Quaestiones Infinitae

Publications of the Z∃NO Institute of Philosophy

CRITERIA OF EMPIRICAL SIGNIFICANCE

Volume LXX
Utrecht University

In the electronic edition, internal references are linked and URLs are hyperlinked.

Printed by Wöhrmann Print Service, Zutphen
ISBN 978-94-6103-023-8

CRITERIA
OF
EMPIRICAL SIGNIFICANCE

Foundations, Relations, Applications

Criteria voor Empirische Significantie:
Grondslagen, Relaties, Toepassingen
(met een samenvatting in het Nederlands)

Proefschrift
ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G. J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op maandag 18 juni 2012 des ochtends te 10.30 uur
door

Sebastian Lutz

geboren op 30 november 1976 te Kevelaer, Duitsland

Promotor:	Prof. dr. A. Visser
Co-promotoren:	Dr. T. Müller
	Dr. J. H. van Lith

Contents

List of figures

Preface

Origins

This book has four central roots. The first is a little monograph by the Polish model theorist Marian Przełęcki, *The Logic of Empirical Theories* (1969), whose central concepts I presented at a seminar at the University of Hamburg in the winter term of 2000. At that time, I thought of them simply as providing a neat formalism, but it turned out that these concept were extremely well-adapted to the philosophical problems that I encountered.

The second root is a hunch I had about six years ago while at the University of Western Ontario; based on the abstract of an article by Lewis (1988b) to which I did not have full access at that time, it seemed to me that one of the criteria of empirical significance he suggests might be equivalent to one given by Przełęcki (1974a). I decided to compare these and other criteria of empirical significance. In a stroke of luck, the hunch turned out to be correct (see claim 6.13 on page 236), and a host of other criteria turned out to be equivalent or closely related to each other as well.

The third root springs from my personal predilection for analyzing problems with the help of predicate logic, one element of Przełęcki's formalism. It was the source of some dismay that my results would often be dismissed as being based on "the syntactic view on scientific theories", the discredited approach to philosophy of science by the logical empiricists. That I was interested in criteria of empirical significance—the logical empiricists' hobby horse and one of the main reasons for logical empiricism's descent into infamy—did not improve matters. Thus I was more or less forced to defend the use of predicate logic in the philosophy of science. I started out with the intent to avoid the mistakes of the logical empiricists, and the first and most obvious step was not to restrict predicate logic to first order. It was (and more or less still is) common knowledge in philosophy that this restriction is central to logical empiricism. As stated in the *Philosophical Lexicon* (Steglich-Petersen 2008):

> **hempel**, adj. (only in the idiom *hempel-minded*) Said of one who insists on recasting the problem in the first order logic.

I was baffled to discover that at no point did the logical empiricists actually demand that theories be formalized in first order logic (for an example of Hempel not being hempel-minded, see page 105). Thus I began to attempt to identify the logical empiricists' real positions on formal methods in philosophy of science, which turned out to be rather defensible. What is more, it became clear that many later developments were either anticipated by the logical empiricists or can find a natural home in their approach.

The fourth root is my reflection on my own philosophical methodology. At one point I realized that when trying to pin down a concept, I would first try to find out what I want to use it for, and then develop a concept that does that.[1] For example, when thinking about free will and responsibility, I would look at the concepts' purported roles in ethics, and try to develop new concepts that would fulfill these roles as much as possible. I was not too surprised when this method turned out to be already described by Carnap as 'explication' (I am rather sure that I had read about Carnap's methodology earlier, and then just forgotten about it for a while). Looking closer at Carnap's elucidations, I was struck by the prominent role that he assigns to empirical results in the formation of concepts, without, though, declaring concept formation an empirical endeavor. This observation was the basis for my conviction that philosophy can be empirically informed without thereby making synthetic claims, which luckily became a central topic in philosophy with the rise of experimental philosophy.

The structure of this book came about by presenting these topics in reverse chronological order (with the exception of Przełęcki's formalism): I would first argue for explication as a philosophical methodology and outline Przełęcki's formalism, then defend the logical empiricists' approach to philosophy of science, and finally analyze criteria of empirical significance. Possible applications for the criteria came up naturally when thinking about their relations, so these would form the final part of the book.

Ideals

Since this book is both a defense and an application of artificial language philosophy, I will not try to defend my conceptual suggestions by reference to ordinary language use, language intuitions, or intuitions about the way the world really is. For what it is worth, my intuitions are realistic: I think of electrons as little hard spheres and of numbers as platonic objects.[2] However, I know that my intuition about electrons is inconsistent with quantum mechanics, which provides me with all my information about electrons, and I know of no reason to treat my intuitions about numbers with more respect. This has all been better expressed by Creath

[1] I am somewhat embarrassed to admit that this was indeed more of a realization than a conscious decision.

[2] Specifically, objects with a specific geometric shape that are added by fitting into each other.

(1991, 349):

> If we are not to give up our pretensions to have reasonable beliefs, something must be added to observation and inference which provides the requisite justification. But what? A classical answer to this goes back to Kant and Descartes and beyond them to the Greeks and has dominated serious philosophers down to and including Frege and (sometimes) Russell. That answer can be given in a single word: 'in- tuition'. On this view we have direct metaphysical insight into a domain of truths independent of ourselves. This is how we know the axioms of arithmetic, geometry, logic, and so on as well as how we know the true essences of things and any other necessary truth that we think we know. This doctrine of intuition, however, is a scandal. Intuitions notoriously differ, and there is no plausible way of resolving these differences. For that matter there is no plausible account of why intuitions should be right even where intuitions agree. [...] I can well appreciate that [the need to justify our beliefs], in the absence of an alternative, would drive someone to appeal to intuition. That, however, must not hide the fact that such an appeal is an act of desperation.

Accordingly, I do not think of the definitions that I give for, e. g., 'abstraction', '$\mathscr{B}$-determinacy', or 'confirmation' as attempts to capture my intuition or that of anyone else. Hence if the definition turns out to be inadequate, either because it fails to be a definition on technical grounds or it does not meet some other condition of adequacy, I cannot reply that even if the definition is inadequate, the underlying intuition is sound (this, of course, still allows one to search for a similar but better definition). In general, I do not consider an appeal to intuitions a sufficient defense. In §9.3, for example, I argue that Elliott Sober's definition of 'testability' renders intelligent design testable under specific circumstances. More than once, the reply to this argument has been that in the circumstances that I consider, "one would not say that" intelligent design has been tested, and thus intelligent design remains untestable. But this claim is unacceptable, since *testability* is not a clear enough concept to decide the cases under consideration, and accordingly the testability of intelligent design cannot be decided by reference to intuition. Other terms whose use as primitives I have avoided are 'explanation', 'causation', and 'natural kind', since I consider none of them to be clear enough to solve the problems under discussion here. Accordingly, inference to the best explanation plays no role in my arguments.[3]

Relatedly, I will also refrain from endowing perfectly well-defined concepts implicitly with additional features. In one especially puzzling response to my

[3]The following rule of thumb has served me well: Whenever a philosopher uses 'explanation', 'causation', or 'natural kind' as a primitive term, hold on to your wallet.

discussion of Sober's criterion, for example, intelligent design was claimed to be untestable *and* Sober's definition was defended as adequate because under the circumstances that I discuss, one would not say that intelligent design was being tested. Sober's intuition was silently being augmented by intuition, which, it seems to me, renders the definition itself pointless. I will accordingly also not silently augment the formalism of higher order logic with non-formal restrictions. For instance, I will assume that $\exists x(x = x) \vDash \exists X_1 X_2(X_1 \neq X_2)$, where X_1 and X_2 are predicate variables. This might seem to be a trivial point, but in fact has considerable implications for philosophical methodology (see page 51), the foundations of philosophy of science (see page 168), and structural realism (see page 389).

Acknowledgments

Throughout the development of this book, and indeed throughout my life, I have been in the very fortunate position to receive all the support I had hoped for, and often more (the somewhat unsettling corollary of which is that every shortcoming of this work is solely my own fault). I have profited enormously from discussions with my advisers Thomas Müller, Janneke van Lith, Albert Visser, and many others at a number of conferences, at Utrecht University, and at the University of Tilburg, where I was supported by a Visiting Fellowship from the Center for Logic and Philosophy of Science. I especially thank Thomas Müller for commenting on and improving almost every page that I have written, and both Thomas Müller and my wife Alana Yu for their tireless help in clarifying and structuring my writings. I admit that it is not for lack of trying on their part that they failed to make me a graceful writer. When I write and rewrite, I try to imagine what their comments would be.[4] I will not even try to list my personal debts beyond those to Alana Yu, my godparents Gabriele Walther and Walter Latsch, and my parents Eva-Maria and Hans-Robert Lutz, whose unwavering support keeps amazing me. I have been fortunate indeed.

[4]Since I have already thanked those who taught me how to write clearly, I should also thank my high school teacher Andreas Engelmann, who taught me how to read clearly.

Chapter 1

Introduction

The interlocking main goals of this book are, first, a defense of artificial language philosophy; second, an analysis of criteria of empirical significance on the basis of artificial language philosophy; and third, an application of the criteria of empirical significance to different problems in philosophy. As it will turn out, one such application is the analysis and improvement of the formalism of artificial language philosophy itself.

1.1 Overview

Part I is a defense of an artificial language methodology in philosophy and a historical and systematic defense of the logical empiricists' application of an artificial language methodology to scientific theories. These defenses provide a justification for the presumptions of a host of criteria of empirical significance, which I will analyze, compare, and develop in part II. On the basis of this analysis, in part III I will use a variety of criteria to evaluate the scientific status of intelligent design, and further discuss confirmation, reduction, and concept formation.

I. Foundations

Foundations of philosophy

As a foundation, I argue in §2.4 and §2.6 that artificial language philosophy (also called 'ideal language philosophy') has advantages over the competing methodologies of philosophy based on intuitions (§2.1), ordinary language philosophy (§2.2), experimental philosophy, and naturalized philosophy (§2.5). Artificial language philosophy relies on language choice to solve philosophical problems; the choice of language determines which sentences are analytically true, but itself has no factual implications, and thus the results of artificial language philosophy are analytic. Although the choice of a language can involve the choice of a logic, it is sufficient for my purposes to assume a fixed logic and a fixed basic vocabulary, and to assume that only the meaning postulates for the terms of an auxiliary vocabulary can be chosen. In this case, language choice amounts to concept formation. I discuss the semantics of concept formation as developed by Przełęcki (1969, §§4–6) in §2.8, and argue that artificial language philosophy is methodologically naturalistic, in that the sciences rely extensively on language choice as well (§2.10). Because of this, artificial language philosophy justifies many observed fruitful interactions between the sciences and philosophy of science (§2.11). Artificial language philosophy furthermore allows for the reinterpretation and productive use of results from the other five mentioned philosophical methodologies (§2.7). As an example of artificial language philosophy in use, I explicate the concepts of abstraction and idealization (§2.12).

Foundations of philosophy of science

The Received View on scientific theories by Carnap, Hempel, and Feigl is an application of artificial language philosophy to scientific theories. The heavy criticisms that led to its downfall, I argue, rest on false assumptions. Specifically, the Received View does not rely on exhaustive axiomatizations (§3.4) in first order logic (§3.3) and in fact has often relied on the same formalization of scientific theories as its critics (§3.5). The Received View also does not oversimplify the actual relation between theories and observations, nor does it introduce needless complexities in this respect (§3.6). It is also not hostile towards the use of models in science (§3.7). Furthermore, the Received View has not failed in its attempt to make the notion of a theory more precise, because it is not intended to do so. Rather, first, it is intended as a generalizable framework in which one can explicate *specific* theories (§3.8.2). Second, explication differs from precisification (§3.8.1).

Apart from this historical defense, I argue that syntactic approaches in general are as powerful as competing semantic ones (§4.1). The Received View in particular can capture and generalize van Fraassen's notions of a theory and of empirical adequacy (§4.2). It can furthermore capture the core notions of the partial structures approach in two different ways (§4.3).

II. Relations

Criteria of empirical significance are meant to distinguish between the sentences or terms that are connected to empirical claims or states and those that are not. Such criteria can be divided into criteria for the empirical significance of sentences and criteria for the empirical significance of terms. Along another dimension, one can distinguish between deductive criteria, which assume deductive inferences, and analogously, probabilistic criteria.

I show how all major deductive criteria for sentences are equivalent to falsifiability (§6.2), verifiability (§6.3), their disjunction (§6.6), or supervenience (§6.5). Based on these equivalences, I provide comparative criteria of empirical significance (§6.7.1), provide a very general notion of empirical claim and empirical fact (§6.8.1), and develop a notion of supplementary assumptions that avoids the problems of Ayer's criterion and its successors (§6.8.2). I further show that the major deductive criteria for terms are equivalent or similar to explicit definability, supervenience of terms (§7.2), reducibility (§7.3), or a criterion suggested independently by William Rozeboom and by Carnap (§7.4). Based on these results, I generalize some of these criteria to criteria for sets of terms (§7.5). I argue that those criteria for terms that provide something other than a mere means of reformulating criteria for sentences cannot be used to identify empirically significant sentences at all. However, some criteria for sets of terms can be used to distinguish theoretical concepts from mathematical ones (§7.6). I further show that Sober's (probabilistic) criterion of testability is inadequate according to his own conditions of adequacy and argue instead for two different criteria: the relevance criterion of probabilistic significance, which is the probabilistic analogue of falsifiability, and the Bayesian criterion, the analogue of the disjunction of falsifiability and verifiability (§8).

III. Applications

I apply the results of parts I and II to four distinct areas. In a straightforward application of deductive and probabilistic criteria for sentences, I analyze the structure of the theory of intelligent design (ID) and show that, in its most plausible formulations, ID is neither falsifiable nor probabilistically relevant. This result suggests that ID is not a science (§9).

Through the relation between probabilistic criteria of empirical significance and the notion of testability, the failure of likelihoodism to provide a criterion of empirical significance leads to a criticism of likelihoodism as an explication of confirmation. Analogously, the success of the relevance criterion and the Bayesian criterion of empirical significance suggests the relevance criterion of confirmation and, respectively, Bayesianism itself (§10). Furthermore, the deductive criteria for sentences and some criteria for terms can be used in the explication of different notions of reduction (§11). Finally, the criteria of empirical significance themselves can be used to analyze, generalize, and improve upon the formalism of artificial language philosophy. I map out such a development and apply some of the results to a question of concept formation in two ethical theories (§12).

1.2 Conventions

Newly introduced terms will be identified by *italics* when they are defined or assigned some weaker meaning postulates. I will use "quotation marks" to identify quotations, and 'single quotation marks' to identify names of words and phrases. In all quotes, I will silently change double quotation marks to single quotation marks whenever a word is mentioned. Scare quotes are accordingly always double quotes. ⌜Corners⌝ identify quasi-quotations, names of words and phrases in which variables may be substituted (so that '$\neg A$' is an instance of ⌜$\neg\sigma$⌝). When there is little danger of confusion, I will often suppress quotation and quasi-quotation marks and thus use symbols autonymously. A *concept* is the intension of a term, so that a definition or some weaker set of meaning postulates for a term (e. g., 'cause') determines the concept itself (e. g., *cause*).

Whenever I discuss a well-defined language like predicate logic, a *term* is any non-logical relation, function, or constant symbol of a language. While this goes against the typical terminology in logic, it is in line with the standard terminology in philosophy of science, where one speaks of 'observational terms', 'basic terms', 'definition in terms of', 'terminology', etc.

As logical constant symbols, I will use λ for generalized predicates (propositional functions), ι for definite descriptions, and '$\wedge$', '$\vee$', '$\rightarrow$', '$\leftrightarrow$', '$\forall$', '$\exists$' as usual; in quotes, I will silently change logical notation accordingly. I will change set theoretic notation from '$\subsetneq$', '$\supsetneq$', '$\subset$', and '$\supset$' to '$\subset$', '$\supset$', '$\subseteq$', and '$\supseteq$', respectively. Quantifiers and operators are always assumed to have minimal scope. To reduce the number of parentheses, I will assume that '$\neg$' binds more strongly than '$\vee$' and '$\wedge$', which binds more strongly than '$\leftrightarrow$' and '$\rightarrow$'. Unless stated otherwise, I will in the following assume a vocabulary $\mathscr{V}$ of m_i-place relation symbols P_i, n_j-place function symbols F_j, and constant symbols c_k. The function and relation symbols can be of any order. $\{m_i\}_{i\in I}$, $\{n_j\}_{j\in J}$, and the respective types of the individual relation, function, and constant symbols are collectively called their *similarity type*.

For model theoretic matters, I will rely on the standard notation as used by Chang and Keisler (1990, §1.3) and, more loosely, by Hodges (1993, §§1.2–1.3). Specifically, a *structure* $\mathfrak{A}$ is a pair $\langle A, \mathscr{I}\rangle$ consisting of a domain A and a function $\mathscr{I}$ from $\mathscr{V}$ to relations, functions, and constants on A of their respective types. Hence P_i is mapped to an m_i-ary relation, F_j to an n_j-ary function, and c_k to a constant on A. Sometimes I use indexed sets of structures $\mathfrak{M}_i$ instead of $\mathfrak{A}$, $\mathfrak{B}$, etc. A will always be the domain $|\mathfrak{A}|$ of $\mathfrak{A}$, $B = |\mathfrak{B}|$ etc. If $\mathfrak{A} = \langle A, \mathscr{I}\rangle$, I write $P_i^{\mathfrak{A}}$ instead of $\mathscr{I}(P_i)$, and analogously for functions and constants. $P_i^{\mathfrak{B}}$ is the relation in $\mathfrak{B}$ that *corresponds* to relation $P_i^{\mathfrak{A}}$ in $\mathfrak{A}$, and analogously for functions and constants. In displayed form, I write a structure $\mathfrak{A}$ as $\langle A, P_1^{\mathfrak{A}}, \dots, P_s^{\mathfrak{A}}, F_1^{\mathfrak{A}}, \dots, F_t^{\mathfrak{A}}, c_1^{\mathfrak{A}}, \dots, c_u^{\mathfrak{A}}\rangle$ or, for possibly infinite index sets, $\langle A, P_i^{\mathfrak{A}}, F_j^{\mathfrak{A}}, c_k^{\mathfrak{A}}\rangle_{i\in I, j\in J, k\in K}$. If it is necessary to give the vocabulary of the structure explicitly, I will write $\langle A, \langle P_1, P_1^{\mathfrak{A}}\rangle, \dots, \langle P_s, P_s^{\mathfrak{A}}\rangle, \langle F_1, F_1^{\mathfrak{A}}\rangle, \dots, \langle F_t, F_t^{\mathfrak{A}}\rangle, \langle c_1, c_1^{\mathfrak{A}}\rangle, \dots, \langle c_u, c_u^{\mathfrak{A}}\rangle\rangle$, even though that is far from graceful. If two structures $\mathfrak{A}$ and $\mathfrak{B}$ are isomorphic (Hodges 1993, 5), I will sometimes write this as '$\mathfrak{A} \simeq \mathfrak{B}$'. Sometimes, it will be convenient to identify the vocabulary of a structure in the structure's name by a subscript. A structure in a vocabulary $\mathscr{I}$, for example, may be identified by $\mathfrak{A}_{\mathscr{I}}$.

I will sometimes use '$\bar{x}$' as a string of the appropriate length and thus let, e. g., '$P_i\bar{x}$' stand for '$P_i x_1 \ldots x_{n_i}$'. Sometimes, '$\bar{x}$' will also stand for a tuple of the appropriate length, so that, e. g., '$\bar{x} \in P_i^{\mathfrak{A}}$' stands for '$\langle x_1, \ldots x_{n_i}\rangle \in P_i^{\mathfrak{A}}$'. '$\bar{x}_{-s}$' with $s \leq n_i$ stands for the string '$x_1 \ldots x_{s-1}x_{s+1} \ldots x_{n_i}$' or, respectively, the tuple '$\langle x_1, \ldots, x_{s-1}, x_{s+1}, \ldots, x_{n_i}\rangle$'.

When two or more phrases are separated by slashes ('/') in definitions, claims, or proofs, each phrase leads to a new sentence. If in close proximity there are multiple instances of n phrases separated by slashes, choosing uniformly the i^{th} phrase ($i \leq n$) each time leads to a new sentence. Phrases in parentheses can be uniformly left out of the definition, claim, or proof, again leading to a new sentence. For example, 'I am (sometimes) hungry/tired and thirsty/disheveled' is, when occurring in a definition, equivalent to 'I am sometimes hungry and thirsty and I am sometimes tired and disheveled and I am hungry and thirsty and I am tired and disheveled'. In some cases, phrases in parentheses can be uniformly substituted for the phrase preceding the parenthesis, so that, for example, 'I am always (sometimes) hungry/tired and thirsty/disheveled when I am traveling (at home)' is equivalent to 'I am always hungry/tired and thirsty/disheveled when I am traveling and I am sometimes hungry/tired and thirsty/disheveled when I am at home.' The different use of parentheses will be clear from the grammar of the sentence.

I will use 'inference' to refer to any rule that allows one to go from one statement to another. As usual, 'deduction' refers to syntactic inference ('$\vdash$'), and 'entailment' to semantic inference ('$\vDash$'). '$\Sigma \Dashv\vDash \Lambda$' stands for '$\Sigma \vDash \Lambda$ and $\Lambda \vDash \Sigma$'. For functions f and g, '$f \circ g$' stands for the concatenation of f and g, that is, for all x, $f \circ g(x) = f(g(x))$.

The years given in the references for quotations are to the original texts, with the bibliography containing a reference to the translation used, to which the page numbers refer.[1] I do not explicitly identify my own translations, but give the original text in a footnote; in that case the bibliography does not contain a reference to a translation.

[1]This avoids the use of rather arbitrary and historically misleading references like 'Carnap (1998)' and of cumbersome double numbers like 'Carnap (1928/1998)', although having the year refer to a different publication than the page numbers is admittedly awkward as well.

