

$$\text{If}\quad E'\vdash E,\quad\text{then}\quad f(T,E',B)\geq f(T,E,B).$$

Theorem 6.2 (G Supports Gathering Evidence) $G(\cdot,\cdot,\cdot)$, $G(\cdot,\cdot,\cdot) : \mathcal{T}\times\mathcal{E}\times\mathcal{B}\rightarrow\Re$, supports gathering evidence, if its definition is based on proper classes of i-facts, and if $C_{B-repr}(E)\subseteq C_{B-repr}(E')$.

More precisely, for every theory T, any evidences E and E', every background knowledge B, and every confirmational domain D_i of T and E:

$$\begin{array}{l}\text{If}\quad E'\vdash E\quad\text{and}\quad C_{B-repr}(E)\subseteq C_{B-repr}(E'),\\ \text{then}\quad G(T,E',B;D_i)\geq G(T,E,B;D_i).\end{array}$$

This additional condition is superfluous, if $C_{B-repr}(E)$ is restricted to those constant terms for which the evidence E and the background knowledge B entail that they are different, i.e. if $E\cup B\vdash t_1\neq t_2$, for '$t_1^i$', '$t_2^i$' $\in C_{B-repr}(E)$. The latter seems to be in accordance with our intuitive understanding and implicitely assumed of the B-representative of $C(E)$ as the set of those constant terms which represent the individuals the evidence is talking about.

For instance, if $E=\{Fa_1,\ldots,Fa_n\}$, $n\gg 1$, is enriched to $E'=\{Fa_1,\ldots,Fa_n,a_1=a_2,\ldots,a_{n-1}$ then what we learn in going from E to E' is that, after all, we know of just one single entitity that it is an F. The proviso can also be dropped if it is stipulated that in describing the data she is observing, the scientist uses a new name, only if she is investigating a new entity.

Definition 6.7 (Liking Supporters of Gathering Evidence) Let $\mathcal{LO}$ be a power searcher, let $\mathcal{LI}$ be a truth indicator, and let $\mathcal{C}_{\mathcal{LO},\mathcal{LI}}$ be sensitive to loveliness and likeliness in the sense of $\mathcal{LO}$ and $\mathcal{LI}$.

A function $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T}\times\mathcal{E}\times\mathcal{B}\rightarrow\Re$, *likes supporters of gathering evidence in the sense of* $\mathcal{C}_{\mathcal{LO},\mathcal{LI}}$ iff it holds for every theory T, any evidences E and E', and every background knowledge B:

> If $E'\vdash E$ and $\mathcal{C}_{\mathcal{LO},\mathcal{LI}}(T,E,B)=\mathcal{C}_{\mathcal{LO},\mathcal{LI}}(T,E',B)$, then $f(T,E',B)\geq f(T,E,B)$.

C^* likes supporters of gathering evidence in the sense of $C=C_{\mathcal{P},\mathcal{LI}}$, which means that the above holds for *every* confirmational domain D_i of T and E, and of T and E'.

Observation 6.3 (C^* Likes Supporters of Gathering Evidence) For every theory T, any evidences E and E', every background knowledge B, and every confirmational domain D_i of T and E, and of T' and E:

> If $E' \vdash E$, $C(T, E, B, D_i) = C(T, E', B; D_i)$, and $C_{B-repr}(E) \subseteq C_{B-repr}(E')$, then $C^*(T, E', B; D_i) \geq C^*(T, E, B; D_i)$,

provided $C(T, E, B; D_i)$ and $C(T, E', B; D_i)$ are defined, and the definition of G is based on proper classes of i-facts.

6.7 The Ravens Paradox

The famous ravens paradox, to which attention has been drawn by Hempel (1945), can be stated as follows: The *Nicod Criterion* (NC) says that conjunctive instances $A[\boldsymbol{a}/\boldsymbol{x}] \wedge C[\boldsymbol{a}/\boldsymbol{x}]$ of universal hypotheses of the form $\forall \boldsymbol{x}(A[\boldsymbol{x}] \rightarrow C[\boldsymbol{x}])$ confirm the latter ('$\boldsymbol{x}$' is a vector of individual variables, and '$\boldsymbol{a}$' is a vector of constant terms of the same length). For instance, according to the Nicod Criterion the statement 'a is a black raven', $Ra \wedge Ba$, confirms the ravens hypothesis 'All ravens are black', $\forall x(Rx \rightarrow Bx)$.

With the following intuitively plausible condition of adequacy, the *Equivalence Condition* (EC),

> (EC) If a (set of) wff(s) E confirms a (set of) wff(s) T, and if a (set of) wff(s) T' is logically equivalent to T, then E confirms T'.

one arrives at the allegedly paradoxical result that the statement 'a is neither black nor a raven', $\neg Ba \wedge \neg Ra$, confirms the ravens hypothesis.[6]

Furthermore, if the Nicod criterion (NC) is replaced by the *Instance Confirmation Condition* (ICC),

> (ICC) A universal hypothesis $\forall \boldsymbol{x} A[\boldsymbol{x}]$ is confirmed by any of its substitution instances $A[\boldsymbol{a}]$.

and if one adopts the *Reversed Consequence Condition* (RCC),

[6](NC) yields that 'a is neither black nor a raven', $\neg Ba \wedge \neg Ra$, confirms 'All non-black things are non-ravens', $\forall x(\neg Bx \rightarrow \neg Rx)$, which is logically equivalent to 'All ravens are black', $\forall x(Rx \rightarrow Bx)$. By (EC), '$a$ is neither black nor a raven', $\neg Ba \wedge \neg Ra$, confirms 'All ravens are black', $\forall x(Rx \rightarrow Bx)$.

> (RCC) If a (set of) wff(s) E confirms a (set of) wff(s) T, and if a (set of) wff(s) E' logically implies E and does not disconfirm T, then E' confirms T.

then it follows that the statement 'a is no raven', $\neg Ra$, and the statement 'a is black', Ba, each confirm the ravens hypothesis.[7]

Can the account presented here resolve this "paradox"? First, let me note that we do not observe that something, say a, is a non-black non-raven. What we observe is that a is, for instance, a white swan. If the evidence contains the statement 'a is a white swan' so that $E = \{Sa, Wa\}$, it follows that the likeliness of the ravens hypothesis for E relative to an appropriate background knowledge B (containing the information that no swan is a raven, and that nothing white is black) is maximal, because

$$Sa, Wa \vdash Dev_{\{`a'\}}(B) \to Dev_{\{`a'\}}(T),$$

i.e.

$$Sa, Wa \vdash (Sa \to \neg Ra) \wedge (Wa \to \neg Ba) \to (Ra \to Ba).$$

But the power of $T = \{\forall x\,(Rx \to Bx)\}$ for E relative to $B = \{\forall x\,(Sx \to \neg Rx),$ $\forall x\,(Wx \to \neg Bx)\}$ is minimal, since T does not account for a in E relative to B, because

$$\forall x\,(Rx \to Bx), \forall x\,(Sx \to \neg Rx), \forall x\,(Wx \to \neg Bx), Sa \nvdash Wa,$$

and

$$\forall x\,(Rx \to Bx), \forall x\,(Sx \to \neg Rx), \forall x\,(Wx \to \neg Bx), Wa \nvdash Sa.$$

However, T accounts for a in E relative to B, if $E = \{\neg Ra, Wa\}$ or $E = \{\neg Ra, \neg Ba\}$, for

$$\forall x\,(Rx \to Bx), \forall x\,(Wx \to \neg Bx), Wa \vdash \neg Ra,$$

and

$$\forall x\,(Rx \to Bx), \neg Ba \vdash \neg Ra,$$

[7] By (ICC), 'All ravens are black', $\forall x\,(Rx \to Bx)$, is confirmed by 'If a is a raven, then a is black', $Ra \to Ba$. The latter is logically implied by 'a is no raven', $\neg Ra$, and by 'a is black', Ba, both of which are assumed to be not disconfirming 'All ravens are black', $\forall x\,(Rx \to Bx)$.

whence here both loveliness and likeliness are maximal. Nevertheless, even this can be handled by the refined measure of confirmation C^*.

Before turning to this let me state what I think is “paradoxical” about the ravens paradox: It is *not* that a statement reporting that something is neither black nor a raven does not confirm the ravens hypothesis, but that the degree to which the ravens hypothesis is confirmed by such a statement is *smaller* than the degree to which it is confirmed by a statement reporting that something is a black raven. In other words, if the evidence E says, among others, that a is a black raven and that b is neither black nor a raven, $\{Ra, Ba, \neg Rb, \neg Bb\} \subseteq E$, then both $E_1 = \{Ra, Wa\}$ and $E_2 = \{\neg Rb, \neg B\}$ confirm the ravens hypothesis, but E_1 is confirming it more than E_2.

The reason for this is that there are by far more non-black non-ravens than there are black ravens, whence the addition of E_1 to the available evidence E yields a greater boost in the degree of confirmation of the ravens hypothesis than does the addition of E_2, provided this information – that there are by far more non-black non-ravens than there are black ravens – is part of the available evidence E; for only with the latter is it the case that the ravens hypothesis is more confirmed by E_1 than by E_2. This is exactly what the refined measure of confirmation C^* yields.

First, the theory in question is $T = \{\forall x\,(Rx \rightarrow Bx)\}$; second, the background knowledge B^* contains, among others, the information that nothing white is black, that no swan is a raven, that nothing green is black, that no avocado is a raven, and so on; i.e.

$$\begin{aligned} B \quad = \quad & \{\forall x\,(Wx \rightarrow \neg Bx)\,, \forall x\,(Sx \rightarrow \neg Rx)\,, \\ & \forall x\,(Gx \rightarrow \neg Bx)\,, \forall x\,(Ax \rightarrow \neg Rx)\} \quad \subseteq \quad B^*; \end{aligned}$$

third, the available evidence E^* contains data to the effect that there are by far more non-black non-ravens – as white swans, green avocados, and so on – than there are black ravens:

$$\begin{aligned} E \quad = \quad & \{Ra_1, Ba_1, \ldots, Ra_p, Ba_p, Sb_1, Wb_1, \ldots, Sb_q, Wb_q, \\ & Ac_1, Gc_1, \ldots, Ac_r, Gc_r\} \quad \subseteq \quad E^*, \quad q + r \gg p.^8 \end{aligned}$$

Now suppose the degree of C-confirmation the ravens-hypothesis receives from ‘b is a white swan’, $Sb \wedge Wb$, or from ‘c is neither black nor a raven’, $\neg Rc \wedge \neg Bc$,

[8] Instead of considering white swans and green avocados one can also consider non-black non-ravens $\neg Rb_1, \neg Bb_1, \ldots, \neg Rb_q, \neg Rb_q$, where $q \gg p$. I have chosen this way of dealing with the ravens paradox, because it is more realistic (we usually do not observe non-black non-ravens, but infer that the green avocados we had for dinner are – or were – non-black non-ravens).

is the same as the one it receives from 'a is a black raven', $Ra \wedge Ba$ – each time relative to the same appropriate background knowledge. In other words,

$$C\left(T, E \cup \{Sb, Wb\}, B\right) = C\left(T, E \cup \{Ra, Ba\}, B\right).$$ [9]

Still,

$$C^*\left(T, E \cup \{Ra, Ba\}, B\right) > C^*\left(T, E \cup \{Sb, Wb\}, B\right),$$

for $E \cup \{Ra, Ba\}$ is better evidence relative to T and B than $E \cup \{Sb, Wb\}$, i.e.

$$G\left(T, E \cup \{Ra, Ba\}, B\right) > G\left(T, E \cup \{Sb, Wba\}, B\right).$$

This holds independently of basing G on proper classes of facts or on maximal classes of facts.

The set of predicates $PR = PR_{ess}(T)$ is $\{$'Rx', 'Bx'$\}$, which is the same as the set of one-place predicates PR_1 generated by PR and $C_{B-repr}(E)$. The (non-empty and consistent) sets of negated or unnegated one-place predicates are

$$\begin{array}{l} C_1 = \{\text{`}Rx\text{'}\}, \quad C_2 = \{\text{`}Bx\text{'}\}, \quad C_3 = \{\text{`}\neg Rx\text{'}\}, \quad C_4 = \{\text{`}\neg Bx\text{'}\}, \\ C_5 = \{\text{`}Rx\text{'}, \text{`}Bx\text{'}\}, \quad C_6 = \{\text{`}Rx\text{'}, \text{`}\neg Bx\text{'}\}, \\ C_7 = \{\text{`}\neg Rx\text{'}, \text{`}Bx\text{'}\}, \quad \text{and} \quad C_8 = \{\text{`}\neg Rx\text{'}, \text{`}\neg Bx\text{'}\}, \end{array}$$

which induce the following two non-empty maximal classes of facts relative to T, E, and B:

$$\begin{array}{l} CF_5 = \{\text{`}t\text{'} \in C_{B-repr}(E) : E \cup B \vdash Rt \wedge Bt\} = \{a_1, \ldots, a_p\}, \quad \text{and} \\ CF_8 = \{\text{`}t\text{'} \in C_{B-repr}(E) : E \cup B \vdash \neg Rt \wedge \neg Bt\} = \{b_1, \ldots, b_q, c_1, \ldots, c_r\}. \end{array}$$

The maximal classes of facts CF_6 and CF_7 are empty, because it holds for every 't' $\in C_{B-repr}(E)$:

$$E \cup B \nvdash \neg Rt \wedge Bt \quad \text{and} \quad E \cup B \nvdash Rt \wedge \neg Bt.$$

Finally, the proper classes of facts T, E, and B give rise to are the non-empty maximal classes of facts, for CF_5 and CF_8 are both accounted for by T in E relative to B, and there is no set of negated or unnegated predicates $C \subset C_5$ or $C \subset C_8$ such that T accounts for the class of facts CF induced by C relative to T, E, and B.

That evidence E contains data to the effect that there are by far more non-black non-ravens than there are black ravens means that it mentions by far more

[9] The case of $E \cup \{\neg Rc, \neg Bc\}$ is dealt with in the same way.

white swans or green avocados and so on (or just non-black non-ravens) than black ravens. That is, the class of facts CF_8 consists of much more individuals than the class of facts CF_5. This finds its expression in the numbers p, q, and r, which are such that

$$|CF_8| = q + r \gg p = |CF_5| .$$

So we get the following:[10]

$$\begin{aligned} g(T, E, B) &= |C_1 \triangle C_4| \cdot \left[1 - \frac{1}{\log(|CF_1| + 1) + \log(|CF_4| + 1) + 1}\right] \\ &= 4 \cdot \left[1 - \frac{1}{\log(p + 1) + \log(q + r + 1) + 1}\right] . \end{aligned}$$

Although the ravens hypothesis is assumed to be equally C-confirmed by $E \cup \{Ra, Ba\}$ and by $E \cup \{Sb, Wb\}$ (relative to the appropriate background knowledge), it still holds that $E \cup \{Ra, Ba\}$ provides more C^*-confirmation than $E \cup \{Sb, Wb\}$. For

$$\begin{aligned} C^*(T, E \cup \{Ra, Ba\}, B) &> C^*(T, E \cup \{Sb, Wb\}, B) \\ &\text{iff} \\ g(T, E \cup \{Ra, Ba\}, B) &> g(T, E \cup \{Sb, Wb\}, B) , \end{aligned}$$

which holds just in case

$$4 \cdot \left[1 - \frac{1}{\log(1 + p + 1) + \log(q + r + 1) + 1}\right] >$$

$$> 4 \cdot \left[1 - \frac{1}{\log(p + 1) + \log(1 + q + r + 1) + 1}\right] .$$

The latter is the case if and only if

$$q + r > p,$$

i.e. if and only if there are more non-black non-ravens than black ravens. (Note, this means also that the ravens hypothesis is *less* confirmed by a black raven than by a non-black non-raven, if there are *fewer* non-black non-ravens than black ravens – which is as it should be.)

[10]The confirmational domain D of T and E is suppressed, where D is the domain (variable) corresponding to the variable 'x' and the individual constants 'a_i', 'b_j', and 'c_k'.

6.8 Reliable Inquiry

The topic of the last section is the reliability of measures $f_P(\cdot,\cdot,\cdot)$ as indicators of truth[11], where P is some property of theories T in relation to evidences E and background knowledges B.

Let P be any such property that is assumed to be indicating truth in some model $\mathcal{M}^* = \langle Dom^*, \varphi^* \rangle$, and suppose $f_P(\cdot,\cdot,\cdot)$, $f_P(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, is a function such that $f_P(T,E,B)$ measures the degree to which property P is exhibited by T, E, and B, for all theories T, evidences E, and background knowledges B.

As argued in chapter 4, P respectively its measure f_P cannot be indicative of truth in $\mathcal{M}^*$, if its applying to T, E, and B is independent of the model $\mathcal{M}^*$ whose truth in one is interested in. Although P is a property of theories T in relation to evidences E and background knowledges B – and not a relation between theories and models – P may be indicating truth in $\mathcal{M}^*$ by means of the assumption that $\mathcal{M}^* \in mod(E)$ (or $\mathcal{M}^* \in mod(E) \cap mod(B)$).

Suppose this is the case. Then $f_P(\cdot,E,\cdot)$ indicates truth in all models $\mathcal{M} \in mod(E)$. As the model of interest $\mathcal{M}^*$ is only one among them, the reliability of P respectively $f_P(\cdot,E,\cdot)$ as indicator of truth in $\mathcal{M}^*$ depends on how many models there are in $mod(E)$. Therefore the value $f_P(T,E,B)$ is not the only parameter that must be considered in determining the degree to which truth of T in $\mathcal{M}^*$ is indicated.

An example illustrating this for $\mathcal{LI}$, $\mathcal{P}$, and C is one from chapter 5.[12] Let $E = \{Fa_1, \ldots, Fa_n\}, n \geq 1$, $B = \emptyset$, and $T = \{\forall x Fx\}$. Then

$$C(T,E,B) = \mathcal{P}(T,E,B) = \mathcal{LI}(T,E,B) = 1.^{13}$$

In case of $\mathcal{LI}$, the reliability of the indication of truth in $\mathcal{M}^* \in mod(E)$ depends on the number of individuals mentioned in E.

But the size of the evidence E (in the sense of $|C_{B-repr}(E)|$) is not the only property of importance for the reliability of $\mathcal{LI}(T,E,B)$ as indicator of truth of T in $\mathcal{M}^*$ relative to B. Let

$$E = \{F_1a_1, G_1a_1, \ldots G_na_1, \ldots, F_1a_m, G_1a_m, \ldots, G_na_m\},$$

[11] More generally, as indicators of some epistemically distinguished property of (sets of) statements in relation to models.

[12] In case of $\mathcal{P}$ the epistemically distinguished property of theories T in relation to models $\mathcal{M}$ is not truth of T in $\mathcal{M}$, but power of T for $\mathcal{M}$; in case of C it is the concatenation of power for and truth in $\mathcal{M}$.

[13] Once more, the confirmational domain is suppressed.

$$\begin{aligned}
E' &= \{F_1a_1, \ldots, F_na_1, G_1a_1, \ldots, G_na_1, \ldots, \\
&\quad \ldots, F_1a_m, \ldots, F_na_m, G_1a_m, \ldots, G_na_m\}, \\
B &= \emptyset, \quad \text{and} \\
T &= \{\forall x\,(F_1x \rightarrow G_1x), \ldots, \forall x\,(F_nx \rightarrow G_nx)\}.
\end{aligned}$$

Then

$$\mathcal{P}(T, E, B) = \mathcal{P}(T, E', B) = \mathcal{LI}(T, E, B) = \mathcal{LI}(T, E', B) = 1.$$

Although the number of individuals mentioned in E equals the number of individuals mentioned in E', the indication of truth of T in $\mathcal{M}^* \in mod(E) \cap mod(E')$ relative to B by means of $\mathcal{LI}(T, E', B)$ is more reliable than that by means of $\mathcal{LI}(T, E, B)$.

The reason for this may be seen in the fact that the set of all possible worlds or models in which E' is true is a subset of the set of all models in which E is true. Put differently, the more models evidence E excludes as candidates for the model of interest $\mathcal{M}^*$, the more reliable the indication of truth of T in $\mathcal{M}^*$ relative to B by means of $f_P(T, E, B)$, provided $f_P(T, E, B)$ is held constant. As a consequence, the indication of truth in $\mathcal{M}^*$ by means of $f_P(\cdot, E, \cdot)$ is maximally reliable, if $mod(E) = \{\mathcal{M}^*\}$.

If evidence E is considered as a test of theory T relative to background knowledge B, and T is taken to be the more severely tested by E relative to B, the less models there are in which E (and T) are true, then the above may be put as follows: The indication of truth of T in $\mathcal{M}^*$ relative to B by means of $f_P(T, E', B)$ is more reliable than by means of $f_P(T, E, B)$, because E' provides a more severe test of T relative to B than E. The degree of severity of the test provided by some evidence E for some theory T relative to some background knowledge B could then be defined as a function of some measure function $m(\cdot)$ defined on the powerset of the set of all models.[14]

[14]For reasons discussed below (arbitrariness), I do not pursue the question of how the severity of a test can be measured in this way. Let me just refer to Mayo (1996), p. 180, where it is argued that

> [p]assing a test T (with [result] e) counts as a good test of or as good evidence for [hypothesis] H just to the extent H fits e and T is a *severe test* of H,

and the criterion for severe tests is that

> [t]here is a high probability that test procedure T would not yield such a passing result, if H is false[,]

Under the assumption that evidence E is true in the model of interest $\mathcal{M}^*$, one may argue that E excludes the more models $\mathcal{M}$ as candidates for $\mathcal{M}^*$, the more E is similar to a (correct and) complete description $D_{\mathcal{M}^*}$ of $\mathcal{M}^*$, or the more $mod\,(E)$ is similar to $\mathcal{M}^*$. Given this, an alternative of this position says that the indication of truth in $\mathcal{M}^*$ by means of $f_P\,(\cdot, E, \cdot)$ is the more reliable, the greater the similarity of E to a (correct and) complete description $D_{\mathcal{M}^*}$ of $\mathcal{M}^*$ (the greater the similarity of $mod\,(E)$ to $\mathcal{M}^*$).[15]

However, one may also be of the opinion that the indication of truth of T in $\mathcal{M}^*$ by means of $f_P\,(T, E', B)$ is more reliable than that by means of $f_P\,(T, E, B)$, because evidence E' is (not smaller and) more varied than evidence E, i.e. because E' is better than E.

The question is whether these two positions – namely (1) exclusion of many models as candidates for $\mathcal{M}^*$ by E, and (2) goodness (size plus diversity) of E – coincide.

Suppose the first position is right. In order to get a measure for the reliability of $f_P\,(\cdot, E, \cdot)$ as indicator of truth in $\mathcal{M}^*$, one has to measure how many models $\mathcal{M}$ there are excluded by E as candidates for $\mathcal{M}^*$, or – in terms of verisimilitude and under the assumption that $\mathcal{M}^* \in mod\,(E)$ – how similar E is to a (correct and) complete description of $\mathcal{M}^*$.

By assumption, every evidence E is true in $\mathcal{M}^*$, whence every (correct and) complete description $D_{\mathcal{M}^*}$ is an extension of E in the sense that $D_{\mathcal{M}^*}$ logically implies E. But then it either holds that (i) the reliability of $f_P\,(\cdot, E, \cdot)$ as indicator of truth in $\mathcal{M}^*$ coincides with the logical content of E, because the way E is made logically stronger respectively logically weaker does not matter; or else (ii) one is in need of some (correct and) complete description $D_{\mathcal{M}^*}$ of $\mathcal{M}^*$ so that one can determine the similarity of E to $D_{\mathcal{M}^*}$.

In the first case the question is how such a measure could look like. To take some measure function $m\,(\cdot)$ defined on the powerset of the set of all models, and to define the reliability of $f_P\,(\cdot, E, \cdot)$ as indicator of truth in $\mathcal{M}^* \in mod\,(E)$ as a function of $m\,(\cdot)$ is problematic. There are uncountably many measures $m\,(\cdot)$, but no criterion for choosing the right one, whence defining such a function *non-arbitrarily* seems to be impossible.

---

where such a passing result is

one that accords at least as well with H as e does.

[15]For the topic of verisimilitude, truthlikeness, or likeness to truth cf. Kuipers (1987), Niiniluoto (1987), and Oddie (1986); for a survey article see Niiniluoto (1998).

So one has to be more modest, and be satisfied with a comparative concept of reliability. The indication of truth in $\mathcal{M}^*$ by means of $f_P(\cdot, E', \cdot)$ can then be defined to be more reliable than that by means of $f_P(\cdot, E, \cdot)$, if evidence E' logically implies evidence E. However, a consequence of this is that hardly any two functions $f_P(\cdot, E, \cdot)$ and $f_P(\cdot, E', \cdot)$ can be compared with respect to their reliability as truth indicators. Furthermore, theorem 6.2 tells us that every such ordinal measure is inferior to the goodness measure G, provided $C_{B-repr}(E) \subseteq C_{B-repr}(E')$.

In the second case the problem is that we do not have a (correct and) complete description $D_{\mathcal{M}^*}$ of $\mathcal{M}^*$ – at least, if $\mathcal{M}^*$ is some model of the actual world – whence we will never be in the position to determine the reliability of $f_P(\cdot, E, \cdot)$ as indicator of truth in $\mathcal{M}^* \in mod(E)$. Moreover, if we had such a (correct and) complete description $D_{\mathcal{M}^*}$ of $\mathcal{M}^*$ (in $\mathcal{L}_{PL1=}$), then every indicator of truth in $\mathcal{M}^*$ would be superfluous, for in this case we would know how the truth, the whole truth, and nothing but the truth in $\mathcal{M}^*$ looks like.

I conclude that even if the similarity of E to $D_{\mathcal{M}^*}$, for some (correct and) complete description $D_{\mathcal{M}^*}$ of $\mathcal{M}^*$ is the reason for the reliability of $f_P(\cdot, E, \cdot)$ as indicator of truth in $\mathcal{M}^*$, this line of argument is not worth being pursued, because we simply do not have a (correct and) complete description $D_{\mathcal{M}^*}$ of the model of interest $\mathcal{M}^*$ – and if we did, we would not be in need of an indicator of truth in $\mathcal{M}^*$.

Furthermore, the only practically applicable approach to reliability of truth indicators based on the exclusion-of-many-models-claim – namely to identify it with the logical content of E – is not only not promising, but is also inferior to the goodness measure G.

Thus the second approach – reliability of $f_P(T, E, B)$ as indicator of truth of T in $\mathcal{M}^* \in mod(E)$ relative to B is goodness of E relative to T and B – is superior to the first one, because it does not presuppose a (correct and) complete description $D_{\mathcal{M}^*}$ of $\mathcal{M}^*$ (respectively the model of interest $\mathcal{M}^*$ itself), and because it takes into account the *ratio* behind the first approach in the sense that G supports gathering evidence.

Chapter 7

In Conclusion

Concerning the combination of likeliness and loveliness, let me remark that the epistemically distinguished properties (of theories in relation to worlds or models) behind them are truth for the former, and power for the latter. For long, truth has been considered *the* epistemically distinguished property. This monograph should, among others, show that this exclusive focus on truth is not warranted. Having true theories is nice, but it is not all we want our theories to be – the theories we aim at should also be informative.

Finally, some prospects: As mentioned in chapter 3, confirmation has been a hot topic in the philosophy of science for more than a half century; but despite great efforts, there is still no generally accepted definition of (degree of) confirmation. This may be surprising. However, another observation strikes me as even more surprising: To the best of my knowledge, no-one has ever dealt with – let alone answered – the question what confirmation is good for, why we should stick to theories that are well confirmed. There are many theories of confirmation, but – as far as I know – there is no argument to the effect that confirmation is worth being pursued. Until now, there is no justification of confirmation!

One obvious reply is that confirmation by evidence from the actual world is indicative of truth in the actual world. However, if I am right, this is not the only feature we are after. Confirmation should not only lead to true theories (those are easy to obtain); it should lead to theories that are both true and informative.

A future project I am working on is therefore to investigate whether (probabilistic and non-probabilistic) theories of confirmation can be justified. In my opinion, the framework best suited for dealing with this question is formal learning theory (cf. Kelly 1996): Roughly speaking, the idea is to consider the long-run behaviour of a method (of discovery or assessment) that obeys the methodological recommendations of a given theory of confirmation. The question is whether and in which sense such a method is *reliable* (for which class of hypotheses does the method converge to the correct answer; and in which sense of convergence does it do so?). For the approach presented here, the questions concern the performance of a lovely learner, a likely learner, and a learner which is sensitive to loveliness and likeliness.[1]

[1] I am grateful to Vincent F. Hendricks for his suggestions concerning this justification of confirmation.

Appendix A

Proofs for Chapter 1

A.1 Proof of Theorem 1.1

Theorem A.1 (Domains of Proper Investigation) Let T be a scientific theory with domain $Dom_T = \langle D_1, \ldots, D_r \rangle$ and $D_{k_1}, \ldots, D_{k_n}$ as its domains of proper investigation, $1 \leq k_l \leq r$, for every $l, 1 \leq l \leq n$.

Then there is at least one finite axiomatization A_T of T with at least one occurrence of a k_l-variable, and without occurrences of k_l-constants, for every $l, 1 \leq l \leq n$.

Proof.
Let T be a theory with domain $Dom_T = \langle D_1, \ldots, D_r \rangle$ and $D_{k_1}, \ldots, D_{k_n}$ as its domains of proper investigation, $1 \leq k_l \leq r$, for every $l, 1 \leq l \leq n$. This means that for every D_{k_l} there is at least one finite axiomatization A_{k_l} of T with at least one essential occurrence of a k_l-variable, and without occurrences of k_l-constants.

Obviously, the A_{k_l} are not logically determined, for otherwise they cannot contain essential occurrences of a variable.

Consider A_{k_1}. From the interpolation-theorem (and the compactness of $PL1 =$) it follows that there is at least one finite set of wffs I_1 such that

$$A_{k_1} \vdash I_1 \vdash A_{k_2}, \quad C\left(I_1\right) \subseteq C\left(A_{k_1}\right) \cap C\left(A_{k_2}\right).$$

As $A_{k_2} \vdash A_{k_1}$, it follows that $A_{k_1} \dashv\vdash I_1 \dashv\vdash A_{k_2}$, which means that I_1 is a finite axiomatization of T with at least one essential occurrence of a k_1-variable, with at least one essential occurrence of a k_2-variable, and without occurrences of k_1- or k_2-constants. In particular, this means that $I_1 \vdash A_{k_3}$.

A second application of the interpolation theorem (in combination with the compactness of $PL1 =$) yields the existence of a finite set of wffs I_2 with

$$I_1 \vdash I_2 \vdash A_{k_3}, \quad C(I_2) \subseteq C(I_1) \cap C(A_{k_3}) \subseteq C(A_{k_1}) \cap C(A_{k_2}) \cap C(A_{k_3}),$$

which is a finite axiomatization of T with at least one essential occurrence of a k_1-variable, with at least one essential occurrence of a k_2-variable, with at least one essential occurrence of a k_3-variable, and without occurrences of k_1-, k_2-, or k_3-constants, because $A_{k_3} \vdash I_1$.

By continuing in this manner one arrives (after $n-1$ steps) at a finite set of wffs I_{n-1} with

$$I_{n-2} \vdash I_{n-1} \vdash A_{k_n}, \quad C(I_{n-1}) \subseteq C(I_{n-2}) \cap C(A_{k_n}) \subseteq \bigcap_{1 \le l \le n} C(A_{k_l}),$$

which is a finite axiomatization of T with at least one essential occurrence of a k_l-variable, and without occurrences of k_l-constants, for every $l, 1 \le l \le n$, because $A_{k_n} \vdash I_{n-2}$. $\square$

A.2 Proof of Theorem 1.2

Theorem A.2 (Strict Probabilities) Let $p(\cdot)$, $p(\cdot) : \mathcal{L}_{prop} \to \Re$, be a strict (unconditional) probability, and let $p(\cdot \mid \cdot)$ be the conditional probability based on $p(\cdot)$. Then it holds for any wffs $A, B \in \mathcal{L}_{prop}$ with $p(A) > 0$:

$$p(B \mid A) = 1, \quad \text{only if} \quad A \vdash B.$$

Proof.
Let $p(\cdot)$, $p(\cdot) : \mathcal{L}_{prop} \to \Re$, be a strict (unconditional) probability, and let $p(\cdot \mid \cdot)$ be the conditional probability based on $p(\cdot)$. Let $A, B \in \mathcal{L}_{prop}$ with $p(A) > 0$, and suppose $p(B \mid A) = 1$, i.e. $p(B \wedge A) = p(A)$, which holds iff

$$\begin{array}{ccc} p(B \wedge A) + p(\neg A) & = & 1 \\ & \text{iff} & \\ p((B \wedge A) \vee \neg A) & = & 1 \quad A \wedge B \vdash \neg\neg A \\ & \text{only if} & \\ & \vdash (B \wedge A) \vee \neg A & \quad p(\cdot) \text{ is strict} \\ & \text{iff} & \\ & A \vdash B. & \end{array}$$

$\square$

Appendix B

Proofs for Chapter 2

B.1 Proof of (Non-) Arbitrariness Claim

Application B.1 (Arbitrariness)

1. Every set of Bayesian relevance measures is arbitray; in particular, this holds of d, r, l, s, and c.
3. The uncountable set F of functions $f_a(\cdot)$, $f_a(\cdot) : \Re_0^+ \to \Re$, with

$$\begin{aligned} f_a(x) = x^a, \quad & x \in \Re_0^+ = \{x : x \in \Re, x \geq 0\}, \\ & a \in \Re^+ = \{x : x \in \Re, x > 0\}, \end{aligned}$$

is not arbitrary.

Proof.
1. It suffices to construct two probability spaces with four events, where the background knowledge K is set to $\top$, H is some hypothesis, and E is some evidence. Let

$$p_1(H \wedge E) = 0.5, p_1(\neg H \wedge E) = 0.25, p_1(H \wedge \neg E) = p_1(\neg H \wedge \neg E) = 0.125,$$

and let

$$p_2(H \wedge E) = 0.25, p_2(\neg H \wedge E) = 0.5, p_2(H \wedge \neg E) = p_2(\neg H \wedge \neg E) = 0.125.$$

$$p_1(H \mid E) = \frac{16}{24} > \frac{15}{24} = p_1(H) \quad \text{and} \quad p_1(\neg H \mid E) = \frac{8}{24} < \frac{9}{24} = p_1(\neg H),$$

whence it holds for every set of relevance measures m:

$$m_{p_1}(H,E) > 0 > m_{p_1}(\neg H,E).$$

$$p_2(H \mid E) = \frac{8}{24} < \frac{9}{24} = p_2(H) \quad \text{and} \quad p_2(\neg H \mid E) = \frac{16}{24} > \frac{15}{24} = p_2(\neg H),$$

whence it holds for every set of relevance measures m:

$$m_{p_2}(H,E) < 0 < m_{p_2}(\neg H,E).$$

3. There are no $x, y \in \Re_0^+$ and $a, b \in \Re^+$ such that

$$f_a(x) < f_a(y) \quad \text{and} \quad f_b(y) < f_b(x),$$

since it holds for any $x, y \in \Re_0^+$ and every $a \in \Re^+$:

$$x < y \quad \text{iff} \quad x^a < y^a.$$

□

B.2 Calculations

B.2.1 Calculation 1

$$\begin{array}{rcl} hu_p(T,E,B) & > & 0 \\ & \text{iff} & \\ p(E \mid T \wedge (B-E)) > p(E \mid B-E) & \text{and} & p(E \mid B) > p(E \mid B-E) \\ & \text{or} & \\ p(E \mid T \wedge (B-E)) < p(E \mid B-E) & \text{and} & p(E \mid B) < p(E \mid B-E), \end{array}$$

provided $p(T \wedge (B-E)) > 0$, $1 > p(E \mid B-E) > 0$, and $p(B) > 0$.

Calculation:

$$\begin{array}{rcl} hu_p(T,E,B) & = & p(T \mid (B-E) \wedge E) \cdot p(E \mid B) + \\ & & +p(T \mid (B-E) \wedge \neg E) \cdot p(\neg E \mid B) - p(T \mid B-E) \end{array}$$

is positive, i.e. >0, iff

$$\begin{array}{rcl}
\frac{p\left(T\wedge(B-E)\wedge E\right)\cdot p\left(E\mid B\right)}{p\left((B-E)\wedge E\right)}+ & & \\
+\frac{p\left(T\wedge(B-E)\wedge\neg E\right)\cdot p\left(\neg E\mid B\right)}{p\left((B-E)\wedge\neg E\right)} & > & \frac{p\left(T\wedge(B-E)\right)}{p\left(B-E\right)} \\
 & \text{iff} & \\
p\left(E\mid T\wedge(B-E)\right)\cdot\frac{p\left(E\mid B\right)}{p\left(E\mid B-E\right)}+ & & \\
+p\left(\neg E\mid T\wedge(B-E)\right)\cdot\frac{p\left(\neg E\mid B\right)}{p\left(\neg E\mid B-E\right)} & > & 1 \\
 & \text{iff} & \\
p\left(E\mid T\wedge(B-E)\right)\cdot & & \\
\cdot p\left(E\mid B\right)\cdot\left(1-p\left(E\mid B-E\right)\right)+ & & \\
+\left(1-p\left(E\mid T\wedge(B-E)\right)\right)\cdot & & \\
\cdot\left(1-p\left(E\mid B\right)\right)\cdot p\left(E\mid B-E\right) & > & p\left(E\mid B-E\right)\cdot \\
 & & \cdot\left(1-p\left(E\mid B-E\right)\right) \\
 & \text{iff} & \\
p\left(E\mid T\wedge(B-E)\right)\cdot & & \\
\cdot\left(p\left(E\mid B\right)-p\left(E\mid B-E\right)\right) & > & p\left(E\mid B-E\right)\cdot \\
 & & \cdot\left(p\left(E\mid B\right)-p\left(E\mid B-E\right)\right) \\
 & \text{iff} & \\
p\left(E\mid T\wedge(B-E)\right)>p\left(E\mid B-E\right) & \text{and} & p\left(E\mid B\right)>p\left(E\mid B-E\right) \\
 & \text{or} & \\
p\left(E\mid T\wedge(B-E)\right)<p\left(E\mid B-E\right) & \text{and} & p\left(E\mid B\right)<p\left(E\mid B-E\right).
\end{array}$$

Note that this result holds also in case '$p\left(B-E\right)$' is replaced by '$p\left(B\wr E\right)$'.

B.2.2 Calculation 2

$$p_2\left(T\mid(B_2\wr E)\wedge E\right)=p_1\left(T\mid(B_1\wr E)\wedge E\right),$$

provided $B_1\wr E\ \dashv\vdash\ B_2\wr E$, $p_2\left(E\mid B_2\right)>0$, $p_2\left((B_2\wr E)\wedge E\right)>0$, and $p_1\left((B_1\wr E)\wedge E\right)>0$.

Calculation:

$$
\begin{aligned}
p_2\left(T \mid (B_2 \wr E) \wedge E\right) &= \frac{p_2\left(T \wedge (B_2 \wr E) \wedge E\right)}{p_2\left((B_2 \wr E) \wedge E\right)} \\
&= \frac{p_1\left(T \wedge (B_1 \wr E) \wedge E \mid E\right) \cdot p_2\left(E \mid B_2\right) +}{p_1\left((B_1 \wr E) \wedge E \mid E\right) \cdot p_2\left(E \mid B_2\right) +} \\
&\quad \frac{+p_1\left(T \wedge (B_1 \wr E) \wedge E \mid \neg E\right) \cdot p_2\left(\neg E \mid B_2\right)}{+p_1\left((B_1 \wr E) \wedge E \mid \neg E\right) \cdot p_2\left(\neg E\right)} \\
&\quad JC, \quad \text{and} \quad B_1 \wr E \dashv\vdash B_2 \wr E \quad ^1 \\
&= \frac{p_1\left(T \wedge (B_1 \wr E) \wedge E \mid E\right) \cdot p_2\left(E \mid B_2\right)}{p_1\left((B_1 \wr E) \wedge E \mid E\right) \cdot p_2\left(E \mid B_2\right)} \\
&= p_1\left(T \mid (B_1 \wr E) \wedge E\right),
\end{aligned}
$$

where Jeffrey conditionalisation is replaced by strict conditionalisation, if $p_1(E) = 1$, and so $p_2(E) = p_2(E \mid B_2) = 1$.

In the same way one shows that

$$p_2\left(T \mid (B_2 \wr E) \wedge \neg E\right) = p_1\left(T \mid (B_1 \wr E) \wedge \neg E\right),$$

provided $B_1 \wr E \dashv\vdash B_2 \wr E$, $p_2(\neg E \mid B_2) > 0$, $p_2((B_2 \wr E) \wedge \neg E) > 0$, and $p_1((B_1 \wr E) \wedge \neg E) > 0$, where Jeffrey conditionalisation is replaced by strict conditionalisation, if $p_1(\neg E) = 1$, and so $p_2(\neg E) = p_2(\neg E \mid B_2) = 1$.

Note that this result holds also for counterfactual Jeffrey conditionalisation, i.e. if '$p_2(\pm E \mid B_2)$' is replaced by '$p_2(\pm E \mid B_2 \wr E)$'.

B.2.3 Calculation 3

$$
\begin{aligned}
p_2\left(T \mid B_2 \wr E\right) &= \frac{p_1\left(T \mid B_1 \wr E\right)}{p_1(E) \cdot \left(1 - p_1(E)\right)} \cdot \\
&\quad \cdot \left[p_1\left(E \mid T \wedge (B_1 \wr E)\right) \cdot \left(p_2\left(E \mid B_2\right) - p_1(E)\right) + \right. \\
&\quad \left. + p_1(E) \cdot \left(1 - p_2\left(E \mid B_2\right)\right)\right],
\end{aligned}
$$

provided $B_1 \wr E \dashv\vdash B_2 \wr E$, $p_1(E \mid B_1 \wr E) = p_1(E)$, $p_2(B_2) > 0$, $p_2(B_2 \wr E) > 0$, $p_1(B_1 \wr E) > 0$, and $1 > p_1(E) > 0$.

[1]The reader should be so kind to read this as one single fraction over two lines.

Calculation:
If $p_1\left(T \wedge (B_1 \wr E)\right) > 0$, then

$$
\begin{aligned}
p_2\left(T \mid B_2 \wr E\right) &= \frac{p_2\left(T \wedge (B_2 \wr E)\right)}{p_2\left(B_2 \wr E\right)} \\
&= \frac{p_1\left(T \wedge (B_1 \wr E) \mid E\right) \cdot p_2\left(E \mid B_2\right) +}{p_1\left(B_1 \wr E \mid E\right) \cdot p_2\left(E \mid B_2\right) +} \\
&\quad \frac{+p_1\left(T \wedge (B_1 \wr E) \mid \neg E\right) \cdot p_2\left(\neg E \mid B_2\right)}{+p_1\left(B_1 \wr E \mid \neg E\right) \cdot p_2\left(\neg E \mid B_2\right)} \\
&\quad JC, \quad \text{and} \quad B_1 \wr E \dashv\vdash B_2 \wr E \quad \text{[2]} \\
&= \frac{p_1\left(T \wedge (B_1 \wr E)\right)}{p_1\left(B_1 \wr E\right)} \cdot \\
&\quad \cdot \left(\frac{p_1\left(E \mid T \wedge (B_1 \wr E)\right) \cdot p_1\left(\neg E\right) \cdot p_2\left(E \mid B_2\right) +}{p_1\left(E \mid B_1 \wr E\right) \cdot p_1\left(\neg E\right) \cdot p_2\left(E \mid B_2\right) +} \right. \\
&\quad \left. \frac{+p_1\left(\neg E \mid T \wedge (B_1 \wr E)\right) \cdot p_1\left(E\right) \cdot p_2\left(\neg E \mid B_2\right)}{+p_1\left(\neg E \mid B_1 \wr E\right) \cdot p_1\left(E\right) \cdot p_2\left(\neg E \mid B_2\right)} \right) \\
&= \frac{p_1\left(T \mid B_1 \wr E\right)}{p_1\left(E\right) \cdot \left(1 - p_1\left(E\right)\right)} \cdot \\
&\quad \cdot \left[p_1\left(E \mid T \wedge (B_1 \wr E)\right) \cdot p_1\left(\neg E\right) \cdot p_2\left(E \mid B_2\right) + \right. \\
&\quad \left. + p_1\left(\neg E \mid T \wedge (B_1 \wr E)\right) \cdot p_1\left(E\right) \cdot p_2\left(\neg E \mid B_2\right)\right] \\
&\quad p_1\left(E \mid B_1 \wr E\right) = p_1\left(E\right) \\
&= \frac{p_1\left(T \mid B_1 \wr E\right)}{p_1\left(E\right) \cdot \left(1 - p_1\left(E\right)\right)} \cdot \\
&\quad \cdot \left[p_1\left(E \mid T \wedge (B_1 \wr E)\right) \cdot \left(p_2\left(E \mid B_2\right) - p_1\left(E\right)\right) + \right. \\
&\quad \left. + p_1\left(E\right) \cdot \left(1 - p_2\left(E \mid B_2\right)\right)\right]
\end{aligned}
$$

which is equal to

$$
p_1\left(T \mid B_1 \wr E\right) \cdot \frac{p_2\left(E \mid B_2\right)}{p_1\left(E\right)},
$$

if $T \vdash E$. In case $p_1\left(T \wedge (B_1 \wr E)\right) = 0$, $p_2\left(T \mid B_2 \wr E\right) = 0$.

If counterfactual Jeffrey conditionalisation is used and '$p_2\left(E \mid B_2\right)$' is re-

[2] The reader should be so kind to read this and the following as one single fraction over two lines.

placed by '$p_2(E \mid B_2 \wr E)$', calculation 3 yields that

$$\begin{aligned} p_2(T \mid B_2 \wr E) &= \frac{p_1(T \mid B_1 \wr E)}{p_1(E) \cdot (1 - p_1(E))} \cdot \\ &\quad [p_1(E \mid T \wedge (B_1 \wr E)) \cdot (p_2(E \mid B_2 \wr E) - p_1(E)) + \\ &\quad p_1(E) \cdot (1 - p_2(E \mid B_2 \wr E))], \end{aligned}$$

which is greater than $p_1(T \mid B_1 \wr E)$ just in case

$$\begin{array}{ccc} p_1(E \mid T \wedge (B_1 \wr E)) > p_1(E \mid B_1 \wr E) & \text{and} & p_2(E \mid B_2 \wr E) > p_1(E \mid B_1 \wr E) \\ & \text{or} & \\ p_1(E \mid T \wedge (B_1 \wr E)) < p_1(E \mid B_1 \wr E) & \text{and} & p_2(E \mid B_2 \wr E) < p_1(E \mid B_1 \wr E); \end{array}$$

and $p_2(T \mid B_2 \wr E) = p_1(T \mid B_1 \wr E)$ iff

$$p_1(E \mid T \wedge (B_1 \wr E)) = p_1(E \mid B_1 \wr E) \quad \text{or} \quad p_2(E \mid B_2 \wr E) = p_1(E \mid B_1 \wr E).$$

B.2.4 Calculation 4

Suppose

$$B_1 \wr E \dashv\vdash B_2 \wr E \quad \text{and} \quad p_1(E \mid B_1 \wr E) = p_1(E).$$

If $p_2(T \mid B_2 \wr E)$ is the result of Jeffrey conditioning on E, then

$$\begin{array}{ccc} hu_{p_1}(T, E, B_1) & > & hu_{p_2}(T, E, B_2) \\ & \text{iff} & \\ p_1(E \mid T \wedge (B_1 \wr E)) > p_1(E \mid B_1 \wr E) & \text{and} & p_1(E \mid B_1) > p_1(E \mid B_1 \wr E) \\ & \text{or} & \\ p_1(E \mid T \wedge (B_1 \wr E)) < p_1(E \mid B_1 \wr E) & \text{and} & p_1(E \mid B_1) < p_1(E \mid B_1 \wr E), \end{array}$$

provided

$$\begin{array}{l} p_2(B_2) > 0, \quad p_2((B_2 \wr E) \wedge E) > 0, \quad p_2((B_2 \wr E) \wedge \neg E) > 0, \\ p_1(B_1 \wr E) > 0, \quad 1 > p_1(E) > 0, \quad \text{and} \quad p_1(T \mid B_1 \wr E) > 0. \end{array}$$

Calculation:
Let

$$x := \frac{p_1(E \mid T \wedge (B_1 \wr E)) \cdot (p_2(E \mid B_2) - p_1(E)) + p_1(E) \cdot (1 - p_2(E \mid B_2))}{p_1(E) \cdot (1 - p_1(E))}.$$

Then

$$
\begin{aligned}
p_1(T \mid (B_1 \wr E) \wedge E) \cdot p_1(E \mid B_1) + & \\
+p_1(T \mid (B_1 \wr E) \wedge \neg E) \cdot & \\
\cdot p_1(\neg E \mid B_1) - p_1(T \mid B_1 \wr E) &> p_2(T \mid (B_2 \wr E) \wedge E) \cdot p_2(E \mid B_2) + \\
& \quad +p_2(T \mid (B_2 \wr E) \wedge \neg E) \cdot \\
& \quad \cdot p_2(\neg E \mid B_2) - p_2(T \mid B_2 \wr E)
\end{aligned}
$$

iff

$$
\begin{aligned}
p_1(T \mid (B_1 \wr E) \wedge E) \cdot p_1(E \mid B_1) + & \\
+p_1(T \mid (B_1 \wr E) \wedge \neg E) \cdot & \\
\cdot p_1(\neg E \mid B_1) - p_1(T \mid B_1 \wr E) &> p_1(T \mid (B_1 \wr E) \wedge E) \cdot p_2(E \mid B_2) + \\
& \quad +p_1(T \mid (B_1 \wr E) \wedge \neg E) \cdot \\
& \quad \cdot p_2(\neg E \mid B_2) - p_1(T \mid B_1 \wr E) \cdot x \\
& \quad B_1 \wr E \dashv\vdash B_2 \wr E, \\
& \quad \text{calculations 2 and 3}
\end{aligned}
$$

iff

$$
\begin{aligned}
p_1(T \mid (B_1 \wr E) \wedge \neg E) \cdot & \\
\cdot (p_1(\neg E \mid B_1) - p_2(\neg E \mid B_2)) + & \\
+p_1(T \mid B_1 \wr E) \cdot (x-1) &> p_1(T \mid (B_1 \wr E) \wedge E) \cdot \\
& \quad \cdot (p_2(E \mid B_2) - p_1(E \mid B_1)).
\end{aligned}
$$

The latter holds just in case

$$
\begin{aligned}
x - 1 &> (p_2(E \mid B_2) - p_1(E \mid B_1)) \cdot \\
& \quad \cdot \left(\frac{p_1(E \mid T \wedge (B_1 \wr E))}{p_1(E \mid B_1 \wr E)} - \right. \\
& \quad \left. - \frac{p_1(\neg E \mid T \wedge (B_1 \wr E))}{p_1(\neg E \mid B_1 \wr E)} \right)
\end{aligned}
$$

iff

$$
\begin{aligned}
\frac{p_1(E \mid T \wedge (B_1 \wr E)) \cdot}{1 \cdot} & \\
\frac{\cdot (p_2(E \mid B_2) - p_1(E)) +}{\cdot p_1(E) \cdot} & \\
\frac{+p_1(E) \cdot (1 - p_2(E \mid B_2))}{\cdot (1 - p_1(E))} - &
\end{aligned}
$$

$$
\begin{array}{rcl}
-\dfrac{p_1(E)\cdot(1-p_1(E))}{p_1(E)\cdot(1-p_1(E))} & > & (p_2(E \mid B_2) - p_1(E \mid B_1)) \cdot \\
 & & \cdot\dfrac{p_1(E \mid T \wedge (B_1 \wr E)) - p_1(E)}{p_1(E)\cdot(1-p_1(E))} \\
 & & p_1(E \mid B_1 \wr E) = p_1(E) \quad {}^{3} \\
 & \text{iff} & \\
(p_1(E \mid T \wedge (B_1 \wr E)) - p_1(E)) \cdot & & \\
\cdot\, (p_2(E \mid B_2) - p_1(E)) & > & (p_2(E \mid B_2) - p_1(E \mid B_1)) \cdot \\
 & & \cdot\, (p_1(E \mid T \wedge (B_1 \wr E)) - p_1(E)),
\end{array}
$$

which holds if and only if

$$
\begin{array}{ccc}
p_1(E \mid T \wedge (B_1 \wr E)) > p_1(E \mid B_1 \wr E) & \text{and} & p_1(E \mid B_1) > p_1(E \mid B_1 \wr E) \\
 & \text{or} & \\
p_1(E \mid T \wedge (B_1 \wr E)) < p_1(E \mid B_1 \wr E) & \text{and} & p_1(E \mid B_1) < p_1(E \mid B_1 \wr E).
\end{array}
$$

B.3 Proof of Theorem 2.1

Theorem B.1 (NecSuff) Given

$$B_1 \wr E \dashv\vdash B_2 \wr E, \quad p_1(E \mid B_1 \wr E) = p_1(E), \quad \text{and} \quad p_1(T \mid B_1 \wr E) > 0,$$

the equality

$$p_1(T \mid B_1 \wr E) = p_2(T \mid B_2 \wr E)$$

is necessary and sufficient for the equivalence

$$
\begin{array}{ccc}
hu_{p_2}(T, E, B_2) & > & hu_{p_1}(T, E, B_1) \\
 & \text{iff} & \\
p_1(E \mid T \wedge (B_1 \wr E)) > p_1(E \mid B_1 \wr E) & \text{and} & p_2(E \mid B_2) > p_1(E \mid B_1) \\
 & \text{or} & \\
p_1(E \mid T \wedge (B_1 \wr E)) < p_1(E \mid B_1 \wr E) & \text{and} & p_2(E \mid B_2) < p_1(E \mid B_1),
\end{array}
$$

provided

$$
\begin{array}{l}
p_2(B_2) > 0, \quad p_2((B_2 \wr E) \wedge E) > 0, \\
p_2((B_2 \wr E) \wedge \neg E) > 0, \quad \text{and} \quad 1 > p_1(E) > 0.
\end{array}
$$

[3] The reader should be so kind to read this as one single fraction over three lines.

With counterfactual Jeffrey condition this means that

$$p_1(E \mid T \wedge (B_1 \wr E)) = p_1(E \mid B_1 \wr E) \quad \text{or} \quad p_1(E \mid B_1 \wr E) = p_2(E \mid B_2 \wr E)$$

is necessary and sufficient for this equivalence.

Proof.
It is easily seen that the equation

$$p_1(T \mid B_1 \wr E) = p_2(T \mid B_2 \wr E)$$

is sufficient for the above equivalence. That it is also necessary is seen as follows. Suppose the equivalence holds. Now

$$\begin{aligned}
&p_2(T \mid (B_2 \wr E) \wedge E) \cdot p_2(E \mid B_2) + \\
&+p_2(T \mid (B_2 \wr E) \wedge \neg E) \cdot \\
&\cdot p_2(\neg E \mid B_2) - p_2(T \mid B_2 \wr E) \quad > \quad p_1(T \mid (B_1 \wr E) \wedge E) \cdot p_1(E \mid B_1) + \\
&\qquad +p_1(T \mid (B_1 \wr E) \wedge \neg E) \cdot \\
&\qquad \cdot p_1(E \mid B) - p_1(T \mid B_1 \wr E)
\end{aligned}$$

iff

$$\begin{aligned}
&p_1(T \mid (B_1 \wr E) \wedge E) \cdot \\
&\cdot (p_2(E \mid B_2) - p_1(E \mid B_1)) + \\
&+p_1(T \mid (B_1 \wr E) \wedge \neg E) \cdot \\
&\cdot (p_2(\neg E \mid B_2) - p_1(\neg E \mid B_1)) \quad > \quad p_2(T \mid B_2 \wr E) - \\
&\qquad -p_1(T \mid B_1 \wr E)
\end{aligned}$$

calculation 2

iff

$$\begin{aligned}
&\frac{p_1(E \mid T \wedge (B_1 \wr E))}{p_1(E \mid B_1 \wr E)} \cdot \\
&\cdot (p_2(E \mid B_2) - p_1(E \mid B_1)) \quad > \quad \frac{p_1(\neg E \mid T \wedge (B_1 \wr E))}{p_1(\neg E \mid B_1 \wr E)} \cdot \\
&\qquad \cdot (p_2(E \mid B_2) - p_1(E \mid B_1)) + \\
&\qquad + \frac{p_2(T \mid B_2 \wr E)}{p_1(T \mid B_1 \wr E)} - 1,
\end{aligned}$$

which, by assumption, holds just in case

$$p_1(E \mid T \wedge (B_1 \wr E)) > p_1(E \mid B_1 \wr E) \quad \text{and} \quad p_2(E \mid B_2) > p_1(E \mid B_1)$$

$$\text{or}$$
$$p_1(E \mid T \wedge (B_1 \wr E)) < p_1(E \mid B_1 \wr E) \quad \text{and} \quad p_2(E \mid B_2) < p_1(E \mid B_1).$$

Since

$$\begin{array}{rcl} \dfrac{p_1(E \mid T \wedge (B_1 \wr E))}{p_1(E \mid B_1 \wr E)} & > & \dfrac{p_1(\neg E \mid T \wedge (B_1 \wr E))}{p_1(\neg E \mid B_1 \wr E)} \\ & \text{iff} & \\ p_1(E \mid T \wedge (B_1 \wr E)) & > & p_1(E \mid B_1 \wr E), \end{array}$$

the following has to hold:

$$p_1(T \mid B_1 \wr E) = p_2(T \mid B_2 \wr E).$$

Finally, counterfactual Jeffrey conditionalisation yields that

$$\begin{array}{rcl} p_2(T \mid B_2 \wr E) & = & p_1(T \mid B_1 \wr E) \\ & \text{iff} & \\ p_1(E \mid T \wedge (B_1 \wr E)) \cdot & & \\ \cdot (p_2(E \mid B_2 \wr E) - p_1(E)) + & & \\ \cdot p_1(E) \cdot (1 - p_2(E \mid B_2 \wr E)) & = & p_1(E) \cdot (1 - p_1(E)) \\ & & \text{calculation 3} \\ & \text{iff} & \\ p_1(E \mid T \wedge (B_1 \wr E)) \cdot & & \\ \cdot (p_2(E \mid B_2 \wr E) - p_1(E)) & = & p_1(E) \cdot (p_2(E \mid B_2 \wr E) - p_1(E)) \\ & \text{iff} & \\ p_1(E \mid T \wedge (B_1 \wr E) = p_1(E \mid B_1 \wr E) & \text{or} & p_1(E \mid B_1 \wr E) = p_2(E \mid B_2 \wr E). \end{array}$$

□

Appendix C

Proofs for Chapter 3

C.1 Proof of Theorem 3.1

Theorem C.1 (Power Searcher and Truth Indicator) Let T, E, and B range over wffs of $\mathcal{L}_{prop}$ (instead of theories, evidences, and background knowlegdes, respectively, which are sets of wffs of $\mathcal{L}_{PL1=}$) in the definitions of searching power and indicating truth. Then it holds for every contingent wff E and every strict (unconditional) probability $p(\cdot)$:

1. $p(\cdot \mid E \wedge \cdot)$ is indicating truth in $mod(E)$.
2. $i(\cdot, E, \cdot) := 1 - p(\cdot \wedge \cdot \mid \neg E)$ is searching power for $mod(E)$.
3. $i'(\cdot, E, \cdot) := 1 - p(\cdot \mid \neg E \wedge \cdot)$ is searching power for $mod(E)$, if it is defined, i.e. if $\neg E \wedge B \nvdash \bot$.

Proof.
Let T, T', E, and B be four wffs of $\mathcal{L}_{prop}$, E being contingent, and let $p(\cdot \mid \cdot)$ be the conditional probability based on some strict (unconditional) probability $p(\cdot)$. So $0 < p(E), p(\neg E) < 1$. Suppose $E \wedge B \nvdash \bot$, whence $p(E \wedge B) > 0$.
(1.1)

$$p(T \mid E \wedge B) = \frac{p(T \wedge E \wedge B)}{p(E \wedge B)} \geq 0.$$

(1.2) If $E \wedge B \vdash T$, then

$$p(T \mid E \wedge B) = \frac{p(T \wedge E \wedge B)}{p(E \wedge B)} = \frac{p(E \wedge B)}{p(E \wedge B)} = 1.$$

(1.3) If $T' \vdash T$, then

$$p\left(T' \mid E \wedge B\right) = \frac{p\left(T' \wedge E \wedge B\right)}{p\left(E \wedge B\right)} \leq \frac{p\left(T \wedge E \wedge B\right)}{p\left(E \wedge B\right)} = p\left(T \mid E \wedge B\right).$$

(2.1)

$$i\left(T, E, B\right) = 1 - p\left(T \wedge B \mid \neg E\right) \geq 1 - 1 = 0.$$

(2.2) If $T \wedge B \vdash E$, then $p\left(T \wedge B \wedge \neg E\right) = 0$, whence

$$i\left(T, E, B\right) = 1 - p\left(T \wedge B \mid \neg E\right) = 1 - \frac{p\left(T \wedge B \wedge \neg E\right)}{p\left(\neg E\right)} = 1 - 0 = 1.$$

(2.3) If $T' \vdash T$, then $p\left(T' \wedge B \mid \neg E\right) \leq p\left(T \wedge B \mid \neg E\right)$, whence

$$i\left(T', E, B\right) = 1 - p\left(T' \wedge B \mid \neg E\right) \geq 1 - p\left(T \wedge B \mid \neg E\right) = i\left(T, E, B\right).$$

(3) is shown in a similar way. □

C.2 Proof of Theorem 3.2

Theorem C.2 (Truth Indicating Power Searchers Are Constant) Let E be an evidence, and let $f\left(\cdot, E, \cdot\right)$, $f\left(\cdot, E, \cdot\right) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, be searching power for $mod\left(E\right)$.

If $f\left(\cdot, E, B\right)$ is indicating truth in $mod\left(E\right)$, then it holds for every theory T and every background knowledge B with $E \cup B \nvdash \bot$: $f\left(T, E, B\right) = 1$.

Proof.
Let E be an evidence from $D_1, \ldots, D_k$, and let $f\left(\cdot, E, \cdot\right)$, $f\left(\cdot, E, \cdot\right) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, be searching power for $mod\left(E\right)$. Suppose $f\left(\cdot, E, \cdot\right)$ is indicating truth in $mod\left(E\right)$, and let B be a background knowledge with $E \cup B \nvdash \bot$.

E is contingent, for otherwise it cannot contain an essential occurrence of an i-constant. Let T_E be defined as follows:

$$T_E := \left\{\exists \mathbf{x}^1 \ldots \exists \mathbf{x}^k A\left[\mathbf{x}^1/\mathbf{a}^1, \ldots, \mathbf{x}^k/\mathbf{a}^k\right] : A \in E\right\},$$

where

$$\begin{aligned}&\exists \mathbf{x}^1 \ldots \exists \mathbf{x}^k A\left[\mathbf{x}^1/\mathbf{a}^1, \ldots, \mathbf{x}^k/\mathbf{a}^k\right] := \\ &= \exists x_1^1 \ldots \exists x_{l_1}^1 \ldots \exists x_1^k \ldots \exists x_{l_k}^k A\left[x_1^1/a_1^1, \ldots, x_{l_1}^1/a_{l_1}^1, \ldots, x_1^k/a_1^k, \ldots, x_{l_k}^k/a_{l_k}^k\right],\end{aligned}$$

and $a_1^i, \ldots, a_{l_i}^i$ are all i-constants essentially occurring in E, for every $i, 1 \leq i \leq k$. T_E is a theory with $D_1, \ldots, D_k$ as its domains of proper investigation, and such that

$$T_E \cup B \nvdash \perp, \quad \text{and} \quad E \cup B \vdash T_E,$$

whence $f(T_E, E, B) = 1$.

Let T be any theory with $T \cup T_E \cup B \nvdash \perp$. As $T \cup T_E \vdash T_E$; as $f(\cdot, E, \cdot)$ is indicating truth in $mod(E)$; and as $f(\cdot, E, \cdot)$ is searching power for $mod(E)$,

$$\begin{array}{l} f(T \cup T_E, E, B) \leq f(T_E, E, B) = 1, \quad \text{and} \\ f(T \cup T_E, E, B) \geq f(T_E, E, B) = 1, \end{array}$$

i.e. $f(T \cup T_E, E, B) = 1$. In the same way it follows from $T \cup T_E \vdash T$ that

$$\begin{array}{l} f(T, E, B) \geq f(T \cup T_E, E, B) = 1, \quad \text{and} \\ f(T, E, B) \leq f(T \cup T_E, E, B) = 1, \end{array}$$

i.e. $f(T, E, B) = 1$. □

C.3 Proof of Theorem 3.3

Theorem C.3 (SensSimplCons and Unimpressability) Let $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \to \Re$, be a function.

1. If f is sensitive to simplicity considerations in the very strong sense, then f is sensitive to simplicity considerations in the strong sense.

2. If f is sensitive to simplicity considerations in the strong sense, then f is sensitive to simplicity considerations in the weak sense.

3. If f is sensitive to simplicity considerations in the weak sense, then f cannot be impressed by redundancies.

Proof.
(1) Let $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \to \Re$, be a function that is sensitive to simplicity considerations in the very strong sense. Then there is at least one power searcher $\mathcal{LO}$ such that it holds for any theories T and T', every evidence E, and every background knowledge B:

If $T' \vdash T$ and $\mathcal{LO}(T, E, B) = \mathcal{LO}(T', E, B)$, then $f(T, E, B) \geq f(T', E, B)$.

Let T be a theory, let E be an evidence, let B be a background knowledge, and suppose $h \in T$ is a $\mathcal{LO}$-superfluous part of T for E and B, i.e.

$$\mathcal{LO}\left(T \setminus \{h\}, E, B\right) = \mathcal{LO}\left(T, E, B\right).$$

As $T \vdash T \setminus \{h\}$, and as f is sensitive to simplicity considerations in the very strong sense,

$$f\left(T \setminus \{h\}, E, B\right) \geq f\left(T, E, B\right).$$

So there is at least one power searcher $\mathcal{LO}$ such that it holds for every theory T, every evidence E, and every background knowledge B:

> If h is a $\mathcal{LO}$-superfluous part of T for E and B, then $f\left(T \setminus \{h\}, E, B\right) \geq f\left(T, E, B\right)$,

which just means that f is sensitive to simplicity considerations in the strong sense.

(2) Let $f\left(\cdot,\cdot,\cdot\right)$, $f\left(\cdot,\cdot,\cdot\right) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, be a function that is sensitive to simplicity considerations in the strong sense. Then there is at least one power searcher $\mathcal{LO}$ such that it holds for every theory T, every evidence E, every background knowledge B, and every wff $h \in T$:

> If h is a $\mathcal{LO}$-superfluous part of T for E and B, then $f\left(T \setminus \{h\}, E, B\right) \geq f\left(T, E, B\right)$,

i.e. which is such that it holds for every theory T, every evidence E, every background knowledge B, and every wff $h \in T$:

> If $\mathcal{LO}\left(T \setminus \{h\}, E, B\right) = \mathcal{LO}\left(T, E, B\right)$, then $f\left(T \setminus \{h\}, E, B\right) \geq f\left(T, E, B\right)$.

Let T be a theory, and suppose $h \in T$ is a $\mathcal{LO}$-superfluous part of T. Then it holds for every evidence E, and every background knowledge B:

$$\mathcal{LO}\left(T \setminus \{h\}, E, B\right) = \mathcal{LO}\left(T, E, B\right).$$

From the above it follows for every evidence E, and every background knowledge B:

$$f\left(T \setminus \{h\}, E, B\right) \geq f\left(T, E, B\right).$$

So there is at least one power searcher $\mathcal{LO}$ such that it holds for every theory T, and every wff $h \in T$:

> If $\mathcal{LO}(T \setminus \{h\}, E, B) = \mathcal{LO}(T, E, B)$, for every evidence E, and every background knowledge B, then $f(T \setminus \{h\}, E, B) \geq f(T, E, B)$, for every evidence E, and every background knowledge B,

which just means that f is sensitive to simplicity considerations in the weak sense.

(3) Let $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \to \Re$, be a function that is sensitive to simplicity considerations in the weak sense. Then there is at least one power searcher $\mathcal{LO}$ such that it holds for every theory T, every background knowledge B, and every wff $h \in T$:

> If h is a $\mathcal{LO}$-superfluous part of T, then $f(T \setminus \{h\}, E, B) \geq f(T, E, B)$,

i.e. which is such that it holds for every theory T, and every wff $h \in T$:

> If $\mathcal{LO}(T \setminus \{h\}, E, B) = \mathcal{LO}(T, E, B)$, for every evidence E, and every background knowledge B, then $f(T \setminus \{h\}, E, B) \geq f(T, E, B)$, for every evidence E, and every background knowledge B.

Let T be a theory, and let $h \in T$ be a redundant part of T. Then $T \setminus \{h\} \dashv\vdash T$, whence

$$\mathcal{LO}(T \setminus \{h\}, E, B) = \mathcal{LO}(T, E, B),$$

for every evidence E, and every background knowledge B, because $\mathcal{LO}$ is closed under equivalence transformations of T. As f is sensitive to simplicity considerations i.w.s.,

$$f(T \setminus \{h\}, E, B) \geq f(T, E, B),$$

for every evidence E, and every background knowledge B, which just means that f is sensitive to redundancy considerations.

Please note that if '$\geq$' is replaced by '$>$' in the definitions of sensitivity to simplicity considerations in any sense and unimpressability by redundancies, then theorem 3.5 still holds. The proof is obtained by substituting '$>$' for all occurrences of '$\geq$' in this proof. $\square$

C.4 Proof of Theorem 3.4

Theorem C.4 (SensSimplCons i.s.s. Does Not Imply InvEquTrans) Let $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \to \Re$, be a function. If f is sensitive to simplicity considerations in the strong sense, then f need not be closed under equivalence transformations of T in the sense that

$$f(T, E, B) = f(T', E, B), \quad \text{if} \quad T \dashv\vdash T',$$

for any theories T and T', every evidence E, and every background knowledge B.

Proof.
It suffices to give an example of a function $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \to \Re$, which is sensitive to simplicity considerations in the strong sense, but not closed under equivalence transformations of T in the sense that

$$f(T,E,B) = f(T',E,B), \quad \text{if} \quad T \dashv\vdash T',$$

for any theories T and T', every evidence E, and every background knowledge B. The following one is a case in point:

$$f(T,E,B) := \mathcal{LO}(T,E,B) \cdot \frac{|T|+2}{|T|+1},$$

for every theory T, every evidence E, and every background knowledge B, where $\mathcal{LO}$ is a power searcher (theorem 3.1 guarantuees that there are such).

Obviously f is not closed under equivalence transformations. That f is sensitive to simplicity considerations i.s.s. is seen as follows. Let T be a theory, let E be an evidence, let B be a background knowledge, and suppose $h \in T$ is a $\mathcal{LO}$-superfluous part of T for E and B. Then

$$\mathcal{LO}(T \setminus \{h\}, E, B) = \mathcal{LO}(T,E,B),$$

whence

$$\begin{aligned} f(T \setminus \{h\}, E, B) &= \mathcal{LO}(T \setminus \{h\}, E, B) \cdot \frac{|T \setminus \{h\}|+2}{|T \setminus \{h\}|+1} \\ &> \mathcal{LO}(T,E,B) \cdot \frac{|T|+2}{|T|+1} \\ &= f(T,E,B), \end{aligned}$$

which just means that f is (even strictly) sensitive to simplicity considerations in the strong sense. $\square$

C.5 Proof of Theorem 3.5

Theorem C.5 (InvEquTrans Implies SensSimplCons i.w.s.) If f is closed under equivalence transformations of T, then f is sensitive to simplicity considerations in the weak sense.