Part I

Foundations

Chapter 2

Artificial language philosophy

In the 20th century, the methods of traditional philosophy came under scrutiny by proponents of linguistic philosophy. Linguistic philosophy is partitioned into ordinary and artificial language philosophy.[1] According to the former, philosophical problems are best solved or dissolved by investigating the ordinary language we presently use (Rorty 1967a, §2; Kauppinen 2007). According to the latter, philosophical problems are best solved or dissolved by the development of new languages and concepts, and by the regimentation of existing languages and concepts (Rorty 1967a, §2), which has also been called their "improvement" (Carnap 1963c, §19), "reform" (Maxwell and Feigl 1961), or "explication" (Carnap 1950b, §§2–5).[2]

According to linguistic philosophy, language analysis or reform alone are sufficient to either provide solutions to philosophical problems, or avoid the problems completely. At first sight, this is an implausible position, for the answers to many questions traditionally considered philosophical quite obviously need more than language reform or analysis. For example, it seems that whether the

[1]Parts of this chapter have been published under the title "Ideal language philosophy and experiments on intuitions" (Lutz 2009). I thank an anonymous referee for *Studia Philosophica Estonica* for helpful comments. Other parts have been presented at Herman Philipse's *Dutch Research Seminar in Analytic Philosophy* at Utrecht University, the *Future of Philosophy of Science* conference at Tilburg University, The Netherlands, April 16, 2010 under the title "Ideal language philosophy of science" and have been published under the title "Artificial language philosophy of science" (Lutz 2011a). I thank the audience and two anonymous referees for the *European Journal for Philosophy of Science* for helpful comments.

[2]Rorty (1967a) popularized the term 'ideal language philosophy', but as Matteo Collodel, Eric Schliesser, and an anonymous referee for the *European Journal for Philosophy of Science* have pointed out, the name suggests the existence of a unique ideal language. The term's originator, Bergmann (1949, 439), similarly assumed that there can be one ideal language for all contexts. Since expositors (e. g. Rorty 1967a, Lutz 2009) have applied the term to the works of Carnap and philosophers of a similar inclination, while Carnap (1963c, 938) himself spoke of the construction of "artificial languages", the term 'artificial language philosophy' seems more apt.

world is deterministic should depend on what the world is like and thus be informed by our theories of the world, and the existence of exactly one god will be more or less plausible depending on alternative theories meant to account for empirical phenomena.

To be plausible, the position of linguistic philosophy has to be understood to pertain to philosophy's *contribution* to the solution of problems. To the extent that empirical investigation alone suffices for solving a problem, philosophy does not have a role to play. Philosophy's role is to solve or dissolve what is left of a problem when its empirical components (those questions that can be answered by empirical research) are excluded. As the extensive use and discussion of empirical results within logical empiricism (arguably the main application of artificial language philosophy) suggest, this position is neither in competition with empirical science, nor does it render empirical science irrelevant for philosophical questions (see Carnap 1934a, §72). It just divides the labor between empirical and linguistic work, the latter being the domain of philosophy. As Carnap (1934a, foreword) puts it, equating linguistic work with logical analysis (cf. Carus 2007, 256–259):

> In our "Vienna Circle" [. . .] the conviction has grown, and is steadily increasing, that metaphysics can make no claim to possessing a scientific character. That part of the work of philosophers which may be held to be scientific in its nature—excluding the empirical questions which can be referred to empirical science—consists of logical analysis.

Linguistic philosophy can be seen as a reaction to what I will call, for lack of a better term, '*traditional philosophy*'. In the paradigmatic case, traditional philosophy shares with empirical research the aim of arriving at truths about the world, but without direct recourse to empirical methods (cf. Cohnitz and Häggqvist 2009, 9). As an illustration, consider a point made by Williamson (2007, §2.1) in defense of traditional philosophy. He argues that "Was Mars always either dry or not dry?" is a question about Mars and is philosophical because the vague word 'dry' occurs crucially. Therefore philosophical questions are not always about language. A lot hinges here on the concept of *aboutness*, but quite simply, a positive answer to the question plausibly entails the existence of Mars. Whether there is such a thing is an empirical question and thus clearly cannot be answered by language analysis or reform alone. The question may therefore seem like a counterexample to the position of linguistic philosophy. But in linguistic philosophy not every traditional philosophical question as a whole is considered a philosophical problem: According to linguistic philosophy, the philosophical problem of the original question about Mars is the problem that, first, must be solved in order to answer the question, and second, cannot be solved by empirical research. If Mars does not exist or if it was always dry, there is no philosophical problem, because the original question can be answered on empirical grounds. But if Mars first was not dry and then gradually became dry, the philosophical problem of the original question is whether or under what conditions something that is first φ and then gradually becomes not φ is during the whole process φ or not φ. In traditional philosophy, one might use intuition as a guide for better understanding the nature of dryness to answer this question, but it is far from clear that this is the only option. Since the problem is non-empirical, it may also just be a problem of language.[3]

Apart from this division of labor between the empirical sciences and philosophy, linguistic philosophy goes one step further and claims that non-empirical problems are always problems of language, that is, they cannot be solved in any other way but by language analysis or reform (and thus, for example, the use of intuitions about the nature of dryness is not an option). Specifically, according to ordinary language philosophy, in Williamson's question about Mars the philosophical problem can only be solved by analyzing the use of the terms 'or' and 'not' in connection with vague terms; if the rules governing their use turn out to be too imprecise to answer the question, there is no fact of the matter. According to artificial language philosophy, the use of the terms 'or' and 'not' can be stipulated so that Williamson's question is answered in the positive or negative, or the problem can be avoided altogether. This could be achieved by developing a language without the vague term 'dry', and instead with terms to describe the amount of water on a planet. Let me call problems that cannot be solved, questions that cannot be answered, and statements that cannot be confirmed by empirical or linguistic means or a combination thereof '*trans-empirical*'.[4] Linguistic philosophy then claims that trans-empirical problems cannot be solved at all. That so far no trans-empirical question has been answered to almost everyone's satisfaction is seen as evidence for this (see Rorty 1967a, §1).

In large part as a reaction to artificial language philosophy, Quine (1969a) suggested to naturalize epistemology (and in fact all of philosophy that consists of the explication of concepts). *Naturalized philosophy* is more widely accepted as a philosophical methodology than both kinds of linguistic philosophy, but not as easily circumscribed. Roughly, it is the position that philosophical problems are best solved or dissolved through empirical research (Giere 1985; Feldman 2008, §2). A naturalized approach to the problem of the dryness of Mars might involve an empirical investigation into the neurological, psychological, or sociological phenomena connected to the use of the concept *dry*. In this abstract description, naturalized philosophy is a close kin to *experimental philosophy*, which uses experimental methods to investigate intuitions relevant for philosophical analysis. An experimental approach may consist, for example, in polling a selection of people about their use of the term 'dry'.

[3] This treatment of the example was developed in the reading group of the Theoretical Philosophy Unit at Utrecht University in the winter term of 2008/2009, in particular by Harmen Ghijsen and Jesper Tijmstra.

[4] In the old debate about linguistic philosophy, these problems, questions, and statements were often called 'metaphysical' (see the Carnap quote above) or 'speculative', but the first term is often used differently in current philosophical discussions, and the second seems unnecessarily pejorative.

There are other methodologies than ordinary language, artificial language, traditional, naturalized, and experimental philosophy. Although the methodologies are also not mutually exclusive, they diverge in important ways. It would be surprising if all of them led to the same results (i. e., solutions or dissolutions of problems in ways that are satisfactory within the respective methodology). It thus becomes important to determine which, if any, of these methodologies are correct.

In defense of experimental philosophy, Knobe (2007) suggests that linguistic philosophy is too restricted in its topics, and the philosophically interesting questions should be answered with the aid of experimental philosophy. In defense of traditional philosophy, Williamson (2007) argues for a metaphysics based on counterfactual reasoning that leaves linguistic philosophy behind. Ordinary language philosophy has been defended by Kauppinen (2007), who argues that intuitions about ordinary language are best elucidated without recourse to experiment. Proponents of artificial language philosophy, however, are conspicuously absent in the current debate about philosophical method. This may be because "logical empiricism [has lost] its status as a philosophical project to be pursued" (Richardson 2007, 346), and since artificial language philosophy was mainly applied within logical empiricism, it, too, has lost this status (Kuipers 2007, §1).

The neglect of artificial language philosophy in the current debate on intuitions and experimental philosophy is an oversight because, first, the old debates between traditional, ordinary, and artificial language philosophy and between artificial and naturalized philosophy were never conclusively decided. Artificial language philosophy is therefore as much a contender now as it ever was.

Second, in the current debate, the philosophical relevance of experimental philosophy is sometimes challenged through an analogical consideration that, on further thought, supports artificial language philosophy (if it supports anything at all, given that it is only an analogy). Nadelhoffer and Nahmias (2007, 129) describe the consideration in the following way:

> Upon first hearing of experimental philosophy, many philosophers conclude from the start that empirical data concerning folk intuitions are irrelevant to philosophical debates because the folk intuitions *themselves* are irrelevant to such debates. After all, scientists and mathematicians tend not to worry about whether their theories settle with the intuitions of lay-persons. Why should philosophers be any different? On this view, even if our own "expert" intuitions correctly come into play when we're doing philosophy, the untutored and uninformed intuitions of the "person on the street" have no similar role to play.

Under the assumption that "so-called intuitions are simply judgments (or dispositions to judgment)" (Williamson 2007, 3), Williamson (2007, 191) gives an elaboration of this consideration:

> Much of the evidence for cross-cultural variation in judgments on thought experiments concerns verdicts by people without philosophical training. Yet philosophy students have to learn how to apply general concepts to specific examples with careful attention to the relevant subtleties, just as law students have to learn how to analyze hypothetical cases.

In jurisprudence it may be even easier to note than in natural science or mathematics that the concepts the experts apply are not those of ordinary language (cf. Carus 2007, 275). The concepts are highly refined and have been developed in science or jurisprudence itself, sometimes but not always with a basis in ordinary language. It is this kind of concept that is applied in artificial language philosophy. Therefore, this consideration is no defense of ordinary language philosophy against criticisms relying on experiments, but it may be a defense of artificial language philosophy.

There are many discussions about philosophical methodology (e. g., Williamson 2007, Papineau 2009b), some of which even take into account artificial language philosophy (Rorty 1967b). But, arguably, no conclusive defense or criticism has been established about any of the five divergent methodologies. In such a case, a cumulative strategy is often pursued: The individual advantages and disadvantages of each methodology are compared and weighed. For instance, a methodology that yields more results than another in a specific domain has a clear advantage. In the domain of philosophy, naturalized philosophy has been charged with a complete lack of results (cf. Kim 1988), putting it at an immediate disadvantage in any such comparison. Williamson's analogy amounts to the same charge with respect to experimental philosophy. Rorty (1967a, 3) discusses the argument that both ordinary and artificial language philosophy have yielded more philosophical results than traditional philosophy, and Carnap (1963c, 939–940) and Maxwell and Feigl (1961, 491–492) argue that artificial language philosophy has yielded more philosophical results than ordinary language philosophy. The following discussion of the status of philosophical results and the relation of science and philosophy of science according to artificial language philosophy suggest another advantage of artificial language philosophy: The methodology leads to results in the domain of *meta*-philosophy (rather than philosophy) that are not obvious for the other methods.

Those meta-philosophical results are that, first, artificial language philosophy interacts fruitfully with the sciences (which is not obvious for ordinary language and traditional philosophy), and second, it answers philosophical questions (which is not obvious for naturalized and experimental philosophy). Furthermore, I will argue that results from experimental philosophy support artificial language philosophy, and that artificial language philosophy is methodologically naturalistic.

2.1 An outline of traditional philosophy

Traditional philosophy can be considered an investigation of facts about the world. However, a straightforward empirical investigation of, say, hydrochloric acid differs from the philosophical investigation of causation in that hydrochloric acid has a specific density, decomposes at a specific temperature, and generally has properties that can be determined with a certain degree of confidence through measurements and, ultimately, through observations of the outcomes of these measurements. Accordingly, there are experimental ways to determine whether an unknown substance is hydrochloric acid. On the other hand, there is no measurement to determine how causation connects to explanation, and there is no experimental way of determining whether a specific situation or process is an instance of explanation or causation.[5]

In traditional philosophy, intuition is often the analogue of experiments and observations (cf. Sosa 2007, 105). Papineau (2009b, 14), for example, states that

> Gettier appealed to the intuition that a belief whose truth is accidental relative to its method of justification is not knowledge; Kripke appealed to the intuition that something that is not the causal origin of a name is not its bearer; and so on. On my account, all these intuitions are synthetic claims about the relevant kind of scenario.

In other words, these intuitions are about the world. But there are many meanings of 'intuition'; for example, Feigl (1958, 2) distinguishes seven of them, where

> [t]he common core in the many connotations of 'intuition' is, of course, *immediacy*. Intuition has thus been contrasted, traditionally and quite generally, with indirect, mediate, relational, or inferential knowledge. Intuition is often identified with direct insight or immediate apprehension.

Since Feigl is mainly interested in trans-empirical claims, his discussion focuses on the justifiability of "mystical or trans-empirical intuition", where "the target or object of the intuition is claimed to be something that is absolutely beyond the reach of ordinary experience and reasoning, something which cannot be checked empirically" (Feigl 1958, 6). Note the distinction between an intuition and its target, the target being the statement for which the intuition is purported evidence. Using intuitions to answer trans-empirical questions goes beyond what is accepted in linguistic philosophy, and such trans-empirical intuitions must be distinguished

[5] Of course, once the philosophical investigation of, say, causation has resulted in a clear definition (e. g., in the form of a causal search algorithm or via the concept of mark transmission), the definition can be applied to determine whether a situation or process is an instance of causation. I thank Jan Sprenger for asking about the status of causal search algorithms.

from another kind of intuition that Feigl (1958, 6) calls, for lack of a better word, 'hunch':

> We can define the "hunch", then, as "a product of learning from past experience, which learning is not made explicit at the moment of the use of judgment".

If one has a hunch, one follows an empirical rule of which one is not aware. Hunches are therefore unproblematic, but must not be confused with trans-empirical intuitions, because unlike the target of a hunch, the target of a trans-empirical intuition cannot be tested empirically (Feigl 1958, 6–7). In general, "[i]f 'intuition' in one of its many senses designates a way of knowing, it need not, and indeed does not, designate such a way in some of the other senses" (Feigl 1958, 1).

If traditional philosophy were to rely on hunches, it would be amenable to empirical test. However, according to Papineau (2009b, 18) it relies on trans-empirical intuitions:

> If my judgemental procedures decide who is a knower by assuming, *inter alia*, that accidentally true believers are not knowers, then clearly there isn't any question of my meeting up with a case where I judge such an accidentally true believer to be a knower after all. Again, if my judgemental procedures decide what thing bears some name by noting the causal origin of the use of the name, then I'm not going to come across cases where I judge that some name is borne by something other than its causal origin. But this impossibility of direct falsification does not mean that the relevant general assumptions are analytic. They may yet have a substantial synthetic content [. . .].

For both Papineau (2009b, §IV) and Williamson (2007, §6), thought experiments are a core method in traditional philosophy. But while Papineau considers thought experiments to elicit intuitive judgments, Williamson does not distinguish this type of judgment and judgments simpliciter. Williamson (2007, 3) holds that "so-called intuitions are simply judgments (or dispositions to judgment)", but thereby brushes over the important distinction between judgments that can be supported by explicit argument or observation and those that cannot be, or at least not completely. Judgments that do not have sufficient support are intuitive, and are used analogously to observations. This is actually demonstrated in a brief overview given by Williamson (2007, §7.2) of other philosophers' positions on intuition, in which intuitions are always used as premises (rather than conclusions).

It is clear that intuitions in individual thought experiments about, say, the presence or absence of causation or explanation cannot entail a general rule about the relation of causation and explanation. Such a rule has to be either postulated like a scientific hypothesis, or gathered from more general intuitions about the relation itself. The general rule can then be tested against the intuitions in the individual thought experiments, and in the case of inconsistency, the intuitions about either the general rule or the thought experiments have to be modified. This method of testing and revision may be repeated and may eventually lead to a reflective equilibrium, where the intuitions about the general rule and the thought experiments agree (Daniels 2008, §1). Testing these intuitions for consistency is a matter of rigorous derivation,[6] which is used in every philosophical methodology here discussed. How an inconsistency is resolved, however, will rely crucially on the intuitions themselves.

An obvious problem for the use of trans-empirical intuitions is that it is not clear why their targets should be true. One could argue for the truth of the targets of intuitions that are shared among many people with an inference to the best explanation. Against this, Feigl (1958, 12–13) points out that intuitions can be treated by empirical psychology just like other mental states. It might then be possible to account for the occurrence of shared intuitions, that is, to explain why different people have intuitions with the same targets without having to assume their truth.

This line of research into the source of intuitions is taken up by Hare (1975, §I), who rhetorically asks: "[H]ow do we know whether what we feel inclined to say [about some example of a moral conflict] has any secure ground? May we not feel inclined to say it just because of the way we were brought up to think?" For Hare, our intuitions may simply be the result of our upbringing, and will differ accordingly. Given the lack of consensus on trans-empirical questions and Feigl's hope of explaining trans-empirical intuitions without assuming the truth of their targets, it seems that language analysis or reform may be the only means of rationally solving problems that cannot be answered empirically (see also Bohnert 1963, 410).

2.2 An outline of ordinary language philosophy

Ordinary language philosophy claims that trans-empirical problems only occur when language is not used in ordinary ways; therefore the analysis of ordinary language is enough to solve philosophical problems (Rorty 1967a, 12).[7] Hare (1975, §III) gives an example of this when he rejects the use of the term 'person' in the discussion of the problem of abortion, because

> 'person' [. . .] is not a fully determinate concept[.] It is no use looking more closely at the fetus to satisfy ourselves that it is *really* a person

[6]Like in mathematics, a rigorous derivation need not be completely formalized. Of central importance is that it does not rely on unarticulated assumptions.

[7]According to Williamson (2007, 13–14), the linguistic turn towards language has been largely superseded by a conceptual turn towards the concepts of thought. While I will discuss only ordinary language philosophy, this discussion is also applicable to methodologies that are based on analyses of the concepts of thought.

> [. . .]; we already have all the information that we need about the fetus. [. . .]
>
> To say that the fetus is (or is not) a person gives *by itself* no moral reason for or against killing it; it merely incapsulates any reasons we may have for including the fetus within a certain category of creatures that it is, or is not, wrong to kill (i. e. persons or nonpersons). The word 'person' is doing no work here (other than that of bemusing us).

Hare claims that because the concept of a person is vague, and specifically not determinate in the case of a fetus, one should avoid the question of whether a fetus is a person altogether. Trying to answer it in order to solve the problem of abortion would be to use language in a non-ordinary way, and would therefore only introduce trans-empirical problems because neither language nor empirical science can decide the personhood question. If Hare now solves the problem of abortion in some other way, he thereby dissolves the problem of the personhood of the fetus. Similarly, one might argue that 'not' and 'or' are too vague to determine whether 'Mars was always dry or not dry' is true. In this case, however, it may be that nothing further hinges on the answer (since the question was only introduced to make a philosophical point), and thus the problem may already be dissolved.

In traditional philosophy, intuitions are taken to provide information about the world, for example about the dryness of Mars; in the terminology of Fedyk (2009, §4), they are interpreted as world-directed. In ordinary language philosophy, on the other hand, the same intuitions are interpreted as providing information about the language in which the world is described; they are interpreted as meaning-directed. If ordinary language philosophy relies on the actual language use of some group, or the actual linguistic intuitions of some group, then one difference between traditional and ordinary language philosophy is straightforward: The claims of traditional philosophy are about unobservable states of the world, while the claims of ordinary language philosophy are about observable states (in the case of actual language use) or states that can be empirically determined with some certainty (in the case of linguistic intuitions).

Like the reliance on intuitions in thought experiments in traditional philosophy, the exclusive reliance on actual language use, that is, individual speech acts, cannot establish general rules about language use. Such general rules may be either stipulated like any other empirical theory or gathered from intuitions about the general rules themselves. Once spelled out explicitly, the general rules may be incompatible with individual speech acts. As in the case of traditional philosophy, such inconsistencies can be established through rigorous derivation. If the method of reflective equilibrium is used to resolve inconsistencies by excluding certain instances of language use or certain general rules, then ordinary language philosophy has a normative component (Carnap 1939, §4). In this case, any intuitions about how language *should* be used are trans-empirical, and the discussion in §2.1 applies.

2.3 An outline of artificial language philosophy

Like ordinary language philosophy, artificial language philosophy considers philosophical problems to be problems of language. Unlike ordinary language philosophy, however, artificial language philosophy contends that philosophical problems are best solved or dissolved by the conventional prescription of a new language, not by the analysis of actual language use. For there are so many vague concepts in ordinary language that many philosophical problems cannot be solved in ordinary language at all. Other problems with ordinary language include the ambiguity of words even in ordinary contexts, and the possible embedding of false beliefs in the rules of ordinary language, as Maxwell and Feigl (1961, 496) argue. When applied to philosophical questions, these problems worsen, and so artificial language philosophers claim that philosophical problems are best solved by the development of new languages. Such a new language must be clear enough that an answer to the original question can be rigorously derived (thus leading to a solution of the problem), or be such that the problem cannot be formulated in the first place (amounting to its dissolution) (Bergmann 1957, 326). An artificial language philosopher may even suggest language reforms for contexts in which there are no problems, either because the resulting language has some pragmatic advantage like greater simplicity or precision, or because this change helps to avoid problems in other contexts (Maxwell and Feigl 1961, 491).

2.3.1 Language choice

In principle, there is no restriction on the choice of language. As Carnap (1934a, §17) states in his "Principle of Tolerance", even the logic of a language is conventional.[8] For what follows, however, it will suffice to rely on a very simple case of language choice. I will assume that in addressing the philosophical problems at hand, the logic of the language is taken as fixed, and there is some set $\mathscr{B}$ of *basic terms* ($\mathscr{B}$-terms) whose application is taken as unproblematic. It may be taken as unproblematic because the terms apply more or less immediately to observations (cf. Chang 2005), but more generally, the basic terms simply refer to concepts that are not themselves under investigation (Reichenbach 1951, 49; cf. Lewis 1970, 428). The choice of a language then amounts to *concept formation*, that is, the choice of meaning postulates for the terms not in $\mathscr{B}$, which are the *auxiliary terms* ($\mathscr{A}$-terms) with the corresponding set '$\mathscr{A}$'.

$\mathscr{A}$ can contain terms for pre-theoretically understood concepts whose explication is intended to solve or dissolve the problems at hand. Carnap (1962, §2) describes this process:

> The task of *explication* consists in transforming a given more or less inexact concept into an exact one or, rather, in replacing the

[8] Carnap (1928a, §107) held this view of the conventionality of logic somewhat earlier.

> first by the second. We call the given concept (or the term used for it) the *explicandum*, and the exact concept proposed to take the place of the first (or the term proposed for it) the *explicatum*. The explicandum may belong to everyday language or to a previous stage in the development of scientific language.

Finding an exact explicatum for an inexact explicandum is not a straightforward matter. The first step is to describe the explicandum as precisely as possible. This description forms the basis for the explication itself, which Carnap (1962, §3) describes as follows:

> If a concept is given as explicandum, the task consists in finding another concept as its explicatum which fulfils the following requirements to a sufficient degree.
>
> 1. The explicatum is to be *similar to the explicandum* in such a way that, in most cases in which the explicandum has so far been used, the explicatum can be used; however, close similarity is not required, and considerable differences are permitted.
> 2. The characterization of the explicatum, that is, the rules of its use (for instance, in the form of a definition), is to be given in an *exact* form, so as to introduce the explicatum into a well-connected system of scientific concepts.
> 3. The explicatum is to be a *fruitful* concept, that is, useful for the formulation of many universal statements (empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept).
> 4. The explicatum should be as *simple* as possible; this means as simple as the more important requirements (1), (2), and (3) permit.

The difference between artificial and ordinary language philosophy is clear: According to artificial language philosophy, the analysis of a concept carried out in ordinary language philosophy is only the preliminary step needed for an explication in Carnap's sense. This explication may result not only in a more precise concept, but also in a concept that conflicts with clear cases of the explicandum. As Carnap (1962, §3) notes, "one might perhaps think that the explicatum should be as close to or as similar with the explicandum as the latter's vagueness permits", but he claims that "this requirement would be too strong" because of requirement (3). That the requirement of fruitfulness sometimes leads to conflicts between an explicatum and clear cases of its explicandum can be seen from "the actual procedure of scientists", for example in zoology's explication of 'fish': "The prescientific term 'fish' was meant in about the sense of 'animal living in water'", while the zoologists' explicatum means "animals which live in water, are cold-blooded vertebrates, and have gills throughout life" (Carnap 1962, §3). Oftentimes, one must explicate several terms at once, because only then are the resulting explicata fruitful.

The explicata also have to fulfill conditions of adequacy (cf. Tarski 1944, §4), which identify what problems the newly formed concepts should solve or dissolve, and in what contexts they should be applicable. The contexts are suggested by the pre-theoretic uses of the explicanda (Kuipers 2007, §2), and thus can be used to determine when explicata are similar enough to their explicanda.[9] The terms 'dry' and 'wet', for instance, may be replaced by the explicatum 'aridity', defined as 'the ratio of the volume of water on the surface and the surface area'. Depending on the intended application, claims involving the dryness of Mars can then often be more precisely expressed by claims involving the aridity of Mars. In this example, 'ratio', 'volume', 'water', etc. are assumed to be in $\mathscr{B}$, but each of these terms can be in $\mathscr{A}$ in other contexts. In general, since the bipartition of terms into $\mathscr{B}$ and $\mathscr{A}$ is context dependent, a term P explicated in one context with the help of a term Q could in another context be used to explicate Q.

This is different from *circular definitions*, in which Q occurs essentially in the definiens of P and *vice versa* in the *same* context; an example is the interventionist definition of causation (Woodward 2008, §7). If both terms are in $\mathscr{B}$, circular definitions are unproblematic, since both of the terms are clear and the two definitions just follow from the terms' precise meanings. If one term is in $\mathscr{B}$ and the other in $\mathscr{A}$, the interpretation of the $\mathscr{B}$-term determines the interpretation of the $\mathscr{A}$-term. If both terms are in $\mathscr{A}$, circular definitions amount to postulates for the two terms, which determine the terms' interpretations only insofar as they connect the terms to $\mathscr{B}$-terms (see §2.8.2). The terms may also be somewhat interpreted but vague, so that the two definitions amount to additional meaning postulates that precisify one or both terms. I will discuss vague terms and circular definitions in §2.8.3.

$\mathscr{A}$ can also contain terms for entirely new concepts, which do not act as explicata. One may, for example, introduce concepts like *mark transmission* or *lawlike generalization* simply because they solve some problems in philosophy, not because there are corresponding explicanda. In this case, the conditions of adequacy can be chosen freely and may, for instance, amount to a list of problems that the new concepts should solve. Since explicata are new concepts as well, there is, in a sense, a continuum between the explication of terms and the development of wholly new concepts. Luckily, nothing hinges on the status of a new concept as an explicatum. It is only important that it fulfills the conditions of adequacy

[9] Therefore, if preserving a specific previous use of the explicanda by the explicata is deemed expedient, this has to be made explicit in the conditions of adequacy. This avoids the implication of Carnap's similarity demand that more than half of current uses must be captured, which suggests that there is something sacred about the current usage of a term, and has justly been criticized by Laudan (1986, 120).

and is fruitful.

The choice of the conditions of adequacy is ultimately pragmatic because it depends on the problems that the concepts are meant to solve. The conditions should be precise enough, however, to determine reliably whether they are fulfilled by the concepts. Of two sets of concepts that fulfill all of the conditions, the more fruitful is to be preferred, where Carnap judges fruitfulness by the number of results one can establish about the concepts, to which Kemeny (1963, 76) adds the number of new research questions that they suggest. This evaluation itself is, of course, deeply pragmatic, since neither every result nor every new research question should count equally.

Since the auxiliary terms are taken to be problematic, their interpretation is determined solely by the meaning of the basic terms, empirical results about the extensions of the basic terms, and the set Π of *meaning postulates*. The meaning postulates have been chosen to be true in the process of concept formation. For it to be possible to *choose* Π to be true, Π must be analytic (that is, devoid of empirical import). This identification of analyticity with conventionality finds its strongest expression in Carnap's Principle of Tolerance, in which logic—the paradigmatic example of analyticity—is taken to be conventional.

2.3.2 Solving and dissolving problems

There are now three methods of determining the truth or falsity of a sentence φ involving auxiliary terms. All of them rely on rigorous derivation, thereby sharing a core aspect with traditional and ordinary language philosophy. In the first method, the analytic truth (or falsity) of φ is derived from the meaning postulates Π. If this is impossible, it may still be possible to determine empirically that φ is true (or false), given the meaning of the basic terms. Finally, if neither method is applicable, new meaning postulates for the terms in φ must be developed so that one of the first two methods *becomes* applicable. This third method thus involves concept formation, making the truth or falsity of φ a matter of convention. Since the derivations rely on the meaning postulates and the logic, which have to be chosen by convention, φ can be true (or false) for only two reasons: language convention (which determines the meaning of the basic terms and the meaning postulates) or empirical research (which determines empirical results phrased in basic terms). If philosophy does not engage in empirical research, this means that all philosophical results are analytic, consisting of language conventions or rigorous derivations that rely on them. Note that rigorous derivations can be important for language choice, since they can reveal otherwise hidden features of a language.

The difference between the first two methods and the last method of determining the status of φ illustrates the distinction between "internal" and "external" questions introduced by Carnap (1950a, §2). The internal questions are those that rely on a chosen language (in Carnap's terms, a "linguistic framework"), that is, a chosen logic and a chosen set of meaning postulates. Within this language, the investigation of the status of φ is objective. For, whether φ is true or false depends solely on the state of the world and on the language, which is fixed.[10] The third method, that of concept formation, provides a means of answering external questions. Here, the truth or falsity of φ is not determined objectively, but rather by convention, and a claim about the status of φ cannot be right or wrong, but only more or less practical. There is, in this sense, no fact of the matter.

The situation is different if one discovers an inconsistency in a set of postulates (say, for some specific concept). In this case, one cannot add more postulates, but must rather remove some until the postulates become consistent and the $\mathscr{B}$-sentences true. Of course, one may add new postulates as well. Conversely, one may remove some postulates even if there is no inconsistency, simply because the resulting concepts are more fruitful.

The dissolution of problems always involves the removal of postulates Θ. In Hare's case, dissolving the problem of the personhood of the fetus involves the removal of the postulates for *person*. More importantly, the dissolution of a problem requires the introduction of postulates Ξ for new terms that avoid the original problems. If the postulates Θ have not been shown to be false, Ξ should be at least as expressive as Θ with respect to $\mathscr{B}$. Ξ should also be at least as fruitful as Θ, although this, of course, is much harder to determine.

Since both solving and dissolving problems can involve the removal and the addition of postulates, there is no clear cut distinction between the two endeavors. As in the case of the continuum between the explication of old terms and the introduction of completely new concepts, nothing philosophically important hinges on the distinction. It is just important that—whether through solution or dissolution—the problem has been resolved.

2.4 The old debate within linguistic philosophy

It has frequently been pointed out by proponents of ordinary and artificial language philosophy that the distinction between their two approaches is only a matter of degree (see for example Carnap 1955, §1, 1963c, §19, Hare 1960, 158, and Kemeny 1963, 71, 74), for one because the preliminary step of an explication consists in describing the natural language concept. Rorty (1967b, 12) puts it this way: "As has often been (somewhat crudely, but fairly accurately) said, the only difference between Ideal Language Philosophers and Natural Language Philosophers is a disagreement about which language is Ideal". Still, each side considers its own approach to be more appropriate for solving philosophical problems, as the discussions collected by Rorty (1967b), the criticism of artificial language philosophy by Strawson (1963), and the response by Carnap (1963c, §19) make

[10] *Within* a given language, the main tenets of realism would therefore seem true (I thank an anonymous referee for the *European Journal of Philosophy of Science* for this point).

clear.

The central charge by Strawson (1963, 504–505) is that artificial language philosophy's "claim to clarify will seem empty, unless the results achieved have some bearing on the typical philosophical problems and difficulties which arise concerning the concepts to be clarified". Bergmann (1949) counters this kind of criticism, arguing that these problems arising in ordinary language use need not be solved in the first place, but can be avoided without loss. For, the goal of an artificial language is precisely to express and analyze empirical claims without leading to such problems. The artificial language can use explicata rather than the explicanda of ordinary language philosophy, and if some statement in the artificial language helps to answer a question phrased in ordinary language, say, because the explicata involved are sufficiently close to their explicanda, this is nice but not necessary. A supporting position is taken up by Bohnert (1963, §II) and arguably Neurath (1932, 206), who argue that an artificial language does not need to be translated into ordinary language to be understood, because it can be learned by itself like a natural language (cf. Carnap 1963c, 938–939). Rorty (1967b, 16) notes the possible pragmatic response that a philosophical approach can only be considered viable insofar as it leads to results, and so far, ordinary language has not shown itself to be very helpful for solving philosophical problems (see also Maxwell and Feigl 1961, 491–492).

In a critique of ordinary language philosophy, Maxwell and Feigl (1961, 490–491, emphasis removed) point out that "[a] large portion of philosophical problems arise from consideration of unusual cases", and they see "absolutely no reason to believe that examination of ordinary use in the 'paradigm', normal cases can provide us with definitive rules for 'proper' use in the unusual and novel cases". In other words, Strawson's insistence on the use of ordinary language in order to clarify philosophical problems may be exactly why there is continuing disagreement. Maxwell and Feigl (1961, 491) state further that the "consideration of atypical cases often points up possible inadequacies and may suggest improvements in our conceptualization of the 'normal' cases". This clearly marks the move from ordinary to artificial language philosophy: If a concept is, for example, too vague to be applied in atypical but philosophically interesting cases, this necessitates its explication, which may lead to a concept that is applied differently even in typical cases.

Mates (1958) discusses ordinary language philosophy from a point of view that will be particularly useful in what follows: He treats the apparently factual statements about ordinary language that the proponents of ordinary language philosophy make like any other empirical hypotheses, and accordingly asks how such statements could be tested. First, Mates (1958, 165) considers the claim that

> the average adult has already amassed such a tremendous amount of empirical information about the use of his native language, that he can depend upon his own intuition or memory and need not undertake a

> laborious questioning of other people, even when he is dealing with the tricky terms which are central in philosophical problems. Such a assertion is itself an empirical hypothesis, of a sort which used to be invoked in favor of armchair psychology, and it is not born out by the facts.

Mates goes on to state that many authors are not even reliable when it comes to their own linguistic behavior, and after noting a disagreement between Ryle and Austin about the use of 'voluntary', concludes: "If agreement about usage cannot be reached within so restricted a sample as the class of Oxford Professors of Philosophy, what are the prospects when the sample is enlarged?"

Mates then suggests that there are essentially two empirical methods for determining the meaning and use of a word, which he calls extensional and intensional. The extensional method consists in observing a certain number of applications of a word and finding out what these applications have in common. Mates saw this method used almost exclusively in the ordinary language philosophy of his time, and laments the neglect of the intensional method. In the intensional method, the subjects are asked how they use or what they mean by a given word, and, "in Socratic fashion", are subsequently presented with apparent counterexamples and borderline cases, then are asked to revise their initial response, and so on, until a fairly stable account is reached (Mates 1958, 165–166). However, Mates observes that there is no guarantee these two methods will yield the same results, and the only way to solve this problem may be to make do with the different meanings of words that result from each method.

Furthermore, Mates argues, both methods have internal difficulties. In the extensional method, it is unclear which occurrences of a word are under consideration, and what the relevant features are of any object to which the word is applied. Since any set of objects will have infinitely many things in common, it is, for example, not obvious when a word has more than one meaning. Two words might also, just by happenstance, apply to the same objects in the domain under investigation (Mates 1958, 167–168). The problem with the intensional method is that "it does not seem possible to differentiate in a practical way between *finding out* what someone means by a word, and *influencing* his linguistic behavior relative to that word" (Mates 1958, 169–170). Mates suggests we test this by trying to devise "Socratic questionnaires" that make the definitions from different subjects converge, and others that make them diverge. If it is possible to construct the latter, the Socratic method cannot be considered a reliable means of finding out the meanings of words (Mates 1958, 171, n. 11).

This doubt about the reliability of the intensional method is also voiced by Maxwell and Feigl (1961, 489), who "know of no decision procedure for classifying each particular case [as one of finding out or of influencing], and [. . .] strongly suspect that many cases of putative ordinary-usage analysis are, in fact, disguised reformations". Accordingly, artificial language philosophy cannot be dismissed on the grounds that a change of language introduces insurmountable problems, because the intensional, Socratic method of ordinary language philosophy may very well lead to as much of a change of meaning as the process of explication. The difference between artificial and ordinary language philosophy then would be mainly that, while explication is done with very specific, explicitly stated goals in mind, it is not clear how or why the change of language is effected in the intensional method of ordinary language philosophy.

But even though artificial language philosophers are doubtful of the possibility of solving philosophical problems through ordinary language analysis, there is no doubt that the ordinary use of terms can be determined in some cases and that the construction of an artificial language can be inspired by ordinary language. Like Mates, artificial language philosophers consider claims about ordinary language to be empirical hypotheses. Carnap (1963c, §15.C), for example, states when discussing an article of his on meaning and synonymy in natural languages:

> "The sentence S_1 is analytic in language L for person X" [...] is an empirical hypothesis which can be tested by observations of the speaking behavior of X. If anyone is still sceptical about this possibility, I should like to refer him to a recent book by Arne Naess, which shows by numerous examples how hypotheses about the synonymy of expressions can be tested by empirical procedures.

The book, which Carnap identifies in a footnote, is *Interpretation and Preciseness: A Contribution to the Theory of Communication* (Naess 1953). It is an empirical study of natural language, much as demanded by Mates. In his own article on meaning and synonymy, Carnap (1955, §1) notes how such a study of natural language (which he calls "pragmatics") may be useful for the logician's development of an artificial language (which he calls "semantics"):

> If he wishes to find out an efficient form for a language system to be used, say, in a branch of empirical science, he might find fruitful suggestions by a study of the natural development of the language of scientists and even of the everyday language. Many of the concepts used today in pure semantics were indeed suggested by corresponding pragmatical concepts which had been used for natural languages by philosophers or linguists, though usually without exact definitions. Those semantical concepts were, in a sense, intended as explicata for the corresponding pragmatical concepts.

Carnap then goes on to describe an experimental procedure for determining the meaning of terms.

So the empirical study of ordinary language use can be of much value for artificial language philosophers. They do object, however, to the claim that the results of such empirical research can show their constructed artificial languages wrong. As Popper (1963, 201, n. 44) recalls, Naess began his research for an earlier book, *"Truth" as Conceived by those who are not Professional Philosophers* (1938), "in the hope to refute Tarski" (that is, Tarski's explication of 'truth'). Carnap (1948, 29, §7) replies that

> Arne Ness [sic] has expressed some doubts about the assertion [that Tarski's explication is in agreement with the ordinary use of the word 'true'], based on systematic questioning of people. At any rate, this question is of a pragmatical (historical, psychological) nature and has not much bearing on the questions of the method and result of semantics.

Carnap's opinion is shared by Popper (1963, 213, n. 64), who describes the reply as "a just dismissal of the relevance of Arne Ness' questionnaire method". This makes sense, given that an explication is not meant to capture ordinary language use.

In general, artificial language philosophers see the analysis of ordinary language as a straightforwardly empirical endeavor whose results are complementary to their own. The empirical results can serve as a starting point for explication or as inspiration for the construction of an artificial language, and in this respect, artificial language philosophy can profit from the analysis of ordinary language. But results from this analysis will not contradict those from artificial language philosophy, because artificial languages need not capture every feature of ordinary language. As Maxwell and Feigl (1961, 491) put it, "ordinary language *is* (often) the first word—but, quite often, this is all that it can do".

2.5 An outline of experimental and naturalized philosophy

There are a multitude of interpretations of 'naturalized philosophy' and 'naturalism' (Papineau 2009a, Feldman 2008), but for the sequel, it will be convenient to distinguish between *replacement naturalism*, the position that philosophical questions are best answered by science, *cooperative naturalism*, the position that empirical results are essential or useful for making progress in addressing philosophical questions, and *substantive naturalism*, the position that all philosophical facts are natural facts (Feldman 2008).[11] The first two kinds of naturalism are instances of *methodological naturalism*, the position that "philosophy and science [are] engaged in essentially the same enterprise, pursuing similar ends and using similar methods", and the last kind can also be called *ontological naturalism* (Papineau 2009a). Since I am interested here in philosophical methodologies, I will

[11] This is a generalization of Feldman's exposition, who only defines the different kinds of naturalism for epistemology.

focus on the two kinds of methodological naturalism.

Papineau's definition of methodological naturalism is ambiguous between the claim that *good* philosophy is engaged in the same enterprise as science and the claim that philosophy *as currently practiced* is so engaged. I will criticize Papineau's argument for the latter claim in §2.9. For now, I will consider methodological naturalism as interpreted by the former claim.

Quine (1969b) has given a highly influential argument for replacement naturalism, which he develops in the context of epistemology as a reaction to Carnap's method of explication, or, as it is also called, "rational reconstruction". He distinguishes between the "doctrinal side" of traditional epistemology (the attempt to justify all knowledge from sense experience) and its "conceptual side" (the attempt to explain all our concepts in sensory terms) (Quine 1969b, 71–74) and presents Carnap's work *Der logische Aufbau der Welt* (Carnap 1928a) as the most successful but still failed attempt at completing the conceptual side of epistemology by defining all concepts in sensory terms. Quine (1969b, 75) identifies the concepts that Carnap (1928a) introduces as explicata, and wonders about the relevance of explication:

> But why all this creative reconstruction, all this make-believe? The stimulation of his sensory receptors is all the evidence anybody has had to go on, ultimately, in arriving at his picture of the world. Why not just see how this construction really proceeds? Why not settle for psychology?

Since this is a rhetorical question, Quine then proceeds to outline how psychology should replace epistemology. Of course, the question could be asked of any explication of a concept, not only epistemic ones. Therefore this line of reasoning leads to a naturalization of all of philosophy, not only epistemology.

However, given the preceding discussion, the answer to Quine's rhetorical question is clear: Explication cannot be replaced by psychology because the goal of explication is not to find out about the actual concepts that humans have (the explicanda), but rather to find explicata that are *better* than the actual concepts. A naturalized philosophy as described by Quine, and indeed any naturalized philosophy that only determines which concepts humans have, would only provide the first step of artificial language philosophy.

Naturalized philosophy *could* be considered a replacement for ordinary language philosophy, however, since ordinary language philosophy ostensibly aims at determining the concepts that people actually have. To the extent that these concepts are revealed by people's judgments about trans-empirical matters, naturalized philosophy has a close connection to experimental philosophy, "whose participants use the methods of experimental psychology to probe the way people make judgments that bear on debates in philosophy" (Nadelhoffer and Nahmias 2007, 123). In fact, Weinberg and Crowley (2009, 227) categorize the possible relations between science and philosophy that make experimental philosophy possible into a *replacement thesis*, according to which science and philosophy offer competing explanations, and a *continuity thesis*, according to which "science and philosophy constitute generally overlapping areas of inquiry". In the first case, a proponent of experimental philosophy would assume replacement naturalism, and in the second case cooperative naturalism.[12]

In an overview of experimental philosophy, Nadelhoffer and Nahmias (2007, §2) identify its three strains of research: experimental restrictionism, experimental analysis, and experimental descriptivism, which can roughly be described as determining *that* people's intuitions differ, *how* they differ, and *why* they differ in the way they do. In the context of replacement naturalism, *experimental descriptivism* (Nadelhoffer and Nahmias 2007, 127, footnote removed) amounts to Quine's specific kind of naturalized philosophy, and it is

> important to try to determine how [...] intuitions are generated. [They] explore human psychology by testing how various manipulations to scenarios influence the intuitions people express. One goal of this project is to better understand the nature of the underlying psychological processes and cognitive mechanisms that produce our intuitions and explore the relevance of this research to philosophical questions.