Proof.
Let $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, be a function which is closed under equivalence transformations of T. That f is sensitive to simplicity considerations in the weak sense means that there is at least one power searcher $\mathcal{LO}$ such that it holds for every theory T, and every wff $h \in T$:

> If $\mathcal{LO}(T \setminus \{h\}, E, B) = \mathcal{LO}(T, E, B)$, for every evidence E, and every background knowledge B, then $f(T \setminus \{h\}, E, B) \geq f(T, E, B)$, for every evidence E, and every background knowledge B.

Let such a function $\mathcal{LO}$ be defined as follows:

$$\mathcal{LO}(T, E, B) = \begin{cases} 1, & \text{if} \quad T \cup B \vdash E, \\ 0 & \text{otherwise, i.e. if} \quad T \cup B \nvdash E, \end{cases}$$

for every theory T, every evidence E, and every background knowledge B.

That $\mathcal{LO}$ is a power searcher is seen as follows. Let T and T' be theories, let E be an evidence, and let B and B' be background knowledges. Obviously, $\mathcal{LO}(T, E, B) \geq 0$, and $\mathcal{LO}(T, E, B) = 1$, if $T \cup B \vdash E$.

If $T' \cup B' \vdash T \cup B$ and $T \cup B \nvdash E$, then

$$\mathcal{LO}(T', E, B') \geq \mathcal{LO}(T, E, B) = 0,$$

and if $T' \cup B' \vdash T \cup B$ and $T \cup B \vdash E$, then

$$\mathcal{LO}(T', E, B') = \mathcal{LO}(T, E, B) = 1.$$

Let me now show for every theory T, every function $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, which is closed under equivalence transformations of T, and every wff $h \in T$:

> If $\mathcal{LO}(T \setminus \{h\}, E, B) = \mathcal{LO}(T, E, B)$, for every evidence E, and every background knowledge B, then $f(T \setminus \{h\}, E, B) \geq f(T, E, B)$, for every evidence E, and every background knowledge B.

Let T be a theory, let f be a function which is closed under equivalence transformations of T, and let h be a wff of T. Suppose $\mathcal{LO}(T \setminus \{h\}, E, B) = \mathcal{LO}(T, E, B)$, for every evidence E, and every background knowledge B.

One has to show that $f(T \setminus \{h\}, E, B) \geq f(T, E, B)$, for every evidence E, and every background knowledge B.

It suffices to show that $T \setminus \{h\} \vdash T$, for then $f(T \setminus \{h\}, E, B) = f(T, E, B)$, for every evidence E, and every background knowledge B, because f is closed under equivalence transformations of T. Suppose that $T \setminus \{h\} \nvdash T$.

Let 'P' be an n-ary predicate which does not occur in (any wff of) T, and let 'a_1', ..., 'a_n' be n individual constants not occurring in (any wff of) T. Then

$$T \vdash P(a_1, \ldots, a_n), \quad \text{only if} \quad T \vdash \bot.$$

As $T \setminus \{h\} \nvdash h$, $T \setminus \{h\} \nvdash \bot$, whence

$$T \setminus \{h\} \nvdash P(a_1, \ldots, a_n).$$

It follows that

$$\mathcal{LO}(T \setminus \{h\}, \{P(a_1, \ldots, a_n)\}, \{h \to P(a_1, \ldots, a_n)\}) = 0, \quad \text{and}$$
$$\mathcal{LO}(T, \{P(a_1, \ldots, a_n)\}, \{h \to P(a_1, \ldots, a_n)\}) = 1,$$

whence there is an evidence E and a background knowledge B such that

$$\mathcal{LO}(T \setminus \{h\}, E, B) \neq \mathcal{LO}(T, E, B).$$

– a contradiction. So there is at least one power searcher $\mathcal{LO}$ such that it holds for every theory T, every function $f(\cdot, \cdot, \cdot)$, $f(\cdot, \cdot, \cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \to \Re$, which is closed under equivalence transformations of T, and every wff $h \in T$: If

> $\mathcal{LO}(T \setminus \{h\}, E, B) = \mathcal{LO}(T, E, B)$, for every evidence E, and every background knowledge B,

then $T \setminus \{h\} \dashv\vdash T$, and thus

> $f(T \setminus \{h\}, E, B) = f(T, E, B)$, for every evidence E, and every background knowledge B,

which means (even something stronger than) that f is sensitive to simplicity considerations in the weak sense, if f is closed under equivalence transformations of T. □

C.6 Proof of Theorem 3.6

Theorem C.6 (InvEquTrans Does Not Imply SensSimplCons i.s.s.) If f is closed under equivalence transformations of T, then f need not be sensitive to simplicity considerations in the strong sense.

Proof.
It suffices to give an example of a function $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot):\mathcal{T}\times\mathcal{E}\times\mathcal{B}\rightarrow\Re$, which is closed under equivalence transformations of T, and which is not sensitive to simplicity considerations in the strong sense. That f is not sensitive to simplicity considerations in the strong sense means that there is no power searcher $\mathcal{LO}$ such that it holds for every theory T, every evidence E, every background knowledge B, and every wff $h \in T$:

> If $\mathcal{LO}(T\setminus\{h\},E,B) = \mathcal{LO}(T,E,B)$, then $f(T\setminus\{h\},E,B) \geq f(T,E,B)$.

In other words, one has to show that there is at least one function $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot):\mathcal{T}\times\mathcal{E}\times\mathcal{B}\rightarrow\Re$, which is closed under equivalence transformations of T, and such that for every power searcher $\mathcal{LO}$ there are theories T, evidences E, background knowledges B, and wffs $h \in T$ with:

$$\begin{aligned}&\mathcal{LO}(T\setminus\{h\},E,B) = \mathcal{LO}(T,E,B), \quad \text{and}\\ &f(T\setminus\{h\},E,B) < f(T,E,B).\end{aligned}$$

A consequence of the latter is that $T\setminus\{h\} \nvdash T$, because f is closed under equivalence transformations of T.

The following one is a case in point:

$$f(T,E,B) = \begin{cases} 1, & \text{if} \quad T \vdash B\cup E, \\ 0 & \text{otherwise, i.e. if} \quad T \nvdash B\cup E,\end{cases}$$

for every theory T, every evidence E, and every background knowledge B.

Obviously, f is closed under equivalence transformations of T. Let $\mathcal{LO}$ be a power searcher, let $E = \{Ga_1,\ldots,Ga_n\}$, for some $n \geq 1$, let $T = \{\forall x(Fx\rightarrow Gx), \forall xFx\}$, let $B = \{\forall x(Fx\rightarrow Gx)\}$, and let $h = \forall x(Fx\rightarrow Gx)$. As $T\setminus\{h\} \nvdash E\cup B$ and $T \vdash E\cup B$,

$$f(T\setminus\{h\},E,B) = 0 < 1 = f(T,E,B).$$

Furthermore, $T\setminus\{h\}\cup B \vdash E$ and $T\cup B \vdash E$, whence

$$\mathcal{LO}(T\setminus\{h\},E,B) = \mathcal{LO}(T,E,B) = 1.$$

So there is at least one function $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot):\mathcal{T}\times\mathcal{E}\times\mathcal{B}\rightarrow\Re$, which is closed under equivalence transformations of T, and for which there are theories

T, evidences E, background knowledges B, and wffs $h \in T$ such that it holds for every power searcher $\mathcal{LO}$:

$$\begin{aligned}&\mathcal{LO}\left(T \setminus \{h\}, E, B\right) = \mathcal{LO}\left(T, E, B\right), \quad \text{and}\\&f\left(T \setminus \{h\}, E, B\right) < f\left(T, E, B\right),\end{aligned}$$

which means (even something stronger than) that f is closed under equivalence transformations of T, but not sensitive to simplicity considerations in the strong sense. □

C.7 Proof of Theorem 3.7

Theorem C.7 (SensLoveLike Implies SensSimplCons i.v.s.s.) Let $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, be a function. If f is sensitive to loveliness and likeliness in the sense of some power searcher $\mathcal{LO}$ and some truth indicator $\mathcal{LI}$, then f is sensitive to simplicity considerations in the very strong sense.

Proof.
Let $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, be a function which is sensitive to loveliness and likeliness in the sense of some power searcher $\mathcal{LO}^*$ and some truth indicator $\mathcal{LI}^*$. Then it holds for any theories T and T', any evidences E and E', and any background knowledges B and B', where $X = T, E, B$ and $X' = T', E', B'$:

1. If $\mathcal{LI}^*(X) = \mathcal{LI}^*(X') \neq 0$, then $f(X) \geq f(X')$ iff $\mathcal{LO}^*(X) \geq \mathcal{LO}^*(X')$,
2. if $\mathcal{LO}^*(X) = \mathcal{LO}^*(X') \neq 0$, then $f(X) \geq f(X')$ iff $\mathcal{LI}^*(X) \geq \mathcal{LI}^*(X')$,
3. $f(X) = 0$ iff $\mathcal{LI}^*(X) = 0$ or $\mathcal{LO}^*(X) = 0$, and
4. $f(X) = 1$ iff $\mathcal{LI}^*(X) = 1$ and $\mathcal{LO}^*(X') = 1$.

It has to be shown that there is at least one power searcher $\mathcal{LO}$ such that it holds for any theories T and T', every evidence E, and every background knowledge B:

> If $T' \vdash T$ and $\mathcal{LO}(T, E, B) = \mathcal{LO}(T', E, B)$, then $f(T, E, B) \geq f(T', E, B)$.

I show that $\mathcal{LO}^*$ is such a function.

Let T and T' be theories, let E be an evidence, let B be a background knowledge, and suppose $T' \vdash T$. Then $\mathcal{LI}^*(T, E, B) \geq \mathcal{LI}^*(T', E, B)$. Suppose

$$\mathcal{LO}^*(T, E, B) = \mathcal{LO}^*(T', E, B) \neq 0.$$

As f is sensitive to loveliness and likeliness in the sense of $\mathcal{LO}^*$ and $\mathcal{LI}^*$, it follows that

$$f(T, E, B) \geq f(T', E, B) \quad \text{iff} \quad \mathcal{LI}^*(T, E, B) \geq \mathcal{LI}^*(T', E', B'),$$

whence $f(T, E, B) \geq f(T', E, B)$.

Suppose

$$\mathcal{LO}^*(T, E, B) = \mathcal{LO}^*(T', E, B) = 0.$$

As f is sensitive to loveliness and likeliness in the sense of $\mathcal{LO}^*$ and $\mathcal{LI}^*$, it follows that

$$f(T, E, B) = f(T', E, B) = 0,$$

whence again $f(T, E, B) \geq f(T', E, B)$.

So there is at least one power searcher $\mathcal{LO}$ such that it holds for any theories T and T', every evidence E, and every background knowledge B:

> If $T' \vdash T$ and $\mathcal{LO}(T, E, B) = \mathcal{LO}(T', E, B)$, then $f(T, E, B) \geq f(T', E, B)$,

which just means that f is sensitive to simplicity considerations in the very strong sense. □

Appendix D

Proofs for Chapter 4

D.1 Proof of Theorem 4.1

Theorem D.1 ($\succeq$ Is Arbitrary)) The ordinal measure of coherence $\succeq$ of Hartmann/Bovens (2000) is arbitrary.

Proof.
The ordinal measure of coherence $\succeq$ of Hartmann/Bovens (2000) is defined as follows:

> For any two information sets S, S':
> S is more coherent than or equally coherent as $\mathrm{S}', \mathrm{S} \succeq \mathrm{S}'$, iff $f_x(\mathrm{S}, \mathrm{S}') \geq 0$, for all values of $x \in (0, 1)$.

An information set S is a set of finitely many propositions R_1, …, R_n. The function f_x is defined for pairs of information sets S, S' in the following way:

$$f_x(\mathrm{S}, \mathrm{S}') = c_x(\mathrm{S}) - c_x(\mathrm{S}').$$

c_x measures the impact of the coherence of an information set $\mathrm{S} = \{\mathrm{R}_1, \ldots, \mathrm{R}_\mathrm{n}\}$ on the degree of confidence in S,

$$P^*(\mathrm{R}_1, \ldots, \mathrm{R}_\mathrm{n}) = P(\mathrm{R}_1, \ldots, \mathrm{R}_\mathrm{n} \mid \mathrm{Repr}_1, \ldots, \mathrm{Repr}_\mathrm{n}),$$

and is defined as follows:

$$\begin{aligned} c_x(S) &= c_x(\mathrm{R}_1, \ldots, \mathrm{R}_\mathrm{n}) \\ &= P^*(\mathrm{R}_1, \ldots, \mathrm{R}_\mathrm{n}) / P^{*max}(\mathrm{R}_1, \ldots, \mathrm{R}_\mathrm{n}) \\ &= \frac{a_0 + (1 - a_0) \cdot x^n}{\sum_{i=0}^{n} a_i \cdot x^i}. \end{aligned}$$

$\mathrm{Repr_i}$ is the proposition that after consultation with the proper source, there is a report to the effect that $\mathrm{R_i}$ is the case, and P is a joint probability for the propositional variables $R_1, \ldots, R_n$, $Repr_1$, ..., $Repr_n$. The information sources are assumed to be independent in the sense that the variable $Repr_i$ is probabilistically independent (under P) of the variables $R_1, Repr_1$, ..., $R_{i-1}, Repr_{i-1}$, $R_{i+1}, Repr_{i+1}, \ldots R_n, Repr_n$ given R_i, for every $i, 1 \leq i \leq n$. The propositional variable R_i can take on the two values $\mathrm{R_i}$ and $\overline{\mathrm{R_i}}$, i.e. $\mathrm{not\text{-}R_i}$, and the propositional variable $Repr_i$ can take on the two values $\mathrm{Repr_i}$ and $\overline{\mathrm{Repr_i}}$; the latter saying that after consultation with the proper source, there is no report to the effect that $\mathrm{R_i}$ is the case. Note that for the ordinal measure of coherence $\succeq$ it suffices that P is defined over $R_1, \ldots, R_n$.

Furthermore, a_i is the sum of the joint probabilities of all combinations of the variables $R_1, \ldots, R_n$ that have i negative values and $n - i$ positive values, i.e.

$$a_i = \sum_{neg(\pm\mathrm{R_1}, \ldots, \pm\mathrm{R_n}) = i} P\left(\pm\mathrm{R_1}, \ldots, \pm\mathrm{R_n}\right),$$

where

$$neg\left(\pm\mathrm{R_1}, \ldots, \pm\mathrm{R_n}\right) := \left|\left\{\overline{\mathrm{R_j}} : 1 \leq \mathrm{j} \leq \mathrm{n}\right\}\right|,$$

and $\pm\mathrm{R_j}$ is either $\mathrm{R_j}$ or $\overline{\mathrm{R_j}}$, for every $\mathrm{j}, 1 \leq \mathrm{j} \leq \mathrm{n}$, whence in particular

$$a_0 = P\left(\mathrm{R_1}, \ldots, \mathrm{R_n}\right).$$

The variable

$$x = q/p = P\left(\mathrm{Repr_i} \mid \overline{\mathrm{R_i}}\right) / P\left(\mathrm{Repr_i} \mid \mathrm{R_i}\right),$$

is assumed to be equal for every $\mathrm{i}, 1 \leq \mathrm{i} \leq \mathrm{n}$, and expresses the reliability of the information source i which reports by means of $\mathrm{Repr_i}$ that $\mathrm{R_i}$ is the case.

Finally, for a given probability distribution P, the probability distribution P^{max} is defined as

$$P^{max}\left(\mathrm{R_i}\right) = P\left(\bigcap_{\mathrm{i}=0}^{\mathrm{n}} \mathrm{R_i}\right) = P\left(\mathrm{R_1}, \ldots, \mathrm{R_n}\right),$$

and

$$P^{max}\left(\mathrm{R_i} \mid \mathrm{R_j}\right) = 1, \quad \text{for every i and j}, 1 \leq \mathrm{i}, \mathrm{j} \leq \mathrm{n}.$$

In order to prove that the ordinal measure of coherence $\succeq$ of Hartmann/Bovens (2000) is arbitrary, I have to rely on the strict ordinal measure of coherence $\succ$ which is induced by $\succeq$ on the set of all information sets, and which is defined as follows:

For any two information sets S, S':
S is more coherent than $\mathrm{S}', \mathrm{S} \succ \mathrm{S}'$, iff $S \succeq S'$ and $S' \not\succeq S$,

i.e.

For any two information sets S, S':
S is more coherent than $\mathrm{S}', \mathrm{S} \succ \mathrm{S}'$, iff $f_x(\mathrm{S}, \mathrm{S}') \geq 0$, for all values of $x \in (0,1)$, and $f_x(\mathrm{S}', \mathrm{S}) < 0$, for at least one value of $x \in (0,1)$.

When proving $\mathrm{S} \succ \mathrm{S}'$, for some information sets S and S', it will be shown that

$$f_x(\mathrm{S}, \mathrm{S}') > 0, \text{ for all values of } x \in (0,1),$$

which is something stronger than $\mathrm{S} \succ \mathrm{S}'$.

It suffices to give an example of two information sets $\mathrm{S} = \{\mathrm{R}_1, \ldots, \mathrm{R}_\mathrm{m}\}$ and $\mathrm{S}' = \{\mathrm{R}_1', \ldots, \mathrm{R}_\mathrm{n}'\}$, two probability distributions P_1 and P_2 over $R_1, \ldots, R_m$, and two probability distributions P_1' and P_2' over $R_1', \ldots, R_m'$ such that $\mathrm{S}' \succ \mathrm{S}$ according to P_1 and P_1', and $\mathrm{S} \succ \mathrm{S}'$ according to P_2 and P_2'. The following example does the job. Let

$$\begin{aligned}
&\mathrm{S} = \{\mathrm{R}_1, \mathrm{R}_2\}, \mathrm{S}' = \{\mathrm{R}_1', \mathrm{R}_2'\},\\
&P_1(\mathrm{R}_1, \mathrm{R}_2) = P_1\left(\overline{\mathrm{R}_1}, \mathrm{R}_2\right) = P_1\left(\mathrm{R}_1, \overline{\mathrm{R}_2}\right) = P_1\left(\overline{\mathrm{R}_1}, \overline{\mathrm{R}_2}\right) = 0.25,\\
&P_1'(\mathrm{R}_1', \mathrm{R}_2') = 0.125, P_1'\left(\overline{\mathrm{R}_1'}, \mathrm{R}_2'\right) = P_1'\left(\mathrm{R}_1', \overline{\mathrm{R}_2'}\right) = 0.25, \text{ and}\\
&P_1'\left(\overline{\mathrm{R}_1'}, \overline{\mathrm{R}_2'}\right) = 0.375.
\end{aligned}$$

Then

$$\begin{aligned}
a_{1_0} &= P_1(\mathrm{R}_1, \mathrm{R}_2) = 0.25,\\
a_{1_1} &= P_1\left(\overline{\mathrm{R}_1}, \mathrm{R}_2\right) + P_1\left(\mathrm{R}_1, \overline{\mathrm{R}_2}\right) = 0.25 + 0.25 = 0.5,\\
a_{1_2} &= P_1\left(\overline{\mathrm{R}_1}, \overline{\mathrm{R}_2}\right) = 0.25,\\
a_{1_0}' &= P_1'(\mathrm{R}_1, \mathrm{R}_2) = 0.125,\\
a_{1_1}' &= P_1'\left(\overline{\mathrm{R}_1}, \mathrm{R}_2\right) + P_1'\left(\mathrm{R}_1, \overline{\mathrm{R}_2}\right) = 0.25 + 0.25 = 0.5, \text{ and}\\
a_{1_2}' &= P_1'\left(\overline{\mathrm{R}_1}, \overline{\mathrm{R}_2}\right) = 0.375.
\end{aligned}$$

It follows that

$$\begin{aligned}
f_{1_x}(\mathrm{S}, \mathrm{S}') &= c_{1_x}(\mathrm{S}) - c_{1_x}(\mathrm{S}')\\
&= \frac{a_{1_0} + (1 - a_{1_0}) \cdot x^2}{\sum_{i=0}^{2} a_{1_i} \cdot x^i} -
\end{aligned}$$

$$
\begin{aligned}
& -\frac{a'_{1_0}+\left(1-a'_{1_0}\right)\cdot x^2}{\sum_{i=0}^{2} a'_{1_i}\cdot x^i} \\
= & \quad \frac{0.25+(1-0.25)\cdot x^2}{0.25\cdot x^0+0.5\cdot x^1+0.25\cdot x^2} \\
& -\frac{0.125+(1-0.125)\cdot x^2}{0.125\cdot x^0+0.5\cdot x^1+0.375\cdot x^2},
\end{aligned}
$$

whence

$$
\begin{aligned}
f_{1_x}\left(\mathrm{S},\mathrm{S}'\right) &> 0 \\
&\text{iff} \\
\frac{0.25+0.75\cdot x^2}{0.25+0.5\cdot x+0.25\cdot x^2} &> \frac{0.125+0.875\cdot x^2}{0.125+0.5\cdot x+0.375\cdot x^2} \\
&\text{iff} \\
0.25\cdot 0.125+0.25\cdot 0.5\cdot x+ & \\
+(0.25\cdot 0.375+0.75\cdot 0.125)\cdot x^2+ & \\
+0.75\cdot 0.5\cdot x^3+0.75\cdot 0.375\cdot x^4 &> 0.125\cdot 0.25+0.125\cdot 0.5\cdot x+ \\
& \quad +(0.125\cdot 0.25+0.875\cdot 0.25)\cdot x^2+ \\
& \quad +0.875\cdot 0.5\cdot x^3+0.875\cdot 0.25\cdot x^4 \\
&\text{iff} \\
2/32\cdot x-2/32\cdot x^2 &> 2/32\cdot x^3-2/32\cdot x^4 \\
&\text{iff} \\
x\cdot(1-x) &> x^3\cdot(1-x),
\end{aligned}
$$

which holds for all values of $x\in(0,1)$. Thus $\mathrm{S}\succ_1\mathrm{S}'$. Let

$$
\begin{aligned}
& P_2\left(\mathrm{R}_1,\mathrm{R}_2\right)=P_2\left(\overline{\mathrm{R}_1},\mathrm{R}_2\right)=P_2\left(\mathrm{R}_1,\overline{\mathrm{R}_2}\right)=P_2\left(\overline{\mathrm{R}_1},\overline{\mathrm{R}_2}\right)=0.25, \\
& P'_2\left(\mathrm{R}'_1,\mathrm{R}'_2\right)=0.375, P'_2\left(\overline{\mathrm{R}'_1},\mathrm{R}'_2\right)=P'_2\left(\mathrm{R}'_1,\overline{\mathrm{R}'_2}\right)=0.25, \text{ and} \\
& P'_2\left(\overline{\mathrm{R}'_1},\overline{\mathrm{R}'_2}\right)=0.125.
\end{aligned}
$$

Then

$$
\begin{aligned}
a_{2_0} &= P_2\left(\mathrm{R}_1,\mathrm{R}_2\right)=0.25, \\
a_{2_1} &= P_2\left(\overline{\mathrm{R}_1},\mathrm{R}_2\right)+P_2\left(\mathrm{R}_1,\overline{\mathrm{R}_2}\right)=0.25+0.25=0.5, \\
a_{2_2} &= P_2\left(\overline{\mathrm{R}_1},\overline{\mathrm{R}_2}\right)=0.25, \\
a'_{2_0} &= P'_2\left(\mathrm{R}_1,\mathrm{R}_2\right)=0.375, \\
a'_{2_1} &= P'_2\left(\overline{\mathrm{R}_1},\mathrm{R}_2\right)+P'_2\left(\mathrm{R}_1,\overline{\mathrm{R}_2}\right)=0.25+0.25=0.5, \text{ and} \\
a'_{2_2} &= P'_2\left(\overline{\mathrm{R}_1},\overline{\mathrm{R}_2}\right)=0.125.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
f_{2_x}(\mathrm{S}', \mathrm{S}) &= c_{2_x}(\mathrm{S}') - c_{2_x}(\mathrm{S}) \\
&= \frac{a'_{2_0} + \left(1 - a'_{2_0}\right) \cdot x^2}{\sum_{i=0}^{2} a'_{2_i} \cdot x^i} - \\
&\quad - \frac{a_{2_0} + (1 - a_{2_0}) \cdot x^2}{\sum_{i=0}^{2} a_{2_i} \cdot x^i} \\
&= \frac{0.375 + (1 - 0.375) \cdot x^2}{0.375 \cdot x^0 + 0.5 \cdot x^1 + 0.125 \cdot x^2} - \\
&\quad - \frac{0.25 + (1 - 0.25) \cdot x^2}{0.25 \cdot x^0 + 0.5 \cdot x^1 + 0.25 \cdot x^2},
\end{aligned}
$$

whence

$$
\begin{aligned}
f_{2_x}(\mathrm{S}', \mathrm{S}) &> 0 \\
&\text{iff} \\
\frac{0.375 + 0.625 \cdot x^2}{0.375 + 0.5 \cdot x + 0.125 \cdot x^2} &> \frac{0.25 + 0.75 \cdot x^2}{0.25 + 0.5 \cdot x + 0.25 \cdot x^2} \\
&\text{iff} \\
0.375 \cdot 0.25 + 0.375 \cdot 0.5 \cdot x + & \\
+ (0.375 \cdot 0.25 + 0.625 \cdot 0.25) \cdot x^2 + & \\
+ 0.625 \cdot 0.5 \cdot x^3 + 0.625 \cdot 0.25 \cdot x^4 &> 0.25 \cdot 0.375 + 0.25 \cdot 0.5 \cdot x + \\
&\quad + (0.25 \cdot 0.125 + 0.75 \cdot 0.375) \cdot x^2 + \\
&\quad + 0.75 \cdot 0.5 \cdot x^3 + 0.75 \cdot 0.125 \cdot x^4 \\
&\text{iff} \\
2/32 \cdot x - 2/32 \cdot x^2 &> 2/32 \cdot x^3 - 2/32 \cdot x^4 \\
&\text{iff} \\
x \cdot (1 - x) &> x^3 \cdot (1 - x),
\end{aligned}
$$

which holds for all values of $x \in (0, 1)$. Thus $\mathrm{S}' \succ_2 \mathrm{S}$.

Put together, these two results yield that $\mathrm{S} \succ_1 \mathrm{S}'$ and $\mathrm{S}' \succ_2 \mathrm{S}$, which just means that the strict ordinal measure of coherence $\succ$, which is induced by $\succeq$, is arbitrary. $\square$

D.2 Proof of Theorem 4.2

Theorem D.2 ($ECHO$ Is Arbitrary) The computer program $ECHO$, which models the theory of explanatory coherence TEC of Thagard (1989), is arbitrary.

Proof.
The definition of the measure $H(S,t)$ of the global coherence of a system S of n propositions at time t runs as follows:

$$H(S,t) = \sum_{0 \leq i \leq n} \sum_{0 \leq j \leq n} w_{ij} \cdot a_i(t) \cdot a_j(t).$$

w_{ij} is the weight of the excitatory or inhibitory link from unit i to unit j, $a_i(t)$ is the activation of unit i at time t, and n is the number of propositions in the system S which are represented by the units $1, \ldots, n$.

An excitatory link between two units i and j represents a coherence relation between the two propositions the units i and j stand for, whereas an inhibitory link represents an incoherence relation. The activation $a_i(t)$ of unit i at time t expresses the degree of acceptance of the proposition represented by unit i at time t.

An input in form of (values for the) activations $a_i(0)$ of the units i of some system of propositions S at time 0 is used to set up a network which includes – besides the units $1, \ldots, n$ – a special unit 0 with activation $a_0(t) = 1$, for every time t. Then the network is run in cycles that synchronously update all the units so that the activation streams from the special unit 0 over units representing data (evidences) to units representing hypotheses which are explanatorily linked to these data.

The activation $a_i(\cdot)$, $a_i(\cdot) : \mathcal{N} \rightarrow [-1,1]$, of any unit i is a continuous function of all units j linked to it. The contribution of each such unit j depends on the weight w_{ij} of the link from i to j. These weights w_{ij} – expressing the strength of the (in)coherence relation between the propositions P_k and Q, which are represented by the units i and j, respectively – have to obey the equation

$$weight(P_k, Q) = \frac{\textit{default weight}}{(\textit{number of cohypotheses of } P_k)^{\textit{simplicity impact}}},$$

where Q is explained by $P_1, \ldots, P_k, \ldots, P_m, 1 \leq k \leq m$.

Despite this, the w_{ij} can be chosen in an arbitrary way, for both the default weight and the simplicity impact can be freely chosen. The number of cohypotheses of proposition P_k is the number $m-1$ of propositions that occur in the explanation of Q by $P_1, \ldots, P_k, \ldots, P_m$ apart from P_k. The activation $a_i(\cdot)$ of unit i is updated by the following equation:

$$a_i(t+1) = \begin{cases} a_i(t) \cdot (1-\theta) + net_i(t+1) \cdot (\max - a_i(t)), \\ \quad \text{if } net_i(t+1) > 0, \\ a_i(t) \cdot (1-\theta) + net_i(t+1) \cdot (a_i(t) - \min), \\ \quad \text{if } net_i(t+1) \leq 0. \end{cases}$$

θ is a decay parameter decrementing each unit i at every cycle, $\min = -1$ is the minimum activation, $\max = 1$ is the maximum activation, and $net_i(t+1)$ is the net input to unit i at time $t+1$, which is given as

$$net_i(t+1) \quad = \sum_{0 \leq j \leq n} w_{ij} \cdot a_j(t),$$

where n is again the number of propositions (respectively units without the special unit 0) in the system of propositions S. By repeating updating cycles some units get activated, whereas others get deactivated.

In order to prove the above claim it suffices to give an example of two sets of propositions S_1 and S_2, and two measures $H(\cdot,\cdot)$ and $H'(\cdot,\cdot)$ of the global coherence of a system of propositions for which there is a time t' such that it holds for every time $t \geq t'$:

$$H(S_1, t) > H(S_2, t) \quad \text{and} \quad H'(S_1, t) < H'(S_2, t).$$

Let $S_1 = \{E_1, P_2, P_3\}$, where evidence E_1 is supposed to be explained by each of the two hypotheses P_2 and P_3. E_1 is represented by 1, P_2 by 2, and P_3 by 3; the special unit with activation 1 is represented by 0. Let

$$w_{10} = w_{01} = 1, \quad w_{12} = w_{21} = 1 = w_{13} = w_{31}, \quad \theta = 0.$$

Thus the strength of the explanatory relation between P_2 and E_1 is assumed to be equal to the strength of the explanatory relation between P_3 and E_1; and both are supposed to be equal to the degree of acceptance which E_1 has *qua* being an evidence.

$$a_1(1) = a_1(0) \cdot (1-\theta) + net_1(1) \cdot (\max - a_1(0)) = 0 \cdot (1-0) + 1 \cdot (1-0) = 1,$$

for

$$net_1(1) = w_{10} \cdot a_0(0) + w_{12} \cdot a_2(0) + w_{13} \cdot a_3(0) = 1 \cdot 1 + 1 \cdot 0 + 1 \cdot 0 = 1 > 0.$$

$a_2(1) = a_2(0) \cdot (1-\theta) + net_2(1) \cdot (a_2(0) - \min) = 0 \cdot (1-0) + 0 \cdot (0-(-1)) = 0,$ for

$$net_2(1) = w_{21} \cdot a_1(0) = 1 \cdot 0 = 0 \leq 0.$$

$a_3(1) = a_3(0) \cdot (1-\theta) + net_3(1) \cdot (a_3(0) - \min) = 0 \cdot (1-0) + 0 \cdot (0-(-1)) = 0,$ for

$$net_3(1) = w_{31} \cdot a_1(0) = 1 \cdot 0 = 0 \leq 0.$$

$a_1(2) = a_1(1) \cdot (1-\theta) + net_1(2) \cdot (\max - a_1(1)) = 1 \cdot (1-0) + 1 \cdot (1-1) = 1,$ for

$$net_1(2) = w_{10} \cdot a_0(1) + w_{12} \cdot a_2(1) + w_{13} \cdot a_3(1) = 1 \cdot 1 + 1 \cdot 0 + 1 \cdot 0 = 1 > 0.$$

$a_2(2) = a_2(1) \cdot (1-\theta) + net_2(2) \cdot (\max - a_2(1)) = 0 \cdot (1-0) + 1 \cdot (1-0) = 1,$ for

$$net_2(2) = w_{21} \cdot a_1(1) = 1 \cdot 1 = 1 > 0.$$

$a_3(2) = a_3(1) \cdot (1-\theta) + net_3(2) \cdot (\max - a_3(1)) = 0 \cdot (1-0) + 1 \cdot (1-0) = 1,$ for

$$net_3(2) = w_{31} \cdot a_1(1) = 1 \cdot 1 = 1 > 0.$$

So

$$\begin{aligned}
H_1(S_1, 2) &= \sum_{0 \leq i \leq 3} \sum_{0 \leq j \leq 3} w_{ij} \cdot a_i(2) \cdot a_j(2) \\
&= w_{01} \cdot a_0(2) \cdot a_1(2) + w_{10} \cdot a_1(2) \cdot a_0(2) + \\
&\quad + w_{12} \cdot a_1(2) \cdot a_2(2) + w_{13} \cdot a_1(2) \cdot a_3(2) + \\
&\quad + w_{21} \cdot a_2(2) \cdot a_1(2) + w_{31} \cdot a_3(2) \cdot a_1(2) \\
&= 1 \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 1 + \\
&\quad + 1 \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 1 \\
&= 6.
\end{aligned}$$

It will be shown (by induction on time t) that $a_i(t) = 1$, for every i and $t, 1 \leq i \leq 3, t \geq 2$. Let $t \geq 2$, and suppose the induction hypothesis holds.

$$\begin{aligned}
a_1(t+1) &= a_1(t) \cdot (1-\theta) + net_1(t+1) \cdot (\max - a_1(t)) \\
&\qquad net_1(t+1) = 3 > 0 \\
&= 1 \cdot (1-\theta) + net_1(t+1) \cdot (\max - 1) \\
&\qquad \text{by induction hypothesis} \\
&= 1 \cdot (1-0) + 3 \cdot 0 \quad \theta = 0, \quad net_1(t+1) = 3 \\
&= 1,
\end{aligned}$$

for

$$\begin{aligned} net_1(t+1) &= w_{10} \cdot a_0(t) + w_{12} \cdot a_2(t) + w_{13} \cdot a_3(t) \\ &= w_{10} \cdot 1 + w_{12} \cdot 1 + w_{13} \cdot 1 \quad \text{by induction hypothesis} \\ &= 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 = 3 > 0. \end{aligned}$$

$$\begin{aligned} a_2(t+1) &= a_2(t) \cdot (1-\theta) + net_2(t+1) \cdot (\max - a_2(t)) \\ &\qquad net_2(t+1) = 1 > 0 \\ &= 1 \cdot (1-\theta) + net_2(t+1) \cdot (\max - 1) \\ &\qquad \text{by induction hypothesis} \\ &= 1 \cdot (1-0) + 0 \cdot (1-1) \quad \theta = 0, \quad net_2(t+1) = 1 \\ &= 1, \end{aligned}$$

for

$$\begin{aligned} net_2(t+1) &= w_{21} \cdot a_1(t) \\ &= w_{21} \cdot 1 \quad \text{by induction hypothesis} \\ &= 1 \cdot 1 = 1 > 0. \end{aligned}$$

$$\begin{aligned} a_3(t+1) &= a_3(t) \cdot (1-\theta) + net_3(t+1) \cdot (\max - a_3(t)) \\ &\qquad net_3(t+1) = 1 > 0 \\ &= 1 \cdot (1-\theta) + net_3(t+1) \cdot (\max - 1) \\ &\qquad \text{by induction hypothesis} \\ &= 1 \cdot (1-0) + 0 \cdot (1-1) \quad \theta = 0, \quad net_3(t+1) = 1 \\ &= 1, \end{aligned}$$

for

$$\begin{aligned} net_3(t+1) &= w_{31} \cdot a_1(t) \\ &= w_{31} \cdot 1 \quad \text{by induction hypothesis} \\ &= 1 \cdot 1 = 1 > 0. \end{aligned}$$

Thus for every $t \geq 2$:

$$H_1(S_1, t) = \sum_{0 \leq i \leq 3} \sum_{0 \leq j \leq 3} w_{ij} \cdot a_i(t) \cdot a_j(t)$$

$$\begin{aligned}
&= w_{01} \cdot a_0(t) \cdot a_1(t) + w_{10} \cdot a_1(t) \cdot a_0(t) + \\
&\quad + w_{12} \cdot a_1(t) \cdot a_2(t) + w_{13} \cdot a_1(t) \cdot a_3(t) + \\
&\quad + w_{21} \cdot a_2(t) \cdot a_1(t) + w_{31} \cdot a_3(t) \cdot a_1(t) \\
&= 1 \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 1 + \\
&\quad + 1 \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 1 \\
&= 6.
\end{aligned}$$

Let $S_2 = \{E_1, E_2\}$, E_1 and E_2 being evidences. E_1 is represented by 1, E_2 by 2, and the special unit with activation 1 is represented by 0. Let

$$w_{10} = w_{01} = w_{20} = w_{02} = 1, \quad \theta = 0.$$

Thus the degree of acceptance which E_1 has *qua* being an evidence is supposed to be equal to the degree of acceptance which E_2 has *qua* being an evidence.