The goal of these investigations is the evaluation of philosophical theories about these concepts. The main goal of *experimental restrictionists* (Nadelhoffer and Nahmias 2007, 128, footnote removed), on the other hand,

> is to show that some of the methods and techniques that philosophers working in the analytic tradition have taken for granted are threatened by the gathering empirical evidence concerning both the diversity and the unreliability of folk intuitions. More specifically, they argue that if our intuitions about a particular topic vary cross-culturally or socio-economically and we don't have independent grounds for privileging our own intuitions to those of others, these particular intuitions will be insufficient for philosophical theory building.

Some intuitions may turn out to be cross-culturally invariant, however, and these provide the basis for *experimental analysis*, whose

> primary goal is to explore in a controlled and systematic manner what intuitions ordinary people tend to express and examine their relevance to philosophical debates. Hence, [experimental analysts] aim to test philosophers' claims that their positions align with common sense and to challenge those claims that are not supported by

[12] This commonality in categorization is unsurprising given that Weinberg and Crowley (2009, 227) follow the categorization of Kornblith (1994) for naturalized epistemology.

> the evidence. On this view, philosophical theories that most closely accord with and account for ordinary beliefs and practices should enjoy "squatters' rights" until they are shown to be defective for other reasons. In this respect, [experimental analysts] essentially agree with many traditional philosophers [. . .].

Nadelhoffer and Nahmias (2007, 126, footnote removed) here include the proponents of both traditional and of ordinary language philosophy in the class of "traditional philosophers". These traditional philosophers thus differ from experimental analysts only in that they rely on their own hunches to determine which trans-empirical intuitions are widely shared.

Taking into account Feigl's distinction between an intuition and its target, it is, strictly speaking, not the trans-empirical intuition that features in a traditional philosophical argument, but a description of its target or, following the terminology by Fedyk (2009, §2), the propositional content of the intuition. When intuitions are considered as psychological phenomena, it is of interest how their occurrence can be established (Feigl 1958, 8–11). If the intuitions under examination belong to a specific class of people (e. g., some group of philosophers, ordinary people, or scientists), then the best method of determining the content of those intuitions seems to be statistical. The relevant intuitions may also be those that a specific class of people *would* have, if presented with some class of facts, considerations, or examples.[13] Then the best and probably only method of determining the intuitions' content is empirical psychology, which could establish that people in fact usually develop these intuitions. Some parts of traditional philosophy therefore may have to be naturalized.

Ordinary language philosophers who rely on their own intuitions to determine the language use or linguistic intuitions of others also rely on hunches in Feigl's sense. Of course, whether these hunches are accurate is itself an empirical question, and eventually, the truth of their propositional content has to be established empirically (Feigl 1958, 6–7; Mates 1958, 165; Nadelhoffer and Nahmias 2007, 129; Sytsma 2010, §1). This also holds for intuitions about the language use or linguistic intuitions that people *would* have if presented with some class of facts, considerations, or examples.[14] Some parts of ordinary language philosophy therefore have to be naturalized (Mates 1958).[15] Experimental analysis does not only have to be applied to intuitions about concepts in ordinary language, of course. Stotz (2009), for example, provides a short overview of research on the concepts used in biology.

Experimental philosophy does not have to be pursued solely in the context of

[13]This may be what Williamson (2007, 191, 216) has in mind when he speaks of philosophical judgments that require "philosophical training" leading to specific "skills".

[14]This may be what Kauppinen (2007, §5) has in mind when he claims that "(philosophical) dialogue and reflection" lead to a convergence of linguistic intuitions.

[15]Arguably, this naturalization leads to experimental philosophy (see §2.6 and Sytsma 2010, §§1–3).

replacement naturalism. As Nadelhoffer and Nahmias (2007, 126–127) put it:

> In addition to reporting the results of their studies, experimental philosophers also explore background issues such as the nature and sources of intuitions, the role that they should play in philosophy, how best to explore them, and what responses are available to theorists whose views do not settle with folk intuitions. [Some] work in experimental philosophy [. . .] has included and inspired numerous arguments about how and why the data are significant to [the respective philosophical debates] and how best to interpret the data in light of various philosophical theories.

These discussions are not decided by experiments, and experimental philosophy pursued in this vein thus assumes cooperative naturalism.

2.6 The new debate about intuitions and experiments

I am going to argue in the following that some results of experimental philosophy, if correct, support linguistic philosophy, some support artificial over ordinary language philosophy, and some provide a friendly starting point for artificial language philosophy. In the old debate about and within linguistic philosophy, experimental philosophy thus strengthens the position of artificial language philosophy. As suggested by the analogy between technical terms in philosophy, the sciences, and jurisprudence (page 12), I will then argue that artificial language philosophy can be pursued almost completely independently of experimental research on intuitions. As a historical aside, I will also show how many of the arguments in the old debate apply to the current debate about the role of intuitions and experimental results in philosophy. Specifically, I claim that the arguments by artificial language philosophers support experimental philosophy.

Since the discussion by Feigl (1958), the use of intuition has had a remarkable renaissance in philosophy. Hintikka (1999) spells out how Chomsky was perceived to have based his approach to linguistics on the intuitions of the linguist and how his success contributed to an increased use of intuition in philosophy. Symons (2008) describes how G. E. Moore's conception of common sense, embraced by ordinary language philosophers, became a tool in Kripke's trans-empirical philosophy, and hence far removed from the original idea of ordinary language as a restriction on trans-empirical claims. This jumbled heritage of contemporary applications of intuition has led to two distinct forms of use. Sometimes, an intuition is considered to be a judgment of common sense, and sometimes, intuition has an evidential role analogous to that of perception because of its immediacy (cf. Feigl 1958).

Although the historical connections between trans-empirical and ordinary language philosophy and the contemporary uses of intuition are fascinating, they will not be my main focus in this section. Rather, I aim to show that conceptually, experimental and artificial language philosophy are complementary in two very distinct ways. First, the assumptions of each approach are supported by the other: Artificial language philosophers, in their critiques of trans-empirical philosophy and of non-empirical approaches to ordinary language, provide arguments in favor of experimental philosophy. The results of experimental philosophy support the empirical premises of the artificial language philosophers' critiques. Second, the two approaches are independent in their application in the sense that the results of one approach cannot prove or disprove the results of the other.

Experimental restrictionism provides a confirmation of Mates's hypothesis that the disagreement over ordinary language use among Oxford professors of philosophy is only one case of a wider disagreement in the general population. When intuitions lead to differing moral judgments, experimental restrictionism is very much in line with Hare's contention that there is no reason to assume our intuitions will agree in difficult moral situations. Therefore, experimental restrictionism is a problem for the common sense conception of intuition, and this holds whether intuitions disagree on linguistic (and therefore empirical) matters, or on non-linguistic matters (be they empirical or trans-empirical). It also presents a prima facie problem with the use of intuition in analogy to perception, because when there is disagreement between intuitions, at least some of these intuitions must be wrong, and sometimes they might all be. Analogously to perception, then, the more that experimental philosophy restricts the domain of agreement between intuitions, the less useful intuitions are as evidence for their targets.

Experimental analysis adds to experimental restrictionism, because its results suggest that intuitions depend on social status and cultural background. These dependencies are a concretization of Hare's rather general suggestion that our intuitions are a result of our upbringing.

Liao (2008) has pointed out that the results of experimental analysis have shown *some* intuitions to be robust. Recalling Maxwell and Feigl's (1961) objection to ordinary language philosophy, one can see that for Liao's point to be a defense of the methods of ordinary language philosophy, he must further show that the robust intuitions are also philosophically relevant, do not embody factually false assumptions, and do not involve concepts that should be reformed for other reasons. However, even then, this would not defend the wide applicability of common sense intuitions as evidence for trans-empirical claims, that is, it would not undermine linguistic philosophy itself. For, trans-empirical intuitions cannot simply be *assumed* to be evidence for their targets, and whether they *are* evidence cannot be tested independently. As discussed earlier, shared intuitions might justify the belief in their targets by an inference to the best explanation, but as Feigl notes, such an argument would be weakened by the existence of other explanations for shared intuitions. Experimental descriptivism aims at providing such an alternative explanation: Since experimental descriptivism relies only on empirical claims, it would, if successful, provide an explanation of shared intuitions that does not rely on the truth of their trans-empirical targets.

Experimental descriptivism hence may eventually come to support linguistic over non-linguistic philosophy by explaining shared trans-empirical intuitions. Experimental analysis and experimental restrictionism support those arguments against trans-empirical and ordinary language philosophy that are based on the systematic disagreement of people's intuitions.

Of course, trans-empirical and ordinary language philosophy have both been defended against criticisms from experimental philosophy. In defense of trans-empirical philosophy, Sosa (2007, 101) proposes an account of intuitions that gives them an evidential status analogous to perception (though he disavows the perceptual model of intuition), applicable to any kind of statement. He bases his analogy between intuition and perception on competence: On his proposal, "to intuit that p is to be attracted to assent simply through entertaining that representational content. The intuition is *rational* if and only if it derives from competence, and the content is explicitly modal". There is "no very deep reason [for the restriction to modal propositions]. It's just that it seems the proper domain for philosophical uses of intuition". An intuition is thus the (possibly irrational) inclination to agree with a proposition. The rationality of an intuition, that is, the justification for believing its target to be true, stems from competence. Referring to Sosa's conception of intuition, Symons (2008, 87–88) argues that competence can be established by empirical research: "[T]he lasting significance of experimental philosophy is not that it undermines appeals to consensus, but that it opens a fertile field of inquiry into our commonsense or intuitive capacities". Specifically, "determining the boundaries of our competence is the most fruitful task that lies ahead for experimental philosophy".

The important question then is the source of the competence claim, and here Feigl's distinction between trans-empirical intuitions and hunches becomes important. For hunches, competence can be established by empirically testing the statement for which the hunch is supposed to be evidence. Claiming competence then amounts to claiming a correlation between the occurrence of a hunch and the truth of its target. It is this correlation that can be the object of empirical study, and thus of experimental philosophy. Experiments on hunches have demonstrated systematic mistakes among children, laypersons, and experts, as Nadelhoffer and Nahmias (2007, 125, 129) point out, and optical illusions cause systematic mistakes in the case of perception, so competence cannot simply be assumed. Therefore, even *if* there is a successful analogical argument from competence in the case of hunches or perceptions to competence in the case of trans-empirical intuitions, the latter suffer, according to the analogy, from systematic mistakes as well.

Since trans-empirical intuitions cannot be tested like hunches or perceptions, competence claims are outside the realm of experimental philosophy and empirical research in general, and have to be established in some other way. This needs to be done in order to show that trans-empirical intuitions are indeed evidence for their targets in the same way perceptions are, and that therefore linguistic philosophy is mistaken. If competence claims cannot be established, there may also be no other justification to prefer one person's intuitions over another's. Without such a justification, empirical restrictionism's results cannot be rendered irrelevant by considering only the preferred intuitions. Further justifications pending, the results of experimental philosophy therefore support linguistic philosophy and pose a problem for trans-empirical philosophy.

In a critique of experimental philosophy, Kauppinen (2007) argues that ordinary language philosophy can only be pursued by what Mates calls the intensional method, and claims that experimental philosophy is in principle restricted to the extensional method. The latter claim is decisively criticized by Nadelhoffer and Nahmias (2007), who raise the possibility of devising the Socratic questionnaires that Mates suggests for the intensional method. Kauppinen's criticism hence loses its force against experimental philosophy. It is still noteworthy, though, that with this claim Kauppinen moves away from the historical practice of overly relying on the extensional method, which was lamented by Mates, to the other extreme of excluding it completely.

Kauppinen further argues that the intensional method can be expected to yield converging results because people can communicate. However, as Mates has pointed out, this argument does not establish that people use words with the same meaning they would settle on via the intensional method. Whether there is such a convergence of meaning is very much an empirical question, to be tested, for example, by the method Mates proposes. And even then, that agreement exists in *some* cases does not imply agreement in the difficult ones, as Maxwell and Feigl have noted. In their reply to Kauppinen, Nadelhoffer and Nahmias (2007, 144, n. 36) state as much.

Concluding his critique, Kauppinen (2007, 110) claims that "assessing the truth of intuition claims can remain a relatively armchair business [. . .]. We are entitled to have confidence in such reflection, since we take a lot of real-life experience of using concepts to the armchair with us". That is, the intensional method can be replaced by recourse to the investigator's intuitions about the use of her native language. The argument is rejected by Nadelhoffer and Nahmias (2007, 129) once again with Mates's point that Kauppinen's claim is an empirical one that must be tested.

Without having established the possibility for the ordinary language philosopher to rely on her intuitions alone, Kauppinen is forced to accept Mates's intensional method as the only viable one. Therefore, Mates's worry about this method becomes acute: It is not clear how to distinguish between finding out and influencing what someone takes a word's meaning to be. Kauppinen (2007, §5.1) himself notes that one is "never free of the danger of leading the witness in the direction favored by the questioner", but does not suggest a way to avoid this influence. As detailed above, Maxwell and Feigl turn this into an argument for artificial language philosophy by suggesting that there is no such way, while Mates at least thinks that one can test this empirically.

In their argument for artificial language philosophy, Maxwell and Feigl go beyond simply arguing that in the intensional method, ordinary language philosophy cannot help but reform language. They hold that it *should* reform language, because ordinary language may not be good enough to solve or dissolve philosophical problems. Philosophical language must contain explicata, not explicanda. It is this reliance on explicata, and constructed languages more generally, that ensures the independence of artificial language philosophy from folk intuition, just as the analogical consideration mentioned in the beginning suggests. Underlying the analogy is a general statement that holds for natural sciences, mathematics, and jurisprudence, as well as for artificial language philosophy: They apply constructed languages, and many of their concepts are explicata for the explicanda of ordinary language. Even if a word occurs in both constructed and ordinary language, it will therefore typically have different meanings in each. For this reason, Carnap and Popper can dismiss Naess's experiments on ordinary language as irrelevant to Tarski's explication of 'true', and artificial language philosophy can be pursued largely independently of the results of contemporary experimental philosophy.

This dismissal of folk intuitions does not simply shift the authority to the experts' intuitions, though. Their intuitions about the application of an explicatum can be checked by using the rules for an explicatum's use, which must be laid down precisely. This was already remarked upon very early by Carnap (1967a, §100) in a discussion of the rational reconstruction of concepts in philosophy and the sciences:

> The fact that the synthesis of cognition, namely, the object formation and the recognition of, or classification into, species, takes place intuitively, has the advantage of ease, speed, and obviousness. But intuitive recognition (e. g., of a plant) can become useful for further scientific work only because it is possible to give, in addition, the indicators (of the particular species of plant), to compare them with the perception and thus to give a rational justification of intuition.

Experts' intuitions about how an explicandum should be explicated can be checked against Carnap's requirements for explication, and in principle, anyone may suggest and use a new explicatum according to expedience, as long as this new concept is clearly distinguished from existing ones.

Experimental philosophy is not useless for artificial language philosophy, however. To the extent that experimental restrictionism establishes actual disagreement in the application of concepts, it identifies areas where an explication is clearly needed, and experimental analysis can help by identifying an explicandum as a starting point for such an explication. In general, experiments on ordinary language can bear all the fruits for artificial language philosophers that Carnap (1955) lists in the quotation in section 2.4. Experimental philosophy is relevant for ordinary language philosophy, and to the extent that ordinary language philosophy is relevant for artificial language philosophy, so is experimental philosophy.

Artificial language philosophy can be challenged, of course. Kauppinen (2007, 96) and Nadelhoffer and Nahmias (2007, 130) cut to the core of the debate between ordinary and artificial language philosophy when they echo the claim by Strawson (1963) that artificial language philosophy does not solve the right problems if it does not capture the concepts of ordinary language. However, first, the preceding discussion shows that even ordinary language philosophy may not capture ordinary language, and second, the responses by Bergmann (1949) and Maxwell and Feigl (1961) to Strawson's argument in the old debate within linguistic philosophy show that this may not be a problem in the first place. At this point, I do not want to claim that Kauppinen, Nadelhoffer, Nahmias, and Strawson are wrong. But I do want to claim that Nadelhoffer, Nahmias, and Kauppinen are jumping to conclusions, given that the discussion between ordinary and artificial language philosophy starts, but does not end, with Strawson's criticism. In the next section, I will claim that they are wrong.

2.7 Philosophical methodologies according to artificial language philosophy

Like artificial language philosophy, naturalized and experimental philosophy have also been charged with changing the subject—that they do not, in fact, address philosophical problems at all (cf. Kim 1988). That naturalized methodologies cannot by themselves solve philosophical problems is also suggested by the possibility of using them as a proper part of traditional and ordinary language philosophy. I now want to argue that artificial language philosophy can be defended against this charge.

A straightforward defense of artificial language philosophy would require a precise definition of 'philosophical problem' and 'solution to a philosophical problem'. Since neither term has, so far, been defined to general satisfaction, I only aim to show that artificial language philosophy succeeds in capturing many philosophers' posited solutions to philosophical problems, that is, it captures much of philosophical practice. In particular, I will argue that artificial language philosophy can capture many of the applications of the five philosophical methodologies described.

Artificial language philosophy trivially captures its own applications. And as the old discussion within linguistic philosophy (§2.4) has shown, there is a general consensus that the results of ordinary language philosophy can be a starting point for the choice of a language, and there are good reasons to assume that exclusive reliance on actual language use would make for an inordinately weak philosophical methodology. Especially the arguments by Maxwell and Feigl (1961) suggest that in many cases in which ordinary language philosophy leads to clarifications, it does so by regimenting the language. Because of this normative part, ordinary language philosophy therefore threatens to collapse into artificial language philosophy.

The practices of traditional philosophy can be captured in artificial language philosophy by interpreting alleged discoveries of facts as inventions of new concepts or whole new languages. In the terminology of Carnap (1934b, 13–17, 19), this means switching from the "material" or "connotative mode of expression" to the "formal mode of expression". When intuitions are used as evidence, they are interpreted as meaning-directed rather than world-directed, and their successive development in the method of reflective equilibrium is interpreted as a method of explication (cf. Kuipers 2007, xiv). However, while traditional philosophy faces the challenge of justifying its claims as *discoveries*, and thus of explaining how philosophers gain cognitive access to those facts that are the subjects of these claims (e. g., the targets of the philosophers' intuitions), artificial language philosophy can simply justify them as pragmatic language choices.

Indeed, the descriptions of the methods of traditional philosophy by its practitioners sometimes already read like descriptions of artificial language philosophy. Reviewing a critique of traditional metaphysics by Ladyman and Ross (2007), Dorr (2010) describes a tentative consensus among metaphysicians about methodology:

> It is not enough simply to announce that Xs are more fundamental than Ys: if I want to defend this claim, I am supposed, at a minimum, to (i) introduce a language in which I can talk about Xs without even seeming to talk about Ys; and (ii) make some kind of adequacy claim about this language, e. g., that it can express all the genuine facts that we can express using Y-talk, or that all the Y-facts supervene on the facts stateable in the language. For example, if I want to maintain that spacetime is less fundamental than the spatiotemporal relations between bodies, I must describe a language for characterizing these relations, and explain how it can adequately capture, e. g., claims about the global topological structure of spacetime.

Furthermore, Dorr states that one "earn[s] the right" to consider a philosophical problem "dissolved [. . .] by describing a fundamental language within which no corresponding questions can be formulated". If the "genuine facts that we can express using Y-talk" are taken to determine the contexts in which the X-language should be applicable, Dorr essentially describes the conditions of adequacy on an artificial language as discussed in §2.3.1. Dorr (2010) also emphasizes the importance of language choice:

> The whole approach [by Ladyman and Ross (2007)] reflects an exag-

> gerated sense of the importance of argument in metaphysics, and a corresponding underestimation of the difficulty of merely crafting a view coherent and explicit enough for arguments to get any grip.

From the perspective of artificial language philosophy, this crafting of a "coherent and explicit" view is nothing but the search for a language in which philosophical problems can be solved by rigorous derivation.

The practices of naturalized philosophy are hard to circumscribe because naturalized philosophy itself is hard to circumscribe, but to the extent that it complements traditional and ordinary language philosophy, that is, in the context of replacement naturalism, artificial language philosophy can capture its practices as well. And to the extent that naturalized philosophy *relies* on empirical results rather than establishes them, it is engaged in language choice and rigorous derivation. When it thereby addresses philosophical problems, it is pursued in the context of cooperative naturalism, and naturalized philosophy is then artificial language philosophy. For the same reason the non-experimental part of experimental philosophy, which assumes cooperative naturalism, is artificial language philosophy. The one aspect of naturalized philosophy that artificial language philosophy cannot accommodate is empirical research into a non-linguistic phenomenon. For instance, a philosopher who determines the angles of a triangle of light rays over great distances does not describe an explicandum or engage in language choice or rigorous derivation. Such research, however, is often charged with not being philosophy at all.

As an illustration of the reinterpretation of philosophical practice in artificial language philosophy, consider Sosa's response (Sosa 2007, 104) to the claim by Nichols and Knobe (2007) that the usage of 'responsible' in ordinary language is inconsistent due to a performance error:

> [T]here is an alternative explanation that will cast no affect-involving doubt on the intuitions in play. This other possibility came to mind on reading their paper, and was soon confirmed in the article on moral responsibility in the Stanford Encyclopedia of Philosophy, where we are told that at least two different senses of 'moral responsibility' have emerged: the attributability sense, and the accountability sense.[. . .]
>
> So, here again, quite possibly the striking divergence reported above is explicable mainly if not entirely through verbal divergence.

For Sosa (2007, 100), the "use of intuitions in philosophy should not be tied exclusively to conceptual analysis. [. . .] Some such questions concern an ethical or epistemic subject matter, and not just our corresponding concepts". This is presumably how he interprets the "emergence" of two kinds of responsibility: They both exist, but are referred to with the same word, leading to "verbal divergence".

Sosa probably refers to the fall 2004 edition of the *Stanford Encyclopedia of*

Philosophy, in which Eshleman (2004, §2.2) writes that

> at least some disagreements about the most plausible overall theory of responsibility might be based on a failure to distinguish between different aspects of the concept of responsibility, or perhaps several distinguishable but related concepts of responsibility.
>
> Broadly speaking, a distinction has been drawn between responsibility understood as attributability and responsibility as accountability.

Eshleman's formulation differs from Sosa's paraphrase in that Eshleman considers the disagreements to stem from confusion over "distinguishable but related *concepts* of responsibility". In other words, there are pre-existing concepts (not pre-existing kinds) that get confused, and an analysis of the concepts of ordinary language would resolve the inconsistency. Regarding such a line of reasoning, Maxwell and Feigl (1961, 489) note that there is little reason to think that the two concepts allegedly being confused are somehow already present in ordinary thinking. Certainly, the ordinary user of the term 'responsibility' is not aware of them—otherwise there would be no confusion. And if the ordinary user were to agree with the distinction between the two concepts, Maxwell and Feigl argue, this agreement would amount to a change of language.

According to artificial language philosophy, then, the introduction of the distinction between responsibility as attributability and responsibility as accountability into the philosophical discourse is a conventional change of language—it is not the discovery of a fact about the world or the meaning of the term 'responsibility'.

* * *

I have argued above that in the context of cooperative naturalism, the non-empirical part of experimental philosophy is artificial language philosophy. But the 'is' here refers to the membership relation, not identity. Artificial language philosophy encompasses more than the non-empirical part of experimental philosophy because, as described in §2.3.1, it is empirically informed by experimental results in general, not experimental results about people's intuitions. In direct contradiction to this position, the difference between experimental philosophy and empirically informed philosophy is sometimes only or mainly attributed to the conduction of experiments by the philosophers themselves. In a critique of experimental philosophy, Sosa (2007, 100) puts it this way:

> Mining the sciences is not in itself novel, of course. [. . .] Just think of how 20$^{\text{th}}$-century physics bears on the philosophy of space and time, or split-brain phenomena on issues of personal identity, to take just two examples. Perhaps the novelty is rather that experimental

> philosophers do not so much borrow from the scientists as that they become scientists. This they do by designing and running experiments aimed to throw light on philosophically interesting issues.

In a defense of experimental philosophy, Nadelhoffer and Nahmias (2007, 124) come to the same result:

> [W]hat distinguishes experimental philosophers from not only similarly-minded naturalistic and empirically informed philosophers but also from experimental psychologists? Though the boundary here is blurry, the primary difference is that experimental philosophers actually run their own studies to get at the data they need and then show why these data are philosophically interesting.

If Sosa, Nadelhoffer, and Nahmias were right, every article in empirically informed philosophy (that is, artificial language philosophy) could be transformed into an article in experimental philosophy if only the philosophers themselves conducted the experiments that form the basis of their analyses. Considering that, say, philosophy of physics relies on results from, for example, particle accelerators, this consideration throws the difference between experimental and artificial language philosophy into sharp relief. Experimental philosophy (at least as currently practiced) restricts its experiments to people's language use, language intuitions, and trans-empirical intuitions. Artificial language philosophy, on the other hand, can rely on *any kind* of experimental results to suggest new concepts, and assigns no special value to people's actual usage and intuitions.

There is thus a difference between cooperatively understood experimental philosophy and artificial language philosophy that goes beyond who gathers the data. It consists in what kind of data are gathered and what they are used for. Experimental philosophy takes as data the actual use of concepts and contributes empirical generalizations about them; this holds even if experimental philosophers discuss the relations of philosophical theories about some concept on the one hand and people's intuitions about that concept on the other. Artificial language philosophy in general takes any data whatsoever, and contributes concepts that structure them well. For instance, when investigating the *wave function* or *gene* concepts, relevant data are experimental results for electrons or DNA, not scientists' or laymen's use of the words 'wave function' or 'gene'.

2.8 Formal semantics for concept formation

So far, I have developed and defended artificial language philosophy without relying on a specific semantics. The arguments in the rest of this chapter and much of the rest of this book rely on the formal semantics presented next. The semantics assumes a first or higher order predicate logic, which will be convenient; for one, the logic and its standard semantics are fairly simple and well understood. It furthermore relates in a straightforward way to classical mathematics. Finally, all of the criteria of empirical significance that I will discuss assume classical predicate logic, which to a certain extent forces my hand—any other logic would have to contain predicate logic at least as a special case. As already pointed out in §2.3.1, when the logic of a language is fixed, language choice becomes concept formation. This will keep things comparably simple.

2.8.1 Basic terms

Carnap (1939, §24) provides a general outline of the semantics for a language bipartitioned into basic and auxiliary terms that I will discuss in chapter 3. Przełęcki (1969, ch. 5–6) gives what can be seen as an elaboration of Carnap's account in formal semantics, and much of what follows will rely on a slight generalization of Przełęcki's account. Because the basic terms are unproblematic, Przełęcki can assume that their meaning determines a set $\mathbf{M}_{\mathscr{B}}$ of *possible $\mathscr{B}$-structures*. [16] On pain of triviality, $\mathbf{M}_{\mathscr{B}}$ cannot contain all $\mathscr{B}$-structures (Przełęcki 1969, ch. 4), and thus may lead to a set of meaning postulates $\Pi_{\mathscr{B}}$ for the basic terms, where $\Pi_{\mathscr{B}}$ is the set of $\mathscr{B}$-sentences that are true in every $\mathfrak{M}_{\mathscr{B}} \in \mathbf{M}_{\mathscr{B}}$.[17] Here and in the following, I will assume that the closure of $\mathbf{M}_{\mathscr{B}}$ under isomorphism is the class of all models of $\Pi_{\mathscr{B}}$; the motivation for this assumption is that the role of $\mathbf{M}_{\mathscr{B}}$ consists only in distinguishing between isomorphic structures. If two non-isomorphic structures have to be distinguished, this is either achieved on the level of the language or not at all. Following the terminology of Fine (1975, §3), one can give

Definition 2.1. A set Σ of sentences is *supertrue/superfalse* in a set $\mathbf{A}$ of structures if and only if Σ is true/false in every $\mathfrak{A} \in \mathbf{A}$.

A set of sentences is *true* in a single structure $\mathfrak{A}$ if and only if all of its elements are true in $\mathfrak{A}$ according to the standard definition (Hodges 1993, 12–13; Leivant 1994, §3.1).[18] A set of sentences is *false* in $\mathfrak{A}$ if and only if it is not true in $\mathfrak{A}$. Here and in the following, a definition for sets of sentences holds for a single sentence if and only if it holds for the sentence's singleton set.

Przełęcki (1969, 20–21) suggests that supertruth should be a sufficient condition for truth in $\mathbf{A}$, and superfalsity a sufficient condition for falsity in $\mathbf{A}$. I will also assume the converse, so that *truth in* $\mathbf{A}$ is supertruth in $\mathbf{A}$, and *falsity in* $\mathbf{A}$ is superfalsity in $\mathbf{A}$. Thus $\Pi_{\mathscr{B}}$ is the set of sentences that are true in $\mathbf{M}_{\mathscr{B}}$. If the

[16] For mnemonic purposes, '$\mathbf{M}_{\mathscr{B}}$' could stand for 'meaning of the $\mathscr{B}$-terms'.

[17] Carnap (1952) describes how to treat meaning postulates for basic terms on a syntactic level, Przełęcki (1969, §10.II) gives a method for introducing meaning postulates for basic terms, and Kyburg (1990) discusses a method for choosing between different sets of meaning postulates for basic terms in probabilistic theories.

[18] For higher order logic, I will usually assume Tarski semantics, not Henkin semantics (Leivant 1994, §5.4). I will discuss the implications of this assumption in §4.1.1 and §4.1.3.

truth value of a $\mathscr{B}$-sentence β differs for different possible structures in $\mathbf{M}_{\mathscr{B}}$, its truth value can only be determined by restricting $\mathbf{M}_{\mathscr{B}}$ through empirical research to a proper subset $\mathbf{N}_{\mathscr{B}} \subset \mathbf{M}_{\mathscr{B}}$ of *intended $\mathscr{B}$-structures*.[19]

Przełęcki (1969, 42) assumes that in general $|\mathfrak{M}_{\mathscr{B}}| = |\mathfrak{M}'_{\mathscr{B}}|$ for all $\mathfrak{M}_{\mathscr{B}}, \mathfrak{M}'_{\mathscr{B}} \in \mathbf{M}_{\mathscr{B}}$, but I will not make this assumption. For one, this restriction will usually not be necessary for my purposes, and so there is no need to curtail the expressiveness of the formalism. Furthermore, it requires some contortions in the application of the formalism to typical situations in the sciences and even in everyday life. A theory about chickens, for example, may be phrased so that it applies only to individual flocks, not all of them at once. One could, of course, quantify over all flocks, but this does not seem necessary. Furthermore, one may want to apply an already interpreted theory to a new domain (flocks of geese, say), and it would be convenient if this did not require the modification of the semantics for any of the previous applications. Finally, assuming different domains seems to allow one to formalize the actual applications of theories much more directly. For a theory will typically be applied to only one small part of the world (say, fifteen flocks of chickens), even though it is meant to hold for a much bigger part (all flocks of chickens). All of these considerations are only considerations of convenience, however. It does not seem to be impossible in principle to formalize the applications of theories by way of a single domain (cf. Przełęcki 1974c, 103).

2.8.2 Auxiliary terms

The meaning postulates Π for the whole language $\mathscr{V}$ contain the meaning postulates $\Pi_{\mathscr{B}}$ for the $\mathscr{B}$-terms. To arrive at a formal definition, let $\mathfrak{A}|_{\mathscr{B}}$ refer to the *reduct* of $\mathfrak{A}$ to $\mathscr{B}$, that is, the structure that results from eliminating the interpretations of all $\mathscr{A}$-terms from $\mathfrak{A}$. For a $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$, a structure $\mathfrak{B}$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$ is called an *expansion* of $\mathfrak{A}_{\mathscr{B}}$ (Hodges 1993, 9).

For it to be possible to *choose* the analytic sentences Π to be true, it must be possible to expand every structure of $\mathbf{M}_{\mathscr{B}}$ to a model of Π. Otherwise, empirical research could show that there is no model of Π that respects the meaning of the $\mathscr{B}$-terms; that is, Π could be empirically false. The expansions of $\mathbf{M}_{\mathscr{B}}$ to models of Π then are the *possible $\mathscr{V}$-structures* $\mathbf{M}$. Structures of subsets for $\mathscr{V}$ are possible only if they are reducts of a possible $\mathscr{V}$-structure:

Definition 2.2. A *$\mathscr{V}$-structure* $\mathfrak{A}$ is *possible* ($\mathfrak{A} \in \mathbf{M}$) if and only if $\mathfrak{A} \models \Pi$ and $\mathfrak{A}|_{\mathscr{B}} \in \mathbf{M}_{\mathscr{B}}$. A *structure* is *possible* if and only if it can be expanded to some $\mathfrak{M} \in \mathbf{M}$.

Now the following holds:

[19] $\mathbf{N}_{\mathscr{B}}$ is called $\mathbf{M}^*_{\mathrm{O}}$ by Przełęcki (1969, 42), and $\mathbf{M}_{\mathscr{B}}$ comes closest to what Przełęcki (1969, 43) calls 'the characterization of $\mathbf{M}^*_{\mathrm{O}}$'. I thank Antje Rumberg and Tom Sterkenburg for helpful discussions on this point. As a mnemonic, one might think of $\mathbf{M}_{\mathscr{B}}$ as requiring one (empirical/alphabetical) step to $\mathbf{N}_{\mathscr{B}}$.

Claim 2.1. *A structure can be expanded to a model of Π if and only if it is isomorphic to a possible structure.*

Proof. '$\Leftarrow$': Trivial

'$\Rightarrow$': Assume that $\mathfrak{A}$ can be expanded to a model $\mathfrak{B}$ of Π. The reduct $\mathfrak{B}|_{\mathscr{B}}$ is a model of $\Pi_{\mathscr{B}}$, and thus, by assumption, there is an isomorphism f from $\mathfrak{B}|_{\mathscr{B}}$ to some $\mathfrak{C}_{\mathscr{B}} \in \mathbf{M}_{\mathscr{B}}$. Then f is also an isomorphism from $\mathfrak{A}$ to a structure that can be expanded to the model $\mathfrak{C}$ of Π with $\mathfrak{C}|_{\mathscr{B}} = \mathfrak{B}_{\mathscr{B}} \in \mathbf{M}_{\mathscr{B}}$. □

Therefore $\mathbf{M}$, just as $\mathbf{M}_{\mathscr{B}}$, only distinguishes between isomorphic structures, and the closure of the class of possible structures under isomorphism is the class of all structures that can be expanded to a model of Π. For sentences, one can give

Definition 2.3. A *set Σ of $\mathscr{V}$-sentences* is *possible* if and only if Σ is true in some possible structure. Σ is *analytically true/analytically false* if and only if Σ is true/false in $\mathbf{M}$. Σ is *analytically determined* if and only if it is true in $\mathbf{M}$ or false in $\mathbf{M}$. Σ is *analytically contingent* if and only if it is not analytically determined. Σ *analytically entails Λ* if and only if Λ is true in all possible structures in which Σ is true.

I will sometimes call an analytically true set of sentences simply 'analytic'. Definition 2.3 leads to

Claim 2.2. *A set Σ of $\mathscr{V}$-sentences is possible if and only if $\Sigma \cup \Pi$ has a model. Σ is analytic if and only if $\Pi \vDash \Sigma$. All and only sets of sentences incompatible with Π are analytically false; all and only analytically true sets and analytically false sets are analytically determined. Σ analytically entails Λ if and only if $\Sigma \cup \Pi \vDash \Lambda$.*

Proof. From claim 2.1 because the truth/falsity of a sentence is invariant under isomorphisms between structures. □

Π is thus sufficient to determine which $\mathscr{V}$-sentences can, cannot, or must be true within the chosen language and for questions of analyticity and analytic entailment, so that $\mathbf{M}_{\mathscr{B}}$ and $\mathbf{M}$ will not play a major role in the formal analyses in the following. Put somewhat counterintuitively, analyticity is a syntactic, not a semantic, notion.

In general, a set of postulates for auxiliary terms may be such that not every possible $\mathscr{B}$-structure can be expanded to a model of the postulates. The postulates thus restrict $\mathbf{M}_{\mathscr{B}}$, and hence are not completely analytic. The demarcation criteria for science are a good example of this in philosophy. Such postulates for the term 'science', S, are sometimes given by a set Θ of one sufficient and one necessary condition, $\Theta \mathrel{\text{⊨⊨}} \{\forall x[\varphi(x) \to Sx], \forall x[Sx \to \psi(x)]\}$, which entails the $\mathscr{B}$-sentence $\forall x[\varphi(x) \to \psi(x)]$ (see §9.6). An example from physics is Ohm's law, which contains an analytic component—the meaning postulates for the auxiliary terms 'voltage' and 'internal resistance'—and an empirical component that establishes a relation between the basic terms ‘resistance’ and ‘current intensity’ (Simon 1970, §§2–3). For $\Pi_{\mathscr{B}} = \varnothing$ and a single postulate $\sigma(\bar{B},\bar{A})$ that contains the basic terms $\bar{B}$ and auxiliary terms $\bar{A}$,[20] Carnap (1958, §4) proposes identifying its empirical component with its *Ramsey sentence*

$$\mathsf{R}_{\mathscr{B}}(\sigma) := \exists \bar{X}\, \sigma\left(\bar{B},\bar{X}\right) , \tag{2.1}$$

which results from σ by existentially generalizing on all $\mathscr{A}$-terms in σ. Ramsey sentences have some easily established properties:

Claim 2.3. *For a $\mathscr{B}$-sentence β, a $\mathscr{V}$-sentence σ, and a set $\mathscr{J} \subseteq \mathscr{A}$ of terms, the following holds:*

$$\mathsf{R}_{\mathscr{B}}(\beta) \mathrel{\text{⟚}} \beta \tag{2.2a}$$

$$\mathsf{R}_{\mathscr{B}}\left(\mathsf{R}_{\mathscr{B}}(\sigma)\right) \mathrel{\text{⟚}} \mathsf{R}_{\mathscr{B}}(\sigma) \tag{2.2b}$$

$$\mathsf{R}_{\mathscr{B}}\left(\mathsf{R}_{\mathscr{B}\cup\mathscr{J}}(\sigma)\right) \mathrel{\text{⟚}} \mathsf{R}_{\mathscr{B}}(\sigma) \tag{2.2c}$$

Proof. Formula (2.2a) holds because β contains no $\mathscr{A}$-terms. Formula (2.2b) follows because $\mathsf{R}_{\mathscr{B}}(\sigma)$ is a $\mathscr{B}$-sentence. Formula (2.2c) holds because existential quantifiers commute. □

For finite sets Σ of sentences, I will sometimes write $\mathsf{R}_{\mathscr{B}}(\Sigma)$ instead of $\mathsf{R}_{\mathscr{B}}(\bigwedge \Sigma)$. $\mathsf{R}_{\mathscr{B}}(\Sigma)$ entails the same $\mathscr{B}$-sentences as Σ (Rozeboom 1962, 291–293),[21] which makes the choice of the Ramsey sentence as the synthetic component plausible. Note that the Ramsey sentence may increase the order of the theory. Specifically, a first order theory with auxiliary predicate or function symbols has a second order Ramsey sentence. If $\mathscr{B}$-sentences can be of any order, $\mathsf{R}_{\mathscr{B}}(\sigma)$ is a $\mathscr{B}$-sentence. Carnap further suggests identifying the analytic component of σ with what is now called the ‘*Carnap sentence*’

$$\mathsf{C}_{\mathscr{B}}(\sigma) \mathrel{\text{⟚}} \mathsf{R}_{\mathscr{B}}(\sigma) \rightarrow \sigma , \tag{2.3}$$

where, as for the Ramsey sentence, I will sometimes suppress the conjunction for finite sets of sentences.

In the example of the demarcation criteria, $\mathsf{R}_{\mathscr{B}}(\Theta) = \exists X \left(\forall x[\varphi(x) \rightarrow Xx] \wedge \forall x[Xx \rightarrow \psi(x)]\right) \mathrel{\text{⟚}} \forall x[\varphi(x) \rightarrow \psi(x)]$. I will argue in §12.1 that the Carnap sentence is just the weakest of a number of possible meaning postulates corresponding to a set of postulates. For the example of Θ, $\Pi = \{\forall x[\varphi(x) \wedge \psi(x) \rightarrow Sx], \forall x[Sx \rightarrow \psi(x)]\}$ is another possibility (Przełęcki and Wójcicki 1969, 391; cf. Przełęcki 1969, §7.III).

[20] In general, if I refer to a set Θ of sentences by $\ulcorner\Theta(B_1,\ldots,B_m,A_1,\ldots,A_n)\urcorner$ or $\ulcorner\Theta(\bar{B},\bar{A})\urcorner$, I will assume that $B_1,\ldots,B_m \in \mathscr{B}$ and $A_1,\ldots,A_n \in \mathscr{A}$ are the only terms that occur in Θ.

[21] Hence, unless Σ is empirically false, $\mathbf{N}_{\mathscr{B}} \subseteq \{\mathfrak{A} \mid \mathfrak{A} \in \mathbf{M}_{\mathscr{B}} \text{ and } \mathfrak{A} \vDash \mathsf{R}_{\mathscr{B}}(\Sigma)\}$.

The set $\mathbf{N}$ of the *intended* $\mathscr{V}$*-structures* that interpret *all* terms ($\mathscr{V} = \mathscr{B} \cup \mathscr{A}$) contains all and only those models of the meaning postulates Π that expand the structures in $\mathbf{N}_{\mathscr{B}}$ to $\mathscr{V}$; that is, $\mathbf{N} = \{\mathfrak{A} \mid \mathfrak{A}|_{\mathscr{B}} \in \mathbf{N}_{\mathscr{B}} \text{ and } \mathfrak{A} \vDash \Pi\}$. More generally, a structure is *intended* if and only if it can be expanded to an intended $\mathscr{V}$-structure. If Π is a singleton set containing the Carnap sentence of some set Σ of postulates, the meaning postulates restrict the interpretation of the $\mathscr{A}$-terms only if the postulates' Ramsey sentence $\mathsf{R}_{\mathscr{B}}(\Sigma)$ is true. The Carnap sentence thus formalizes the assumption that the meaning postulates are motivated by the empirical implications of the postulates in Σ and have no relevance for the interpretation of $\mathscr{A}$-terms if the postulates turn out to be empirically false. In other words, while empirical results can show the postulates but not the meaning postulates to be false, they *can* show the meaning postulates to be useless. The stronger meaning postulates accordingly formalize the assumption that the meaning postulates are motivated by the empirical implications of the postulates, but may be applicable even if the postulates turn out to be false. Which assumption is correct will depend on the postulates and the context, as Przełęcki and Wójcicki (1969, 386) note (cf. Przełęcki 1969, 76).

It is now possible to distinguish clearly between analytic and synthetic components of the language: $\mathbf{M}_{\mathscr{B}}$, $\Pi_{\mathscr{B}}$, and Π are purely analytic, and since they determine $\mathbf{M}$, $\mathbf{M}$ is purely analytic as well. The restriction from $\mathbf{M}_{\mathscr{B}}$ to $\mathbf{N}_{\mathscr{B}}$ is purely synthetic, since it relies only on empirical information. The restriction of $\mathbf{M}$ to $\mathbf{N}$, on the other hand, relies on both $\mathbf{N}_{\mathscr{B}}$ and Π, and is thus neither purely analytic nor purely synthetic (while of course it is determined by the purely synthetic restriction from $\mathbf{M}_{\mathscr{B}}$ to $\mathbf{N}_{\mathscr{B}}$ and the purely analytic Π). An analytically contingent set Σ of sentences may thus turn out to be empirically true (if it is true in the intended structures $\mathbf{N}$) or empirically false (if it is false in $\mathbf{N}$).