$a_1(1) = a_1(0) \cdot (1-\theta) + net_1(1) \cdot (\max - a_1(0)) = 0 \cdot (1-0) + 1 \cdot (1-0) = 1,$

for

$$net_1(1) = w_{10} \cdot a_0(0) = 1 \cdot 1 = 1 > 0.$$

$a_2(1) = a_2(0) \cdot (1-\theta) + net_2(1) \cdot (\max - a_2(0)) = 0 \cdot (1-0) + 1 \cdot (1-0) = 1,$

for

$$net_2(1) = w_{20} \cdot a_0(0) = 1 \cdot 1 = 1 > 0.$$

$a_1(2) = a_1(1) \cdot (1-\theta) + net_1(2) \cdot (\max - a_1(1)) = 1 \cdot (1-0) + 1 \cdot (1-1) = 1,$

for

$$net_1(2) = w_{10} \cdot a_0(1) = 1 \cdot 1 = 1 > 0.$$

$a_2(2) = a_2(1) \cdot (1-\theta) + net_2(2) \cdot (\max - a_2(1)) = 1 \cdot (1-0) + 1 \cdot (1-1) = 1,$

for

$$net_2(2) = w_{20} \cdot a_0(1) = 1 \cdot 1 = 1 > 0.$$

So

$$\begin{aligned}
H_1(S_2, 2) &= \sum_{0 \le i \le 2} \sum_{0 \le j \le 2} w_{ij} \cdot a_i(2) \cdot a_j(2) \\
&= w_{01} \cdot a_0(2) \cdot a_1(2) + w_{02} \cdot a_0(2) \cdot a_2(2) + \\
&\quad + w_{10} \cdot a_1(2) \cdot a_0(2) + w_{20} \cdot a_2(2) \cdot a_0(2) \\
&= 1 \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 1 \\
&= 4.
\end{aligned}$$

It will be shown (by induction on time t) that $a_i(t) = 1$, for every i and t, $1 \leq i \leq 2, t \geq 2$. Let $t \geq 2$, and suppose the induction hypothesis holds.

$$\begin{aligned}
a_1(t+1) &= a_1(t) \cdot (1-\theta) + net_1(t+1) \cdot (\max - a_1(t)) \\
&\qquad net_1(t+1) = 1 > 0 \\
&= 1 \cdot (1-\theta) + net_1(1) \cdot (\max - 1) \\
&\qquad \text{by induction hypothesis} \\
&= 1 \cdot (1-0) + 1 \cdot (1-1) \quad \theta = 0, \quad net_1(t+1) = 1 \\
&= 1,
\end{aligned}$$

for

$$net_1(t+1) = w_{10} \cdot a_0(t) = 1 \cdot 1 = 1 > 0.$$

$$\begin{aligned}
a_2(t+1) &= a_2(t) \cdot (1-\theta) + net_2(t+1) \cdot (\max - a_2(t)) \\
&\qquad net_2(t+1) = 1 > 0 \\
&= 1 \cdot (1-\theta) + net_2(t+1) \cdot (\max - 1) \\
&\qquad \text{by induction hypothesis} \\
&= 1 \cdot (1-0) + 1 \cdot (1-1) \quad \theta = 0, \quad net_2(t+1) = 1 \\
&= 1,
\end{aligned}$$

for

$$net_2(t+1) = w_{20} \cdot a_0(t) = 1 \cdot 1 = 1 > 0.$$

So it holds for every $t \geq 2$:

$$\begin{aligned}
H_1(S_2, t) &= \sum_{0 \leq i \leq 2} \sum_{0 \leq j \leq 2} w_{ij} \cdot a_i(t) \cdot a_j(t) \\
&= w_{01} \cdot a_0(t) \cdot a_1(t) + w_{02} \cdot a_0(t) \cdot a_2(t) + \\
&\quad + w_{10} \cdot a_1(t) \cdot a_0(t) + w_{20} \cdot a_2(t) \cdot a_0(t) \\
&= 1 \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 1 \\
&= 4.
\end{aligned}$$

It follows for every $t \geq 2$:

$$H_1(S_1, t) = 6 > 4 = H_1(S_2, t).$$

Consider again $S_1 = \{E_1, P_2, P_3\}$, where evidence E_1 is supposed to be explained by each of P_2 and P_3. This time let

$$w'_{10} = w'_{01} = 1, \quad w'_{12} = w'_{21} = 1/10 = w'_{13} = w'_{31}, \quad \theta' = 0.$$

Thus the strength of the explanatory relation between P_2 and E_1 is again assumed to be equal to the strength of the explanatory between P_3 and E_1; but this time they are both supposed to be smaller than the degree of acceptance which E_1 has *qua* being an evidence.

$$a_1(1) = a_1(0) \cdot (1-\theta) + net'_1(1) \cdot (\max - a_1(0)) = 0 \cdot (1-0) + 1 \cdot (1-0) = 1,$$

for

$$\begin{aligned} net'_1(1) &= w'_{10} \cdot a_0(0) + w12' \cdot a_2(0) + w'_{13} \cdot a_3(0) \\ &= 1 \cdot 1 + (1/10) \cdot 0 + (1/10) \cdot 0 = 1 > 0. \end{aligned}$$

$$a_2(1) = a_2(0) \cdot (1-\theta) + net'_2(1) \cdot (a_2(0) - \min) = 0 \cdot (1-0) + 0 \cdot (0-(-1)) = 0,$$

for

$$net'_2(1) = w'_{21} \cdot a_1(0) = (1/10) \cdot 0 = 0 \leq 0.$$

$$a_3(1) = a_3(0) \cdot (1-\theta) + net'_3(1) \cdot (a_3(0) - \min) = 0 \cdot (1-0) + 0 \cdot (0-(-1)) = 0,$$

for

$$net'_3(1) = w'_{31} \cdot a_1(0) = (1/10) \cdot 0 = 0 \leq 0.$$

$$a_1(2) = a_1(1) \cdot (1-\theta) + net'_1(2) \cdot (\max - a_1(1)) = 1 \cdot (1-0) + 1 \cdot (1-1) = 1,$$

for

$$\begin{aligned} net'_1(2) &= w'_{10} \cdot a_0(1) + w12' \cdot a_2(1) + w'_{13} \cdot a_3(1) \\ &= 1 \cdot 1 + (1/10) \cdot 0 + (1/10) \cdot 0 = 1 > 0. \end{aligned}$$

$$\begin{aligned} a_2(2) &= a_2(1) \cdot (1-\theta) + net'_2(2) \cdot (\max - a_2(1)) \\ &= 0 \cdot (1-0) + (1/10) \cdot (1-0) = 1/10, \end{aligned}$$

for

$$net'_2(2) = w'_{21} \cdot a_1(1) = (1/10) \cdot 1 = 1/10 > 0.$$

$$\begin{aligned} a_3(2) &= a_3(1) \cdot (1-\theta) + net'_3(2) \cdot (\max - a_3(1)) \\ &= 0 \cdot (1-0) + (1/10) \cdot (1-0) = 1/10, \end{aligned}$$

for

$$net'_3(2) = w'_{31} \cdot a_1(1) = (1/10) \cdot 1 = 1/10 > 0.$$

So

$$\begin{aligned}
H'(S_1, 2) &= \sum_{0\le i\le 3}\sum_{0\le j\le 3} w'_{ij}\cdot a_i(2)\cdot a_j(2)\\
&= w'_{01}\cdot a_0(2)\cdot a_1(2) + w'_{10}\cdot a_1(2)\cdot a_0(2) +\\
&\quad +w'_{12}\cdot a_1(2)\cdot a_2(2) + w'_{13}\cdot a_1(2)\cdot a_3(2) +\\
&\quad +w'_{21}\cdot a_2(2)\cdot a_1(2) + w'_{31}\cdot a_3(2)\cdot a_1(2)\\
&= 1\cdot 1\cdot 1 + 1\cdot 1\cdot 1 +\\
&\quad + (1/10)\cdot 1\cdot (1/10) + (1/10)\cdot 1\cdot (1/10)\\
&\quad + (1/10)\cdot (1/10)\cdot 1 + (1/10)\cdot (1/10)\cdot 1\\
&= 2.04.
\end{aligned}$$

As before it holds (by induction on time t) that $a_1(t) = 1$, for every $t \ge 2$. From this one gets for every $t \ge 2$:

$$\begin{aligned}
H'(S_1, t) &= \sum_{0\le i\le 3}\sum_{0\le j\le 3} w'_{ij}\cdot a_i(t)\cdot a_j(t)\\
&= w'_{01}\cdot a_0(t)\cdot a_1(t) + w'_{10}\cdot a_1(t)\cdot a_0(t) +\\
&\quad w'_{12}\cdot a_1(t)\cdot a_2(t) + w'_{13}\cdot a_1(t)\cdot a_3(t) +\\
&\quad +w'_{21}\cdot a_2(t)\cdot a_1(t) + w'_{31}\cdot a_3(t)\cdot a_1(t)\\
&= 1\cdot 1\cdot 1 + 1\cdot 1\cdot 1 +\\
&\quad + (1/10)\cdot 1\cdot a_2(t) + (1/10)\cdot 1\cdot a_3(t) +\\
&\quad + (1/10)\cdot a_2(t)\cdot 1 + (1/10)\cdot a_3(t)\cdot 1\\
&= 1 + 1 + (2/10)\cdot (a_2(t) + a_3(t))\\
&< 4\\
&\text{iff}\\
a_2(t) + a_3(t) &< 10.
\end{aligned}$$

The latter holds for every $t \ge 2$, since $a_i(t) \le 1$, for every i and $t, 2 \le i \le 3, t \ge 2$, because the range of $a_i(\cdot)$ is the closed interval $[-1, 1]$.

Consider again $S_2 = \{E_1, E_2\}$, E_1 and E_2 being evidences. As before the degree of acceptance which E_1 has *qua* being an evidence is supposed to be equal to the degree of acceptance which E_2 has *qua* being an evidence. Again, let

$$w'_{10} = w'_{01} = w'_{20} = w'_{02} = 1, \quad \theta' = 0.$$

As

$$w'_{10} = w_{10}, \quad w'_{01} = w_{01}, \quad w'_{02} = w_{02}, \quad w'_{20} = w_{20}, \quad \theta' = \theta,$$

it follows that

$$H'(S_2, 2) = H(S_2, 2) = 4,$$

and – by the same reasoning (induction) as above – that it holds for every $t \geq 2$:

$$H'(S_2, t) = H(S_2, t) = 4.$$

Thus

$$H'(S_1, t) = 2.04 < 4 = H'(S_2, t), \quad \text{for every } t \geq 2.$$

Put together these results yield that there is at least one t' (any $t' \geq 2$ does the job) such that it holds for every $t \geq t'$:

$$H(S_1, t) > H(S_2, t) \quad \text{and} \quad H'(S_1, t) < H'(S_2, t).$$

$\square$

D.3 Proof of Theorem 4.3

Theorem D.3 (Fuzzy Measure V Is Arbitrary) The fuzzy measure V for explanatory coherence of Schoch (2000) is arbitrary.

Proof.
As mentioned in the chapter on coherence w.r.t. the evidence, the fuzzy measure $V(x_1, \ldots, x_n)$ for explanatory coherence of Schoch (2000) is arbitrary in two respects.

First Respect: On the one hand, the partition of the set of signed propositions $\mathcal{E}$, $\mathcal{E} = \mathcal{PR} \cup \{\neg P : P \in \mathcal{PR}\}$ into two disjoint sets of accepted and rejected propositions – by optimizing the explanatory coherence of some rule system $\mathcal{R}$ on $\mathcal{E}$ – is dependent on the weight factors of incoherence $c_{\mathcal{P}}$ of the incoherent constituents $\mathcal{P} \in \mathbf{I}$, whose choice is arbitrary.

In order to show this one has to find a rule system $\mathcal{R}$ for which there are at least two functions $V(x_1, \ldots, x_n)$ and $V'(x_1, \ldots, x_n)$ – differing from each other at most in the weight factors $c_{\mathcal{P}}$ of the incoherent constituents $\mathcal{P} \in \mathbf{I}$ – such that the truth value assignment φ which maximizes the explanatory coherence of $\mathcal{R}$ according to $V(x_1, \ldots, x_n)$ differs from the truth value assignment φ' which maximizes the explanatory coherence of $\mathcal{R}$ according to $V'(x_1, \ldots, x_n)$. The following example does the job.

Let $\mathcal{PR} = \{P_1, P_2, E_1, E_2\}$ be the set of propositions over which the set of signed propositions $\mathcal{E}$ is defined, and let the rule system $\mathcal{R}$ consist of the following rules:

'$\{P_1\}$ explains E_1', '$\{P_2\}$ explains E_2', 'E_1 is a fact',
'E_2 is a fact', and '$\{P_1\} \cup \{E_2\}$ is competing'.

The weight factor $c_{\{P_1,E_1\}}$ of the coherent constituent $\{P_1, E_1\}$ is $2 = 2^1$, since $N_{\mathcal{R}}(\{P_1\}) = 1$; similarly, the weight factor $c_{\{P_2,E_2\}}$ of the coherent constituent $\{P_2, E_2\}$ is $2 = 2^1$, since $N_{\mathcal{R}}(\{P_2\}) = 1$. The weight factors $c_{\{E_1\}}$ and $c_{\{E_2\}}$ of the coherent constituents $\{E_1\}$ and $\{E_2\}$, respectively, are $4 = 2^2$, because $N_{\mathcal{R}}(\emptyset) = 2$.

For $V(x_1, \ldots, x_4)$, let the weight factor $c_{\{P_1,E_2\}}$ of the incoherent constituent $\{P_1, E_2\}$ be 1. The truth values of the propositions in $\mathcal{PR}$ are supposed to be in $\{0, 1\}$. In the following the propositions in $\mathcal{E}$ will be identified with their truth values.

$$\begin{aligned}
V_{\mathcal{R}}(x_1, \ldots, x_4) &= c_{\{P_1,E_1\}} \cdot P_1 \cdot (2 \cdot E_1 - 1) + \\
&\quad + c_{\{P_2,E_2\}} \cdot P_2 \cdot (2 \cdot E_2 - 1) + c_{\{E_1\}} \cdot (2 \cdot E_1 - 1) + \\
&\quad + c_{\{E_2\}} \cdot (2 \cdot E_2 - 1) - c_{\{P_1,E_2\}} \cdot P_1 \cdot E_2 \\
&= 2 \cdot P_1 \cdot (2 \cdot E_1 - 1) + 2 \cdot P_2 \cdot (2 \cdot E_2 - 1) + \\
&\quad + 4 \cdot (2 \cdot E_1 - 1) + 4 \cdot (2 \cdot E_2 - 1) - 1 \cdot P_1 \cdot E_2,
\end{aligned}$$

which is maximal (= 11) iff[1] $P_1 = P_2 = E_1 = E_2 = 1$.

For $V'(x_1, \ldots, x_4)$, let the weight factor $c'_{\{P_1,E_2\}}$ of the incoherent constituent $\{P_1, E_2\}$ be 10.

$$\begin{aligned}
V'_{\mathcal{R}}(x_1, \ldots, x_4) &= c_{\{P_1,E_1\}} \cdot P_1 \cdot (2 \cdot E_1 - 1) + \\
&\quad + c_{\{P_2,E_2\}} \cdot P_2 \cdot (2 \cdot E_2 - 1) + c_{\{E_1\}} \cdot (2 \cdot E_1 - 1) + \\
&\quad + c_{\{E_2\}} \cdot (2 \cdot E_2 - 1) - c'_{\{P_1,E_2\}} \cdot P_1 \cdot E_2 \\
&= 2 \cdot P_1 \cdot (2 \cdot E_1 - 1) + 2 \cdot P_2 \cdot (2 \cdot E_2 - 1) + \\
&\quad + 4 \cdot (2 \cdot E_1 - 1) + 4 \cdot (2 \cdot E_2 - 1) - 10 \cdot P_1 \cdot E_2,
\end{aligned}$$

which is maximal (= 10) iff[2] $P_1 = 0$, and $P_2 = E_1 = E_2 = 1$.

Second Respect: On the other hand, in comparing two rule systems $\mathcal{R}_1$ and $\mathcal{R}_2$ on

---

[1] The 'only if' holds only if the truth values of the propositions in $\mathcal{PR}$ are restricted to $\{0, 1\}$.

[2] Cf. the preceding footnote.

a common set of signed propositions $\mathcal{E}$ (over some set of propositions $\mathcal{PR}$) with respect to their explanatory coherence, the answer to the question which of the two systems $\mathcal{R}_1$ and $\mathcal{R}_2$ is more coherent depends (apart from the weight factors $c_{\mathcal{P}}$) on the truth value assignment to the propositions in $\mathcal{PR} = \{P_1, \ldots, P_n\}$, i.e. on the values of the variables $x_1, \ldots, x_n$. The choice of the latter is again arbitrary, since there is no criterion telling one which truth value assignment to adopt and to base one's coherence judgement on.

It will be shown that there are rules systems $\mathcal{R}_1$ and $\mathcal{R}_2$ on some common set of signed propositions $\mathcal{E}$ such that

$V_{\mathcal{R}_1}(\varphi_1) > V_{\mathcal{R}_2}(\varphi_1)$, $\quad V_{\mathcal{R}_1}(\varphi_2) < V_{\mathcal{R}_2}(\varphi_2)$, and
$V_{\mathcal{R}_1}(\varphi) = V_{\mathcal{R}_2}(\varphi) = 0$, for any other assignment φ of truth values in $\{0, 1\}$ to the propositions in $\mathcal{PR}$, $\varphi \neq \varphi_1, \varphi \neq \varphi_2$,

where φ_1 is the (uniquely determined[3]) truth value assignment to the propositions in $\mathcal{PR}$ which maximizes the explanatory coherence of $\mathcal{R}_1$; φ_2 is the (uniquely determined[4]) truth value assignment to the propositions in $\mathcal{PR}$ which maximizes the explanatory coherence of $\mathcal{R}_2$; and φ is any assignment of truth values in $\{0, 1\}$ to the propositions in $\mathcal{PR}$, which means that the truth values are again restricted to $\{0, 1\}$. This result holds despite the fact that the weight factors $c_{\mathcal{P}}$ of the constituents $\mathcal{P} \in \mathcal{E}$ occurring in the rules in $\mathcal{R}_1$ and $\mathcal{R}_2$ are assumed to be fixed.

Let $\mathcal{PR} = \{P, E\}$ be the set of propositions over which the set of signed propositions $\mathcal{E}$ is defined; let the rule system $\mathcal{R}_1$ consist of the rule '$\{P\}$ explains E'; and let the rule system $\mathcal{R}_2$ consist of the rule '$\{P\}$ explains $\neg E$'. In the rule system $\mathcal{R}_1$, the weight factor $c_{\{P,E\}}$ of the coherent constituent $\{P, E\}$ is $2 = 2^1$, since $N_{\mathcal{R}_1}(\{P\}) = 1$. In the rule system $\mathcal{R}_2$, the weight factor $c_{\{P,\neg E\}}$ of the coherent constituent $\{P, \neg E\}$ is $2 = 2^1$, since $N_{\mathcal{R}_2}(\{P\}) = 1$.

$$V_{\mathcal{R}_1}(x_1, x_2) = c_{\{P,E\}} \cdot P \cdot (2 \cdot E - 1) = 2 \cdot P \cdot (2 \cdot E - 1),$$

which is maximal (= 2) iff[5] $P = E = 1$, whence $\varphi_1(P) = \varphi_1(E) = 1$.

$$V_{\mathcal{R}_2}(x_1, x_2) = c_{\{P,\neg E\}} \cdot P \cdot (2 \cdot (1 - E) - 1) = 2 \cdot P \cdot (2 \cdot (1 - E) - 1),$$

which is maximal (= 2) iff[6] $P = 1$ and $E = 0$, whence $\varphi_2(P) = 1$ and $\varphi_2(E) = 0$.

[3]Under the assumption that the truth values of the propositions in $\mathcal{PR}$ are restricted to $\{0, 1\}$.
[4]Cf. the preceding footnote.
[5]The 'only if' holds only if the truth values of the propositions in $\mathcal{PR}$ are restricted to $\{0, 1\}$.
[6]Cf. the preceding footnote.

For φ_1 one gets

$$\begin{aligned} V_{\mathcal{R}_1}(\varphi_1) = V_{\mathcal{R}_1}(1,1) &= 2 \cdot 1 \cdot (2 \cdot 1 - 1) \\ &= 2 \\ &> -2 \\ &= 2 \cdot 1 \cdot (2 \cdot (1-1) - 1) = V_{\mathcal{R}_2}(1,1) = V_{\mathcal{R}_2}(\varphi_1); \end{aligned}$$

for φ_2 one gets

$$\begin{aligned} V_{\mathcal{R}_1}(\varphi_2) = V_{\mathcal{R}_1}(1,0) &= 2 \cdot 1 \cdot (2 \cdot 0 - 1) \\ &= -2 \\ &< 2 \\ &= \cdot 1 \cdot (2 \cdot (1-0) - 1) = V_{\mathcal{R}_2}(1,0) = V_{\mathcal{R}_2}(\varphi_2); \end{aligned}$$

and for any other truth value assignment $\varphi, \varphi \neq \varphi_1, \varphi \neq \varphi_2$, one gets

$$\begin{aligned} V_{\mathcal{R}_1}(\varphi) = V_{\mathcal{R}_1}(0,x) &= 2 \cdot 0 \cdot (2 \cdot x - 1) \\ &= 0 \\ &= 2 \cdot 0 \cdot (2 \cdot (1-x) - 1) = V_{\mathcal{R}_2}(0,x) = V_{\mathcal{R}_2}(\varphi) \end{aligned}$$

for every $x \in \{0,1\}$, i.e. for every φ, $\varphi \neq \varphi_1$, $\varphi \neq \varphi_2$.

Finally, for the combined rule system $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$ one gets for every truth value assignment φ (including φ_1 and φ_2)

$$V_{\mathcal{R}}(\varphi) = V_{\mathcal{R}}(x,y) = 2 \cdot x \cdot (2 \cdot y - 1) + 2 \cdot x \cdot (2 \cdot (1-y) - 1) = 0$$

for all $x, y \in \{0,1\}$, i.e. for every $\varphi \in \{\langle x,y \rangle : x,y \in \{0,1\}\} = \{0,1\} \times \{0,1\}$.

$\square$

D.4 Proof of Theorem 4.4

Theorem D.4 (Surplus) Let T and B be (not necessarily finite) sets of wffs, and let E be an evidence.

1. $S(T,E,B) = \emptyset$, if T is infinite,
2. $S(\emptyset,E,B) = A(\emptyset,E,B) = A(B,E,B) = A(T,E,B)$, if $B \vdash T$,
3. $S(B,E,B) = \emptyset$, if $B \neq \emptyset$,

4. $S(T, E, B) = \emptyset$, if $T \neq \emptyset$ and $B \vdash T$, and
5. $S(h_T, E, B) = A(h_T, E, B) = A(T, E, B)$, for every single wff h_T with $h_T \dashv\vdash T$, if $A(\emptyset, E, B) = \emptyset$.

Proof.
Let T, B be (not necessarily finite) sets of wffs, let E be an evidence, and let 't' be a constant term occurring in E.
(1) Suppose T is infinite. If T accounts for 't' in E relative to B, i.e.

$$T \cup B \cup (D \setminus \{A\}) \vdash A,$$

for some finite and non-redundant $D \subseteq (D_E(t))$ and some $A \in D$, then the compactness of $PL1=$ yields that there is a finite set $T_B \subseteq T \cup B$ such that

$$T \cup B \vdash T_B \vdash \bigwedge_{h \in D \setminus \{A\}} h \rightarrow A.$$

Consider $T_{fin} := T \cap T_B$. T_{fin} is finite, because T_B is finite. Furthermore, $T_B \subseteq T_{fin} \cup B$, for if $h \in T_B$, then (i) $h \in T$ or (ii) $h \in B$.
(i): If $h \in T$, then $h \in T \cap T_B = T_{fin}$, whence $h \in T_{fin} \cup B$.
(ii): If $h \in B$, then $h \in T_{fin} \cup B$.
Therefore

$$T_{fin} \cup B \cup (D \setminus \{A\}) \vdash A,$$

which means that T_{fin} accounts for 't' in E relative to B. As T_{fin} is finite and T is infinite, $T_{fin} \subset T$. So for every constant term 't' accounted for by T in E relative to B, there is a finite and thus proper subset T' of T such that T' accounts for 't' in E relative to B. Hence

$$A(T, E, B) \subseteq \bigcup_{T' \subset T} A(T', E, B),$$

and thus

$$S(T, E, B) = A(T, E, B) \setminus \bigcup_{T' \subset T} A(T', E, B) = \emptyset.$$

(2) Suppose that $B \vdash T$, and that B accounts for 't' in E relative to B. Then

$$B \cup B \cup (D \setminus \{A\}) \vdash A,$$

for some finite and non-redundant $D \subseteq (D_E(t))$ and some $A \in D$. This holds just in case

$$\emptyset \cup B \cup (D \setminus \{A\}) \vdash A,$$

for some finite and non-redundant $D \subseteq (D_E(t))$ and some $A \in D$, which holds again just in case
$$T \cup B \cup (D \setminus \{A\}) \vdash A,$$
for some finite and non-redundant $D \subseteq (D_E(t))$ and some $A \in D$, since $B \vdash T$. So
$$A(B, E, B) = A(\emptyset, E, B) = A(T, E, B),$$
if $B \vdash T$. As
$$S(\emptyset, E, B) = A(\emptyset, E, B) \setminus \bigcup_{T' \subset \emptyset} A(T', E, B) = A(\emptyset, E, B),$$
it follows that
$$S(\emptyset, E, B) = A(\emptyset, E, B) = A(B, E, B) = A(T, E, B).$$
(3) If $B \neq \emptyset$, then $A(B, E, B) = A(B', E, B)$, for every B' with $B' \subseteq B$, since $B \vdash B'$ for every such B'. So
$$A(B, E, B) = \bigcup_{B' \subset B} A(B', E, B),$$
and thus
$$S(B, E, B) = A(B, E, B) \setminus \bigcup_{B' \subset B} A(B', E, B) = \emptyset.$$
(4) If $T \neq \emptyset$ and $B \vdash T$, then $B \vdash T'$ for every T' with $T' \subseteq T$, whence
$$A(\emptyset, E, B) = A(B, E, B) = A(T', E, B),$$
for every T' with $T' \subseteq T$. So
$$A(T, E, B) = \bigcup_{T' \subset T} A(T', E, B),$$
and therefore
$$S(T, E, B) = A(T, E, B) \setminus \bigcup_{T' \subset T} A(T', E, B) = \emptyset.$$
(5) Let h_T be a single wff such that $h_T \dashv\vdash T$. Then
$$T \cup B \cup (D \setminus \{A\}) \vdash A,$$

for some finite and non-redundant $D \subseteq (D_E(t))$ and some $A \in D$, iff

$$h_T \cup B \cup (D \setminus \{A\}) \vdash A,$$

for some finite and non-redundant $D \subseteq (D_E(t))$ and some $A \in D$. So $A(T, E, B) = A(h_T, E, B)$. Assume $A(\emptyset, E, B) = \emptyset$. Then it holds for every single wff h:

$$\begin{aligned} S(h, E, B) &= A(h, E, B) \setminus \bigcup_{T' \subset \{h\}} A(T', E, B) \\ &= A(h, E, B) \setminus A_{E,B}(\emptyset) \\ &= A(h, E, B). \end{aligned}$$

So

$$S(h_T, E, B) = A(h_T, E, B) = A(T, E, B).$$

□

D.5 Proof of Theorem 4.5

Theorem D.5 (Coh Is Formally Handy) $Coh(\cdot, \cdot, \cdot)$,

$$Coh(\cdot, \cdot, \cdot) : \wp_{fin}(\mathcal{L}_{PL1=}) \times \mathcal{E} \times \wp_{fin}(\mathcal{L}_{PL1=}) \to \Re,$$

is non-arbitrary, comprehensible, and computable in the limit, where $\wp_{fin}(\mathcal{L}_{PL1=})$ is the set of all finite sets of wffs of $\mathcal{L}_{PL1=}$.

Proof.
Coh is non-arbitray, because it is a single function without parameters that can be freely chosen. It is comprehensible because its definition is stated in the terms of $PL1=$ and ZF.

Computability in the limit is more involved. Let T and B be finite sets of wffs, and let E be an evidence. Suppose $T \cup B \cup E \nvdash \bot$. In order to determine the correct value $Coh(T, E, B)$ of Coh for T, E, and B, one first has to determine the account of T'' in E relative to B, $A(T'', E, B)$, for all the finitely many subsets T'' of any of the finitely many non-empty subsets T' of T. By means of the latter one can determine the surplus of T' in E relative to B, $S(T', E, B)$, for all the finitely many non-empty subsets T' of T. Next one has to determine the B-representatives $S_{B-repr}(T', E, B)$ (respectively their cardinality) of these surpluses $S(T', E, B)$. Together with the (cardinality of the) B-representative of $C(E)$, $C_{B-repr}(E)$, one can then determine the degree of coherence of T w.r.t. E and B, $Coh\,(T,E,B)$. For the former, it is sufficient to determine the (cardinality of the) B-representative of the account of T' in E relative to B, $A_{B-repr}\,(T',E,B)$, for every (possibly empty) subset T' of T.

I will present a method that stabilizes to the correct value $|C_{B-repr}\,(E)|$ of the cardinality of the B-representative of $C\,(E)$, and to the correct value $|A_{B-repr}\,(T',E,B)|$ of the cardinality of the B-representative of the account of T' in E relative to B, for every subset T' of T. This method can then be used to stabilize to the correct value $Coh\,(T,E,B)$ of Coh for T, E, and B, provided $T \cup B \cup E \nvdash \bot$. In addition with a method stabilizing to 1, if $T \cup B \cup E \nvdash \bot$, and to 0 otherwise, the method conjecturing their product will thus stabilize to the correct value $Coh\,(T,E,B)$.

The method doing most of the work is called α. α's conjectures will then be used by another method α^* which eventually starts to conjecture the correct value $Coh\,(T,E,B)$ and continues to do so forever.

Let 't_1', ..., 't_m' be all constant terms occurring in E, and let $T_1,\ldots,T_N$ be the $N := 2^{|T|}$ subsets of T. One first has to answer the $m \cdot N$ questions Q_{ij}: Does T_j account for 't_i' in E relative to B, $1 \leq i \leq m, 1 \leq j \leq N$?

For each such question Q_{ij} there will be a table ij α uses in conjecturing whether T_j accounts for 't_i' in E relative to B. In addition to these $m \cdot N$ tables, α considers m tables $1,\ldots,m$ in conjecturing whether, for a given constant term 't_i', there is a constant term 't_p' with (i) $p < i$, (ii) $E \cup B \vdash t_i = t_p$, and (iii) '$t_p$' $\in C\,(E')$, for every finite set of wffs E' with $E' \dashv\vdash E$ (i.e. 't_p' $\in C_{ess}\,(E)$); *or* whether there is a finite set of wffs E' with $E' \dashv\vdash E$ and 't_i' $\notin C\,(E')$ (i.e. 't_i' $\notin C_{ess}\,(E)$).[7]

Let us first consider the tables $ij, 1 \leq i \leq m, 1 \leq j \leq N$. For a given subset T_j of T and a given constant term 't_i' occurring in E, the question is whether there is a finite and non-redundant set D of relevant elements of E and a wff $A \in D$ such that 't_i' $\in C\,(A')$ for every wff $A' \in D$, and

$$T_j \cup B \cup (D \setminus \{A\}) \vdash A.$$

More precisely, the question is whether there is a finite set of wffs D and a wff A such that

1. $A \in D$,

[7] It is sufficient to consider finite sets of wffs E', for if there is an infinite set of wffs E' with $E' \dashv\vdash E$ – i.e. $E \vdash E'$ and $E' \vdash E$ – and 't_i' $\notin C\,(E')$, then there is a finite set $E'_{fin} \subseteq E'$ such that $E'_{fin} \vdash E$ – and, of course, also $E \vdash E'_{fin}$ and 't_i' $\notin C\left(E'_{fin}\right)$.

2. 't_i' $\in C(A')$, for every wff $A' \in D$,

3. $E \vdash_{crel} A'$, for every wff $A' \in D$,

4. $T_j \cup B \cup (D \setminus \{A\}) \vdash A$,

5. $D \setminus \{A'\} \nvdash A'$, for every wff $A' \in D$,

6. every wff $A' \in D$ is a normal form,

7. there is no wff $A' \in D$ for which there is an $n \geq 1$ such that $A' \dashv\vdash A_1 \wedge \ldots \wedge A_n$, and every wff $A_i, 1 \leq i \leq n$, is shorter than A', where '$\rightarrow$' is eliminated and brackets are not counted, and

8. every quantifier scope of every wff $A' \in D$ is a conjunction $B_1 \wedge \ldots \wedge B_m$, $m \geq 1$, such that it holds for every conjunct B_k, $1 \leq k \leq m$: there is no $n \geq 1$ such that $B_k \dashv\vdash A_1 \wedge \ldots \wedge A_n$, and every wff $A_i, 1 \leq i \leq n$, is shorter than B_k, where '$\rightarrow$' is eliminated and brackets are not counted.

Let $A_1, \ldots, A_n, \ldots$ be an enumeration of all the countably many wffs (of finite length) of $\mathcal{L}_{PL1=}$, and let $D_1, \ldots, D_n, \ldots$ be an enumeration of all the countably many finite sets of wffs of $\mathcal{L}_{PL1=}$. The following table shows that there are only countably many pairs of wffs A and finite sets of wffs D.

	A_1	A_2	A_3	A_4	A_5	$\cdot$	$\cdot$	$\cdot$	A_n	$\cdot$	$\cdot$	$\cdot$
D_1	1	2	6	7	15	16						
D_2	3	5	8	14	17							
D_3	4	9	13	$\cdot$								
D_4	10	12	$\cdot$									
D_5	11	$\cdot$										
$\cdot$												
$\cdot$												
$\cdot$												
D_n												
$\cdot$												
$\cdot$												
$\cdot$												

Let $DA_1, \ldots, DA_n, \ldots$ be an enumeration of all pairs of finite sets of wffs D and wffs A.