2.8.3 Vague terms

Even in the case of exhaustive empirical information, $\mathbf{N}_{\mathscr{B}}$ may not be a singleton set since the basic terms may be vague (Przełęcki 1969, §5). Vague basic terms pose a particular problem for the bipartition of the language in artificial language philosophy, which is again very easily spelled out with the help of a formalism developed by Przełęcki (1976). The denotation of a relation symbol P_i that is vague over some domain A tripartitions the product domain A^{m_i} into a set P_i^+ of definite instances (the *positive extension* of P_i), a set P_i^- of definite non-instances (the *negative extension*), and a set of borderline cases of P_i, which I will call P_i° (the *neutral extension*). The denotation of a function symbol F_j that is vague over A does not assign a single element $b \in A$ to an n_j-tuple $(a_1, \ldots, a_{n_j}) \in A^{n_j}$, but rather a set $F_j^{+\circ}(a_1, \ldots, a_{n_j}) = B \subseteq A$ (Przełęcki 1976, 375).[22] B can be seen as the

[22]This is a slight generalization of Przełęcki's account, who assumes that B is an interval of reals, which would therefore have to be in A.

set of possible values of the function named by F_j for the arguments $a_1, \ldots, a_{n_j}$, and I will refer to the set $\{(a_1, \ldots, a_{n_j}, b) \mid a_1, \ldots, a_{n_j} \in A, b \in F_j^{+\circ}(a_1, \ldots, a_{n_j})\}$ as the *non-negative extension* $F_j^{+\circ}$ of F_j.[23] If $F_j^{+\circ}(a_1, \ldots, a_{n_j})$ is a singleton set, I will say that F_j has a positive extension for $(a_1, \ldots, a_{n_j})$. Treating constant symbols as 0-place function symbols, this means that the denotation of a constant symbol c_k that is vague over A is a set $c_k^{+\circ} \subseteq A$.

Przełęcki (1976, 376) notes that for a function symbol F_j, $F_j^{+\circ}$ may contain unintended functions. For example, unless F_j has a positive extension over the whole domain, $F_j^{+\circ}$ contains discontinuous functions, which may go against the intended denotation of F_j. Przełęcki therefore allows the denotation of a function symbol F_j to be further determined by a set of "additional conditions" $W(F_j)$, which all intended structures have to fulfill as well. Similarly to Przełęcki's additional conditions are what Fine (1975, 124) calls '*penumbral connections*', sentences that have to be true in all intended structures. However, Fine assumes that these connections are given in the object language, not in the metalanguage determining the denotations, and he does not restrict the penumbral connections to functions only. I will follow Fine in both respects, allowing a set $W(P_i, F_j, c_k)_{i \in I, j \in J, k \in K}$ of penumbral connections for all terms in the object language.

The denotations of the vague terms over A and the penumbral connections now determine a set of intended structures:

Definition 2.4. Let the terms $\{P_i, F_j, c_k\}_{i \in I, j \in J, k \in K}$ be vague over domain A with positive, negative, and non-negative extensions $\{P_i^+, P_i^-, F_j^{+\circ}, c_k^{+\circ}\}_{i \in I, j \in J, k \in K}$, and penumbral connections $W(P_i, F_j, c_k)_{i \in I, j \in J, k \in K}$. Then the terms' *vagueness set* $\mathbf{A}$ for A contains all and only models $\mathfrak{A}$ for which

$$|\mathfrak{A}| = A, \tag{2.4}$$

$$\mathfrak{A} \models W(P_i, F_j, c_k)_{i \in I, j \in J, k \in K}, \tag{2.5}$$

$$P_i^+ \subseteq P_i^{\mathfrak{A}} \subseteq A^{m_i} - P_i^- \text{ for all } i \in I, \tag{2.6}$$

$$F_j^{\mathfrak{A}} \subseteq F_j^{+\circ} \text{ for all } j \in J, \text{ and} \tag{2.7}$$

$$c_k^{\mathfrak{A}} \in c_k^{+\circ} \text{ for all } k \in K\,. \tag{2.8}$$

The elements of the vagueness set are the *precisifications* of the terms' interpretations given the penumbral connections. In the following, I will always assume that the vagueness set for terms and penumbral connections is never empty, that is, the penumbral connections are not in conflict with the positive, negative, and non-negative extensions of the terms. Furthermore, I will assume that the penumbral connections are only used to exclude those structures from vagueness sets

[23] $F_j^{+\circ}$ is the union of a vague relation symbol's positive and neutral extension.

that cannot be excluded with the help of positive, negative, and non-negative extensions. This entails that for any vagueness set $\mathbf{A}$ over domain A for terms with $\{P_i^+, P_i^-, F_j^{+\circ}, c_k^{+\circ}\}_{i\in I, j\in J, k\in K}$ and $W(P_i, F_j, c_k)_{i\in I, j\in J, k\in K}$,

$$P_i^+ = \bigcap \left\{ P_i^{\mathfrak{A}} \mid \mathfrak{A} \in \mathbf{A} \right\} , \tag{2.9}$$

$$P_i^- = \bigcap \left\{ A^{m_i+1} - P_i^{\mathfrak{A}} \mid \mathfrak{A} \in \mathbf{A} \right\} , \tag{2.10}$$

$$F_j^{+\circ} = \bigcup \left\{ F_j^{\mathfrak{A}} \mid \mathfrak{A} \in \mathbf{A} \right\} , \text{ and} \tag{2.11}$$

$$c_k^{+\circ} = \bigcup \left\{ c_k^{\mathfrak{A}} \mid \mathfrak{A} \in \mathbf{A} \right\} \tag{2.12}$$

for all $i \in I, j \in J, k \in K$. Truth in $\mathbf{A}$ is defined as supertruth, falsity as superfalsity.[24]

As mentioned in connection with circular definitions (§2.3.1), two vague terms can be used to make each other more precise. For assume that for domain $\{1,2,3\}$, $P_1^+ = \{2\}$, $P_1^- = \{3\}$, $P_2^+ = \{2\}$, and $P_2^- = \{1\}$. Then the postulate $\forall x(P_1 x \leftrightarrow P_2 x)$ defines P_1 in terms of P_2 and *vice versa*, and gives both terms the perfectly precise interpretation $P_1^+ = P_2^+ = \{2\}$ and $P_1^- = P_2^- = \{1,3\}$.

That $\mathbf{N}_{\mathscr{B}}$, the set of intended $\mathscr{B}$-structures, may not be a singleton set has clear philosophical implications. According to Hare, for example, the concept of a person, when 'person' is taken to be in $\mathscr{B}$,[25] is such that the sentence 'The fetus is a person' is neither true nor false in $\mathbf{N}_{\mathscr{B}}$. In this example $\mathbf{N}_{\mathscr{B}}$ is not a singleton set even up to isomorphism, and hence some $\mathscr{B}$-sentences cannot be decided empirically. The decision of whether to treat the fetus as a person is thus a choice of language, and the sentence becomes analytically true or analytically false accordingly.

For most applications in the following, it will be important that $\mathbf{N}_{\mathscr{B}}$ is at least precise up to isomorphism, so that any restriction on $\mathbf{M}_{\mathscr{B}}$ by a set of $\mathscr{V}$-sentences is indeed an empirical claim (and does not just make the $\mathscr{B}$-terms more precise by restricting $\mathbf{N}_{\mathscr{B}}$ further). For this reason, it will sometimes be necessary to *design* a basic vocabulary that is precise in this sense. I will give examples of such design procedures in §2.10.2 and §4.3.3.

2.9 Papineau against artificial language philosophy

Papineau (2009a, §2.1) argues that philosophy, like science, is about the world, and thus naturalistic (cf. Papineau 2009b, 3). There are no differences between the "aims and methods" of philosophy and science, but only in the specific topics.

[24]Note that with this semantic conception of vagueness, 'Mars was always dry or not dry' is true. Other semantic conceptions are possible (cf. Przełęcki 1969, 18–21). Such a language choice would take place in the metalanguage. (See also §3, n. 3.)

[25]$\mathbf{N}_{\mathscr{B}}$ is thus taken to be determined by empirical results and actual usage.

Typically, philosophical theories are very general (e. g. theories of "spatiotemporal continuants, universals and identity"), and

> unlikely ever to be decided between by some simple experiment, which is no doubt one reason that philosophers do not normally seek out new empirical data. Even so, the naturalist will insist, such theories are still synthetic theories about the natural world, answerable in the last instance to the tribunal of empirical data.
>
> Not all philosophical questions are of great generality. Think of topics like weakness of will, the importance of originality in art, or the semantics of fiction. What seems to identify these as philosophical issues is that our thinking is in some kind of theoretical tangle, supporting different lines of thought that lead to conflicting conclusions. Progress requires an unravelling of premises, including perhaps an unearthing of implicit assumptions that we didn't realise we had, and a search for alternative positions that don't generate further contradictions. Here too empirical data are clearly not going to be crucial in deciding theoretical questions—often we have all the data we could want, but can't find a good way of accommodating them. Still, methodological naturalists will urge, this doesn't mean that cogent empirical theories are not the aim of philosophy. An empirical theory unravelled from a tangle is still an empirical theory, even if no new data went into its construction.

The reply of the artificial language philosopher to this argument is straightforward: How to speak about continuants, universals, and identity is a trans-empirical question, and its answer is thus a matter of choice; it cannot be decided by empirical research. And if the unraveling of a theoretical tangle in an empirical theory indeed cannot be decided by any further empirical information, then it, too, is trans-empirical. That of course does not mean that the theory in which the tangle occurred does not have an empirical component. It just means that the problem could be solved without changing the Ramsey sentence of the theory.

But Papineau (2009b, 9) has a counter to this reply. Using as an example a theory T with one auxiliary term F, he argues that only the Ramsey sentence, which cannot be changed by language choice, is philosophically interesting:

> From the perspective of this approach to concepts, the original theory $T(F)$ can be decomposed into the analytic Carnap sentence and the synthetic "Ramsey sentence" of the theory—$\exists\Phi\, T(\Phi)$. The Ramsey sentence expresses the substantial commitments of the theory—there *is* an entity which ...—while the Carnap sentence expresses the definitional commitment to dubbing that entity 'F'.

According to Papineau, if the Ramsey sentence of a theory is not changed, the purported solution of some theoretical tangle can at best consist in the renaming of concepts. But no tangle has ever been solved by renaming alone. For the Ramsey sentence of a theory states which auxiliary concepts there are, and the Carnap sentence only assigns them labels. Thus the Carnap sentence is very uninteresting (Papineau 2009b, 10):

> [T]he natural assumption is surely that it is the synthetic Ramsey sentences that matter to philosophy, not the analytic Carnap sentences. What makes philosophers interested in investigating further is the pretheoretical supposition that there *are* entities fitting such-and-such specifications, not just the hypothetical specification that *if* there were such entities, *then* they would count as free actions, or intentional states, or whatever.

But Papineau's argument for the irrelevance of the Carnap sentence fails even on technical grounds. For the Carnap sentence in his example has the form

$$\exists \Phi T(\Phi) \rightarrow T(F)\,, \tag{2.13}$$

that is, the variable 'Φ' in the antecedent is bound by the existential quantifiers, and thus the 'F' in '$T(F)$' cannot refer back to 'Φ'. This becomes obvious when looking at the open formula $T(\Phi) \rightarrow T(F)$, for it is not generally the case that Φ and F have the same reference: F is interpreted by some structure $\mathfrak{A}$, but Φ is interpreted by a variable assignment ν. For example, if $T(F)$ is the sentence Fb, with $b \in \mathscr{B}$ and $F \in \mathscr{A}$, then $\exists \Phi \Phi b$ is true in $\mathfrak{A}$ with $A = \{1,2\}$, $b^{\mathfrak{A}} = 1$, and $F^{\mathfrak{A}} = \{1,2\}$, because $\mathfrak{A}, \nu \vDash \Phi b$ for $\nu(\Phi) = \{1\}$. But $F^{\mathfrak{A}} \neq \nu(\Phi)$.

Papineau probably does not have the Ramsey and Carnap sentences in mind after all, for he introduces the above discussion with the claim that "it is open to us to regard the concept F as having its reference fixed via the description 'the Φ such that $T(\Phi)$'. That is, F can be understood as referring to the unique Φ that satisfies the assumptions in T, if there is such a thing, and to fail of reference otherwise" (Papineau 2009b, 8). As Papineau (2009a, §2.3) states in a similar discussion, "the Ramsey sentence corresponding to $T(F)$ is '$\exists ! \Phi(T(\Phi))$'". But this is puzzling, because Papineau (2009b, 9) also claims that the "original theory framed using the concept F is [...] equivalent to the conjunction of the Ramsey and Carnap sentences", which is generally false if the Ramsey sentence is defined in Papineau's way.

Take Papineau's own example of a "simple theory of pain", which is "constituted by the two claims that (a) bodily damage typically causes pains, and (b) pains typically cause attempts to avoid further damage" (Papineau 2009b, 4). Simplifying even more by ignoring the 'typically' and expressing causation by a conditional, one gets

$$T(P) \rightleftharpoons \forall x(Dx \rightarrow Px) \wedge \forall x(Px \rightarrow Ax)\,, \tag{2.14}$$

with D, P, and A standing for 'is damaged', 'feels pain', and 'shows avoidance behavior', respectively, and $\mathscr{A} = \{P\}$. Then

$$\forall x(Dx \to Ax) \vDash \forall x(Dx \to Dx) \land \forall x(Dx \to Ax) \tag{2.15a}$$
$$\vDash \exists Y[\forall x(Dx \to Yx) \land \forall x(Yx \to Ax)] \tag{2.15b}$$
$$\vDash \mathsf{R}_{\mathscr{B}}(T) \tag{2.15c}$$

and accordingly

$$\mathsf{C}_{\mathscr{B}}(T) \mathrel{\vDash\!\dashv} \forall x(Dx \to Ax) \to \forall x(Dx \to Px) \land \forall x(Px \to Ax)\,. \tag{2.16}$$

It is easy to see that, indeed, $\mathsf{R}_{\mathscr{B}}(T) \land \mathsf{C}_{\mathscr{B}}(T) \mathrel{\vDash\!\dashv} T$. On the other hand,

$$\exists! Y\, T(Y) \vDash \exists Y\big(T(Y) \land \forall Z[T(Z) \to Z = Y]\big) \tag{2.17}$$
$$\begin{aligned} \vDash \exists Y\big(&\forall x(Dx \to Yx) \land \forall x(Yx \to Ax) \\ &\land [\forall x(Dx \to Dx) \land \forall x(Dx \to Ax) \to D = Y] \\ &\land [\forall x(Dx \to Ax) \land \forall x(Ax \to Ax) \to A = Y]\big) \end{aligned} \tag{2.18}$$
$$\vDash \exists Y(D = Y \land A = Y) \tag{2.19}$$
$$\vDash \forall x(Dx \leftrightarrow Ax)\,, \tag{2.20}$$

which is not entailed by T. Hence $\exists! Y\, T(Y) \land \mathsf{C}_{\mathscr{B}}(T)$ is strictly stronger than T.

In fact, Papineau has silently switched from the Ramsey sentence to something akin to the Ramsey-Lewis sentence, which led to the inconsistency. Lewis (1970) introduces this sentence to allow for the explicit definition of all auxiliary terms. To achieve this, Lewis (1970, 429) first assumes that all auxiliary terms are constant symbols, by which he claims that no "generality is lost, since names can purport to name entities of any kind: individuals, species, states, properties, substances, magnitudes, classes, relations, or what not." Thus all auxiliary terms can be reformulated as names since, as Lewis (1970, 429) states, $\mathscr{B}$

> provides the needed copulas:
>
> ___ has the property ___
> ___ is in the state ___ at time ___
> ___ has ___ to degree ___
>
> and the like.

Lewis (1970, 430) further assumes a logic in which constant names and definite descriptions without denotations in the domain refer to the same object. This object is not in the domain and thus lies outside of the scope of the normal quantifiers. Therefore identities between denotationless constant names or definite descriptions are true. On this basis, Lewis identifies the auxiliary terms with definite descriptions. That is, for each auxiliary term a_i,

$$a_i = \iota y_i\, \exists \bar{y}_{-i} \forall \bar{x}[T(\bar{x}) \leftrightarrow \bar{y} = \bar{x}]\,. \tag{2.21}$$

Because of Lewis's choice of logic, these equations can still be true even if the definite descriptions occurring within them do not have a unique reference in the domain. For then the definite descriptions refer to an object that is not in the domain, to which the theoretical terms have to refer as well. As is proper for definitions, the equations "do not imply any [$\mathscr{B}$]-sentences except logical truths" (Lewis 1970, 438).

However, Lewis's definitions are based on the downright frivolous assumption that a theory T entails explicit definitions for all of its auxiliary terms, for T entails the equations (2.21) only if T also entails (Lewis 1970, 439)

$$\exists \bar{x} \forall \bar{y} [T(\bar{y}) \leftrightarrow \bar{y} = \bar{x}] \vDash\dashv \exists! \bar{x}\, T(\bar{x}) , \tag{2.22}$$

which is almost Papineau's notion of the Ramsey sentence, except that it does not involve higher order quantifiers and relies on a rewritten theory.[26]

Lewis (1970, 432–434) assumes that scientific theories are *meant* to entail the unique realization of their auxiliary terms. This is at least doubtful for Papineau's simple theory of pain, unless he indeed meant that everyone in pain shows avoidance behavior *and* everyone who shows avoidance behavior is in pain. Real scientific theories do not provide any evidence for Lewis's assumption either—at least if the auxiliary terms are introduced by the theories themselves, so that the only postulates in which the auxiliary terms occur are those of the theory. For instance, Simon (1970, §2) shows in his analysis of Ohm's law that 'voltage' and 'internal resistance' can only be explicitly defined by 'resistance' and 'current intensity' if there are at least two electric circuits. Under Lewis's assumption, Ohm's law therefore entails that there are at least two electric circuits. Simon (1947, 901) also lays out that in Newtonian particle mechanics, the component forces cannot be explicitly defined when a system contains more than four particles. Under Lewis's assumption, Newtonian mechanics thus entails that there are at most four particles. The list could go on.[27]

Thus Papineau's argument for the irrelevance of the Carnap sentence is not sound, because it relies on the false assumption that all $\mathscr{A}$-terms can be defined in $\mathscr{B}$-terms. What is more, Papineau's argument fails even under the implausible assumption that all auxiliary terms can be explicitly defined, for the simple reason that $\exists X\, T(X) \rightarrow T(F)$ can be false in models of $\exists X\, T(X)$—in fact, it can *only* be

[26] When the theoretical terms are uniquely realized, the logic on which Lewis relies does not differ from normal predicate logic.

[27] Considering the prominent role of the Ramsey-Lewis sentence in recent philosophical discourse (cf. Kim 2006, 152–154), it is somewhat puzzling, almost frustrating, to see the assumptions that were considered fatal for logical empiricism—the reliance on first order logic (see §3.3) and the explicit definability of auxiliary terms (see §3.1)—embraced so nonchalantly once logical empiricism had been abandoned (cf. Bohnert 1971a). While always implicit in the Ramsey-Lewis sentence, the assumption of first order logic is explicit, for example, in the discussion by Papineau (1996, §3), even though Papineau (1996, n. 4) considers it "arguable" that the quantification over some theoretical terms is "better represented as second order". The assumption of explicit definability is expressly endorsed by Whittle (2008, 60), for example.

false in those. Perhaps more importantly, some models of $\exists X\, T(X) \to T(F)$ are not even structures of the right type for $\exists X\, T(X) \to T(F)$, since they may have no interpretation for F.

These are no idle technical considerations; rather, they are the technical guise of the philosophical point that empirical results do not determine new concepts. Consider an elaboration of the pain example, with m kinds of damage $\{D_i\}_{1\leq i\leq m} \subseteq \mathscr{B}$, n kinds of avoidance behavior $\{A_j\}_{1\leq j\leq n} \subseteq \mathscr{B}$, and the empirical observations

$$\forall x \Big(\bigvee_{1\leq i\leq m} D_i x \to \bigvee_{1\leq j\leq n} A_j x \Big) . \tag{2.23}$$

The question is now what auxiliary concepts should be introduced. One could, for example, group all of the different kinds of pain and different kinds of avoidance behavior together, which would lead to

$$\forall x \Big(\bigvee_{1\leq i\leq m} D_i x \leftrightarrow Dx \Big) \land \forall x \Big(\bigvee_{1\leq j\leq n} A_j x \leftrightarrow Ax \Big) \land \forall x (Dx \to Ax) . \tag{2.24}$$

But one might also consider this too restrictive, since other kinds of damage or avoidance behavior could be discovered. To allow for this, one could use

$$\forall x \Big(\bigvee_{1\leq i\leq m} D_i x \to Dx \Big) \land \forall x \Big(Ax \to \bigvee_{1\leq j\leq n} A_j x \Big) \land \forall x (Dx \to Ax) .^{28} \tag{2.25}$$

Or one could directly introduce pain as an intermediary between the different kinds of damage and avoidance behavior:

$$\bigwedge_{1\leq i\leq m} \forall x (D_i x \to Px) \land \bigvee_{1\leq j\leq n} \forall x (Px \to A_j x) . \tag{2.26}$$

As a final example, the previous two approaches could be combined, which would lead to Papineau's simple theory of pain:

$$\forall x \Big(\bigvee_{1\leq i\leq m} D_i x \to Dx \Big) \land \forall x \Big(Ax \to \bigvee_{1\leq j\leq n} A_j x \Big) \\ \land \forall x (Dx \to Px) \land \forall x (Px \to Ax) . \tag{2.27}$$

Similar to the proof of (2.15a), the Ramsey sentences of all of these theories can be shown to be equivalent to the original empirical claim (2.23). Since the newly introduced concepts in these theories differ, the Carnap sentence cannot just be a

[28] Note that D and A cannot be explicitly defined in $\mathscr{B}$, so Lewis and Papineau's assumption to the contrary is false in this case.

device for labeling concepts. Incidentally, treating the D_i as instances of D or as sufficient conditions for P may make the empirical claim (2.23) psychologically easier to understand, apply, or simply remember the theory, which provides a pragmatic justification for the introduction of auxiliary terms.

The step from a theory to the conjunction of the theory's Ramsey and Carnap sentences is an equivalence transformation. Thus it is not obvious why equivalence transformations of the Ramsey sentence should be disallowed. But such equivalence transformations can make the existentially quantified variables of the Ramsey sentence disappear, for example in the case of the Ramsey sentence of the simple pain theory (2.15a). Conversely, one could introduce additional existentially quantified variables, say $\exists Y_1 \ldots Y_k [\forall x(Dx \to Y_1) \wedge \bigwedge_{i=1}^{k-1} \forall x(Y_i x \to Y_{i+1} x) \wedge \forall x(Y_k x \to Ax)$, which in Papineau's misinterpretation of the Ramsey sentence would amount to different kinds of pain that beget each other, and finally lead to avoidance behavior.

These different options are all possible because the Ramsey sentence, unlike the Carnap sentence, does *not* introduce new concepts. And for this reason, the Carnap sentence is conceptually interesting, even though it does not have empirical content.

2.10 Language choice in the sciences

> If you wish to learn from the theoretical physicist anything about the methods which he uses, I would give you the following piece of advice: Don't listen to his words, examine his achievements. For to the discoverer in that field, the constructions of his imagination appear so necessary and so natural that he is apt to treat them not as the creations of his thoughts but as given realities.
>
> (Einstein 1934, 163)

> [In Carnap's *Foundations of Logic and Mathematics*, designata] are admitted not only for concrete terms but also, in some cases at least, for abstract symbols and expressions. [...] The reviewer would prefer a still more liberal admission of abstract designata, not on any realistic ground, but on the basis that this is the most intelligible and useful way of arranging the matter—it would apparently be meaningless to ask whether abstract terms really have designata, but it is rather a matter of taste or convenience whether abstract designata shall be postulated.
>
> (Church 1939, 822)

Since artificial language philosophy is a type of linguistic philosophy, it is often considered a quintessential armchair philosophy, and thus anathema to naturalism. Quine suggested the naturalization of epistemology specifically as an alternative to artificial language philosophy, but I will argue that artificial language philosophy is already naturalized, although not in Quine's sense of the word. Specifically, artificial language philosophy is "methodologically naturalistic" (Papineau 2009a, §2), that is, it relies exclusively on methodologies that are prominent in the (natural) sciences.[29]

One motivation for methodological naturalism stems from the perceived gap between the number of results in the sciences and in philosophy. Science has achieved staggering successes in this respect, from enduring insights into the nature of the cosmos to conveniences like digitized music to cancer treatments. Philosophical insights, on the other hand, seem comparably fleeting (being disputed by the next generation of philosophers or even immediate colleagues), seldom convenient, and even less often a matter of life and death.[30] The hope is that by relying on a naturalistic methodology, philosophy can achieve similar successes as the sciences, or at least improve a little.

Previous arguments for methodological naturalism were based on the assumption that scientific claims are synthetic, and a naturalized philosophy therefore has to make synthetic claims as well. Papineau's argument (in part discussed in §2.9), for example, aims to show that philosophy as currently practiced leads to synthetic claims about the world. And experimental philosophy is obviously making empirical claims about intuitions or linguistic behavior. I will argue that these approaches to the naturalization of philosophy are trying to mimic exactly the wrong component of science. Rather than trying to arrive at empirical results in philosophy, philosophers should learn from the conceptual work done in the sciences. Just like philosophers, scientists can find themselves in conceptual tangles, and it seems that scientists are generally good at unraveling such problems. It is my aim in this section to show that, like philosophy, science relies on an artificial language methodology.

The motivation behind artificial language methodology was to mimic concept formation in the sciences. For example, Carnap's discussion of explication very often relies on examples from the sciences. Indeed, "[p]hilosophers, scientists, and mathematicians make explications very frequently" (Carnap 1962, §3), which Hempel (1952, 664) also points out:

> Explication is not restricted to logical and mathematical concepts [...]. Thus, e. g., the notions of purposiveness and of adaptive behavior, whose vagueness has fostered much obscure or inconclusive argumentation about the specific characteristics of biological phenomena, have become the objects of systematic explicatory efforts. [...] Similarly, the controversy over whether a satisfactory definition of personality is attainable in purely psychological terms or requires reference to a cultural setting centers around the question whether a sound explicatory or predictive theory of personality is possible

[29] This section has been presented under the title "Armchair Philosophy Naturalized" at the conference *Salzburgiense Concilium Omnibus Philosophis Analyticis* at the Universität Salzburg, Austria, September 9, 2011. I thank the audience for helpful discussions.

[30] What *counts* as death may of course be a philosophical question. But this would be a matter of 'life' and 'death', or *life* and *death*.

> without the use of sociocultural parameters; thus, the problem is one of explication.

Accordingly, science is teeming with explicata, such as 'temperature' explicating 'warm' (Carnap 1950a, §4; Hempel 1952, §10), and completely new terms like 'phlogiston', 'oxygen', 'gene', and 'hydrochloric acid', which were introduced to account for phenomena described in basic terms like 'breathing', 'fire', 'child', and 'dissolving'. I will show that this view of scientific methodology is correct using examples from empirical research and conceptual arguments.

2.10.1 Empirical arguments

Empirical investigations of scientific methodology have been done with increasing empirical rigor, and the most recent studies support the claim that science relies on an artificial language methodology, Chang (2004), for example, has given his investigation of the formation of the temperature concept the title *Inventing Temperature*, and concludes on the basis of his investigation (Chang 2004, 206–208):

> The first thing we need to do is lose the habit of thinking in terms of simple correctness. It is very tempting to think that the ultimate basis on which to judge the validity of an operationalization should be whether measurements made on its basis yield values that correspond to the real values. But what are "the real values"? Why do we assume that unoperationalized abstract concepts, in themselves, possess any concrete values at all? [. . .] An unoperationalized abstract concept does not correspond to anything definite in the realm of physical operations, which is where values of physical quantities belong. [. . .] Once an operationalization is made, the abstract concept possesses values in concrete situations. But we need to keep in mind that those values are products of the operationalization in question, not independent standards against which we can judge the correctness of the operationalization itself.

Thus there is no thing or property called 'temperature' in the world that is being found out. Rather, scientists develop, and in effect *choose* their concepts.

In the previously mentioned overview of biological concept formation, Stotz (2009, §3, footnote and reference removed) states that she and her colleagues

> have come to appreciate that conceptual change in science is rationally motivated by what scientists are trying to achieve, by their accumulated experience of how to achieve it, and by changes in what they are trying to achieve. Empirical science is a powerhouse of conceptual innovation because scientists use and reuse their terminology in a

> truly "exuberant" way. The gene concept is a case in point: despite its ever-changing definition, the gene remains on the laboratory bench after a whole century because it has proved a flexible tool. This only makes sense if we think of concepts as tools of research, as ways of classifying the experience shaped by experimentalists to meet their specific needs. Necessarily these tools get reshaped as the scientists' needs change.

In other words, scientists choose their concepts according to the concepts' expedience. And specifically in biology, Stotz (2009, §4) comes to the following conclusion:

> It is simply that the molecular gene concept is not a good tool for some kinds of research. The instrumental, Mendelian gene remains the best tool in fields like medical genetics and population genetics. So while a particular scientific concept reflects the scientific knowledge at a point in time, this alone cannot explain the parallel use of several different concepts. For a full understanding of that phenomenon we need to see scientific concepts as tools for research, as much as glassware, microscopes or scales.

Thus there are good reasons to believe that some sciences, at least, rely on an artificial language methodology. Justus (2011, abstract) explicitly makes this connection. On the basis of a case study of the concept of ecological stability, he argues that Carnap's theory of explication describes "how concepts should be characterized".

2.10.2 Conceptual arguments

Conceptual arguments also rely implicitly on empirical assumptions about scientific methodology. In the cases that I present, however, the empirical assumptions are not mine. I will only draw conclusions from others' conceptualizations of science.

There are, of course, explicit defenses of artificial language methodology in general and in the sciences in particular. I have already discussed some arguments in §2.4 and presented some more arguments in §2.6. A more general argument rests on the assumption of *semantic empiricism*, the view that the only possibility of assigning meaning to terms is through observation or with the help of terms that have themselves been assigned meaning through observation.[31] Within semantic empiricism, then, non-observational terms are always auxiliary, since their interpretation depends on the introduction of meaning postulates. On the other hand, observational terms may be basic or auxiliary, since semantic empiricism

[31] Note that in 'semantic empiricism', the term 'empiricism' suggests observations, while my use of 'empirical' only refers to statements or states of affairs that can be decided uncontroversially.

does not entail that observations are *sufficient* to determine the meaning of terms. Rozeboom (1962) gives an elaborate defense of semantic empiricism and thus the claim that non-observational terms can only be given meaning through language choice. If semantic empiricism is true, it is clear that the sciences, like philosophy, have no other way to determine non-observational concepts than to choose them.

Carnap (1966, 187–188) goes beyond semantic empiricism and argues that any new concept of the scientific language must be chosen, whether it is observational or not. Thus concept formation in the sciences proceeds in the same way as in artificial language philosophy:

> A working physicist is constantly coming upon methodological questions. What sort of concepts should he use? What rules govern these concepts? By what logical method can he define his concepts? How can he put his concepts together into statements and the statements into a logically connected system or theory? All these questions he must answer as a philosopher of science; clearly, they cannot be answered by empirical procedures.

Unfortunately, Carnap leaves the categorical statement of the last half sentence without proof. But even so, it can be seen as a shifting of the burden of proof: To defend the position that science discovers rather than invents concepts like *temperature* and *gene*, one has to provide (and justify) an empirical procedure for deciding whether a concept is correct.

Other assumptions about the sciences lead only indirectly to an artificial language methodology. In his argument for naturalized epistemology, Quine (1969b, 81–82) concludes that "one has no choice but to be an empiricist so far as one's theory of linguistic meaning is concerned". He argues for the conventionality of language choice in the context of his argument for the indeterminacy of translation:

> [T]he linguist will end up with unequivocal translations of everything; but only by making many arbitrary choices [...]. By this I mean that different choices would still have made everything come out right that is susceptible in principle to any kind of check.

With this claim of language choice being insusceptible to any kind of check, Quine asserts what I will call epistemic empiricism, the position that the justification of a factual claim can only rely on observation. Epistemic empiricism provides pragmatic support for semantic empiricism, for even if words could non-conventionally apply to unobservable things or properties, whether they in fact do apply in a specific case must rely on observation. Without semantic empiricism, epistemic empiricism renders the truth of non-observation statements unknowable in principle due to lack of justification. Thus at a minimum, Quine's view entails that for all practical purposes, all non-observational terms are auxiliary. Arguably, however, Quine's position entails full-blown artificial language philosophy, since Quine (1951, 41) explicitly embraces conventionalism:

> As an empiricist I continue to think of the conceptual scheme of science as a tool, ultimately, for predicting future experience in the light of past experience. Physical objects are conceptually imported into the situation as convenient intermediaries—not by definition in terms of experience, but simply as irreducible posits comparable, epistemologically, to the gods of Homer. [...] Both sorts of entities enter our conception only as cultural posits. The myth of physical objects is epistemologically superior to most in that it has proved more efficacious than other myths as a device for working a manageable structure into the flux of experience.

Quine does not assert that there is a fixed set of observational terms with their associated concepts. Rather, how experiences are described may depend on the conceptual scheme.

In his constructive empiricism, van Fraassen (1980, §1.3) argues that non-observable objects are inaccessible to science, and thus science has to make do with empirical adequacy (see §4.2.1). The result is that there is no scientific means of determining the applicability or non-applicability of terms to unobservable objects. As far as these objects are concerned, science has to rely on convention. Thus the statements of both philosophy and the sciences, whenever they go beyond observation, are conventions about language use, and there is no other methodology that can achieve rational, intersubjective agreement.

Similarly, Sober (1990, 404) states that according to his position of contrastive empiricism, “science is not in the business of discriminating between empirically equivalent hypotheses”, where Sober considers theories empirically equivalent if they assign the same probabilities to all observation statements (see §8.2). Therefore, in particular the decision between two empirically equivalent theories with different concepts is a matter of choice.

My final example is Andreas (2010, 538), who develops a semantics for scientific theories that is holistic, and claims that

> it is rather misleading to construe relative holism as relying on the analytic-synthetic distinction. This becomes evident in light of the present account of semantic holism. In this account, only sentences qualifying as postulates are assumed to determine the meaning of theoretical terms. And the distinction between postulates and other theoretical sentences must clearly not be equated with the analytic-synthetic distinction. Analyticity is therefore no requirement for a sentence to determine the meaning of nonlogical symbols.

Since postulates can have both analytic and synthetic components, analyticity is clearly no requirement for determining the meaning of terms. The interesting question is whether it is possible to distinguish between the analytic and synthetic component of postulates. I will now show that Andreas's account can be captured completely in the account of concept formation given in §2.3, and thus allows for the introduction of an analytic-synthetic distinction and the notion of conventional concept formation developed above.

Andreas (2010, 529–532) relies on a bipartition of the vocabulary into observational terms $\mathscr{O}$ and theoretical terms $\mathscr{T}$, and assumes a single intended $\mathscr{O}$-structure $\mathfrak{A}_{\mathscr{O}}$.[32] He demands only that the $\mathscr{O}$-terms are primitively interpreted and explicitly bases his semantics on Carnap's view on the interpretation of scientific terms. Thus both technically and in spirit, his $\mathscr{O}$-terms do not have to be connected directly to experience, but rather only have to be somehow uncontroversial. Andreas (2010, 532) further assumes that there is a set D of unobservable objects, which are not in $A = |\mathfrak{A}_{\mathscr{O}}|$. $\mathrm{MOD}(\Theta)$ is the set of models of the postulates Θ, and $\mathrm{EXT}(\mathfrak{A}_{\mathscr{O}}, \mathscr{V}, D)$ the set of expansions to $\mathscr{V}$ of the extensions of $\mathfrak{A}_{\mathscr{O}}$ to D. In other words, $\mathrm{EXT}(\mathfrak{A}_{\mathscr{O}}, \mathscr{V}, D)$ is the set of all structures $\mathfrak{B}$ with domain $A \cup D$ that have $\mathfrak{A}_{\mathscr{O}}$ as a *relativized reduct*, $\mathfrak{B}|A_{\mathscr{O}} = \mathfrak{A}_{\mathscr{O}}$. It is a standard notion in model theory (cf. Hodges 1993, §5.1).[33] This allows (Andreas 2010, 533, my formulations)

Definition 2.5.

$$\mathbf{S} := \begin{cases} \mathrm{MOD}(\Theta) \cap \mathrm{EXT}(\mathfrak{A}_{\mathscr{O}}, \mathscr{V}, D) & \text{if } \mathrm{MOD}(\Theta) \cap \mathrm{EXT}(\mathfrak{A}_{\mathscr{O}}, \mathscr{V}, D) \neq \varnothing \\ \mathrm{EXT}(\mathfrak{A}_{\mathscr{O}}, \mathscr{V}, D) & \text{if } \mathrm{MOD}(\Theta) \cap \mathrm{EXT}(\mathfrak{A}_{\mathscr{O}}, \mathscr{V}, D) = \varnothing \end{cases} \tag{2.28}$$

Considering $\mathbf{S}$ as a vagueness set, Andreas thus assumes that for all $\mathscr{O}$-terms, the (tuples of) elements of A are in their positive or negative extensions, while all (tuples with) elements of D are in their neutral extensions. In other words, the $\mathscr{O}$-terms are completely precise over A and completely vague over D.

Definition 2.6. The truth values $\tau(\varphi)$ of $\mathscr{V}$-sentences are defined as follows:

1. $\tau(\varphi) := T$ if and only if $\mathfrak{B} \models \varphi$ for every $\mathfrak{B} \in \mathbf{S}$.
2. $\tau(\varphi) := F$ if and only if $\mathfrak{B} \not\models \varphi$ for every $\mathfrak{B} \in \mathbf{S}$.
3. $\tau(\varphi)$ is indeterminate if and only if $\mathfrak{B} \models \varphi$ for some $\mathfrak{B} \in \mathbf{S}$ and $\mathfrak{B} \not\models \varphi$ for some $\mathfrak{B} \in \mathbf{S}$.

$\mathfrak{A}_{\mathscr{O}}$ combines empirical and conceptual information, namely the interpretation of the $\mathscr{O}$-terms given the way our world happens to be. Andreas allows for

[32] See §2.8.1.

[33] Hodges (1993, 202–203) defines relativized reducts as those substructures of a reduct that have the extension of some one-place predicate as their domain. I use a slight generalization.

any extension of $\mathfrak{A}_{\mathscr{O}}$ to D, and thus, like Przełęcki (see §3.9), assumes that $\mathscr{O}$-terms are completely vague for unobservable objects. This differs from Carnap's assumption that $\mathscr{O}$-terms do not apply to unobservable objects (see §3.6.2). To connect Andreas's account directly with Carnap's, one thus needs to introduce new observational terms as follows: Let the term O^* be interpreted by A. Then the new, starred terms are conditionally defined by[34]

$$\begin{aligned} \Pi^* :=& \{\forall \bar{x}[O^* x_1 \wedge \cdots \wedge O^* x_{m_i} \to (P_i \bar{x} \leftrightarrow P_i^* \bar{x})] \mid i \in I\} \\ &\cup \{\forall \bar{x}[O^* x_1 \wedge \cdots \wedge O^* x_{n_j} \to (F_j \bar{x} = F_j^* \bar{x})] \mid j \in J\} . \end{aligned} \tag{2.29}$$

With $\mathfrak{A}_{\mathscr{O}}$ and D, one can furthermore define a new structure

$$\begin{aligned} \mathfrak{A}_{\mathscr{B}} = \big\langle A \cup D, \langle O^*, A\rangle, \langle P_1^*, P_1^{\mathfrak{A}_{\mathscr{O}}}\rangle, \ldots, \langle P_s^*, P_s^{\mathfrak{A}_{\mathscr{O}}}\rangle, \\ \langle F_1^*, F_1^{\mathfrak{A}_{\mathscr{O}}}\rangle, \ldots, \langle F_t^*, F_t^{\mathfrak{A}_{\mathscr{O}}}\rangle, \langle c_1, c_1^{\mathfrak{A}_{\mathscr{O}}}\rangle, \ldots, \langle c_u, c_u^{\mathfrak{A}_{\mathscr{O}}}\rangle\big\rangle . \end{aligned} \tag{2.30}$$

$\mathfrak{A}_{\mathscr{B}}$ differs from $\mathfrak{A}_{\mathscr{O}}$ in the symbols naming the relations and functions (which are starred in $\mathfrak{A}_{\mathscr{B}}$ and not starred in $\mathfrak{A}_{\mathscr{O}}$), by interpreting the one-place O^* by A and by including D in its domain. It is then straightforward to show that for every $\mathfrak{A}_{\mathscr{B}\cup\mathscr{O}} \models \Pi^*$ with domain $A \cup D$, the relativized reduct $\mathfrak{A}_{\mathscr{B}\cup\mathscr{O}}|A_{\mathscr{O}} = \mathfrak{A}_{\mathscr{O}}$ if and only if $\mathfrak{A}_{\mathscr{B}\cup\mathscr{O}}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$.

$\mathbf{N}_{\mathscr{B}} = \{\mathfrak{A}_{\mathscr{B}}\}$ is a set of intended $\mathscr{B}$-structures that interprets all observational terms precisely (see §2.8.1). The vague $\mathscr{O}$-terms and the theoretical terms then make up the auxiliary terms $\mathscr{A}$, $\mathscr{A} = \mathscr{O} \cup \mathscr{T}$. The set $\mathbf{N}$ of intended structures for $\mathscr{V} = \mathscr{B} \cup \mathscr{A}$ (see §2.8.2) is then given by those expansions of the intended $\mathscr{B}$-structures that are models of the meaning postulates Π. The full set of postulates is $\Pi^* \cup \Theta$, and thus the Carnap sentence $\mathsf{C}_{\mathscr{B}}(\Pi^* \cup \Theta)$ is analytic, which assumes that Θ and Π^* are finite or at least finitely axiomatizable. However, since Π^* is *chosen* to be analytic as well, the full set of meaning postulates is $\Pi^* \cup \{\mathsf{C}_{\mathscr{B}}(\Pi^* \cup \Theta)\}$.[35]

By phrasing Andreas's account in Carnap's terms, a straightforward relation between $\mathbf{S}$ and $\mathbf{N}$ can be established with the help of

Lemma 2.4. $\mathfrak{A}_{\mathscr{B}} \models \mathsf{R}_{\mathscr{B}}(\Theta)$ *if and only if* $\mathfrak{A}_{\mathscr{B}}$ *has an expansion to a model of* Θ.

Proof. Let $\Theta^\dagger$ be the result of substituting each $\mathscr{A}$-term in Θ by a corresponding variable.

'⇒': Since $\mathfrak{A}_{\mathscr{B}} \models \mathsf{R}_{\mathscr{B}}(\Theta)$, there is a relation V_i for every relation symbol P_i in $\mathscr{A}$, a function G_j for every function symbol F_j in $\mathscr{A}$, and a constant d_k for every

[34] As Andreas's reference to Beth's theorem shows (Andreas 2010, 531), he intends his formalism for first order model theory. For higher order $\mathscr{O}$-terms, the definition would have to be generalized accordingly.

[35] Hence, if $\Pi^* \cup \Theta$ is considered one theory, the full set of meaning postulates is stronger than the Carnap sentence. I discuss the relation of the Carnap sentence to meaning postulates in general in §12.1.

constant symbol c_k in $\mathscr{A}$ such that $\{V_i, G_j, d_k\}$ satisfies $\Theta^\dagger$ in $\mathfrak{A}$. Define $\mathfrak{C}$ so that $P_i^{\mathfrak{C}} = V_i$ for each V_i, $F_j^{\mathfrak{C}} = G_j$ for each G_j, $c_k^{\mathfrak{C}} = d_k$ for every d_k, and $\mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}$. Induction on the complexity of Θ shows that $\mathfrak{C} \vDash \Theta$.