The above table shows also that for a given set S of countably many elements s there are only countably many sequences of such elements of length 2. This result can be generalised (by induction) to sequences of any finite length. For suppose the induction hypothesis holds, i.e. there are only countably many sequences $P^m = \langle s_1, \ldots, s_m \rangle$ of length m, where $s_k \in S$, $1 \leq k \leq m$, for some set S of countably many elements. Let $P_1^m, \ldots, P_n^m, \ldots$ be an enumeration of these countably many sequences of length m, and let $s_1, \ldots, s_n, \ldots$ be an enumeration of the countably many elements of S. Every P_i^m, $i \geq 1$, is of the form $P_i^m = \langle s_{i1}, \ldots, s_{im} \rangle$, where $s_{ik} \in S$, for every $k, 1 \leq k \leq m$. Each sequence of elements of S of length $m+1$ is of the form $P_{ij}^{m+1} = \langle s_{i1}, \ldots, s_{im}, s_j \rangle$, for some $i, j \geq 1$. The above table (with P_i^m instead of A_i, and with s_j instead of D_j) shows that there are only countably many such sequences of length $m+1$.

Let $P_1, \ldots, P_n, \ldots$ be an enumeration of all finite sequences of wffs of $\mathcal{L}_{PL1=}$. Each such finite sequence of wffs P_i may be a proof of some wff from some set of wffs. The reasoning of the preceding paragraph can be used once more to obtain that there are only countably many finite sets of finite sequences of wffs. Let $Pr_1, \ldots, Pr_n, \ldots$ be an enumeration of all the countably many finite sets of finite sequences of wffs.

A final application of the mentioned reasoning shows that there are only countably many pairs PrC of finite sets of finite sequences of wffs Pr and wffs C. Let $PrC_1, \ldots, PrC_n, \ldots$ be an enumeration of these pairs (the wffs are denoted by 'C' instead of 'A' in order to avoid confusion).

For a given pair DA_l of a finite set of wffs D_l and a wff A_l,[8] the question is whether

1. $A_l \in D_l$,
2. 't_i' $\in C\,(A')$, for every wff $A' \in D_l$,
3. $E \vdash_{crel} A'$, for every wff $A' \in D_l$,
4. $T_j \cup B \cup (D_l \setminus \{A_l\}) \vdash A_l$,
5. $D_l \setminus \{A'\} \nvdash A'$, for every wff $A' \in D_l$,
6. every wff $A' \in D_l$ is a normal form,

[8] The index l is inherited from DA_l.

7. there is no wff $A' \in D_l$ for which there is an $n \geq 1$ such that $A' \dashv\vdash A_1 \wedge \ldots \wedge A_n$, and every wff $A_i, 1 \leq i \leq n$, is shorter than A', where '$\rightarrow$' is eliminated and brackets are not counted, and

8. every quantifier scope of every wff $A' \in D_l$ is a conjunction $B_1 \wedge \ldots \wedge B_m$, $m \geq 1$, such that it holds for every conjunct B_k, $1 \leq k \leq m$: there is no $n \geq 1$ such that $B_k \dashv\vdash A_1 \wedge \ldots \wedge A_n$, and every wff $A_i, 1 \leq i \leq n$, is shorter than B_k, where '$\rightarrow$' is eliminated and brackets are not counted.

In answering this question one first has to find out whether (1) $A_l \in D_l$; (2) 't_i' $\in C(A')$, for every wff $A' \in D_l$; whether there is a pair PrC_m of a finite set of finite sequences of wffs Pr_m and a wff C_m[9] such that (3a) for every wff $A' \in D_l$ there is a $P \in Pr_m$ which is a proof of A' from E; and (4) there is a $P \in Pr_m$ which is a proof of A_l from $T_j \cup B \cup (D_l \setminus \{A_l\})$.[10]

$T_j \cup B \cup D_l$ is finite, and a proof of some wff C from some set of wffs S is a finite sequence of wffs $\langle A_1, \ldots, A_n \rangle$ such that $A_n = C$, and for every $i, 1 \leq i \leq n$: (i) A_i is an axiom, (ii) A_i is in S, or (iii) A_i is the result of applying a derivation rule to some wffs A_k, $1 \leq k < i$. So questions (1), (2), (3a), and (4) can be answered in finitely many steps for a given pair PrC_m.

If the answer to at least one of these questions is negative for a given DA_l – called a *block* – and a given PrC_m, α writes a "no" in the m-th column of the 0-line of block l of table ij. Otherwise it writes a "yes" in the m-th column of the 0-line of block l of table ij (see below). So a "yes" in the m-th column of the 0-line of block l of table ij means that PrC_m shows that conditions (1), (2), (3a), and (4) hold of T, 't_i', E, B, D_l and A_l.

In a second step α checks for every wff $A' \in D_l$ whether there are (marked) occurrences of predicates $R_1, \ldots, R_n$ in A' such that the following holds of the wff A'^* which is the result of replacing these marked occurrences in A' by new or starred predicates $R_1^*, \ldots, R_n^*$, respectively: There is at least one finite sequence of wffs $P \in Pr_m$ which is a proof of A'^* from E.

Note that for every wff $A' \in D_l$ there are only finitely many such wffs A'^*: namely $2^{n'} - 1$, where n' is the number of occurrences of predicates in A'. In order for A' to be a relevant consequence of E, it has to hold for all these $2^{n'} - 1$ wffs A'^*: $E \nvdash A'^*$.[11]

[9] The index m is inherited from PrC_m.

[10] If there is a finite set Pr_m and a wff C_m satisfying (3a), and if there is a finite set Pr_n satisfying (4), then there is a finite set Pr_k(e.g. $Pr_m \cup Pr_n$) and a wff C_m satisfying (3a) and (4).

[11] So

If there is at least one such wff A'^* for which there is a $P \in Pr_m$ which is a proof of A'^* from E, α writes a *tentative* "yes" *at the top* of the m-th column of the line corresponding to A' in block l of table ij. Otherwise it writes a *tentative* "no" *at the top* of the m-th column of the line corresponding to A' in block l of table ij. Thus a tentative "yes" at the top means that there is a proof showing that A' is no relevant consequence of E.

In a third step α checks for every wff $A' \in D_l$ whether (a) there is a finite sequence of wffs $P \in Pr_m$ which is a proof of $A' \leftrightarrow C_m$ from $\emptyset$; (b) C_m is of the form $A_1 \wedge \ldots \wedge A_n$, for some $n \geq 1$; and (c) every wff $A_i, 1 \leq i \leq n$, is shorter than A', where '$\rightarrow$' is eliminated and brackets are not counted. α writes a *tentative* "yes" *in the middle* of the m-th column of the line corresponding to A' in block l of table ij, if the answers to questions (a)-(c) are affirmative. Otherwise it writes a *tentative* "no" *in the middle* of the m-th column of the line corresponding to A' in block l of table ij. So a tentative "yes" in the middle means that there is a proof showing that A' is not elementary.

In a fourth step α checks for every $A' \in D_l$ whether there is at least one quantifier scope in A' which is a conjunction $B_1 \wedge \ldots \wedge B_m$, $m \geq 1$, such that it holds for at least one conjunct B_k, $1 \leq k \leq m$: (i) there is at least one $P \in Pr_m$ which is a proof of $B_k \leftrightarrow C_m$ from $\emptyset$; (ii) C_m is of the form $A_1 \wedge \ldots \wedge A_n$, for some $n \geq 1$; and (iii) every wff $A_i, 1 \leq i \leq n$, is shorter than B_k, where '$\rightarrow$' is eliminated and brackets are not counted. If at least one quantifier scope in A' is such a conjunction, α writes a *tentative* "yes" *at the bottom* of the m-th column of the line corresponding A' in block l of table ij. Otherwise it writes a *tentative* "no" *at the bottom* of the m-th column of the line corresponding A' in block l of table ij. So a tentative "yes" at the bottom means that there is a proof showing

---

$$A \vdash_{crel,L} B \text{ iff } A \vdash_L B \text{ and } A \nvdash_L B_1^* \text{ and} \ldots \text{and } A \nvdash_L B_N^*,$$

where $N := 2^n - 1$, n is the number of predicate occurrences in B, and B_i^* is the i-th result of replacing (marked) predicate occurrences in B by new or starred ones, $1 \leq i \leq N$. This shows that if the underlying logic L is decidable, then both $\vdash_{crel,L}$ and $\nvdash_{crel,L}$ are recursively enumerable (r.e.), where $\vdash_{crel,L}$ is defined as $\vdash_{crel}$ except that the consequence-relation of $PL1 =$, $\vdash$, is replaced by the consequence relation of L, $\vdash_L$.

A theorem due to Kit Fine shows that the other direction holds, too, for every r.e. logic L closed under substitution and containing classical propositional logic PC – it says that every such logic L is decidable, if $\vdash_{crel,L}$ is r.e. Cf. Schurz (1991), p. 415.

Since $\vdash_{crel,L}$ is decidable just in case both $\vdash_{crel,L}$ and $\nvdash_{crel,L}$ are r.e., it holds for every r.e. logic L closed under substitution and containing PC:

$$L \text{ is decidable iff } \vdash_{crel,L} \text{ is r.e. iff } \vdash_{crel,L} \text{ is decidable.}$$

that at least one quantifier scope of A' is not a conjunction of elementary wffs.

In a fifth step α checks every $A' \in D_l$ on its being a normal form. If A' is a normal form, α writes a *tentative* "yes" *at the left* of the m-th column of the line corresponding to A' in block l of table ij. Otherwise it writes a *tentative* "no" *at the left* of the m-th column of the line corresponding to A' in block l of table ij. So a tentative "no" at the left means that A' is no normal form.

In sum: If there is a tentative "yes" at the top, in the middle, *or* at the bottom, *or* a tentative "no" at the left of the m-th column of the line corresponding to A' in block l of table ij, then PrC_m shows that $A' \in D_l$ is no relevant element of E. It remains to be determined whether D_l is non-redundant.

Therefore, in a sixth step α checks for every $A' \in D_l$ whether there is a finite sequence of wffs $P \in Pr_m$ which is a proof of A' from $D_l \setminus \{A'\}$. α writes a *tentative* "yes" *at the right* of the m-th column of the line corresponding to A' in block l of table ij, if there is such a $P \in Pr$. Otherwise it writes a *tentative* "no" *at the right* of the m-th column of the line corresponding to A' in block l of table ij. So a tentative "yes" at the right means that there is a proof showing that A' is a redundant part of D_l.

In concluding, α looks at the m-th column of the line corresponding to A' in block l of table ij: If there is a tentative "yes" at the top, in the middle, at the bottom, *or* at the right, *or* a tentative "no" at the left, then α *cleans* the m-th column of the line corresponding to A' in block l of table ij and writes a *definite* "yes". Otherwise it cleans the m-th column of this line and writes a *definite* "no".

A definite "yes" in the m-th column of the line corresponding to A' in block l of table ij therefore means that PrC_m shows that D_l is *not* a non-redundant set of relevant elements of E.

Table ij is of the following form, where d_l is the number of wffs in D_l.

table ij	PrC_1	$\cdot$	$\cdot$	$\cdot$	PrC_m	$\cdot$	$\cdot$	$\cdot$		
DA_1	1	2	6	7	C				0-line of	B
A'_{11}	3	5	8	$\cdot$	O				line 1 corresponding to A'_{11} in	L
$\cdot$	4	9	$\cdot$		L					O
$\cdot$	10	$\cdot$			U					C
$\cdot$	11				M					K
A'_{1d_1}					N				line d_1 corresponding to A'_{1d_1} in	1
$\cdot$										
$\cdot$					m					
$\cdot$										
DA_l					$\downarrow$				0-line of	B
A'_{l1}									line 1 corresponding to A'_{l1} in	L
$\cdot$										O
$\cdot$										C
$\cdot$										K
A'_{ld_l}									line d_l corresponding to A'_{ld_l} in	l
$\cdot$										
$\cdot$										
$\cdot$										

In the end α will have filled every blank with a "yes" or "no". A block l is called *positive in the limit* just in case there is a "yes" in at least one column of the 0-line of block l, and there are only "no"s in every column of any line corresponding any wff $A' \in D_l$. A block l is called *negative in the limit* iff there is a "no" in every column of the 0-line of block l, or there is a "yes" in at least one column of at least one line corresponding to some wff $A' \in D_l$.

A block l is called *positive at step* n iff there is a "yes" in at least one column of the 0-line of block l *which has already been investigated by step* n (i.e. at step n, α has already written down a "yes" in the 0-line), and there are only "no"s in every column *already investigated by step* n of any line corresponding to any wff $A' \in D_l$ (i.e. at step n, α has not yet written down a "yes" in any such line). A block l is called *negative at step* n just in case there is a "no" in every column of the 0-line of block l *which has already been investigated by step* n (i.e. at step n, α has not yet written down a "yes" in the 0-line), or there is a "yes" in at least one column *already investigated by step* n of at least one line corresponding to some wff $A' \in D_l$ (i.e. at step n, α has already written down a "yes" in some such line).

The question each table ij is designed to answer is whether T_j accounts for 't_i' in E relative to B. This holds just in case there is at least one finite and non-redundant set D of relevant elements of E and at least one wff $A \in D$ such that 't_i' $\in C(A')$ for every wff $A' \in D$, and

$$T_j \cup B \cup (D \setminus \{A\}) \vdash A.$$

It is straightforward that this holds if and only if there is at least one block l in table ij which is positive in the limit.

At each step n in table ij, α conjectures "yes" – i.e. T_j accounts for 't_i' in E relative to B – just in case there is at least one block l in table ij which is positive at step n. Otherwise it conjectures "no".

There are $m \cdot N$ tables ij. In addition to these, α considers m tables $1, \ldots, m$ – one for each constant term 't_i' $\in C(E)$. Table $i, 1 \leq i \leq m$, has countably many columns listing all the countably many finite sequences of wffs $P_1, \ldots, P_n, \ldots$. Then there are $i-1$ one-line blocks listing the constant terms 't_1', …, 't_{i-1}'; they are put at the beginning. Furthermore table i has countable many blocks listing all the countably many finite sets of wffs $D_1, \ldots, D_n, \ldots$. Each such block consists of i lines (one for each 't_k', $1 \leq k \leq i$). So table i is of the following form:

table i	P_1	$\cdot$	$\cdot$	$\cdot$	P_n	$\cdot$	$\cdot$	$\cdot$		
‘t_1’	1	2	6	7						
$\cdot$	3	5	8	$\cdot$						
$\cdot$	4	9	$\cdot$							
$\cdot$	10	$\cdot$								
‘t_{i-1}’	11									
D_1, ‘t_1’										
$\cdot$										
$\cdot$										
$\cdot$										
D_1, ‘t_i’										
$\cdot$										
$\cdot$										
$\cdot$										
D_n, ‘t_1’										
$\cdot$										
$\cdot$										
$\cdot$										
D_n, ‘t_i’										
$\cdot$										
$\cdot$										
$\cdot$										

For a given ‘t_k’, $1 \leq k \leq i-1$ – called a *one-line block* – and a given P_m, α checks whether P_m is a proof of $t_i = t_k$ from $E \cup B$. If the answer is yes, α writes a “yes” in the m-th column of one-line block k of table i; otherwise it writes a “no” there. So a “yes” in the m-th column of one-line block $k \leq i-1$ of table i means that P_m is a proof of $t_i = t_k$ from $E \cup B$. $E \cup B \vdash t_i = t_k$ holds just in case there is at least one “yes” in one-line block k of table i.

For a given D_l – again called a *block* – and a given P_m, α checks whether (a) P_m is a proof of $\bigwedge_{e \in E} e \leftrightarrow \bigwedge_{d \in D_l} d$ from $\emptyset$, and whether (b.1) ‘t_1’ occurs in D_l, . . ., (b.i) ‘t_i’ occurs in D_l. If (a) is the case, but (b.p) is not, α writes a “yes” in the m-th column of line p of block $l+i-1$ of table i; otherwise it writes a “no” there. So a “yes” in the m-th column of line p of block $l+i-1$ of table i means that P_m is a proof of $\bigwedge_{e \in E} \leftrightarrow \bigwedge_{d \in D_l} d$, where ‘$t_p$’ $\notin C(D_l)$. ‘t_p’ $\notin C_{ess}(E)$ iff there is a “yes” in at least one column of line p of at least one block $l+i-1$ (corresponding to D_l) of table i.

At a given step n, α conjectures "yes" – i.e. 't_i' $\notin C_{ess}(E)$, *or* $E \cup B \vdash t_i = t_p$, for at least one '$t_p$' $\in C_{ess}(E)$, $p < i$ – iff there is a "yes" in at least one column of line i of at least one block $l + i - 1$ of table i *which has already been investigated by step* n (i.e. at step n, α has already written down a "yes" in line i of some such block); *or* if there is a "yes" in at least one column *already investigated by step* n of some one-line block $k \leq i - 1$ of table i, and there are only "no"s in every column *already investigated by step* n of line k of any block $l + i - 1$. Otherwise it conjectures "no".

If 't_i' $\notin C_{ess}(E)$, there is a "yes" in at least one column of line i of some block $l + i - 1$ of table i; so α will eventually start to conjecture "yes", and it will continue to do so forever. If $E \cup B \vdash t_i = t_p$, for some '$t_p$' $\in C_{ess}(E)$, $p < i$, then there is a "yes" in at least one column of some one-line block $p \leq i - 1$, and there is no "yes" in line p of any block $l + i - 1$ of table i; again, α will eventually start to conjecture "yes", and it will continue to do so forever. If 't_i' $\in C_{ess}(E)$ and $E \cup B \vdash t_i = t_p$, for no '$t_p$' $\in C_{ess}(E)$, $p < i$, then there are only "no"s in every column of line i of any block $l + i - 1$ of table i; and for every $p, 1 \leq p \leq i - 1$: there are only "no"s in every column of one-line block p, or there is a "yes" in at least one column of line p of some block $l + i - 1$ of table i. Therefore α will never conjecture "yes", or it will eventually start to conjecture "no", and it will continue to do so forever. So α stabilizes to the correct answer for every constant term 't_i' occurring in E.

Finally, α uses a table 0 in conjecturing whether $T \cup B \cup E \nvdash \perp$. Table 0 consists of a single line and countably many columns listing all finite sequences of wffs $P_1, \ldots, P_n, \ldots$. For a given P_m, α writes a "yes" in the m-th column of table 0, if P_m is a proof of $\perp$ from $T \cup B \cup E$. Otherwise it writes a "no" there. At step n, α conjectures "yes" – i.e. $T \cup B \cup E$ is consistent – iff there are only "no"s in every column *already investigated by step* n (i.e. at step n, α has not yet written down a "yes"). Otherwise it conjectures "no" – i.e. $T \cup B \cup E \nvdash \perp$. α stabilizes to the correct answer for table 0: It conjectures that $T \cup B \cup E$ is consistent except it has found a proof of the opposite claim which makes it conjecture "no" forever.

In sum there are $m \cdot (N + 1) + 1$ tables ij, i, and 0. α starts with step 1 of table 1, and checks through all the $m \cdot (N + 1) + 1$ first steps; after that it continues with step 2 of table 1, and so on. If α starts to conjecture "yes" at some step n of some table ij, because there is a "yes" in at least one column already investigated by step n of the 0-line of some block l of table ij, and because there are only "no"s in any column already investigated by step n of any line corresponding to any wff $A' \in D_l$, then α *sticks to block* l *until it changes its mind* because of a "yes" in some column of some line corresponding to some wff $A' \in D_l$, in which case it goes back to where it has started deviating from its usual path. In other words, in such a case α investigates only the d_l lines corresponding to the wffs $A' \in D_l$. This guarantuees that α conjectures infinitely many "no"s for table ij, if T_j does not account for 't_i' in E relative to B.

At step n, α conjectures $m - c(n)$ as value for $|C_{B-repr}(E)|$, where $c(n)$ is the number of tables i among $1, \ldots, m$ for which α conjectures "yes" at step n – i.e. there is a 't_p' $\in C_{ess}(E)$ such that $p < i$ and $E \cup B \vdash t_i = t_p$. Clearly α stabilizes to the correct value $|C_{B-repr}(E)|$. It remains to be shown that α stabilizes to the correct value $|A_{B-repr}(T_j, E, B)|$ for any subset T_j of T.

Suppose $T_j \subseteq T$ accounts for 't_i' in E relative to B. Then there is at least one block in table ij which is positive in the limit. In other words, there is a block l which contains a "yes" in at least one column, say the m-th, of the 0-line of block l of table ij, and which contains only "no"s in any column of any line corresponding to any wff $A' \in D_l$. When writing down this "yes" in the m-th column of the 0-line of block l, α starts to conjecture that T_j accounts for 't_i' in E relative to B, and it will continue to conjecture this forever, because there will always be this "yes" in the 0-line, and there will never be a "no" in any column of any line corresponding to any wff $A' \in D_l$. As a consequence, α conjectures only finitely many "no"s.

Suppose T_j does not account for 't_i' in E relative to B. Then there is no block l in table ij which is positive in the limit. So for every block l: Either there is no "yes" in any column of the 0-line of block l, or there is a "yes" in at least one column of at least one line corresponding to some wff $A' \in D_l$. Let l be any block of table ij. If there is no "yes" in any column of the 0-line of block l, α can never take l as reason to conjecture "yes". If, however, there is a "yes" in at least one column of at least one line corresponding to some wff $A' \in D_l$, then α cannot take l as reason to conjecture that T_j accounts for 't_i' in E relative to B after it has written down this "yes".

So for every block l there is a step n such that it holds for all later steps $m \geq n$: α cannot take block l as reason to conjecture "yes" at step m. Unfortunately, this does *not* mean that there is a step n' such that it holds for all later steps $m' \geq n'$: At step m', no block l can be taken as reason to conjecture "yes". (If this were the case, the proof would be finished here.) However, α conjectures infinitely many "no"s.

α will therefore eventually conjecture correctly and forever that T_j accounts for 't_i' in E relative to B, if it does so. However, if T_j does not account for 't_i' in E relative to B, it may happen that α does not stabilize to the correct answer "no": Though it will not wrongly stabilize to "yes", it need not stabilize at all, but may continue to change its mind forever.

In order to overcome this difficulty the improved method α^* is introduced: The input for α^* is the output of α. More precisely, where T_j is any subset of T, $1 \leq j \leq N$, let

$$a_j(`t_i', n) = \begin{cases} 1, & \text{if } \alpha \text{ conjectures "no" for table } ij \text{ at step } n, \\ 0 & \text{otherwise;} \end{cases}$$

and

$$a(`t_i', n) = \begin{cases} 1, & \text{if } \alpha \text{ conjectures "yes" for table } i \text{ at step } n, \\ 0 & \text{otherwise.} \end{cases}$$

So

$$a(T_j, `t_i', n) := \sum_{1 \leq k \leq n} \min \{1, a_j(`t_i', k) + a(`t_i', k)\}$$

is the number of steps up to step n at which α conjectures that T_j does *not* account for 't_i' in E relative to B, that 't_i' $\notin C_{ess}(E)$, *or* that there is a 't_p' $\in C_{ess}(E)$ with $p < i$ and $E \cup B \vdash t_i = t_p$. It is important to note that

$$\lim_{n \to \infty} a(T_j, `t_i', n) = \infty,$$

if T_j does not account for 't_i' in E relative to B, if 't_i' $\notin C_{ess}(E)$, or if there is a 't_p' $\in C_{ess}(E)$ with $p < i$ and $E \cup B \vdash t_i = t_p$; and that

$$\lim_{n \to \infty} a(T_j, `t_i', n) < \omega,$$

if T_j accounts for 't_i' in E relative to B, 't_i' $\in C_{ess}(E)$, and there is no 't_p' $\in C_{ess}(E)$ with $p < i$ and $E \cup B \vdash t_i = t_p$.

For if T_j does not account for 't_i' in E relative to B, then α conjectures infinitely many "no"s for table ij; and if 't_i' $\notin C_{ess}(E)$, or if there is a 't_p' $\in C_{ess}(E)$ with $p < i$ and $E \cup B \vdash t_i = t_p$, then α starts to conjecture "yes" for table i after some time, and continues to do so forever. If, however, T_j accounts for 't_i' in E relative to B, 't_i' $\in C_{ess}(E)$, and there is no 't_p' $\in C_{ess}(E)$ with $p < i$ and $E \cup B \vdash t_i = t_p$, then α conjectures only finitely many "no"s for table ij, and only finitely many "yes"s for table i, if it ever conjectures "yes" for table i.

Let K be any of the finitely many subsets of $C(E)$, and define

$$\overline{K} := C(E) \setminus K, \quad \min(K, T_j, n) := \min \{a(T_j, `t_i', n) : `t_i' \in K\},$$

and

$$\max(K, T_j, n) := \max\{a(T_j, \text{‘}t_i\text{’}, n) : \text{‘}t_i\text{’} \in K\}.$$

At step n, α^* considers

$$d_n(K, T_j) := \min\left(\overline{K}, T_j, n\right) - \max(K, T_j, n),$$

for every $K \subseteq C(E)$, and its conjecture at step n is that $K_n^*(T_j)$ is the set of all constant terms ‘t_i’ $\in C_{ess}(E)$ which are accounted for by T_j in E relative to B, and for which there is no ‘t_p’ $\in C_{ess}(E)$ with $p < i$ and $E \cup B \vdash T_i = t_p$, where $K_n^*(T_j)$ is that subset K^* of $C(E)$ such that

$$d_n(K^*, T_j) > d_n(K, T_j),$$

for every $K \subseteq C(E)$, $K \neq K^*$, if there is such a $K^* \subseteq C(E)$; otherwise $K_n^*(T_j)$ is $C(E)$.

It will be shown that there is a step n such that it holds for all later steps $m \geq n$:

$$d_m(A_{B-repr}(T_j, E, B), T_j) > d_m(K, T_j),$$

for every $K \subseteq C(E)$, $K \neq A_{B-repr}(T_j, E, B)$; i.e. there is an n such that it holds for all $m \geq n$:

$$K_m^*(T_j) = A_{B-repr}(T_j, E, B).$$

Note first that

$$\lim_{n\to\infty} d_n(A_{B-repr}(T_j, E, B), T_j) = \infty,$$

because

$$\lim_{n\to\infty} \min\left(\overline{A_{B-repr}(T_j, E, B)}, T_j, n\right) = \infty,$$

and

$$\lim_{n\to\infty} \max(A_{B-repr}(T_j, E, B), T_j, n) < \omega.$$

The reason is that

$$\lim_{n\to\infty} a(T_j, \text{‘}t_i\text{’}, n) = \infty,$$

for every ‘t_i’ $\in \overline{A_{B-repr}(T_j, E, B)}$, and

$$\lim_{n\to\infty} a(T_j, \text{‘}t_i\text{’}, n) < \omega,$$

for every ‘t_i’ $\in A_{B-repr}(T_j, E, B)$.

Let K be any subset of $C(E)$, $K \neq A_{B-repr}(T_j, E, B) \subseteq C(E)$. Then

$$K \subset A_{B-repr}(T_j, E, B) \quad \text{or} \quad A_{B-repr}(T_j, E, B) \subset K.$$

In the first case there is at least one 't_i' $\in C(E)$ with

$$\text{'}t_i\text{'} \in A_{B-repr}(T_j, E, B) \cap \overline{K},$$

whence

$$\lim_{n\to\infty} \min\left(\overline{K}, T_j, n\right) < \omega, \quad \text{and thus} \quad \lim_{n\to\infty} d_n(K, T_j, n) < \omega,$$

because

$$\max(A_{B-repr}(T_j, E, B), T_j, n) \geq \max(K, T_j, n),$$

for every n. In the second case there is at least one 't_i' $\in C(E)$ with

$$\text{'}t_i\text{'} \in \overline{A_{B-repr}(T_j, E, B)} \cap K,$$

whence

$$\lim_{n\to\infty} \max(K, T_j, n) = \infty,$$

where for every n,

$$\min\left(\overline{A_{B-repr}(T_j, E, B)}, T_j, n\right) \geq \min\left(\overline{K}, T_j, n\right).$$

So in both cases there is a step n such that it holds for all later steps $m \geq n$:

$$d_m(A_{B-repr}(T_j, E, B), T_j) > d_m(K, T_j),$$

for every $K \subseteq C(E)$, $K \neq A_{B-repr}(T_j, E, B)$.

As a consequence, α^* stabilizes to the correct value $|A_{B-repr}(T_j, E, B)|$, for every subset T_j of T. At step n, α^* conjectures

$$s^*(T', E, B, n) := \left| K_n^*(T') \setminus \bigcup_{T'' \subset T'} K_n^*(T'') \right|$$

as value for $|S_{B-repr}(T', E, B)|$, $\emptyset \neq T' \subseteq T$. By conjecturing

$$r^*(T, E, B, n) := \sum_{\emptyset \neq T' \subseteq T} \frac{s^*(T', E, B, n)}{(m - c(n)) \cdot (2^{|T|} - 1)} \cdot x_n$$

at step n, α^* stabilizes to the correct value $Coh(T, E, B)$ of Coh for T, E, and B, where

$$x_n = \begin{cases} 1, & \text{if } \alpha^* \text{ conjectures "yes" for table 0 at step } n, \\ 0 & \text{otherwise.} \end{cases}$$

□

D.6 Proof of Theorem 4.6

Theorem D.6 (No InvEquTrans of T for Coh) For every evidence E, and every set of wffs B there are theories T and T' such that

$$T \dashv\vdash T' \quad \text{and} \quad Coh\,(T, E, B) \neq Coh\,(T', E, B)\,,$$

provided there is at least one theory T with $Coh\,(T, E, B) \neq 0$.

Proof.
Let E be an evidence, and let B be a set of wffs. Suppose there is at least one theory T with $Coh\,(T, E, B) \neq 0$.

Let $h := \forall x^i\,(Fx^i \vee \neg Fx^i)$, where '$F$' is some predicate not occurring in T, E, or B, and T contains at least one essential occurrence of an i-variable. So $T \cup \{h\}$ is a theory with $T \dashv\vdash T \cup \{h\}$. Therefore

$$\begin{aligned}
Coh\,(T, E, B) &= \sum_{\emptyset \neq T' \subseteq T} \frac{|S_{B-repr}\,(T', E, B)|}{|C_{B-repr}\,(E)| \cdot (2^{|T|} - 1)} \\
&= \sum_{\emptyset \neq T' \subseteq T} \frac{|S_{B-repr}\,(T', E, B)|}{|C_{B-repr}\,(E)| \cdot (2^{|T|} - 1)} + \\
&\quad + \sum_{\emptyset \neq T' \subseteq T} \frac{|S_{B-repr}\,(T' \cup \{h\}, E, B)|}{|C_{B-repr}\,(E)| \cdot (2^{|T|} - 1)} \\
&= \sum_{\emptyset \neq T' \subseteq T \cup \{h\}} \frac{|S_{B-repr}\,(T', E, B)|}{|C_{B-repr}\,(E)| \cdot (2^{|T|} - 1)} \\
&> \sum_{\emptyset \neq T' \subseteq T \cup \{h\}} \frac{|S_{B-repr}\,(T', E, B)|}{|C_{B-repr}\,(E)| \cdot (2^{|T \cup \{h\}|} - 1)} \\
&= Coh\,(T \cup \{h\}, E, B)\,.
\end{aligned}$$

The proviso is non-trivial, since there are evidences E and sets of wffs B such that it holds for every theory T: $Coh\,(T, E, B) \neq 0$, *only* if none of the evidential domains of E is among the domains of proper investigation of T – even if $E \cup B \nvdash \bot$ and $A_{B-repr}\,(\emptyset, E, B) \neq C_{B-repr}\,(E)$.

Let $E = \{Fa\}$ and $B = \{\neg\forall x Fx\}$. Then $A_{B-repr}\,(\emptyset, E, B) = \emptyset \neq C_{B-repr}\,(E)$ and $E \cup B \nvdash \bot$.

Suppose T is a theory such that (i) $Coh\,(T, E, B) \neq 0$, and (ii) at least one of the evidential domains of E is a domain of proper investigation of T. (i) yields $T \cup B \vdash Fa$, i.e. $T \vdash \neg\forall x Fx \rightarrow Fa$.

Because of (ii), 'a' cannot occur in T, whence $T \vdash \forall y\,(\neg\forall x Fx \rightarrow Fy)$, i.e. $T \vdash \forall y\,(\forall x Fx \vee Fy)$, and thus $T \vdash \forall x Fx$. But then $T \cup B \vdash \bot$, whence $Coh\,(T, E, B) = 0$ – a contradiction. □

D.7 Proof of Theorem 4.7

Theorem D.7 (Coherence Versus Power) Let T be a finite set of wffs, let E be an evidence, and let B be a set of wffs. If $T \cup B \cup E \nvdash \bot$ and $A_{B-repr}\,(\emptyset, E, B) = \emptyset$, then

$$Coh\,(T, E, B) \leq Coh\left(\bigwedge_{h \in T}, E, B\right) = \mathcal{P}\,(T, E, B)\,,$$

where $\mathcal{P}$ is closed under equivalence transformations of T and B.

Proof.
That $\mathcal{P}$ is closed under equivalence transformations of T and B is an immediate consequence of its definition. Let T be a finite set of wffs, let E be an evidence, and let B be a (not necessarily finite) set of wffs. Suppose $T \cup B \cup E \nvdash \bot$ and $A_{B-repr}\,(\emptyset, E, B) = \emptyset$. If $T = \emptyset$, then $\{\bigwedge_{h \in T} h\} = \emptyset$, and thus

$$Coh\,(T, E, B) = Coh\left(\bigwedge_{h \in T} h, E, B\right) = 0 = \mathcal{P}\,(T, E, B)\,,$$

because $C_{B-repr}\,(E) \neq \emptyset$ and $A_{B-repr}\,(T, E, B) = A_{B-repr}\,(\emptyset, E, B) = \emptyset$.

Suppose $T \neq \emptyset$. As T is finite, and $A_{B-repr}\,(\emptyset, E, B) = \emptyset$, it holds for every $T' \subseteq T$,

$$S\,(T', E, B) \subseteq A\,(T', E, B) \subseteq A\,(T, E, B) = S\left(\bigwedge_{h \in T} h, E, B\right).$$

Since $A \setminus C \subseteq B \setminus C$, if $A \subseteq B$, for any sets A, B, C, it holds for every $T' \subseteq T$,

$$|S_{B-repr}\,(T', E, B)| \leq \left|S_{B-repr}\left(\bigwedge_{h \in T} h, E, B\right)\right|.$$

Let $N := 2^{|T|} - 1$, and let $T_1, \ldots, T_N$ be all the N non-empty subsets of T.

$$\frac{|S_{B-repr}\,(T_1, E, B)| + \ldots + |S_{B-repr}\,(T_N, E, B)|}{|C_{B-repr}\,(E)| \cdot (2^{|T|} - 1)} \quad \leq \quad \frac{N \cdot |S_{B-repr}\,(\bigwedge_{h \in T} h, E, B)|}{|C_{B-repr}\,(E)| \cdot (2^{|T|} - 1)}$$

$$\text{iff}$$

$$\sum_{\emptyset \neq T' \subseteq T} \frac{|S_{B-repr}(T', E, B)|}{|C_{B-repr}(E)| \cdot (2^{|T|} - 1)} \leq \frac{|S_{B-repr}(\bigwedge_{h \in T} h, E, B)|}{|C_{B-repr}(E)|}$$

$$N = 2^{|T|} - 1$$

$$\text{iff}$$

$$Coh(T, E, B) \leq Coh\left(\bigwedge_{h \in T} h, E, B\right).$$

As $S_{B-repr}(\bigwedge_{h \in T} h, E, B) = A_{B-repr}(T, E, B)$, it follows that

$$Coh(T, E, B) \leq Coh\left(\bigwedge_{h \in T} h, E, B\right) = \mathcal{P}(T, E, B).$$

□

D.8 Proof of Theorem 4.8

Theorem D.8 (No SensLoveLike of Coh) For every power searcher $\mathcal{LO}$, every truth indicator $\mathcal{LI}$, and every evidence E there is a theory T_E and a background knowledge B_E such that it holds for any sets of wffs T and B, and every evidence E': If $T \dashv\vdash T_E$, $E' \dashv\vdash E$, and $B \dashv\vdash B_E$, then

1. $T \cup B \vdash E'$, and thus $\mathcal{LO}(T, E', B) = 1$,
2. $E' \cup B \vdash T$, and thus $\mathcal{LI}(T, E', B) = 1$, and
3. $Coh(T, E', B) = 0$.

Proof.
Let $\mathcal{LO}$ be a power searcher, let $\mathcal{LI}$ be a truth indicator, and let E be an evidence. T_E and B_E are defined as follows: $T_E = \{\forall x^i F x^i\}$, and $B_E = E \cup T_E$, where 'F' is some predicate not occurring in E. T_E is a theory, and B_E is a background knowledge.

E is consistent, whence $E \cup B_E$ and $T_E \cup B_E$ are consistent, too. By the definition of B_E, $T_E \cup B_E \vdash E$ and $E \cup B_E \vdash T_E$.

Let E' be an evidence, and let T and B be sets of wffs. Suppose $E' \dashv\vdash E$, $T \dashv\vdash T_E$, and $B \dashv\vdash B_E$. Then $T \cup B \vdash E'$ and $E' \cup B \vdash T$. As a consequence, $\mathcal{LO}(T, E', B) = \mathcal{LI}(T, E', B) = 1$.

Theorem 4.4 (Surplus) yields that $S(T, E', B) = \emptyset$, if $T \neq \emptyset$ and $B \vdash T$, whence it holds for every T', $\emptyset \neq T' \subseteq T$:

$$S_{B-repr}(T', E', B) \subseteq S(T', E', B) = \emptyset.$$

Suppose T is finite. As $C_{B-repr}(E')$ is always non-empty,

$$Coh(T, E', B) = \sum_{\emptyset \neq T' \subseteq T} \frac{|S_{B-repr}(T', E', B)|}{|C_{B-repr}(E')| \cdot (2^{|T|} - 1)} = 0.$$

Suppose T is a set of countably many wffs, and $\lim_{i \to \infty} Coh(T_i, E', B)$ exists, and is the same for every enumeration $h_1, \ldots, h_n, \ldots$ of the wffs in T, where $T_i := \{h_1, \ldots, h_i\}$. Let $h_1, \ldots, h_n, \ldots$ be an enumeration of the wffs in T, and consider $T_i = \{h_1, \ldots, h_i\}$, for any $i \geq 1$. As $B \vdash T$, $B \vdash T_i'$ for every T_i', $\emptyset \neq T_i' \subseteq T_i$, whence

$$S_{B-repr}(T_i', E', B) \subseteq S(T_i', E', B) = \emptyset.$$

As before, it follows that $Coh(T_i, E', B) = 0$. Since this holds for every $i \geq 1$, it follows that $\lim_{i \to \infty} Coh(T_i, E', B)$ exists, and equals 0. By assumption, $\lim_{i \to \infty} Coh(T_i, E', B)$ is the same for every enumeration $h_1, \ldots, h_n, \ldots$ of the wffs in T. So $Coh(T, E', B) = 0$.

Finally, if T is a set of uncountably many wffs, $Coh(T, E', B)$ is not defined, and may be set equal to 0.

Note that this holds in particular for the – by assumption existing – unique canonical formulation F_{T_E} of T_E. □

Appendix E

Proofs for Chapter 5

E.1 Proof of Theorem 5.1

Theorem E.1 ($\mathcal{P}$ Is a Formally Handy Power Searcher) $\mathcal{P}(\cdot,\cdot,\cdot), \mathcal{P}(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \to \Re$, is a power searcher which is non-arbitrary, comprehensible, and computable in the limit, provided for every $E \in \mathcal{E}$ and every 't' $\in C_{ess}(E)$ there is a contingent[1] $A \in RE(E)$ with 't' $\in C(A)$.

More precisely, $\mathcal{P}$ is formally handy, and for any theories T and T', every evidence E, every background knowledge B, and every confirmational domain D_i of T and E, and of T' and E:

1. $\mathcal{P}(T, E, B; D_i) \geq 0$,
2. if $T \cup B \vdash E$, then $\mathcal{P}(T, E, B; D_i) = 1$, and
3. if $T' \vdash T$, then $\mathcal{P}(T', E, B; D_i) \geq \mathcal{P}(T, E, B; D_i)$,

provided for every $E \in \mathcal{E}$ and every 't' $\in C_{ess}(E)$ there is a contingent $A \in RE(E)$ with 't' $\in C(A)$.

Proof.
That $\mathcal{P}$ is non-arbitrary, comprehensible, and computable in the limit is a consequence of theorem 4.5 and the proof of theorem 5.2, where it is shown how to stabilize to the correct answer to the question whether D_i is a confirmational domain of T and E.

[1] Contingency should rule out $t = t$, which is a relevant consequence of any E.

Let T and T' be theories, let E be an evidence, let B be a background knowledge, and let D_i be a confirmational domain of T and E, and of T' and E. Clearly,

$$\mathcal{P}\left(T,E,B;D_i\right)=\frac{\left|A_{B-repr}\left(T,E,B\right)\cap C_i\right|}{\left|C_{B-repr}\left(E\right)\cap C_i\right|}\geq 0.$$

Suppose $T\cup B\vdash E$. By assumption, for every 't' $\in C_{ess}\left(E\right)$ – and thus for every 't^i' $\in C_{B-repr}\left(E\right)\cap C_i$ – there is a contingent $A_t\in RE\left(E\right)$ with 't' $\in C\left(A_t\right)$. So for every 't^i' $\in C_{B-repr}\left(E\right)\cap C_i$ there is a finite and non-redundant $D\subseteq D_E\left(t^i\right)$ – namely $D=\{A_{t^i}\}$ – and a wff $A\in D$ (namely A_{t^i}) such that 't^i' $\in C\left(A'\right)$, for every $A'\in D$, and

$$T\cup B\cup\left(D\setminus\{A\}\right)\vdash A.$$

As a consequence, $A_{B-repr}\left(T,E,B\right)\cap C_i=C_{B-repr}\left(E\right)\cap C_i$, and thus $\mathcal{P}\left(T,E,B;D_i\right)=1$.

Finally, suppose $T'\vdash T$, and let 't^i' $\in A_{B-repr}\left(T,E,B\right)\cap C_i$. This means that there is a finite and non-redundant $D\subseteq D_E\left(t^i\right)$ and a wff $A\in D$ such that 't^i' $\in C\left(A'\right)$, for every $A'\in D$, and

$$T\cup B\cup\left(D\setminus\{A\}\right)\vdash A.$$

But then there is also a finite and non-redundant $D\subseteq D_E\left(t^i\right)$ and a wff $A\in D$ such that 't^i' $\in C\left(A'\right)$, for every $A'\in D$, and

$$T'\cup B\cup\left(D\setminus\{A\}\right)\vdash A.$$

So $A_{B-repr}\left(T,E,B\right)\cap C_i\subseteq A_{B-repr}\left(T',E,B\right)\cap C_i$, and therefore

$$\mathcal{P}\left(T,E,B;D_i\right)\leq\mathcal{P}\left(T',E,B;D_i\right).$$

□

E.2 Proof of Theorem 5.2

Theorem E.2 ($\mathcal{LI}$ Is a Formally Handy Truth Indicator) $\mathcal{LI}\left(\cdot,\cdot,\cdot\right)$, $\mathcal{LI}\left(\cdot,\cdot,\cdot\right):\mathcal{T}\times\mathcal{E}\times\mathcal{B}\rightarrow\Re$, is a truth indicator which is non-arbitrary, comprehensible, and computable in the limit.

More precisely, $\mathcal{LI}$ is formally handy, and for any theories T and T', every evidence E, every background knowledge B, and every confirmational domain D_i of T and E, and of T' and E: If $E\cup B\nvdash\bot$, then

1. $\mathcal{LI}(T, E, B; D_i) \geq 0$,
2. if $E \cup B \vdash T$, then $\mathcal{LI}(T, E, B; D_i) = 1$, and
3. if $T' \vdash T$, then $\mathcal{LI}(T', E, B; D_i) \leq \mathcal{LI}(T, E, B; D_i)$.

Proof.
$\mathcal{LI}$ is non-arbitrary, because it is a single function without parameters that can be freely chosen. It is comprehensible because its definition is stated in the terms of $PL1 =$ and ZF.

For computability in the limit one has to show that there is a method α that stabilizes to the correct value $\mathcal{LI}(T, E, B; D_i)$ of $\mathcal{LI}$ for T, E, B, and every confirmational domain D_i of T and E, for all theories T, evidences E, and background knowledges B. The proof of theorem 4.5 shows how to stabilize to the correct value $|C_{B-repr}(E)|$ – and thus to $|C_{B-repr}(E) \cap C_i(E)| = |C_{B-repr}(E) \cap C_i|$ – for every E and B. It also shows how to stabilize to 1, if $E \cup B \nvdash \bot$, and to 0 otherwise. It remains to be shown how to stabilize to the correct value $\max_{\mathcal{LI}}(T, E, B; D_i)$, i.e.

$$\max\left\{|C \cap C_{B-repr}(E)| : C \subseteq C_{E,B,i}, E \vdash Dev_{C_{E,B,i}}(B) \rightarrow Dev_C(T)\right\},$$

$C_{E,B,i} = C(E \cup B) \cap C_i = C_i(E \cup B)$, for all T, E, B, and every confirmational domain D_i of T and E.

Let T be a theory, let E be an evidence, and let B be a background knowledge. One first has to determine the confirmational domains of T and E. Let D_i be any domain such that T contains an occurrence of an i-variable, but no occurrence of a constant i-term, and E contains an occurrence of a constant i-term, but no occurrence of an i-variable; let 't_1^i', …, 't_m^i' be the constant i-terms occurring in E; let 't_{m+1}^i', …, 't_p^i' be the constant i-terms occurring in B but not in E; and let $K_1, \ldots, K_N$ be the 2^p subsets of $C_i(E \cup B)$.

The question is whether T contains an *essential* occurrence of an i-variable, and whether there is at least one constant i-term *essentially* occurring in E. (We already know that there are no occurrences of constant i-terms in T, and no occurrences of i-variables in E.) In order to answer this question method α uses $m + 1$ tables T and $T'_1, \ldots, T'_m$ all of which consist of countably many columns listing all finite sequence of wffs $P_1, \ldots, P_n, \ldots$ and countable many lines listing all finite sets of wffs $D_1, \ldots, D_n, \ldots$.

For a given P_m and a given C_l, α checks whether (i) P_m is a proof of $\bigwedge_{h \in T} h \leftrightarrow \bigwedge_{d \in D_l} d$ from $\emptyset$, and (ii) D_l contains an occurrence of an i-variable.

If (i) is the case, but (ii) is not, α writes a "no" in the m-th column of line l of table T; otherwise it writes a "yes" there. T contains an essential occurrence of an i-variable just in case there are only "yes"s in every column of every line of table T.

Furthermore, α checks whether (i) P_m is a proof of $\bigwedge_{e \in E} e \leftrightarrow \bigwedge_{d \in D_l} d$ from $\emptyset$, and (ii) D_l contains an occurrence of t^i_k, $1 \leq k \leq m$. If (i) is the case, but (ii) is not, α writes a "no" in the m-th column of line l of table T'_k; otherwise it writes a "yes" there. E contains an essential occurrence of 't^i_k' iff there are only "yes"s in every column of every line of table T'_k.

At step n, α conjectures "yes" – i.e. D_i is a confirmational domain of T and E – iff there are only "yes"s in every column of every line *already investigated by step* n of table T, *and* there are only "yes"s in every column of every line *already investigated by step* n of *at least one* table among $T'_1, \ldots, T'_m$ (i.e. at step n, α has not yet written down a "no" in any column of any line of table T, nor has it written down a "no" in any column of any line of *some* table among $T'_1, \ldots, T'_m$).

α stabilizes to the correct answer: If D_i is a confirmational domain of T and E, then there are only "yes"s in every column of every line of table T and of at least one further table among $T'_1, \ldots, T'_m$, whence α will always conjecture correctly "yes". If D_i is no confirmational domain of T and E, then there is a "no" in at least one column of at least one line of table T, *or* there is a "no" in at least one column of at least one line of *all* tables $T'_1, \ldots, T'_m$. In the first case, α conjectures correctly and forever that D_i is no confirmational domain of T and E after it has written down this "no" in table T; in the second case, α conjectures correctly and forever that D_i is no confirmational domain of T and E after it has written down these "no"s in all tables $T'_1, \ldots, T'_m$.

In addition α uses $N = 2^p$ tables T_j, $1 \leq j \leq N$ – one for each $K_j \subseteq C_i (E \cup B)$ – in conjecturing the correct value $\max_{\mathcal{LI}} (T, E, B; D_i)$.

Table T_j consists of one single line and countably many columns listing all finite sequences of wffs $P_1, \ldots, P_n, \ldots$. For a given P_m, α checks whether P_m is a proof of $Dev_{C_{E,B,i}} (B) \rightarrow Dev_{C_j} (T)$ from E. If the answer is affirmative, α writes a "yes" in the m-th column of table T_j; otherwise it writes a "no" there. At step n, α conjectures "yes" – i.e. E logically implies $Dev_{C_{E,B,i}} (B) \rightarrow Dev_{C_j} (T)$ – iff there is a "yes" in at least one column *already investigated by step* n (i.e. at step n, α has already written down a "yes" in some column of table T_j); otherwise it conjectures "no". α stabilizes to the correct answer: $E \vdash Dev_{C_{E,B,i}} (B) \rightarrow Dev_{C_j} (T)$ holds iff there is a proof of $Dev_{C_{E,B,i}} (B) \rightarrow Dev_{C_j} (T)$ from E, which holds just in case there is a "yes" in at least one column of table T_j.

Furthermore, α considers m tables $1, \ldots, m$ – one for each constant i-term 't_k^i' $\in C_i(E)$ – in conjecturing whether 't_k^i' $\in C_{B-repr}(E)$. α just copies what the α of the proof of theorem 4.5 does. Therefore it stabilizes to the correct answer to the question whether (i) 't_k^i' $\in C_{ess}(E)$, and whether (ii) there is a 't_p^i' $\in C_{ess}(E)$ with $p < k$ and $E \cup B \vdash t_k^i = t_p^i$.

Finally, α uses a table 0 in conjecturing whether $E \cup B \nvdash \perp$. Again, α just copies what the α of the proof of theorem 4.5 does, and thus stabilizes to 1, if $E \cup B \nvdash \perp$, and to 0 otherwise.

As in the proof of theorem 4.5, another method α^* observes the output of α. At step n, α^* conjectures that $\mathcal{LI}(T, E, B; D_i)$ is not defined, if, at step n, α conjectures "no" for table T *or* it conjectures "no" for all tables $T'_1, \ldots, T'_m$ – i.e. D_i is no confirmational domain of T and E; *or* if α conjectures 0 for table 0 at step n – i.e. $E \cup B \vdash \perp$.

If, however, α's conjecture at step n is that D_i is a confirmational domain of T and E; and if, at step n, α conjectures "yes" for tables $k_1, \ldots, k_s$[2] among tables $1, \ldots, m$, $s \geq 1$, $1 \leq k_r \leq m$, for every $r, 1 \leq r \leq s$; and if, at step n, α conjectures that $E \cup B \nvdash \perp$; then α^* conjectures at step n that $\mathcal{LI}(T, E, B; D_i)$ is defined, and that

$$\mathcal{LI}(T, E, B; D_i) = \frac{\left| K_n^* \cap \left\{ \text{`}t_{k_1}^i\text{'}, \ldots, \text{`}t_{k_s}^i\text{'} \right\} \right|}{m - c_n^*},$$

where

1. $c_n^* = m - s$ is the number of tables k among $1, \ldots, m$ for which α conjectures "no" at step n – i.e. 't_k^i' $\notin C_{B-repr}(E)$; and

2. K_n^* is that subset K_j of $C_i(E \cup B)$ such that

   (a) at step n, α conjectures "yes" for table j – i.e. E logically implies $Dev_{C_{E,B,i}}(B) \rightarrow Dev_{C_j}(T)$;

   (b) $\left| K_j \cap \left\{ \text{`}t_{k_1}^i\text{'}, \ldots, \text{`}t_{k_s}^i\text{'} \right\} \right| \geq \left| K_l \cap \left\{ \text{`}t_{k_1}^i\text{'}, \ldots, \text{`}t_{k_s}^i\text{'} \right\} \right|$, for every $l, 1 \leq l \leq N$; and

   (c) there is no K_q, $q < j$, satisfying (b) and (c).

If there is no such K_j, then $K_n^* = \emptyset$.

[2] Index n is suppressed for obvious reasons.

α stabilizes to the correct answer for every table k, $1 \leq k \leq m$. So there is a step n_1 such that at all later steps $n \geq n_1$, α^* conjectures correctly that $\left\{ `t^i_{k_1}', \ldots, `t^i_{k_s}' \right\} = C_{B-repr}(E)$. Furthermore, α stabilizes to the correct answer for every table $T_j, 1 \leq j \leq N$, whence there is a step n_2 such that at all later steps $n \geq n_2$, α^* takes correctly into account all and only those $K_j \subseteq C_i(E \cup B)$ with $E \vdash Dev_{C_{E,B,i}}(B) \rightarrow Dev_{C_j}(T)$.

As α also stabilizes to the correct answer for the tables $0, T, T'_1, \ldots, T'_m$, there is a step n^* such that it holds for all later steps $n \geq n^*$: (1) At step n, α^* conjectures that $\mathcal{LI}(T, E, B; D_i)$ is defined, and that

$$\mathcal{LI}(T, E, B; D_i = \frac{\max_{\mathcal{LI}}(T, E, B; D_i)}{|C_{B-repr}(E) \cap C_i|},$$

where

$$\begin{aligned} \max_{\mathcal{LI}}(T, E, B; D_i) \quad &= \quad \max \{|C \cap C_{B-repr}(E)| : C \subseteq C_{E,B,i}, \\ &\qquad E \vdash Dev_{C_{E,B,i}}(B) \rightarrow Dev_C(T)\Big\}, \end{aligned}$$

if D_i is a confirmational domain of T and E, and $E \cup B \nvdash \perp$. (2) At step n, α^* conjetures that $\mathcal{LI}(T, E, B; D_i)$ is not defined, if D_i is no confirmational domain of T and E, or $E \cup B \vdash \perp$. That much to computability in the limit.

As to truth indicativeness, let T and T' be theories, let E be an evidence, and let B be a background knowledge. Let D_i be a confirmational domain of T and E, and of T' and E (with corresponding i-variables and constant i-terms), and let $D_1, \ldots, D_n$ be the domains of T, E, and B (i.e. there occur variables and constants of n different sorts in T, E, and B). Suppose $E \cup B \nvdash \perp$.

(A) Obviously, $\mathcal{LI}(T, E, B; D_i) \geq 0$.

(B) Suppose $E \cup B \vdash T$. I show that

$$E \vdash Dev_{C_{E,B,i}}(B) \rightarrow Dev_{C_{E,B,i}}(T),$$

for then

$$\mathcal{LI}(T, E, B; D_i) = \frac{|C_{B-repr}(E) \cap C_i(E \cup B)|}{|C_{B-repr}(E) \cap C_i|} = 1.$$

Suppose

$$E \nvdash Dev_{C_{E,B,i}}(B) \rightarrow Dev_{C_{E,B,i}}(T).$$

Then there is at least one model $\mathcal{M} = \langle Dom, \varphi \rangle$, $Dom = \langle D_1, \dots, D_n \rangle$, such that

$$\varphi(E) = \varphi\left(Dev_{C_{E,B,i}}(B)\right) = 1 \quad \text{and} \quad \varphi\left(Dev_{C_{E,B,i}}(T)\right) = 0.$$

It is shown that, under this assumption, there is at least one model $\mathcal{M}^* = \langle Dom^*, \varphi^* \rangle$, $Dom^* = \langle D_1^*, \dots, D_n^* \rangle$, such that

$$\varphi^*(E) = \varphi^*(B) = 1 \quad \text{and} \quad \varphi^*(T) = 0$$

– in contradiction to the assumption that $E \cup B \vdash T$.

Let $D_k^* = D_k$, $1 \leq k \neq i \leq n$, and

$$D_i^* = \{\alpha : \varphi(\text{‘}t^i\text{’}) = \alpha, \text{ for some constant } i\text{-term ‘}t^i\text{’ } \in C_i(E \cup B)\},$$

and note that

$$C_i\left(Dev_{C_{E,B,i}}(B) \to Dev_{C_{E,B,i}}(T)\right) = C_i(E \cup B) = C_{E,B,i}.$$

Let C be the set of all constant terms occurring in E, $Dev_{C_{E,B,i}}(B)$, or $Dev_{C_{E,B,i}}(T)$. Let

$$\varphi^*(\text{‘}a\text{’}) = \varphi(\text{‘}a\text{’}),$$

for every individual constant ‘a’ $\in C$,

$$\varphi^*\left(\text{‘}f^{k_{n+1}}\text{’}\right) = \varphi\left(\text{‘}f^{k_{n+1}}\text{’}\right) \cap \left\langle D_{k_1}^*, \dots, D_{k_{n+1}}^*\right\rangle, \quad D_{k_j}^* = \begin{cases} D_{k_j}, & \text{if } k_j \neq i, \\ D_i^*, & \text{if } k_j = i, \end{cases}$$

for every $(n+1)$-ary k_{n+1}-function symbol ‘$f^{k_{n+1}}$’ = ‘$f^{k_{n+1}}\left(x^{k_1}, \dots, x^{k_n}\right)$’ occurring in E, $Dev_{C_{E,B,i}}(B)$, or $Dev_{C_{E,B,i}}(T)$, and

$$\varphi^*(\text{‘}P^n\text{’}) = \varphi(\text{‘}P^n\text{’}) \cap \left\langle D_{k_1}^*, \dots, D_{k_n}^*\right\rangle, \quad D_{k_j}^* = \begin{cases} D_{k_j}, & \text{if } k_j \neq i, \\ D_i^*, & \text{if } k_j = i, \end{cases}$$

for every n-ary ($k_1, \dots, k_n$-) predicate ‘P^n’ = ‘$P^n\left(x^{k_1}, \dots, x^{k_n}\right)$’ occurring in E, $Dev_{C_{E,B,i}}(B)$, or $Dev_{C_{E,B,i}}(T)$.

Note that n-ary predicate ‘P^n’ occurs in $Dev_{C_{E,B,i}}(B)$ or $Dev_{C_{E,B,i}}(T)$ just in case ‘P^n’ occurs in B respectively T; and that $(n+1)$-ary function symbol ‘$f^{k_{n+1}}$’ occurs in $Dev_{C_{E,B,i}}(B)$ or $Dev_{C_{E,B,i}}(T)$, if (but not only if[3]) ‘$f^{k_{n+1}}$’ occurs in B respectively T.

[3] If ‘$f^{k_{n+1}}$’ is part of some constant i-term ‘t^i’ $\in C_i(E)$, but does not occur in B or T, then ‘$f^{k_{n+1}}$’ occurs in $Dev_{C_{E,B,i}}(B)$ and $Dev_{C_{E,B,i}}(T)$.

Let me first show that $\varphi^*(\text{‘}t\text{’}) = \varphi(\text{‘}t\text{’})$, for every constant term ‘t’ $\in C$, and, given this, that $\varphi^*(A) = \varphi(A)$, for every wff A in E, $Dev_{C_{E,B,i}}(B)$, or $Dev_{C_{E,B,i}}(T)$. It follows that

$$\varphi^*(E) = \varphi(E) = 1 = \varphi^*\left(Dev_{C_{E,B,i}}(B)\right) = \varphi\left(Dev_{C_{E,B,i}}(B)\right), \quad \text{and}$$
$$\varphi^*\left(Dev_{C_{E,B,i}}(T)\right) = \varphi\left(Dev_{C_{E,B,i}}(T)\right) = 0.$$

Finally it is shown that

$$\varphi^*\left(Dev_{C_{E,B,i}}(B)\right) = \varphi^*(B) = 1 \quad \text{and} \quad \varphi^*\left(Dev_{C_{E,B,i}}(T)\right) = \varphi^*(T) = 0,$$

which contradicts the assumption that $E \cup B \vdash T$.

By definition, $\varphi^*(\text{‘}a\text{’}) = \varphi(\text{‘}a\text{’})$, for every individual constant ‘a’ $\in C$. Let ‘t^{k_1}’, ..., ‘t^{k_n}’ be n constant k_j-terms, $1 \leq j \leq n$, let ‘$f^{k_{n+1}}$’ be an n-ary k_{n+1}-function symbol, and suppose ‘$f^{k_{n+1}}\left(t^{k_1}, \ldots, t^{k_1}\right)$’ $\in C$:

$$\begin{aligned}
\varphi^*\left(\text{‘}f^{k_{n+1}}\left(t^{k_1}, \ldots, t^{k_n}\right)\text{’}\right) &= \varphi^*\left(\text{‘}f^{k_{n+1}}\text{’}\right)\left(\varphi^*\left(\text{‘}t^{k_1}\text{’}\right), \ldots, \varphi^*\left(\text{‘}t^{k_n}\text{’}\right)\right) \\
&= \varphi^*\left(\text{‘}f^{k_{n+1}}\text{’}\right)\left(\varphi\left(\text{‘}t^{k_1}\text{’}\right), \ldots, \varphi\left(\text{‘}t^{k_n}\text{’}\right)\right) \\
&\quad \text{by induction hypothesis} \\
&= \varphi\left(\text{‘}f^{k_{n+1}}\text{’}\right)\left(\varphi\left(\text{‘}t^{k_1}\text{’}\right), \ldots, \varphi\left(\text{‘}t^{k_n}\text{’}\right)\right) \\
&\quad \varphi\left(\text{‘}t^{k_i}\text{’}\right) \in D^*_{k_i},\ \text{for every } i, 1 \leq i \leq n,\ \text{and} \\
&\quad \varphi\left(\text{‘}f^{k_{n+1}}\left(t^{k_1}, \ldots, t^{k_n}\right)\text{’}\right) \in D^*_{k_{n+1}},\ \text{because} \\
&\quad \text{‘}f^{k_{n+1}}\left(t^{k_1}, \ldots, t^{k_n}\right)\text{’} \in C \\
&= \varphi\left(\text{‘}f^{k_{n+1}}\left(t^{k_1}, \ldots, t^{k_n}\right)\text{’}\right).
\end{aligned}$$

Let A be a wff in E, $Dev_{C_{E,B,i}}(B)$, or $Dev_{C_{E,B,i}}(T)$. A contains no occurrence of an i-variable x^i (or a corresponding quantifier), for these are eliminated in $Dev_{C_{E,B,i}}(B)$ and $Dev_{C_{E,B,i}}(T)$, and do not occur in E, because D_i is an evidential domain of E.

(1) If A is atomic, i.e. if A is of the form $P^n\left(t^{k_1}, \ldots, t^{k_n}\right)$, ‘$t^{k_j}$’ $\in C$ being a constant k_j-term, $1 \leq j \leq n$, and ‘P^n’ being an n-ary ($k_1, \ldots, k_n$-) predicate, then

$$\begin{aligned}
\varphi^*\left(P^n\left(t^{k_1}, \ldots, t^{k_n}\right)\right) = 1 \quad &\text{iff} \quad \left\langle \varphi^*\left(\text{‘}t^{k_1}\text{’}\right), \ldots, \varphi^*\left(\text{‘}t^{k_n}\text{’}\right)\right\rangle \in \varphi^*\left(\text{‘}P^n\text{’}\right) \\
&\text{iff} \quad \left\langle \varphi\left(\text{‘}t^{k_1}\text{’}\right), \ldots, \varphi\left(\text{‘}t^{k_n}\text{’}\right)\right\rangle \in \varphi^*\left(\text{‘}P^n\text{’}\right)
\end{aligned}$$

$$\begin{aligned}
&\quad \text{by the above[4]}\\
\text{iff}\quad &\left\langle \varphi\left(\text{‘}t^{k_1}\text{’}\right),\ldots,\varphi\left(\text{‘}t^{k_n}\text{’}\right)\right\rangle \in \varphi\left(\text{‘}P^n\text{’}\right)\\
&\varphi^*\left(\text{‘}P^n\text{’}\right)=\varphi\left(\text{‘}P^n\text{’}\right)\cap\left\langle D^*_{k_1},\ldots,D^*_{k_n}\right\rangle=\\
&=\varphi\left(\text{‘}P^n\text{’}\right)^{5}\\
\text{iff}\quad &\varphi\left(P^n\left(t^{k_1},\ldots,t^{k_n}\right)\right)=1.
\end{aligned}$$

(2) If $A=\neg B$, then

$$\begin{aligned}
\varphi^*(A)=1 \quad \text{iff}\quad &\varphi^*(\neg B)=1\\
\text{iff}\quad &\varphi^*(B)=0\\
\text{iff}\quad &\varphi(B)=0 \quad \text{by induction hypothesis}\\
\text{iff}\quad &\varphi(\neg B)=1\\
\text{iff}\quad &\varphi(A)=1.
\end{aligned}$$

(3) If $A=B\wedge C$, then

$$\begin{aligned}
\varphi^*(A)=1 \quad \text{iff}\quad &\varphi^*(B\wedge C)=1\\
\text{iff}\quad &\varphi^*(B)=1 \text{ and } \varphi^*(C)=1\\
\text{iff}\quad &\varphi(B)=1 \text{ and } \varphi(C)=1 \quad \text{by induction hypothesis}\\
\text{iff}\quad &\varphi(B\wedge C)=1\\
\text{iff}\quad &\varphi(A)=1.
\end{aligned}$$

(4)-(5) Similarly for $A=B\vee C$ and $A=B\to C$.
(6) If $A=\forall x^k B\left[x^k\right]$, $k\neq i$, then

$$\begin{aligned}
\varphi^*(A)=1 \quad \text{iff}\quad &\varphi^*\left(\forall x^k B\left[x^k\right]\right)=1\\
\text{iff}\quad &\varphi^{*\prime}\left(B\left[x^k\right]\right)=1,\ \text{for every interpretation function } \varphi^{*\prime}\\
&\text{differing from } \varphi^* \text{ at most in the value for ‘}x^k\text{’}\\
\text{iff}\quad &\varphi'\left(B\left[x^k\right]\right)=1,\ \text{for every interpretation function } \varphi'\\
&\text{differing from } \varphi \text{ at most in the value for ‘}x^k\text{’}\\
&\text{by induction hypothesis, and because } D^*_k=D_k\\
\text{iff}\quad &\varphi\left(\forall x^k B\left[x^k\right]\right)=1\\
\text{iff}\quad &\varphi(A)=1.
\end{aligned}$$

[4] ‘t^{k_j}’ $\in C$, for every $j, 1\leq j\leq n$, whence $\varphi^*\left(\text{‘}t^{k_j}\text{’}\right)=\varphi\left(\text{‘}t^{k_j}\text{’}\right)$, for every $j, 1\leq j\leq n$.
[5] $\varphi\left(\text{‘}t^{k_j}\text{’}\right)\in D^*_{k_j}$, since ‘$t^{k_j}$’ $\in C$, for every $j, 1\leq j\leq n$.

(7) Similarly for $A = \exists x^k B\left[x^k\right]$, $k \neq i$.
Thus

$$\varphi^*(E) = \varphi(E) = 1 = \varphi^*\left(Dev_{C_{E,B,i}}(B)\right) = \varphi\left(Dev_{C_{E,B,i}}(B)\right), \quad \text{and}$$
$$\varphi^*\left(Dev_{C_{E,B,i}}(T)\right) = \varphi\left(Dev_{C_{E,B,i}}(T)\right) = 0.$$

Let me now show by induction on the length of the conjunction $\bigwedge_{h \in B} h$ of all wffs $h \in B$ and the conjunction $\bigwedge_{h \in T} h$ of all wffs $h \in T$ that

$$\varphi^*(B) = \varphi^*\left(Dev_{C_{E,B,i}}(B)\right) \quad \text{and} \quad \varphi^*(T) = \varphi^*\left(Dev_{C_{E,B,i}}(T)\right),$$

where 'B' is short for '$\bigwedge_{h \in B} h$', and 'T' is short for '$\bigwedge_{h \in T} h$'. Let 'S' be 'T' or 'B'.
(1) If S is atomic, i.e. if S is of the form $P^n\left(t^{k_1}, \ldots, t^{k_n}\right)$, '$t^{k_j}$' $\in C$ being a constant k_j-term, $1 \leq j \leq n$, and 'P^n' being an n-ary ($k_1, \ldots, k_n$-) predicate occurring in E, $Dev_{C_{E,B,i}}(B)$, or $Dev_{C_{E,B,i}}(T)$, then

$$\begin{aligned}
\varphi^*(S) = 1 \quad & \text{iff} \quad \varphi^*\left(P^n\left(t^{k_1}, \ldots, t^{k_n}\right)\right) \\
& \text{iff} \quad \varphi^*\left(Dev_{C_{E,B,i}}\left(P^n\left(t^{k_1}, \ldots, t^{k_n}\right)\right)\right) = 1 \quad \text{definition of} \\
& \qquad \text{the development } Dev_C(T) \text{ of (a finite set of) wff(s) } T \\
& \qquad \text{for a finite set of constant } i\text{-terms } C \\
& \text{iff} \quad \varphi^*\left(Dev_{C_{E,B,i}}(S)\right) = 1.
\end{aligned}$$

(2) If $S = \neg A$, then

$$\begin{aligned}
\varphi^*(S) = 1 \quad & \text{iff} \quad \varphi^*(\neg A) = 1 \\
& \text{iff} \quad \varphi^*(A) = 0 \\
& \text{iff} \quad \varphi^*\left(Dev_{C_{E,B,i}}(A)\right) = 0 \quad \text{by induction hypothesis} \\
& \text{iff} \quad \varphi^*\left(\neg Dev_{C_{E,B,i}}(A)\right) = 1 \\
& \text{iff} \quad \varphi^*\left(Dev_{C_{E,B,i}}(\neg A)\right) = 1 \quad \text{definition of the develop-} \\
& \qquad \text{ment } Dev_C(T) \text{ of (a finite set of) wff(s) } T \text{ for a} \\
& \qquad \text{finite set of constant } i\text{-terms } C \\
& \text{iff} \quad \varphi^*\left(Dev_{C_{E,B,i}}(S)\right) = 1.
\end{aligned}$$

(3) If $S = A \wedge B$, then

$$\varphi^*(S) = 1 \quad \text{iff} \quad \varphi^*(A \wedge B) = 1$$

iff $\varphi^*(A) = 1$ and $\varphi^*(B) = 1$
iff $\varphi^*\left(Dev_{C_{E,B,i}}(A)\right) = 1$ and $\varphi^*\left(Dev_{C_{E,B,i}}(B)\right) = 1$ by induction hypothesis
iff $\varphi^*\left(Dev_{C_{E,B,i}}(A) \wedge Dev_{C_{E,B,i}}(B)\right) = 1$
iff $\varphi^*\left(Dev_{C_{E,B,i}}(A \wedge B)\right) = 1$ definition of the development $Dev_C(T)$ of (a finite set of) wff(s) T for a finite set of constant i-terms C
iff $\varphi^*\left(Dev_{C_{E,B,i}}(S)\right) = 1.$

(4)-(5) Similarly for $S = A \vee B$ and $S = A \to B$.
(6) If $S = \forall x^i A[x^i]$, then

$\varphi^*(S) = 1$ iff $\varphi^*\left(\forall x^i A\left[x^i\right]\right) = 1$
iff $\varphi^{*\prime}\left(A\left[x^i\right]\right) = 1$ for every interpretation function $\varphi^{*\prime}$ differing from φ^* at most in the value for 'x^i'
iff $\varphi^*\left(\bigwedge_{\text{'}t^i\text{'} \in C_{E,B,i}} A\left[t^i/x^i\right]\right) = 1$ for every $\alpha \in D_i^*$ there is at least one 't^i' $\in C_{E,B,i}$ such that $\varphi^*(\text{'}t^i\text{'}) = \alpha$
iff $\varphi^*\left(Dev_{C_{E,B,i}}\left(\forall x^i A\left[x^i\right]\right)\right) = 1$ definition of the development $Dev_C(T)$ of (a finite set of) wff(s) T for a finite set of constant i-terms C
iff $\varphi^*\left(Dev_{C_{E,B,i}}(S)\right) = 1.$

(7) If $S = \exists x^i A[x^i]$, then

$\varphi^*(S) = 1$ iff $\varphi^*\left(\exists x^i A\left[x^i\right]\right) = 1$
iff $\varphi^{*\prime}\left(A\left[x^i\right]\right) = 1$ for at least one interpretation function $\varphi^{*\prime}$ differing from φ^* at most in the value for 'x^i'
iff $\varphi^*\left(\bigvee_{\text{'}t^i\text{'} \in C_{E,B,i}} A\left[t^i/x^i\right]\right) = 1$ for every $\alpha \in D_i^*$ there is at least one 't^i' $\in C_{E,B,i}$ such that $\varphi^*(\text{'}t^i\text{'}) = \alpha$

iff $\varphi^*\left(Dev_{C_{E,B,i}}\left(\exists x^i A\left[x^i\right]\right)\right)=1$ definition of the development $Dev_C\left(T\right)$ of (a finite set of) wff(s) T for a finite set of constant i-terms C

iff $\varphi^*\left(Dev_{C_{E,B,i}}\left(S\right)\right)=1.$

(8) If $S=\forall x^k A\left[x^k\right]$, $k\neq i$, then

$\varphi^*\left(S\right)=1$ iff $\varphi^*\left(\forall x^k A\left[x^k\right]\right)=1$

iff $\varphi^{*\prime}\left(A\left[x^k\right]\right)=1$ for every interpretation function $\varphi^{*\prime}$ differing from φ^* at most in the value for 'x^k'

iff $\varphi^{*\prime}\left(Dev_{C_{E,B,i}}\left(A\left[x^k\right]\right)\right)=1$ for every interpretation function $\varphi^{*\prime}$ differing from φ^* at most in the value for 'x^k' by induction hypothesis

iff $\varphi^*\left(\forall x^k Dev_{C_{E,B,i}}\left(A\left[x^k\right]\right)\right)=1$

iff $\varphi^*\left(Dev_{C_{E,B,i}}\left(\forall x^k A\left[x^k\right]\right)\right)=1$ definition of the development $Dev_C\left(T\right)$ of (a finite set of) wff(s) T for a finite set of constant i-terms C

iff $\varphi^*\left(Dev_{C_{E,B,i}}\left(S\right)\right)=1.$

(9) Similarly for $S=\exists x^k A\left[x^k\right]$, $k\neq i$.
It follows that

$$\varphi^*\left(B\right)=\varphi^*\left(Dev_{C_{E,B,i}}\left(B\right)\right)=1 \quad \text{and} \quad \varphi^*\left(T\right)=\varphi^*\left(Dev_{C_{E,B,i}}\left(T\right)\right)=0,$$

since

$$\varphi^*\left(Dev_{C_{E,B,i}}\left(B\right)\right)=\varphi\left(Dev_{C_{E,B,i}}\left(B\right)\right)=1, \quad \text{and}$$
$$\varphi^*\left(Dev_{C_{E,B,i}}\left(T\right)\right)=\varphi\left(Dev_{C_{E,B,i}}\left(T\right)\right)=0.$$

So there is at least one model $\mathcal{M}^*=\langle Dom^*,\varphi^*\rangle$, $Dom^*=\langle D_1^*,\ldots,D_n^*\rangle$, such that

$$\varphi^*\left(E\right)=\varphi^*\left(B\right)=1 \quad \text{and} \quad \varphi^*\left(T\right)=0$$

– in contradiction to the assumption that $E\cup B\vdash T$.

(C) Suppose $T'\vdash T$, and let $D_1,\ldots,D_n$ be the domains of T and T' (i.e. there occur variables and constants of n different sorts in T and T'). It suffices to show that

$$E \vdash Dev_{C_{E,B,i}}(B) \rightarrow Dev_{C_{T'}}(T),$$

for every $C_{T'} \subseteq C_i(E \cup B)$ with

$$E \vdash Dev_{C_{E,B,i}}(B) \rightarrow Dev_{C_{T'}}(T').$$

For then

$$\max_{\mathcal{LI}}(T, E, B; D_i) \geq \max_{\mathcal{LI}}(T', E, B; D_i),$$

and thus

$$\mathcal{LI}(T, E, B; D_i) \geq \mathcal{LI}(T', E, B; D_i).$$

Suppose

$$E \vdash Dev_{C_{E,B,i}}(B) \rightarrow Dev_{C_{T'}}(T'),$$

but

$$E \nvdash Dev_{C_{E,B,i}}(B) \rightarrow Dev_{C_{T'}}(T),$$

for some $C_{T'} \subseteq C_i(E \cup B)$. Then

$$Dev_{C_{T'}}(T') \nvdash Dev_{C_{T'}}(T).$$

So there is at least one model $\mathcal{M} = \langle Dom, \varphi \rangle$, $Dom = \langle D_1, \ldots, D_n \rangle$, such that

$$\varphi\left(Dev_{C_{T'}}(T')\right) = 1 \quad \text{and} \quad \varphi\left(Dev_{C_{T'}}(T)\right) = 0.$$

Once more it is shown that, under this assumption, there is at least one model $\mathcal{M}^* = \langle Dom^*, \varphi^* \rangle$, $Dom^* = \langle D_1^*, \ldots, D_n^* \rangle$, such that

$$\varphi^*(T') = 1 \quad \text{and} \quad \varphi^*(T) = 0$$

– in contradiction to the assumption that $T' \vdash T$.

Let $D_k^* = D_k$, $1 \leq k \neq i \leq n$, and

$$D_i^* = \{\alpha : \varphi(`t^{i}\text{'}) = \alpha, \text{ for some constant } i\text{-term } `t^{i}\text{'} \in C_{T'}\},$$

and note that

$$C_i\left(Dev_{C_{T'}}(T')\right) \cup C_i\left(Dev_{C_{T'}}(T)\right) = C_{T'},$$

because D_i is among the domains of proper investigation of both T and T'.

Let C be the set of all constant terms occurring in $Dev_{C_{T'}}(T')$ or $Dev_{C_{T'}}(T)$. Let

$$\varphi^*(\text{‘}a\text{’}) = \varphi(\text{‘}a\text{’}),$$

for every individual constant ‘a’ $\in C$,

$$\varphi^*\left(\text{‘}f^{k_{n+1}}\text{’}\right) = \varphi\left(\text{‘}f^{k_{n+1}}\text{’}\right) \cap \left\langle D^*_{k_1}, \ldots, D^*_{k_{n+1}}\right\rangle, \quad D^*_{k_j} = \begin{cases} D_{k_j}, & \text{if } k_j \neq i, \\ D^*_i, & \text{if } k_j = i, \end{cases}$$

for every $(n+1)$-ary k_{n+1}-function symbol ‘$f^{k_{n+1}}$’ = ‘$f^{k_{n+1}}\left(x^{k_1}, \ldots, x^{k_n}\right)$’ occurring in $Dev_{C_{T'}}(T')$ or $Dev_{C_{T'}}(T)$, and

$$\varphi^*(\text{‘}P^n\text{’}) = \varphi(\text{‘}P^n\text{’}) \cap \left\langle D^*_{k_1}, \ldots, D^*_{k_n}\right\rangle, \quad D^*_{k_j} = \begin{cases} D_{k_j}, & \text{if } k_j \neq i, \\ D^*_i, & \text{if } k_j = i, \end{cases}$$

for every n-ary ($k_1, \ldots, k_n$-) predicate ‘P^n’ = ‘$P^n\left(x^{k_1}, \ldots, x^{k_n}\right)$’ occurring in $Dev_{C_{T'}}(T')$ or $Dev_{C_{T'}}(T)$.

As before, n-ary predicate ‘P^n’ occurs in $Dev_{C_{T'}}(T')$ or $Dev_{C_{T'}}(T)$ just in case ‘P^n’ occurs in T' respectively T; and $(n+1)$-ary function symbol ‘$f^{k_{n+1}}$’ occurs in $Dev_{C_{T'}}(T')$ or $Dev_{C_{T'}}(T)$, if (but not only if) ‘$f^{k_{n+1}}$’ occurs in T' respectively T.

I first show that $\varphi^*(\text{‘}t\text{’}) = \varphi(\text{‘}t\text{’})$, for every constant term ‘t’ $\in C$; given this, I show that $\varphi^*(A) = \varphi(A)$, for every wff A in $Dev_{C_{T'}}(T')$ or $Dev_{C_{T'}}(T)$. It follows again that

$$\begin{array}{l} \varphi^*\left(Dev_{C_{T'}}(T')\right) = \varphi\left(Dev_{C_{T'}}(T')\right) = 1, \quad \text{and} \\ \varphi^*\left(Dev_{C_{T'}}(T)\right) = \varphi\left(Dev_{C_{T'}}(T)\right) = 0. \end{array}$$

Finally, it is shown that

$$\varphi^*\left(Dev_{C_{T'}}(T')\right) = \varphi^*(T') = 1 \quad \text{and} \quad \varphi^*\left(Dev_{C_{T'}}(T)\right) = \varphi^*(T) = 0,$$

which contradicts the assumption that $T' \vdash T$.

By definition, $\varphi^*(\text{‘}a\text{’}) = \varphi(\text{‘}a\text{’})$, for every individual constant ‘a’ $\in C$. Let ‘t^{k_1}’, ..., ‘t^{k_n}’ be n constant k_j-terms, $1 \leq j \leq n$, let ‘$f^{k_{n+1}}$’ be an n-ary k_{n+1}-function symbol, and suppose ‘$f^{k_{n+1}}\left(t^{k_1}, \ldots, t^{k_1}\right)$’ $\in C$:

$$\varphi^*\left(\text{‘}f^{k_{n+1}}\left(t^{k_1}, \ldots, t^{k_n}\right)\text{’}\right) \quad = \quad \varphi^*\left(\text{‘}f^{k_{n+1}}\text{’}\right)\left(\varphi^*\left(\text{‘}t^{k_1}\text{’}\right), \ldots, \varphi^*\left(\text{‘}t^{k_n}\text{’}\right)\right)$$

$$\begin{aligned}
&= \varphi^*\left(\text{‘}f^{k_{n+1}}\text{’}\right)\left(\varphi\left(\text{‘}t^{k_1}\text{’}\right),\ldots,\varphi\left(\text{‘}t^{k_n}\text{’}\right)\right) \\
&\quad \text{by induction hypothesis} \\
&= \varphi\left(\text{‘}f^{k_{n+1}}\text{’}\right)\left(\varphi\left(\text{‘}t^{k_1}\text{’}\right),\ldots,\varphi\left(\text{‘}t^{k_n}\text{’}\right)\right) \\
&\quad \varphi\left(\text{‘}t^{k_i}\text{’}\right) \in D^*_{k_i},\ \text{for every } i, 1 \leq i \leq n,\ \text{and} \\
&\quad \varphi\left(\text{‘}f^{k_{n+1}}\left(t^{k_1},\ldots,t^{k_n}\right)\text{’}\right) \in D^*_{k_{n+1}},\ \text{because} \\
&\quad \text{‘}f^{k_{n+1}}\left(t^{k_1},\ldots,t^{k_n}\right)\text{’} \in C \\
&= \varphi\left(\text{‘}f^{k_{n+1}}\left(t^{k_1},\ldots,t^{k_n}\right)\text{’}\right).
\end{aligned}$$

Let A be a wff in $Dev_{C_{T'}}(T')$ or $Dev_{C_{T'}}(T)$. A contains no occurrence of an i-variable x^i (or a corresponding quantifier), for these are eliminated.
(1) If A is atomic, i.e. if A is of the form $P^n\left(t^{k_1},\ldots,t^{k_n}\right)$, ‘$t^{k_j}$’ $\in C$ being a constant k_j-term, $1 \leq j \leq n$, and ‘P^n’ being an n-ary ($k_1,\ldots,k_n$-) predicate, then

$$\begin{aligned}
\varphi^*\left(P^n\left(t^{k_1},\ldots,t^{k_n}\right)\right) = 1 \quad &\text{iff} \quad \left\langle \varphi^*\left(\text{‘}t^{k_1}\text{’}\right),\ldots,\varphi^*\left(\text{‘}t^{k_n}\text{’}\right)\right\rangle \in \varphi^*\left(\text{‘}P^n\text{’}\right) \\
&\text{iff} \quad \left\langle \varphi\left(\text{‘}t^{k_1}\text{’}\right),\ldots,\varphi\left(\text{‘}t^{k_n}\text{’}\right)\right\rangle \in \varphi^*\left(\text{‘}P^n\text{’}\right) \\
&\qquad \text{by the above} \\
&\text{iff} \quad \left\langle \varphi\left(\text{‘}t^{k_1}\text{’}\right),\ldots,\varphi\left(\text{‘}t^{k_n}\text{’}\right)\right\rangle \in \varphi\left(\text{‘}P^n\text{’}\right) \\
&\qquad \varphi^*\left(\text{‘}P^n\text{’}\right) = \varphi\left(\text{‘}P^n\text{’}\right) \cap \left\langle D^*_{k_1},\ldots,D^*_{k_n}\right\rangle = \\
&\qquad = \varphi\left(\text{‘}P^n\text{’}\right) \\
&\text{iff} \quad \varphi\left(P^n\left(t^{k_1},\ldots,t^{k_n}\right)\right) = 1.
\end{aligned}$$

(2) If $A = \neg B$, then

$$\begin{aligned}
\varphi^*(A) = 1 \quad &\text{iff} \quad \varphi^*(\neg B) = 1 \\
&\text{iff} \quad \varphi^*(B) = 0 \\
&\text{iff} \quad \varphi(B) = 0 \quad \text{by induction hypothesis} \\
&\text{iff} \quad \varphi(\neg B) = 1 \\
&\text{iff} \quad \varphi(A) = 1.
\end{aligned}$$

(3) If $A = B \wedge C$, then

$$\varphi^*(A) = 1 \quad \text{iff} \quad \varphi^*(B \wedge C) = 1$$

$$\begin{aligned}
&\text{iff} \quad \varphi^*(B) = 1 \text{ and } \varphi^*(C) = 1 \\
&\text{iff} \quad \varphi(B) = 1 \text{ and } \varphi(C) = 1 \quad \text{by induction hypothesis} \\
&\text{iff} \quad \varphi(B \wedge C) = 1 \\
&\text{iff} \quad \varphi(A) = 1.
\end{aligned}$$

(4)-(5) Similarly for $A = B \vee C$ and $A = B \to C$.
(6) If $A = \forall x^k B\left[x^k\right]$, $k \neq i$, then

$$\begin{aligned}
\varphi^*(A) = 1 \quad &\text{iff} \quad \varphi^*\left(\forall x^k B\left[x^k\right]\right) = 1 \\
&\text{iff} \quad \varphi^{*\prime}\left(B\left[x^k\right]\right) = 1, \text{ for every interpretation function } \varphi^{*\prime} \\
&\qquad \text{differing from } \varphi^* \text{ at most in the value for `}x^k\text{'} \\
&\text{iff} \quad \varphi'\left(B\left[x^k\right]\right) = 1, \text{ for every interpretation function } \varphi' \\
&\qquad \text{differing from } \varphi \text{ at most in the value for `}x^k\text{'} \\
&\qquad \text{by induction hypothesis, and because } D_k^* = D_k \\
&\text{iff} \quad \varphi\left(\forall x^k B\left[x^k\right]\right) = 1 \\
&\text{iff} \quad \varphi(A) = 1.
\end{aligned}$$

(7) Similarly for $A = \exists x^k B\left[x^k\right]$, $k \neq i$.
Thus

$$\begin{aligned}
&\varphi^*\left(Dev_{C_{T'}}(T')\right) = \varphi\left(Dev_{C_{T'}}(T')\right) = 1, \quad \text{and} \\
&\varphi^*\left(Dev_{C_{T'}}(T)\right) = \varphi\left(Dev_{C_{T'}}(T)\right) = 0.
\end{aligned}$$

Let me now show by induction on the length of the conjunction $\bigwedge_{h\in T'} h$ of all wffs $h \in T'$ and the conjunction $\bigwedge_{h\in T} h$ of all wffs $h \in T$ that

$$\varphi^*(T') = \varphi^*\left(Dev_{C_{T'}}(T')\right) \quad \text{and} \quad \varphi^*(T) = \varphi^*\left(Dev_{C_{T'}}(T)\right),$$

where 'T'' is short for '$\bigwedge_{h\in T'} h$', and 'T' is short for '$\bigwedge_{h\in T} h$'. Let 'S' be 'T' or 'T''.
(1) If S is atomic, i.e. if S is of the form $P^n\left(t^{k_1}, \ldots, t^{k_n}\right)$, '$t^{k_j}$' $\in C$ being a constant k_j-term, $1 \leq j \leq n$, and 'P^n' being an n-ary ($k_1, \ldots, k_n$-) predicate occurring in $Dev_{C_{T'}}(T')$ or $Dev_{C_{T'}}(T)$, then

$$\begin{aligned}
\varphi^*(S) = 1 \quad &\text{iff} \quad \varphi^*\left(P^n\left(t^{k_1}, \ldots, t^{k_n}\right)\right) \\
&\text{iff} \quad \varphi^*\left(Dev_{C_{T'}}\left(P^n\left(t^{k_1}, \ldots, t^{k_n}\right)\right)\right) = 1 \quad \text{definition of}
\end{aligned}$$

the development $Dev_C(T)$ of (a finite set of) wff(s) T for a finite set of constant i-terms C

iff $\varphi^*\left(Dev_{C_{T'}}(S)\right) = 1.$

(2) If $S = \neg A$, then

$$\begin{aligned}
\varphi^*(S) = 1 \quad & \text{iff} \quad \varphi^*(\neg A) = 1 \\
& \text{iff} \quad \varphi^*(A) = 0 \\
& \text{iff} \quad \varphi^*\left(Dev_{C_{T'}}(A)\right) = 0 \quad \text{by induction hypothesis} \\
& \text{iff} \quad \varphi^*\left(\neg Dev_{C_{T'}}(A)\right) = 1 \\
& \text{iff} \quad \varphi^*\left(Dev_{C_{T'}}(\neg A)\right) = 1 \quad \text{definition of the development } Dev_C(T) \text{ of (a finite set of) wff(s) } T \text{ for a finite set of constant } i\text{-terms } C \\
& \text{iff} \quad \varphi^*\left(Dev_{C_{T'}}(S)\right) = 1.
\end{aligned}$$

(3) If $S = A \wedge B$, then

$$\begin{aligned}
\varphi^*(S) = 1 \quad & \text{iff} \quad \varphi^*(A \wedge B) = 1 \\
& \text{iff} \quad \varphi^*(A) = 1 \text{ and } \varphi^*(B) = 1 \\
& \text{iff} \quad \varphi^*\left(Dev_{C_{T'}}(A)\right) = 1 \text{ and } \varphi^*\left(Dev_{C_{T'}}(B)\right) = 1 \text{ by induction hypothesis} \\
& \text{iff} \quad \varphi^*\left(Dev_{C_{T'}}(A) \wedge Dev_{C_{T'}}(B)\right) = 1 \\
& \text{iff} \quad \varphi^*\left(Dev_{C_{T'}}(A \wedge B)\right) = 1 \quad \text{definition of the development } Dev_C(T) \text{ of (a finite set of) wff(s) } T \text{ for a finite set of constant } i\text{-terms } C \\
& \text{iff} \quad \varphi^*\left(Dev_{C_{T'}}(S)\right) = 1.
\end{aligned}$$

(4)-(5) Similarly for $S = A \vee B$ and $S = A \rightarrow B$.
(6) If $S = \forall x^i A[x^i]$, then

$$\begin{aligned}
\varphi^*(S) = 1 \quad & \text{iff} \quad \varphi^*\left(\forall x^i A\left[x^i\right]\right) = 1 \\
& \text{iff} \quad \varphi^{*\prime}\left(A\left[x^i\right]\right) = 1 \text{ for every interpretation function } \varphi^{*\prime} \text{ differing from } \varphi^* \text{ at most in the value for ‘}x^i\text{’}
\end{aligned}$$

$$\begin{aligned}
&\text{iff} \quad \varphi^* \left(\bigwedge_{\text{‘}t^i\text{’} \in C_{T'}} A\left[t^i/x^i\right] \right) = 1 \quad \text{for every } \alpha \in D_i^* \text{ there} \\
&\qquad \text{is at least one ‘}t^i\text{’} \in C_{T'} \text{ such that } \varphi^*\left(\text{‘}t^i\text{’}\right) = \alpha \\
&\text{iff} \quad \varphi^*\left(Dev_{C_{T'}}\left(\forall x^i A\left[x^i\right]\right)\right) = 1 \quad \text{definition of the} \\
&\qquad \text{development } Dev_C\left(T\right) \text{ of (a finite set of) wff(s) } T \\
&\qquad \text{for a finite set of constant } i\text{-terms } C \\
&\text{iff} \quad \varphi^*\left(Dev_{C_{T'}}\left(S\right)\right) = 1.
\end{aligned}$$

(7) Similarly for $S = \exists x^i A\left[x^i\right]$.
(8) If $S = \forall x^k A\left[x^k\right]$, $k \neq i$, then

$$\begin{aligned}
\varphi^*\left(S\right) = 1 \quad &\text{iff} \quad \varphi^*\left(\forall x^k A\left[x^k\right]\right) = 1 \\
&\text{iff} \quad \varphi^{*\prime}\left(A\left[x^k\right]\right) = 1 \text{ for every interpretation function } \varphi^{*\prime} \\
&\qquad \text{differing from } \varphi^* \text{ at most in the value for ‘}x^k\text{’} \\
&\text{iff} \quad \varphi^{*\prime}\left(Dev_{C_{T'}}\left(A\left[x^k\right]\right)\right) = 1 \text{ for every interpretation} \\
&\qquad \text{function } \varphi^{*\prime} \text{ differing from } \varphi^* \text{ at most in the value for ‘}x^k\text{’} \\
&\qquad \text{by induction hypothesis} \\
&\text{iff} \quad \varphi^{*\prime}\left(\forall x^k Dev_{C_{T'}}\left(A\left[x^k\right]\right)\right) = 1 \\
&\text{iff} \quad \varphi^*\left(Dev_{C_{T'}}\left(\forall x^k A\left[x^k\right]\right)\right) = 1 \quad \text{definition of the} \\
&\qquad \text{development } Dev_C\left(T\right) \text{ of (a finite set of) wff(s) } T \\
&\qquad \text{for a finite set of constant } i\text{-terms } C \\
&\text{iff} \quad \varphi^*\left(Dev_{C_{T'}}\left(S\right)\right) = 1.
\end{aligned}$$

(9) Similarly for $S = \exists x^k A\left[x^k\right]$, $k \neq i$.
It follows that

$$\varphi^*\left(T'\right) = \varphi^*\left(Dev_{C_{T'}}\left(T'\right)\right) = 1 \quad \text{and} \quad \varphi^*\left(T\right) = \varphi^*\left(Dev_{C_{T'}}\left(T\right)\right) = 0,$$

since

$$\begin{aligned}
&\varphi^*\left(Dev_{C_{T'}}\left(T'\right)\right) = \varphi\left(Dev_{C_{T'}}\left(T'\right)\right) = 1, \quad \text{and} \\
&\varphi^*\left(Dev_{C_{T'}}\left(T\right)\right) = \varphi\left(Dev_{C_{T'}}\left(T\right)\right) = 0.
\end{aligned}$$

So there is at least one model $\mathcal{M}^* = \langle Dom^*, \varphi^* \rangle$, $Dom^* = \langle D_1^*, \ldots, D_n^* \rangle$, such that

$$\varphi^*\left(T'\right) = 1 \quad \text{and} \quad \varphi^*\left(T\right) = 0$$

– in contradiction to the assumption that $T' \vdash T$. $\square$

Theories T and background knowledges B have to be finite, for otherwise $Dev_C(T)$ and $Dev_C(B)$ are not defined, for any finite set of constant terms C. The following example shows that – for the definition given (cf. the remark below) – it is also necessary that theories do not contain occurrences of constant i-terms:
(1) Let $E = \{Fa, Gb\}$, $B = \emptyset$, $T = \{\forall x\,(Fx \wedge Fb)\}$, and $T' = \{\forall x Fx\}$. Then

$$T' \vdash T, \quad E \vdash Dev_{\{`a'\}}(B) \rightarrow Dev_{\{`a'\}}(T'), \quad \text{and}$$
$$E \nvdash Dev_{\{`a'\}}(B) \rightarrow Dev_{\{`a'\}}(T),$$

i.e.

$$\forall x Fx \vdash \forall x\,(Fx \wedge Fb), \quad Fa, Gb \vdash Fa, \quad \text{and} \quad Fa, Gb \nvdash Fa \wedge Fb,$$

whence

$$\mathcal{LI}(T', E, B; D) = 1/2 > 0 = \mathcal{LI}(T, E, B; D),$$

which violates the third condition in the definition of indicating truth, where D is the domain corresponding to the variable 'x' and the individual constants 'a' and 'b'.

It is also necessary to consider the constant i-terms occurring in both E and B, as illustrated by the second example.
(2) Let $E = \{Fa\}$, $B = \{Gb\}$, and $T = \{\exists x Gx\}$. Then

$$E \cup B \vdash T, \quad \text{and} \quad E \nvdash Dev_{\{`a'\}}(B) \rightarrow Dev_{\{`a'\}}(T),$$

i.e.

$$Fa, Gb \vdash \exists x Gx, \quad \text{and} \quad Fa \nvdash Gb \rightarrow Ga,$$

whence

$$\mathcal{LI}(T, E, B; D) = 0 < 1,$$

which violates the second condition of the definition of indicating truth.

Only recently – and too late in order to rewrite this dissertation – have I realized that by considering $C_i(E \cup B \cup T)$ in the definition of $\max_{\mathcal{LI}}(T, E, B; D_i)$ one can drop the assumption that theories T do not talk about particular individuals of their domains of proper investigation, i.e. one can drop the condition that T contains no occurrences of constant i-terms, if D_i is among T's domains of proper investigation. Though I still think that this restriction is appropriate, it is, of course, always better to do without some assumptions or with weaker ones.

Moreover, it is not even necessary to demand of an evidence with evidential domains $D_1, \ldots, D_n$ to contain no occurrences of i-variables (or corresponding quantifiers), $1 \leq i \leq n$. Just consider

$$\max\left\{|C \cap C_{B-repr}(E)| : C \subseteq C_{T,E,B,i}, Dev_{C_{T,E,B,i}}(E \wedge B) \vdash Dev_C(T)\right\},$$

where $C_{T,E,B,i} := C_i(T \cup E \cup B)$.

Appendix F

Proofs for chapter 6

F.1 Proof of Theorem 6.1

Theorem F.1 (G Is Formally Handy) $G(\cdot,\cdot,\cdot)$, $G(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, is non-arbitrary, comprehensible, computable in the limit, and closed under equivalence transformations of T.

Proof.
G is non-arbitrary and comprehensible, because it is a single function (without parameters that can be chosen freely) which is defined in the terms of $PL1 =$ and ZF.

G is computable in the limit, because all one has to determine for a given theory T, a given evidence E, and a given background knowledge B are (1) the confirmational domains of T and E; (2) the set of predicates essentially occurring in T, $PR_{ess}(T)$; (3) whether the logical consequence relation holds between various sets of statements; and (4) whether various sets of negated or unnegated one-place i-predicates are subsets of other such sets – the functions $\log$, $\cdot$, and $\div$ preserve computability in the limit, because they are computable. The proofs of theorems 4.5 and 5.2 are sufficient to show how to construct a method that stabilizes to the correct answer to all these questions.

G is closed under equivalence transformations of T, for if $T \dashv\vdash T'$, then (i) $PR_{ess}(T) = PR_{ess}(T')$, and (ii) T accounts for a class of facts CF just in case T' does. $\square$

F.2 Proof of Theorem 6.2

Theorem F.2 (G Supports Gathering Evidence) $G(\cdot,\cdot,\cdot)$, $G(\cdot,\cdot,\cdot) : \mathcal{T} \times \mathcal{E} \times \mathcal{B} \rightarrow \Re$, supports gathering evidence, if its definition is based on proper classes of i-facts, and if $C_{B-repr}(E) \subseteq C_{B-repr}(E')$.

More precisely, for every theory T, any evidences E and E', every background knowledge B, and every confirmational domain D_i of T and E:

> If $E' \vdash E$ and $C_{B-repr}(E) \subseteq C_{B-repr}(E')$, then $G(T, E', B; D_i) \geq G(T, E, B; D_i)$.

Proof.
Let T be a theory, let E and E' be evidences, let B be a background knowledge, and let D_i be a confirmational domain of T and E, and of T and E'. Suppose $E' \vdash E$ and $C_{B-repr}(E) \subseteq C_{B-repr}(E')$.

Let $CF_1^i, \ldots, CF_m^i$ be the classes of i-facts T, E, and B give rise to, and let $C_1^i, \ldots, C_m^i$ be the corresponding sets of negated or unnegated one-place i-predicates which induce $CF_1^i, \ldots, CF_m^i$, respectively, relative to T, E, and B. $C_1^i, \ldots, C_m^i$ are generated by PR_1^i, which is generated by $PR_{ess}(T)$ and $C_{B-repr}(E)$.

By assumption $C_{B-repr}(E) \subseteq C_{B-repr}(E')$, whence the set of one-place i-predicates PR_1^i generated by $PR_{ess}(T)$ and $C_{B-repr}(E)$ is a subset of the set of one-place i-predicates PR_1' generated by $PR_{ess}(T)$ and $C_{B-repr}(E')$.

Let $CF_1^{i\prime}, \ldots, CF_n^{i\prime}$, $n \geq m$, be the classes of i-facts T, E', and B give rise to, and let $C_1^{i\prime}, \ldots, C_n^{i\prime}$ be the corresponding sets of negated or unnegated one-place i-predicates which induce $CF_1^{i\prime}, \ldots, CF_n^{i\prime}$, respectively, relative to T, E', and B ($C_1^{i\prime}, \ldots, C_n^{i\prime}$ are generated by PR_1').

As $PR_1 \subseteq PR_1'$, it follows that for every C_j^i, $1 \leq j \leq m$, there is a $C_{j'}^{i\prime}$, $1 \leq j' \leq n$, such that $C_j^i = C_{j'}^{i\prime}$. Since $C_{B-repr}(E) \subseteq C_{B-repr}(E')$, this implies that for every class of i-facts CF_j^i, $1 \leq j \leq m$, there is a class of i-facts $CF_{j'}^{i\prime}$, $1 \leq j' \leq n$, among $CF_1^{i\prime}, \ldots, CF_n^{i\prime}$ such that $CF_j^i \subseteq CF_{j'}^{i\prime}$ and $C_j^i = C_{j'}^{i\prime}$ (though both CF_j^i and $CF_{j'}^{i\prime}$ may be empty).

Suppose CF_j^i, $1 \leq j \leq m$, is a proper class of i-facts relative to T, E, and B. This means that T accounts for CF_j^i in E relative to B, and that there is no class of i-facts $C_k^i \subset C_j^i$, $1 \leq k \leq m$, such that T accounts for CF_k^i in E relative to B. That T accounts for CF_j^i in E relative to B holds independently of E, whence T also accounts for the class of i-facts $CF_{j'}^{i\prime}$ in E' relative to B.

It remains to be shown that $CF_{j'}^{i\prime}$ is a *proper* class of i-facts relative to T, E', and B. So it has to be shown that there is no class of i-facts $C_l^{i\prime} \subset C_{j'}^{i\prime}$,

$1 \leq l \leq n$, among the classes of i-facts T, E', and B give rise to such that T accounts for $CF_l^{i\prime}$ in E' relative to B, where $C_l^{i\prime}$ is the set of negated or unnegated one-place i-predicates which induces $CF_l^{i\prime}$ relative to T, E', and B.

Suppose there is a such a class of i-facts $CF_l^{i\prime}$, $1 \leq l \leq n$. As $C_{j'}^{i}{}' = C_j^i$, and as $C_l^{i\prime} \subset C_{j'}^{i}{}'$, it follows that $C_l^{i\prime} \subset C_j^i$. But then there is a set of negated or unnegated one-place i-predicates C_p^i, $1 \leq p \leq m$, such that $C_p^i = C_l^{i\prime}$ and $CF_p^i \subseteq CF_l^{i\prime}$, where CF_p^i is the class of i-facts induced by C_p^i relative to T, E, and B.

Since T accounts for $CF_l^{i\prime}$ in E' relative to B, and $C_l^{i\prime} = C_p^i$, T accounts for CF_p^i in E relative B. Therefore there is at least one class of i-facts $C_p^i \subset C_j^i$, $1 \leq p \leq m$, such that T accounts for CF_p^i in E relative to B – in contradiction to the assumption that CF_j^i is a proper class of i-facts relative to T, E, and B.

So for every proper class of i-facts CF_j^i, $1 \leq j \leq m$, among (the not necessarily proper) $CF_1^i, \ldots, CF_m^i$ there is a proper class of i-facts $CF_{j'}^{i}{}'$, $1 \leq j' \leq n$, among (the not necessarily proper) $CF_1^{i\prime}, \ldots, CF_n^{i\prime}$ such that $C_j^i = C_{j'}^{i}{}'$ and $CF_j^i \subseteq CF_{j'}^{i}{}'$.

Let CF_j^i and CF_k^i be two proper classes of i-facts relative to T, E, and B, $1 \leq j, k \leq m$, and let $CF_{j'}^{i}{}'$ and $CF_{k'}^{i}{}'$, respectively, be the two corresponding proper classes of i-facts T, E', and B give rise to, $1 \leq j', k' \leq n$. Then $CF_j^i \subseteq CF_{j'}^{i}{}'$ and $CF_k^i \subseteq CF_{k'}^{i}{}'$, and $C_j^i = C_{j'}^{i}{}'$ and $C_k^i = C_{k'}^{i}{}'$. Because of the latter,

$$CF_{j'}^{i}{}' \neq CF_{k'}^{i}{}', \quad \text{if} \quad CF_j^i \neq CF_k^i.$$

For if $CF_j^i \neq CF_k^i$, then there is at least one 't' $\in C_{B-repr}(E) \cap C_i$ such that 't' $\in CF_j^i$ and 't' $\notin CF_k^i$, or 't' $\notin CF_j^i$ and 't' $\in CF_k^i$. Suppose without loss of generality that 't' $\in CF_j^i$ and 't' $\notin CF_k^i$. As $CF_j^i \subseteq CF_{j'}^{i}{}'$, '$t$' $\in CF_{j'}^{i}{}'$. But then 't' $\notin CF_{k'}^{i}{}'$; otherwise 't' $\in (C_{B-repr}(E') \setminus C_{B-repr}(E)) \cap C_i$, because '$t$' $\notin CF_k^i$ and $C_k^i = C_{k'}^{i}{}'$ – in contradiction to 't' $\in CF_j^i \subseteq C_{B-repr}(E) \cap C_i$.

This means that there is an injective (one-to-one) function from the set of all proper classes of i-facts T, E, and B give rise to into – but not necessarily onto (i.e. the function need not be surjective) – the set of all proper classes of i-facts T, E', and B give rise to. As $C_{B-repr}(E) \subseteq C_{B-repr}(E')$, there is also an injective function from the set of all non-empty proper classes of i-facts T, E, and B give rise to into the set of all non-empty proper classes of i-facts T, E', and B give rise to.

Let $CF_1^i, \ldots, CF_p^i$, $p \leq m$, be the non-empty proper classes of i-facts T, E, and B give rise to, and let $CF_1^{i\prime}, \ldots, CF_p^{i\prime}$ be the corresponding non-empty proper classes of i-facts T, E', and B give rise to. So $CF_1^{i\prime}, \ldots, CF_p^{i\prime}$ are those non-empty proper classes of i-facts among all non-empty proper classes of i-facts T, E', and B give rise to with $C_i^i = C_i^{i\prime}$ and $CF_j^i \subseteq CF_j^{i\prime}$, for every $j, 1 \leq j \leq p$. Finally, let $CF_{p+1}^i{}', \ldots, CF_q^{i\prime}$, $q \geq p$, be the remaining non-empty proper classes of i-facts T, E', and B give rise to.

As $C_j^i = C_j^{i\prime}$, for every $j, 1 \leq j \leq p$,

$$\sum_{1 \leq j \neq k \leq p} \left| C_j^i \bigtriangleup C_k^i \right| = \sum_{1 \leq j \neq k \leq p} \left| C_j^{i\prime} \bigtriangleup C_k^{i\prime} \right|,$$

and as $CF_j^i \subseteq CF_j^{i\prime}$, for every $j, 1 \leq j \leq p$,

$$\sum_{1 \leq j \neq k \leq p} \left[1 - \frac{1}{\log\left(\left|CF_j^i\right| + 1\right) + \log\left(\left|CF_k^i\right| + 1\right) + 1} \right] \leq$$

$$\leq \sum_{1 \leq j \neq k \leq p} \left[1 - \frac{1}{\log\left(\left|CF_j^{i\prime}\right| + 1\right) + \log\left(\left|CF_k^{i\prime}\right| + 1\right) + 1} \right],$$

whence

$$\begin{aligned}
g(T, E, B; D_i) &= \sum_{1 \leq j \neq k \leq p} \left| C_j^i \bigtriangleup C_k^i \right| \cdot \\
&\quad \cdot \left[1 - \frac{1}{\log\left(\left|CF_j^i\right| + 1\right) + \log\left(\left|CF_k^i\right| + 1\right) + 1} \right] \\
&\leq \sum_{1 \leq j \neq k \leq p} \left| C_j^{i\prime} \bigtriangleup C_k^{i\prime} \right| \cdot \\
&\quad \cdot \left[1 - \frac{1}{\log\left(\left|CF_j^{i\prime}\right| + 1\right) + \log\left(\left|CF_k^{i\prime}\right| + 1\right) + 1} \right] \\
&\leq \sum_{1 \leq j \neq k \leq q} \left| C_j^{i\prime} \bigtriangleup C_k^{i\prime} \right| \cdot \\
&\quad \cdot \left[1 - \frac{1}{\log\left(\left|CF_j^{i\prime}\right| + 1\right) + \log\left(\left|CF_k^{i\prime}\right| + 1\right) + 1} \right] \\
&= g(T, E', B; D_i).
\end{aligned}$$

The claim follows, because G is a monotone increasing function of g. $\square$

G does not support gathering evidence, if its definition is based on maximal classes of i-facts, as is shown by the following example (the basis of $\log$ is assumed to be 2). Let

$$\begin{aligned}
E &= \{Fa, Fb, Gb, Fc_1, Gc_1, Hc_1, \ldots, Fc_n, Gc_n, Hc_n\}, \\
E' &= \{Fa, Fb, Gb, Hb, Fc_1, Gc_1, Hc_1, \ldots, Fc_n, Gc_n, Hc_n\} = E \cup \{Hb\}, \\
T &= \{\forall x\,(Fx \wedge Gx \to Hx)\}, \quad \text{and} \\
B &= \emptyset.
\end{aligned}$$

There are the following three non-empty maximal classes of facts relative to T, E, and B:

$$\begin{aligned}
&CF_1 = \{\text{`}t\text{'} \in C_{B-repr}(E) : E \cup B \vdash Ft \wedge Gt \wedge Ht\} = \{c_1, \ldots, c_n\}, [1] \\
&CF_2 = \{\text{`}t\text{'} \in C_{B-repr}(E) : E \cup B \vdash Ft \wedge Gt\} = \{b\}, \quad \text{and} \\
&CF_3 = \{\text{`}t\text{'} \in C_{B-repr}(E) : E \cup B \vdash Ft\} = \{a\},
\end{aligned}$$

which are induced by the sets of negated or unnegated one-place predicates

$$C_1 = \{\text{`}Fx\text{'}, \text{`}Gx\text{'}, \text{`}Hx\text{'}\}, \quad C_2 = \{\text{`}Fx\text{'}, \text{`}Gx\text{'}\}, \quad \text{and} \quad C_3 = \{\text{`}Fx\text{'}\},$$

respectively, relative to T, E, and B. T, E', and B, on the other hand, give rise to the following two non-empty maximal classes of facts:

$$\begin{aligned}
CF_1' &= \{\text{`}t\text{'} \in C_{B-repr}(E') : E' \cup B \vdash Ft \wedge Gt \wedge Ht\} = \{b, c_1, \ldots, c_n\} \\
&= CF_1 \cup \{b\}, \quad \text{and} \\
CF_3' &= \{\text{`}t\text{'} \in C_{B-repr}(E') : E' \cup B \vdash Ft\} = \{a\},
\end{aligned}$$

which are induced by the sets of negated or unnegated one-place predicates

$$C_1' = \{\text{`}Fx\text{'}, \text{`}Gx\text{'}, \text{`}Hx\text{'}\} = C_1 \quad \text{and} \quad C_3' = \{\text{`}Fx\text{'}\} = C_3,$$

respectively, relative to T, E' and B. Although $E \subseteq E'$ (and thus $E' \vdash E$) and $C_{B-repr}(E) \subseteq C_{B-repr}(E')$, it holds for every $n \geq 1$:[2]

$$\begin{aligned}
g(T, E, B) &= |C_1 \triangle C_2| \cdot \\
&\quad \cdot \left[1 - \frac{1}{\log(|CF_1| + 1) + \log(|CF_2| + 1) + 1}\right] + \\
&\quad + |C_1 \triangle C_3| \cdot
\end{aligned}$$

---

[1]The quotation marks are suppressed.

[2]The reference to the confirmational domain D is dropped, where D is the domain (variable) corresponding to the variable ‘x’ and the individual constants ‘a’, ‘b’, and ‘c_m’.

$$
\begin{aligned}
& \cdot \left[1 - \frac{1}{\log\left(|CF_1| + 1\right) + \log\left(|CF_3| + 1\right) + 1}\right] + \\
& + |C_2 \triangle C_3| \cdot \\
& \cdot \left[1 - \frac{1}{\log\left(|CF_2| + 1\right) + \log\left(|CF_3| + 1\right) + 1}\right] \\
= \; & 1 \cdot \left[1 - \frac{1}{\log\left(n + 1\right) + \log\left(1 + 1\right) + 1}\right] + \\
& +2 \cdot \left[1 - \frac{1}{\log\left(n + 1\right) + \log\left(1 + 1\right) + 1}\right] + \\
& +1 \cdot \left[1 - \frac{1}{\log\left(1 + 1\right) + \log\left(1 + 1\right) + 1}\right] \\
> \; & 2 \cdot \left[1 - \frac{1}{\log\left(n + 1 + 1\right) + \log\left(1 + 1\right) + 1}\right] \\
= \; & |C_1' \triangle C_3'| \cdot \\
& \cdot \left[1 - \frac{1}{\log\left(|CF_1'| + 1\right) + \log\left(|CF_3'| + 1\right) + 1}\right] \\
= \; & g\left(T, E', B\right).
\end{aligned}
$$

Bibliography

[1] Akiba, Ken: Shogenji's probabilistic measure of coherence is incoherent, in: *Analysis* 60, 2000, pp. 356-359.

[2] Albert, Max: Bayesian Learning and Expectations Formation: Anything Goes, in: Corfield, David/Williamson, Jon (eds.): Foundations of Bayesianism, Dordrecht: Kluwer Academic Publishers, 2001, pp. 351-372.

[3] Armstrong, D.M.: What Is a Law of Nature?, Cambridge: Cambridge University Press, 1983.

[4] Balzer, Wolfgang: Theoretical Terms: Recent Developments, in: Balzer, Wolfgang/Moulines, C. Ulises (eds.): Structuralist Theory of Science. Focal Issues, New Results, Berlin/New York: Walter de Gruyter, 1996, pp. 139-166.

[5] Balzer, Wolfgang/Moulines, C. Ulises (eds.): Structuralist Theory of Science. Focal Issues, New Results, Berlin/New York: Walter de Gruyter, 1996.

[6] Balzer, Wolfgang/Moulines, C. Ulises: Introduction, in: Balzer, Wolfgang/Sneed, Joseph D./Moulines, C. Ulises (eds.): Structuralist Knowledge Representation. Paradigmatic Examples, Poznan Studies in the Philosophy of the Sciences and the Humanities, vol. 75, Amsterdam/Atlanta: Rodopi, 2000, pp. 5-18.

[7] Balzer, Wolfgang/Sneed, Joseph D./Moulines, C. Ulises (eds.): Structuralist Knowledge Representation. Paradigmatic Examples, Poznan Studies in the Philosophy of the Sciences and the Humanities, vol. 75, Amsterdam/Atlanta: Rodopi, 2000.

[8] Barker, Stephen J.: Causation, Facts and Coherence, in: *Analysis* 54, 1994, pp. 179-182.

[9] Barnes, A./Thagard, Paul R.: Emotional Decisions, in: Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society, Erlbaum, 1996, pp. 426-429.

[10] Barnes, A./Thagard, Paul R.: Empathy and Analogy, in: *Dialogue*: *Canadian Philosophical Review* 36, 1997, pp. 705-720.

[11] Bartelborth, Thomas: Begründungsstrategien. Ein Weg durch die analytische Erkenntnistheorie, Berlin: Akademie Verlag, 1996.

[12] Beckermann, Ansgar/Nimtz, Christian (eds.): Argument & Analysis. Paper contributed to the sections, A selection of papers contributed to the sections of the 4^{th} International Congress of the Society for Analytic Philosophy, Bielefeld, September 2000.

[13] BonJour, Laurence: The Structure of Empirical Knowledge, Cambridge, Mass.: Harvard University Press, 1985.

[14] Bovens, Luc/Olsson, Erik J.: Coherentism, Reliability, and Bayesian Networks, in: *Mind* 109, 2000, pp. 685-719.

[15] Carnap, Rudolph: The Two Concepts of Probability, in: *Philosophy and Phenomenological Research* 5, 1945, pp. 513-532.

[16] Carnap, Rudolf: Remarks on Induction and Truth, in: *Philosophy and Phenomenological Research* 6, pp. 590-602, 1946.

[17] Carnap, Rudolf: Logical Foundations of Probability, Chicago: University of Chicago Press, 1950.

[18] Carnap, Rudolf: The Continuum of Inductive Methods, Chicago: University of Chicago Press, 1952.

[19] Carnap, Rudolf: Logical Foundations of Probability, 2^{nd} ed., Chicago: University of Chicago Press, 1962.

[20] Carnap, Rudolf/Stegmüller, Wolfgang: Induktive Logik und Wahrscheinlichkeit, Wien: Springer, 1959.

[21] Carroll, John W.: Laws of Nature, Cambridge: Cambridge University Press, 1994.

[22] Christensen, David: The Irrelevance of Bootstrapping, in: *Philosophy of Science* 57, 1990, pp. 644-662.

[23] Christensen, David: Measuring Confirmation, in: *Journal of Philosophy* 96, 1999, pp. 437-461.

[24] Corfield, David/Williamson, Jon (eds.): Foundations of Bayesianism, Dordrecht: Kluwer Academic Publishers, 2001.

[25] Cross, Charles B.: Coherence and truth conducive justification, in: *Analysis* 59, 1999, pp. 186-193.

[26] Culler, Madison: Beyond Bootstrapping: A New Account of Evidential Relevance, in: *Philosophy of Science* 62, 1995, pp. 561-579.

[27] Curd, Martin/Cover, J.A.: Philosophy of Science. The Central Issues, New York/London: Norton, 1998.

[28] Cutland, N. J.: Computability: An Introduction to Recursive Function Theory, Cambridge: Cambridge University Press, 1986.

[29] Dalla Chiara, M.-L. et al. (eds.): Logic and Scientific Methods, Dordrecht: Kluwer Academic Publishers, 1997.

[30] Earman, John: Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory, Cambridge, Mass.: The MIT Press, 1992.

[31] Earman, John (ed.): Testing Scientific Theories, Minnesota Studies in the Philosophy of Science, vol. 10, Minnesota: University of Minnesota Press, 1983.

[32] Edidin, Aron: Bootstrapping without Bootstraps, in: Earman, John (ed.): Testing Scientific Theories, Minnesota Studies in the Philosophy of Science, vol. 10, Minnesota: University of Minnesota Press, 1983, pp. 43-54.

[33] Eliasmith, Chris/Thagard, Paul R.: Waves, Particles, and Explanatory Coherence, in: *British Journal for the Philosophy of Science* 48, 1997, pp. 1-19.

[34] Eliasmith, Chris/Thagard, Paul R.: Integrating Structure and Meaning: A Distributed Model of Analogical Mapping, in: *Cognitive Science* 25, 2001, pp. 245-286.

[35] Fitelson, Branden: Studies in Bayesian Confirmation Theory, University of Wisconsin-Madison: Dissertation, 2001.

[36] Friedman, Michael: Explanation and Scientific Understanding, in: *The Journal of Philosophy* 71, 1974, pp. 5-19.

[37] Friedman, Michael: Truth and Confirmation, in: *The Journal of Philosophy* 76, 1979, pp. 361-382.

[38] Friedman, Michael: Erklärung und wissenschaftliches Verstehen. Die Vereinheitlichung der Gesetze, in: Schurz, Gerhard (ed.): Erklären und Verstehen in der Wissenschaft, München: R. Oldenbourg Verlag, 1990, pp. 171-191.

[39] Gabbay, Dov M./Smets, Philippe (eds.): Handbook of Defeasible Reasoning and Uncertainty Management Systems, Vol. 1, Dordrecht: Kluwer Academic Publishers, 1998.

[40] Garber, Daniel: Old Evidence and Logical Omniscience in Bayesian Confirmation Theory, in: Earman, John (ed.): Testing Scientific Theories, Minnesota Studies in the Philosophy of Science, vol. 10, Minnesota: University of Minnesota Press, 1983, pp. 99-131.

[41] Gaifman, Haim/Snir, Marc: Probabilities Over Rich Languages, Testing and Randomness, in: *Journal of Symbolic Logic* 47, 1982, pp. 495-548.

[42] Gemes, Ken: Discussion: Horwich and Hempel on Hypothetico-Deductivism, in: *Philosophy of Science* 57, 1990, pp. 699-702.

[43] Gemes, Ken: Hypothetico-Deductivism, Content, and the Natural Axiomatization of Theories, in: *Philosophy of Science* 60, 1993, pp. 477-487.

[44] Gemes, Ken: Schurz on Hypothetico-Deductivism, in: *Erkenntnis* 41, 1994a, pp. 171-181.

[45] Gemes, Ken: A New Theory of Content I: Basic Content, in: *Journal of Philosophical Logic* 23, 1994c, pp. 595-620.

[46] Gemes, Ken: A New Theory of Content II: Model Theory and Some Alternatives, in: *Journal of Philosophical Logic* 26, 1997a, pp. 449-476.

[47] Gemes, Ken: Hypothetico-Deductivism: The Current State of Play; The Criterion of Significance: Endgame, in: *Erkenntnis* 49, 1998, pp. 1-20.

[48] Gemes, Ken: Carnap-Confirmation, Content-Cutting, and Real Confirmation, Yale: unpublished essay, 1999.

[49] Gentner, D./Holyoak, K.H./Kokinov, B.K. (eds.): The Analogical Mind: Perspectives from Cognitive Science, Cambridge, MA: MIT Press, 2001.

[50] Gillies, Donald: Confirmation Theory, in: Gabbay, Dov M./Smets, Philippe (eds.): Handbook of Defeasible Reasoning and Uncertainty Management Systems, Vol. 1, Dordrecht: Kluwer Academic Publishers, 1998, pp. 135-167.

[51] Gillies, Donald: Philosophical Theories of Probability, London/New York: Routledge, 2000.

[52] Gillies, Donald: Bayesianism and the Fixity of the Theoretical Framework, in: Corfield, David/Williamson, Jon (eds.): Foundations of Bayesianism, Dordrecht: Kluwer Academic Publishers, 2001, p. 373.

[53] Giere, Ronald N. (ed.): Cognitive Models of Science, Minnesota Studies in the Philosophy of Science, vol. 15, Minneapolis: University of Minnesota Press, 1992.

[54] Glymour, Clark: Relevant Evidence, in: *The Journal of Philosophy* 72, 1975, pp. 403-426.

[55] Glymour, Clark: The Epistemology of Geometry, in: *Noûs* 11, 1977, pp. 227-251.

[56] Glymour, Clark: Theory and Evidence, Princeton: Princeton University Press, 1980a.

[57] Glymour, Clark: Explanations, Tests, Unity and Necessity, in: *Noûs* 14, 1980b, pp. 31-50.

[58] Glymour, Clark: Discussion: Hypothetico-Deductivism is Hopeless, in: *Philosophy of Science* 47, 1980c, pp. 320-325.

[59] Glymour, Clark: On Testing and Evidence, in: Earman, John (ed.): Testing Scientific Theories, Minnesota Studies in the Philosophy of Science, vol. 10, Minnesota: University of Minnesota Press, 1983, pp. 3-26.

[60] Good, I. J.: Good Thinking: The Foundations of Probability and its Applications, Minneapolis: University of Minnesota Press, 1983.

[61] Goodman, Nelson: A Query on Confirmation, in: *The Journal of Philosophy* 43, 1946, pp. 383-385.

[62] Grimes, Thomas R.: Discussion: Truth, Content, and the Hypothetico-Deductive Method, in: *Philosophy of Science* 57, 1990, pp. 514-522.

[63] Harré, Rom: Laws of Nature, London: Duckworth, 1993.

[64] Hartmann, Stephan/Bovens, Luc: A Probabilistic Theory of the Coherence of an Information Set, in: Beckermann, Ansgar/Niemtz, Christian (eds.): Argument & Analysis. Papers contributed to the sections, Bielefeld, 2000, pp. 195-206.

[65] Helmer, Olaf/Oppenheim, Paul: A Syntactical Definition of Probability and of Degree of Confirmation, in: *Journal of Symbolic Logic* 10, 1945, pp. 25-60.

[66] Hempel, Carl Gustav: A Purely Syntactical Definition of Confirmation, in: *Journal of Symbolic Logic* 8, 1943, pp. 122-143.

[67] Hempel, Carl Gustav: Studies in the Logic of Confirmation, in: *Mind* 54, 1945, pp. 1-26, pp. 97-121.

[68] Hempel, Carl Gustav: Aspects of Scientific Explanation and Other Essays in the Philosophy of Science, New York/London: The Free Press/Collier-MacMillan Limited, 1965.

[69] Hermes, Hans: Aufzählbarkeit, Entscheidbarkeit, Berechenbarkeit, Berlin/Göttingen/Heidelberg: Springer, 1961.

[70] Holyoak, Keith J./Thagard, Paul R.: The Analogical Mind, in: *American Psychologist* 52, 1997, pp. 35-44.

[71] Horwich, Paul: Probability and Evidence, Cambridge: Cambridge University Press, 1982.

[72] Hosiasson-Lindenbaum, Janina: On Confirmation, in: *Journal of Symbolic Logic* 5, 1940, pp. 133-148.

[73] Howson, Colin: A Logic of Induction, in: *Philosophy of Science* 64, 1997a, pp. 268-290.

[74] Howson, Colin: Error Probabilities in Error, in: *Philosophy of Science* 64, 1997b, pp. S185-S194.

[75] Howson, Colin: Logic and Probability, in: *British Journal for the Philosophy of Science* 48, 1997c, pp. 517-531.

[76] Howson, Colin/Urbach, Peter: Scientific Reasoning: The Bayesian Approach, 2^{nd} ed., Chicago/La Salle: Open Court, 1993.

[77] Jeffrey, Richard C.: The Logic of Decision, 2^{nd} ed., New York: McGraw-Hill, 1967.

[78] Jeffrey, Richard C.: Bayesianism with a Human Face, in: Earman, John (ed.): Testing Scientific Theories, Minnesota Studies in the Philosophy of Science, vol. 10, Minnesota: University of Minnesota Press, 1983, pp. 133-156.

[79] Kelly, Kevin T.: The Logic of Reliable Inquiry, New York/Oxford: Oxford University Press, 1996.

[80] Kemeny, John G.: A Logical Measure Function, in: *Journal of Symbolic Logic* 18, 1953, pp. 289-308.

[81] Kemeny, John G.: Fair Bets and Inductive Probabilities, in: *Journal of Symbolic Logic 20*, 1955, pp. 263-273.

[82] Kitcher, Philip: Explanatory Unification, in: *Philosophy of Science* 48, 1981, pp. 507-531.

[83] Klein, Peter/Warfield, Ted. A.: What Price Coherence?, in: *Analysis* 54, 1994, pp. 129-132.

[84] Klein, Peter/Warfield, Ted. A.: No help for the coherentist?, in: *Analysis* 56, 1996, pp. 118-121.

[85] Kuhn, Thomas S.: The Structure of Scientific Revolutions, 3^rd ed., Chicago: University of Chicago Press, 1996.

[86] Kuipers, Theo F. (ed.): What Is Closer-to-the-Truth? A parade of approaches to truthlikeness, Amsterdam: Rodopi, 1987.

[87] Lehman, R. Sherman: On Confirmation and Rational Betting, in: *Journal of Symbolic Logic* 20, pp. 251-262.

[88] Lehrer, Keith: Theory of Knowledge, London: Routledge, 1990.

[89] Lipton, Peter: Inference to the Best Explanation, London: Routledge, 1993.

[90] Maher, Patrick: Why Scientists Gather Evidence, in: *British Journal for the Philosophy of Science* 41, 1990, pp. 103-119.

[91] Mayo, Deborah G.: Error and the Growth of Experimental Knowledge, Chicago: The University of Chicago Press, 1996.

[92] Merricks, Trenton: On behalf of the coherentist, in: *Analysis* 55, 1995, pp. 306-309.

[93] Merrill, Gary: Confirmation and Prediction, in: *Philosophy of Science* 46, 1979, pp. 98-117.

[94] Millgram, Elijah: Coherence: The Price of the Ticket, in: *The Journal of Philosophy* 97, 2000, pp. 82-93.

[95] Milne, Peter: $log[P(h/eb)/P(h/b)]$ is the One True Measure of Confirmation, in: *Philosophy of Science* 63, 1996, pp. 699-702.

[96] Mitchell, Sam: Toward a Defensible Bootstrapping, in: *Philosophy of Science* 62, 1995, pp. 241-260.

[97] Morrison, Margaret: Unification, Realism and Inference, in: *British Journal for the Philosophy of Science* 41, 1990, pp. 305-332.

[98] Moulines, C. Ulises: Structuralism: The Basic Ideas, in: Balzer, Wolfgang/Moulines, C. Ulises (eds.): Structuralist Theory of Science. Focal Issues, New Results, Berlin/New York: Walter de Gruyter, 1996, pp. 1-13.

[99] Niiniluoto, Ilkka: Truthlikeness, Dordrecht: D. Reidel, 1987.

[100] Niiniluoto, Illkka: Verisimilitude: The Third Period, in: *British Journal for the Philosophy of Science* 49, 1998, pp. 1-29.

[101] Nowak, Greg/Thagard, Paul R.: Copernicus, Ptolemy, and Explanatory Coherence, in: Giere, Ronald N. (ed.): Cognitive Models of Science, Minnesota Studies in the Philosophy of Science, vol. 15, Minneapolis: University of Minnesota Press, 1992, pp. 274-309.

[102] O'Laughlin, C./Thagard, Paul R.: Autism and Coherence: A Computational Model, in: *Mind and Language* 15, 2000, pp. 375-392.

[103] Oddie, Graham: Likeness to Truth, Dordrecht: D. Reidel, 1986.

[104] Olsson, Erik J.: Why coherence is not truth-conducive, in: *Analysis* 61, 2001, pp. 236-241.

[105] Olsson, Erik J.: What Is the Problem of Coherence and Truth?, in: *The Journal of Philosophy* 99, 2002, pp. 246-272.

[106] Popper, Karl R.: Logik der Forschung, 10^{th} ed., Tübingen: J.C.B. Mohr (Paul Siebeck), 1994.

[107] Quine, Willard Van Orman: On What There Is, in: *Review of Metaphysics* 2, 1948, pp. 21-38.

[108] Quine, Willard Van Orman: Two Dogmas of Empiricism, in: *Philosophical Review* 60, 1951, pp. 20-43.

[109] Quine, Willard Van Orman: Speaking of Objects, in: *Proceedings and Addresses of the American Philosophical Association* 31, 1958, pp. 5-22.

[110] Quine, Willard Van Orman: From a Logical Point of View, 2^{nd} ed., Harvard College, 1961.

[111] Ram, A./Leake, D.B. (eds.): Goal-driven Learning, Cambridge, MA: MIT Press, 1995.

[112] Read, S.J./Miller, L.C. (eds.): Connectionist Models of Social Reasoning and Social Behavior, Hillsdale, NJ: Erlbaum, 1998.

[113] Schoch, Daniel: A Fuzzy Measure for Explanatory Coherence, in: *Synthese* 122, 2000, pp. 291-311.

[114] Schoch, Daniel: Wissen und Kohärenz, in: Beckermann, Ansgar/Niemtz, Christian (eds.): Argument & Analysis. Papers contributed to the sections, Bielefeld, 2000, pp. 247-258.

[115] Schurz, Gerhard: Paradoxical Consequences of Balzer's and Gähde's Criterion of Theoreticity. Results of an Application to Ten Scientific Theories, in: *Erkenntnis* 32, 1990, pp. 161-214.

[116] Schurz, Gerhard (ed.): Erklären und Verstehen in der Wissenschaft, München: R. Oldenbourg Verlag, 1990.

[117] Schurz, Gerhard: Relevant Deduction. From Solving Paradoxes Towards a General Theory, in: *Erkenntnis* 35, 1991a, pp. 391-437.

[118] Schurz, Gerhard: Relevant Deductive Inference: Criteria and Logics, in: Schurz, Gerhard/Dorn, Georg J. W. (eds.): Advances in Scientific Philosophy. Essays in Honour of Paul Weingartner on the Occasion of the 60^{th} Anniversary of his Birthday, Poznan Studies in the Philosophy of the Sciences and the Humanities, vol. 24, Amsterdam/Atlanta: Rodopi, 1991b, pp. 57-84.

[119] Schurz, Gerhard: Relevant Deduction and Hypothetico-Deductivism: A Reply to Gemes, in: *Erkenntnis* 41, 1994, pp. 183-188.

[120] Schurz, Gerhard: Relevance in Deductive Reasoning: A Critical Overview, in: Schurz/Uršič (eds.): Beyond Classical Reasoning. Philosophical and Computational Investigations in Deductive Reasoning and Relevance, Conceptus-Studien 13, St. Augustin: Academia Verlag, 1998, pp. 9-56.

[121] Schurz, Gerhard: Explanation as Unification, in: *Synthese* 120, 1999, pp. 95-114.

[122] Schurz, Gerhard/Dorn, Georg J. W. (eds.): Advances in Scientific Philosophy. Essays in Honour of Paul Weingartner on the Occasion of the 60^{th} Anniversary of his Birthday, Poznan Studies in the Philosophy of the Sciences and the Humanities, vol. 24, Amsterdam/Atlanta: Rodopi, 1991.

[123] Schurz, Gerhard/Lambert, Karel: Outline of a Theory of Scientific Understanding, in: *Synthese* 101, 1994, pp. 65-120.

[124] Schurz, Gerhard/Uršič, Marko (eds.): Beyond Classical Reasoning. Philosophical and Computational Investigations in Deductive Reasoning and Relevance, Conceptus-Studien 13, St. Augustin: Academia Verlag, 1998.

[125] Schurz, Gerhard/Weingartner, Paul: Verisimilitude defined by relevant consequence-elements, in: Kuipers, Theo F. (ed.): What Is Closer-to-the-Truth? A parade of approaches to truthlikeness, Amsterdam: Rodopi, 1987, pp. 47-77.

[126] Shimony, Abner: Coherence and the Axioms of Confirmation, in: *Journal of Symbolic Logic* 20, 1955, pp. 1-28.

[127] Shogenji, Tomoji: Is coherence truth conducive?, in: *Analysis* 59, 1999, pp. 338-345.

[128] Shogenji, Tomoji: Reply to Akiba on the probabilistic measure of coherence, in: *Analysis* 61, 2001a, pp. 147-150.

[129] Shogenji, Tomoji: The Role of Coherence in Epistemic Justification, in: *Australasian Journal of Philosophy* 79, 2001b, pp. 90-106.

[130] Skyrms, Brian: Choice and Chance. An Introduction to Inductive Logic, 4^{th} ed., Australia/Belmont, CA: Wadsworth/Thomson Learning, 2000.

[131] Smokler, Howard: Conflicting Conceptions of Confirmation, in: *The Journal of Philosophy* 65, 1968, pp. 300-312.

[132] Sperber, Dan/Wilson, Deirdre: Relevance. Communication and Cognition, 2^{nd} ed., Oxford: Blackwell, 1995.

[133] Suppes, Patrick: Introduction to Logic, New York: D. Van Nostrand Company, 1957.

[134] Sylvan, Richard/Nola, Robert: Confirmation without Paradoxes, in: Schurz, Gerhard/Dorn, Georg J. W. (eds.): Advances in Scientific Philosophy. Essays in Honour of Paul Weingartner on the Occasion of the 60^{th} Anniversary of his Birthday, Poznan Studies in the Philosophy of the Sciences and the Humanities, vol. 24, Amsterdam/Atlanta: Rodopi, 1991, pp. 5-44.

[135] Thagard, Paul R.: Explanatory Coherence, in: *Behavioral and Brain Sciences* 12, 1989, pp. 435-502.

[136] Thagard, Paul R.: Coherent and Creative Conceptual Combinations, in: Ward, T.B./Smith, S.M./Viad, J. (eds.): Creative Thought: An Investigation of Conceptual Structures and Processes, Washington, D.C.: American Psychological Association, 1997, pp. 129-141.

[137] Thagard, Paul R.: Ethical Coherence, in: *Philosophical Psychology* 11, 1999, pp. 405-422.

[138] Thagard, Paul R.: Probabilistic Networks and Explanatory Coherence, in: *Cognitive Science Quarterly* 1, 2000, pp. 91-114.

[139] Thagard, Paul R./Kunda, Z.: Making Sense of People: Coherence Mechanisms, in: Read, S.J./Miller, L.C. (eds.): Connectionist Models of Social Reasoning and Social Behavior, Hillsdale, NJ: Erlbaum, 1998, pp. 3-26.

[140] Thagard, Paul R./Millgram, Elijah: Inference to the Best Plan: A Coherence Theory of Decision, in: Ram, A./Leake, D.B. (eds.): Goal-driven Learning, Cambridge, MA: MIT Press, 1995, pp. 439-454.

[141] Thagard, Paul R./Shelley, C. P.: Abductive Reasoning: Logic, Visual thinking, and Coherence, in: Dalla Chiara, M.-L. et al. (eds.): Logic and Scientific Methods, Dordrecht: Kluwer Academic Publishers, 1997, pp. 413-427.

[142] Thagard, Paul R./Shelley, C. P.: Emotional Analogies and Analogical Inference, in: Gentner, D./Holyoak, K.H./Kokinov, B.K. (eds.): The Analogical Mind: Perspectives from Cognitive Science, Cambridge, MA: MIT Press, 2001, pp. 335-362.

[143] Thagard, Paul R./Verbeurgt, K.: Coherence as Constraint Satisfaction, in: *Cognitive Science* 22, 1998, pp. 1-24.

[144] Ward, T.B./Smith, S.M./Viad, J. (eds.): Creative Thought: An Investigation of Conceptual Structures and Processes, Washington, D.C.: American Psychological Association, 1997.

[145] Waters, C. Kenneth: Relevance Logic Brings Hope to Hypothetico-Deductivism, in: *Philosophy of Science* 54, 1987, pp. 453-464.

[146] Weber, Erik/van Dyck, Maarten: Unification and Explanation. A Comment on Halonen and Hintikka, and Schurz, in: *Synthese* 131, 2002, pp. 145-154.

[147] Weingartner, Paul/Schurz, Gerhard: Paradoxes Solved by Simple Relevance Criteria, in: *Logique et Analyse* 113, 1986, pp. 3-40.

[148] Zoglauer, Thomas: Das Problem der theoretischen Terme. Eine Kritik an der strukturalistischen Wissenschaftstheorie, Braunschweig/Wiesbaden: Vieweg, 1993.

Curriculum Vitae

Particulars

Current Address:	*Philosophy, Probability, and Modeling* Research Group Centre for Junior Research Fellows of the University of Konstanz P.O. Box M 682 D-78457 Konstanz
e-mail:	franz.huber@uni-konstanz.de
homepage:	www.uni-konstanz.de/ppm/Huber

Date of Birth:	November 21st, 1977
Place of Birth:	Vöcklabruck, Austria
Nationality:	Austria

Education

1984 - 1988:	Elementary School in Vöcklamarkt, Austria
1988 - 1996:	Comprehensive School at the Public High School of the Convent Kremsmünster, Austria Living in the boarding school run by the Benedictines
June 3rd, 1996:	School Leaving Exam (with excellence)
1996 - 2000:	University of Salzburg, Austria M.A. studies in Philosophy and German Philology
1997:	additional course of studies in General Linguistics
1998:	additional course of studies in Mathematics
May 22nd, 2000:	M.A. (with excellence) in Philosophy (first course of studies) and selected subjects from German Philology and General Linguistics (second course of studies) with a master’s thesis (“On Opposites”) in

	Logic supervised by Prof.es Paul Weingartner and Johannes Czermak
2000 - 2002:	University of Erfurt, Germany Ph.D. studies in Philosophy
2002 - :	University of Konstanz, Germany Fellow of the research group “Philosophy, Probability, and Modeling” led by Prof. Luc Bovens (University of Colorado at Boulder, USA) and Dr. Stephan Hartmann (University of Konstanz)
April 22nd, 2003:	Ph.D. in Philosophy with a doctoral dissertation in Philosophy of Science (“Assessing Theories. The Problem of a Quantitative Theory of Confirmation”) supervised by Prof. Dr. Gerhard Schurz (University of Erfurt and University of Duesseldorf, Germany)

Grants

1998: - Grant of Honor (Leistungsstipendium) from the Faculty of the Humanities of the University of Salzburg
- Dr. Heinrich Gleißner Grant

1999: - Grant of Honor (Leistungsstipendium) from the Faculty of the Natural Sciences of the University of Salzburg
- Grant of Honor (Leistungsstipendium) from the Faculty of the Humanities of the University of Salzburg
- Dr. Heinrich Gleißner Grant

2000: - Grant of Honor (Leistungsstipendium) from the Faculty of the Humanities of the University of Salzburg
- Dr. Heinrich Gleißner Grant
- KIP (Kepler Internationalisierungsprogramm) Grant from the Upper Austrian Government (2000-2002)

Academic Positions

June 1999 - July 2000:	Scientific collaborator at the Department of Philosophy of Science of the International Research Centre in Salzburg (assistant to Prof.es Paul Weingartner and Georg Kreisel)
August 2000 - July 2002:	Scientific collaborator at the Chair for Philosophy of Science and Logic at the University of Erfurt (assistant to Prof. Dr. Gerhard Schurz)

September 2002 - :	Research Fellow at the Centre for Junior Research Fellows of the University of Konstanz