'$\Leftarrow$': Induction shows that $\{P_i^{\mathfrak{C}}, F_j^{\mathfrak{C}}, c_k^{\mathfrak{C}}\}$ satisfies $\Theta^\dagger$ in $\mathfrak{A}$, so that $\mathfrak{A} \vDash \exists_i X_i \exists_j Y_j \exists_k x_k \Theta^\dagger$. □

Note that the proof also works in higher order logic. In Tarski semantics, lemma 2.4 provides a connection between the model theory of a logic of order n and the quantification of order $n+1$. It will be used again and again in the rest of this book.

Since there are no $\mathscr{B}$-terms in Andreas's account, define $\mathbf{N}|_{\mathscr{A}} := \{\mathfrak{A}|_{\mathscr{A}} \mid \mathfrak{A} \in \mathbf{N}\}$. Then the following holds:

Claim 2.5. *Let $\mathscr{B}$, $\mathscr{A}$, $\mathbf{N}_{\mathscr{B}}$, and Π be defined as above. Then* $\mathbf{N}|_{\mathscr{A}} = \mathbf{S}$.

Proof. First, note that for any Θ and $\mathfrak{C} \vDash \Pi^*$, $\mathfrak{C} \vDash \Pi^* \cup \Theta$ and $\mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$ iff $\mathfrak{C}|_{\mathscr{A}} \vDash \Theta$ and $\mathfrak{C}|A_{\mathscr{O}} = \mathfrak{A}_{\mathscr{O}}$ $(*)$, since $\mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$ iff $\mathfrak{C}|A_{\mathscr{B}} = \mathfrak{A}_{\mathscr{O}}$. Conversely, for any Θ and $\mathfrak{D}_{\mathscr{A}}$, $\mathfrak{D}_{\mathscr{A}} \vDash \Theta$ and $\mathfrak{D}_{\mathscr{A}}|A_{\mathscr{O}} = \mathfrak{A}_{\mathscr{O}}$ iff there is a $\mathfrak{C}$ with $\mathfrak{C} \vDash \Pi^* \cup \Theta$, $\mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$, and $\mathfrak{C}|_{\mathscr{A}} = \mathfrak{D}_{\mathscr{A}}$ $(**)$, since $\mathfrak{D}_{\mathscr{A}}|A_{\mathscr{O}} = \mathfrak{A}_{\mathscr{O}}$ iff every expansion to $\mathscr{B}$ in which Π^* is true is such that its reduct to $\mathscr{B}$ equals $\mathfrak{A}_{\mathscr{B}}$.

Now, $\mathfrak{C} \in \mathbf{N}$ iff $\mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$ and $\mathfrak{C} \vDash \bigwedge \Pi^* \wedge \mathsf{C}_{\mathscr{B}}(\Pi^* \cup \Theta)$. This holds iff $\mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$, $\mathfrak{C} \vDash \Pi^*$, and either $\mathfrak{C} \vDash \Pi^* \cup \Theta$ or (by lemma 2.4) there is no $\mathfrak{D} \vDash \Pi^* \cup \Theta$ with $\mathfrak{D}|_{\mathscr{B}} = \mathfrak{C}|_{\mathscr{B}}$. Because of $(*)$, this holds iff either $\mathfrak{C}|A_{\mathscr{O}} = \mathfrak{A}_{\mathscr{O}}$, $\mathfrak{C} \vDash \Pi^*$, and (by $**$) there is no $\mathfrak{D} \vDash \Theta$ with $\mathfrak{D}|A_{\mathscr{O}} = \mathfrak{A}_{\mathscr{O}}$, or $\mathfrak{C}|A_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$ and $\mathfrak{C} \vDash \Theta$. This is the case iff $\mathfrak{C}|_{\mathscr{A}} \in \mathrm{EXT}(\mathfrak{A}_{\mathscr{O}}, \mathscr{V}, D)$ and $\mathrm{MOD}(\Theta) \cap \mathrm{EXT}(\mathfrak{A}_{\mathscr{O}}, \mathscr{V}, D) = \varnothing$ or $\mathfrak{C}|_{\mathscr{A}} \in \mathrm{MOD}(\Theta) \cap \mathrm{EXT}(\mathfrak{A}_{\mathscr{O}}, \mathscr{V}, D)$ (and thus $\mathrm{MOD}(\Theta) \cap \mathrm{EXT}(\mathfrak{A}_{\mathscr{O}}, \mathscr{V}, D) \neq \varnothing$). This holds iff $\mathfrak{C}|_{\mathscr{A}} \in \mathbf{S}$. □

Corollary 2.6. *Let $\mathscr{B}$, $\mathscr{A}$, $\mathbf{N}_{\mathscr{B}}$, and Π be defined as above. Then for every $\mathscr{A}$-sentence φ,*

1. *$\tau(\varphi) = T$ if and only if φ is true in* $\mathbf{N}$.
2. *$\tau(\varphi) = F$ if and only if φ is false in* $\mathbf{N}$.
3. *$\tau(\varphi)$ is indeterminate if and only if φ is neither true nor false in* $\mathbf{N}$.

Proof. From claim 2.5 and definition 2.6, since φ is true/false in $\mathbf{N}$ iff φ is true/false in $\mathbf{N}|_{\mathscr{A}}$. □

Corollary 2.6 shows that Andreas's semantics, like the other philosophical positions discussed, allows for a distinction between the empirically accessible synthetic component of scientific theories and the analytic component. The analytic component in this case is $\Pi = \Pi^* \cup \{\mathsf{C}_{\mathscr{B}}(\Pi^* \cup \Theta)\}$, where Π^* had to be introduced to ensure that every $\mathscr{B}$-sentence makes an empirical assertion, rather than just reducing vagueness by introducing new conventions about the interpretations of Andreas's $\mathscr{O}$-terms. The synthetic component of the theory is $\mathsf{R}_{\mathscr{B}}(\Pi^* \cup \Theta)$.

In conclusion, Quine argues that the choice of language cannot be restricted with the help of an empirical check, van Fraassen argues that there is no epistemic access to the applicability of terms to unobservables, Sober argues that science does not discriminate between empirically equivalent theories, and Andreas implicitly introduces a set Π that can be chosen to be true without the possibility of an empirical check. In all these cases, then, scientific theories have an empirical component and a conventional component. How extensive this conventional component is can be inferred from an argument due to William Demopoulos.

2.10.3 Demopoulos against artificial language science

Papineau argues, in effect, that the Ramsey sentence of a set of postulates Θ is so strong that there is nothing left for the Carnap sentence to do besides the labeling of concepts, and in particular, there is no relevant conceptual work to be done. Demopoulos (2003, 2007, 2008) on the other hand argues that the Ramsey sentence, with $\mathscr{B}$ as the set of observational terms, is too weak even to capture the synthetic content of a theory.

In his first argument, Demopoulos (2003, 387) constructs an interpretation of Θ's $\mathscr{A}$-terms given a single intended structure $\mathfrak{N}_{\mathscr{B}}$. Given that Θ is consistent, Θ has a model $\mathfrak{A}$, and he assumes that A has the same cardinality as N without "any significant loss of generality or philosophical interest". Demopoulos (2003, 387, my notation) continues:

> It is therefore possible to extend the partial interpretation $[\mathfrak{N}_{\mathscr{B}}]$ to the theoretical vocabulary of Θ by letting each predicate of its theoretical vocabulary denote the image in N of its interpretation in $\mathfrak{A}$ under any one–one correspondence between A and N.[36] For example, suppose T is a binary theoretical relation of Θ. Then the interpretation $T^{\mathfrak{N}}$ of T in $\mathfrak{N}$ is *defined* as the image under φ, φ one–one from A on to N, of its interpretation $T^{\mathfrak{A}}$ in $\mathfrak{A}$. Since by construction $\langle a, b\rangle$ is in $T^{\mathfrak{A}}$ if and only if $\langle \varphi a, \varphi b\rangle$ is in $T^{\mathfrak{N}}$, φ is an isomorphism; and therefore, if $\mathfrak{A}$ is a model of Θ, so is $\mathfrak{N}$.

This is false:[37] Choose $\Theta = \{\forall xy(Oxy \leftrightarrow Txy)\}$, $A = \{a, b\}$ and $O^{\mathfrak{A}} = T^{\mathfrak{A}} = \{\langle a, b\rangle\}$. Then $\mathfrak{A} \models \Theta$. Choose further the $\mathscr{B}$-structure $\mathfrak{N}_{\mathscr{B}}$ with $N = \{1, 2\}$,

[36] Here and in the following, the original text does not distinguish between a structure and its domain (referring, for example, to the "the image in N of its interpretation in A under any one–one correspondence between A and N"). The distinction in my notation thus has to be taken as an interpretation of the quote.

[37] Wagner (2009) was the first to criticize Demopoulos's proof, although the following argument differs from his.

$O^{\mathfrak{N}_{\mathscr{B}}} = \{\langle 1,1\rangle\}$. Then $\mathfrak{N}$ with $\mathfrak{N}|\mathscr{B} = \mathfrak{N}_{\mathscr{B}}$ and $T^{\mathfrak{N}} = \{\langle \varphi a, \varphi b\rangle\}$ is not a model of Θ under any bijection $\varphi : A \longrightarrow N$ (let alone *all* bijections). φ is an isomorphism from $\mathfrak{A}|_{\mathscr{A}}$ to $\mathfrak{N}|_{\mathscr{A}}$, but not from $\mathfrak{A}$ to $\mathfrak{N}$. Of course, if $\mathfrak{N}_{\mathscr{B}}$ was not fixed and the bijection φ could be used to define not only the auxiliary terms but all terms (and thus $\mathfrak{N}$), the proof would work. But this would also go against the point of a partial interpretation, for it would mean that the vocabulary is not interpreted at all.

Demopoulos (2003, 387–388, my notation) goes on to criticize the Carnap sentence, since it entails that Θ is true whenever Θ's Ramsey sentence is true:

> Call the interpretation of Θ's $\mathscr{A}$-vocabulary in $\mathfrak{N}$ that we have just described '$\mathfrak{N}_{\mathscr{A}}$' [$:= \mathfrak{N}|_{\mathscr{A}}$]. Any theory of knowledge and reference that is incapable of distinguishing truth from truth under $\mathfrak{N}_{\mathscr{A}}$ is committed to the implication that Θ is true if Θ is true under $\mathfrak{N}_{\mathscr{A}}$. But modulo our assumption about cardinality, that Θ is true under $\mathfrak{N}_{\mathscr{A}}$ is a matter of model theory. $\mathfrak{N}_{\mathscr{A}}$ is arbitrary; the construction which employs it is clearly unacceptable, since it trivializes the question whether Θ is true. [...] By equating truth with truth under $\mathfrak{N}_{\mathscr{A}}$ we rob our knowledge of the truth of our theoretical claims of its a posteriori character: modulo a single assumption about cardinality, the theoretical statements of an empirically adequate theory come out true as a matter of metalogic.

Given the problem with Demopoulos's construction of $\mathfrak{N}_{\mathscr{A}}$, the criticism is too strong. The last quoted sentence states that *if* a theory is empirically adequate (that is, it entails only true $\mathscr{B}$-sentences), then it is true. The claim that $\mathfrak{N}_{\mathscr{A}}$ is arbitrary, however, is only true if $\mathfrak{N}_{\mathscr{B}}$ is arbitrary as well, and that is not the case. In the counterexample given, $T^{\mathfrak{N}_{\mathscr{A}}} = T^{\mathfrak{N}}$ is fixed completely by $\mathfrak{N}_{\mathscr{B}}$ and the explicit definition of T by O given in Θ.

In his other arguments against the Carnap sentence, (Demopoulos 2008, 376–377; Demopoulos 2011, 186–187[38]) starts from a technical observation in first order logic (cf. Shoenfield 1967, §5, ex. 9.a; Tuomela 1973, theorem III.3): All $\mathscr{B}$-sentences entailed by Θ are true in $\mathfrak{N}_{\mathscr{B}}$ if and only if there is an expansion $\mathfrak{N}$ of an elementary extension of $\mathfrak{N}_{\mathscr{B}}$ such that $\mathfrak{N} \vDash \Theta$, where an *elementary extension* of $\mathfrak{N}_{\mathscr{B}}$ is an extension of $\mathfrak{N}_{\mathscr{B}}$ such that every tuple of from $|\mathfrak{N}_{\mathscr{B}}|$ that satisfies a $\mathscr{B}$-formula β in $\mathfrak{N}_{\mathscr{B}}$ also satisfies β in the extension (Hodges 1993, 54). Together with lemma 2.4, this means that if all basic sentences entailed by Θ are true in $\mathfrak{N}_{\mathscr{B}}$, $\mathsf{R}_{\mathscr{B}}(\Theta)$ is true in an extension of $\mathfrak{N}_{\mathscr{B}}$. (If the basic sentences are assumed to be of higher order, this result is trivial, for then $\mathsf{R}_{\mathscr{B}}(\Theta)$ itself is a basic sentence entailed by Θ.)[39]

[38] I thank F. A. Muller for pointing me to this article.

[39] In this argument, Demopoulos (2011, 187) also mentions an arbitrary mapping function (like φ in the previous argument), but restricts it to the objects in $|\mathfrak{N}| - |\mathfrak{N}_{\mathscr{B}}|$. Thus the arbitrariness of $\mathfrak{N}_{\mathscr{A}}$ is restricted to objects that are newly introduced by the elementary extension.

Demopoulos's conceptual point is the following (Demopoulos 2008, 377): If the Carnap sentence of Θ is taken to be a meaning postulate, then the truth of Θ follows analytically from the truth of the basic sentences entailed by Θ—*modulo* a cardinality assumption (and in higher order logic without this assumption). This is correct. However, Demopoulos (2008, 381) further argues that the Carnap sentence

> is incapable of accurately representing the truth of theoretical claims because it takes their truth to collapse into satisfiability in a sufficiently large domain. This is hardly what we take the truth of theoretical claims to consist in, since we characteristically—and rightly—distinguish them from those of pure mathematics. A reconstruction which fails to acknowledge this, is not merely odd, it misses what is arguably one of the chief desiderata of an adequate philosophy of the exact sciences.

This conclusion is again too strong, for mathematical claims typically differ significantly from theoretical claims according to the Carnap sentence. For one, the truth of mathematical claims does not depend on the truth of any empirical claims, while the truth of theoretical claims does. Extending the above example, define $\Theta' := \Theta \cup \{\exists xy\, Txy\}$. Then $\mathsf{C}_{\mathscr{B}}(\Theta') \mathrel{\vDash\!\dashv} \exists xy\, Oxy \to \Theta'$, so that the truth of any theoretical claim of Θ' depends on the $\mathscr{B}$-sentence $\exists xy\, Oxy$ being true. This is in stark contrast to mathematical claims.[40] Furthermore, as the original example already shows, $\mathscr{A}$-terms can differ significantly from mathematical terms in that they may be explicitly definable in $\mathscr{B}$-terms. Finally, note that Demopoulos starts from the bare intuition that there must be something more substantial to scientific theories than only their empirical implications and their conventions. But this stance only contradicts the basic idea of semantic empiricism, without providing an argument.

Since Demopoulos's argument fails to show anything wrong with taking $\mathsf{R}_{\mathscr{B}}(\Theta)$ as the empirical content of Θ, his results can be turned around: They show how much in a theory is a matter of convention. Possibly *modulo* a cardinality assumption, the conceptual apparatus of a theory can be chosen at will, as long as the $\mathscr{B}$-sentences come out true.

* * *

In conclusion, then, both artificial language philosophy and science rely on an artificial language methodology, at least for non-observational terms. For observational terms, Carnap's shifting of the burden of proof remains: If there is a term that has to be used in a specific way, such that other uses are incorrect (not in the

[40] Indeed, I will discuss in §12 a result by Winnie (1970) according to which mathematical sentences, assuming they are phrased entirely in $\mathscr{A}$-terms, can never be entailed by a Carnap sentence.

sense of going against convention, but somehow in contradiction with the way the world is), then there should be some way of testing this.

This conclusion provides a new perspective on the preceding discussions. Whenever science is used to solve philosophical (that is, non-empirical) problems, it has to rely on language choice like artificial language philosophy. In this way, the application of science to solve philosophical problems escapes Kim's criticism that a naturalized epistemology loses the normative component of traditional epistemology. According to Kim, epistemology determines what, for instance, *justification* should be, while, say, psychology only determines how people think about justification. But when relying on an artificial language methodology, scientists, like philosophers, can develop a new concept called 'justification' that is fruitful in their research. And since this new concept is *suggested*, its development has a normative component.

2.11 The relation between science and artificial language philosophy of science

That traditional or ordinary language philosophy, partially naturalized or not, leads to fruitful interactions between science and philosophy of science is far from certain. It is, for example, not obvious how insights into the use of a term in ordinary language relate to scientific insights. And while the methodology of ordinary language philosophy can be applied to scientific language to reveal inconsistent usage (Philipse 2009, §3), it cannot resolve inconsistencies without threatening to collapse into artificial language philosophy. Traditional philosophy has to establish its own access to facts about the world, besides the scientific route. Williamson (2007, §§6, 8) and Papineau (2009b, §IV) consider thought experiments and ultimately intuitions to provide this access, but they both rely on contentious claims about the workings of the human mind. In the following, I will argue that in contradistinction, the relation between science and artificial language philosophy of science is unstrained.

According to the proponents of artificial language philosophy, science very often engages in language choice in the same way that philosophy does. Therefore it is of interest to establish the scientific language more precisely. Here the methodology of ordinary language philosophy can be very helpful, except that it has to be applied to scientific rather than ordinary language. To avoid exclusive reliance on linguistic hunches, scientists' actual usage of scientific concepts can be determined empirically (cf. Stotz et al. 2004). In this way, the application of the methodology of ordinary language philosophy would help to fulfill what Reichenbach (1938, 3) calls the "descriptive task of epistemology", the search for the rules of scientific language that capture the language intuitions of the scientists. As Waters (2004, §3) argues, however, even this descriptive task goes beyond pure observation, for actual usage is often too vague or inconsistent to establish proper rules.

Reichenbach (1938, §1) identifies two additional tasks of epistemology. One is the "critical task", the identification and evaluation of inferences. In the terminology of artificial language philosophy, this amounts to rigorous derivation. Next is the "advisory task", the proposal of concepts for use in the sciences. As Waters (2004, §§5–6) lays out, Reichenbach sees scientists as the final arbiters of language choice, but such a restriction is not inherent in artificial language philosophy in general. For not all philosophers have the same goals as the scientists on whose research they rely. In principle, concepts suggested by scientists and by philosophers are on a par.

In a helpful overview, Hansson (2008a) describes several ways in which philosophy has been found to relate to scientific disciplines. However, his description conveys only sociological observations about the behavior of philosophers and scientists—even if the observations could be explained on psychological grounds, a *justification* of the observed relations has to rely on some feature of philosophy itself. I will argue that the relations are justified and clarified by the relation of science and philosophy outlined above.

New empirical results provide material for philosophical investigation. Hansson (2008a, 477) describes a host of influences of scientific disciplines on philosophical work, but his examples mix concept formation, rigorous derivation, and empirical results. With respect to the empirical results, Hansson notes the influence of quantum mechanics and evolutionary biology on philosophy, the influence of psychology and neuroscience on the philosophy of mind, and the influence of linguistics on the philosophy of language. These examples show that some philosophical concepts (in the philosophy of mind, philosophy of language, etc.) are chosen to accommodate empirical results, and thus have to change to remain relevant and fruitful in the light of new results. This is a trivial implication of the way in which languages are chosen in philosophy as well as the sciences.

New rigorous derivations provide material for philosophical investigation. Hansson (2008a, 477) states that results in game and decision theory have provided moral philosophy with new problems for ethical analysis. Such results are established by rigorous derivations based on the language. Because they are not empirical, they can also fall within the domain of philosophy, or they can suggest new language choices in philosophy, for example by revealing previously hidden relations between concepts.

New concepts provide material for philosophical investigation. Hansson (2008a, 477) further notes that game and decision theory have also provided new formulations of old problems in moral philosophy. While moral philosophy is outside the scope of this article, it seems clear that, to use Hansson's examples, psychology, neurosciences, linguistics, quantum mechanics, and biology have all engaged in concept formation. Carnap (1966, 187–189) automatically considers such conceptual work philosophical because it does not involve asserting or testing observational claims. But even with a more restrictive view of philosophy, some philosophical concepts rely on scientific ones (by way of conditions of adequacy, for example), and therefore must be updated whenever there are changes in the scientific concepts. Furthermore, completely new scientific concepts provide new ways for philosophical concepts to be fruitful. Since scientific concepts can also be introduced and changed because of new empirical results, the relation between scientific and philosophical language choice provides yet another way for philosophical concepts to change in light of empirical results.

Methods and issues of philosophy are taken up by other sciences. According to Hansson (2008a, 477), some issues and methods of philosophy have been taken up within other disciplines, for example the investigation of structures of concepts and thought processes in computer science. Since concept formation and rigorous derivation occur in both philosophy and the sciences, it is unsurprising that science can join philosophy in these tasks. The particular proximity of computer science to philosophical research may stem from the computer scientists' need for new languages that capture the structure of concepts and thought processes. But close connections have also formed in the case of formal logic and mathematics, and also in the empirical sciences (the explication of 'measurement' in quantum mechanics being an excellent example).

Philosophy is part of the community of interdependent disciplines. Hansson (2008a, §3) notes the growing number of interdisciplinary endeavors and concludes on historical grounds that philosophy is part of the "community of interdependent disciplines". Successful philosophical investigations into natural or social phenomena, he claims, have always relied on results from other disciplines (such as the reliance of the philosophy of space and time on relativity theory).

Given the discussion so far, the interdisciplinary nature of philosophical research seems clear, following from both the conditions of adequacy and the demand that philosophical concepts be fruitful. The large role of language choice in the sciences is probably most evident in space-time physics, for relativity theory not only predicts new empirical phenomena, but also suggests a new language to accommodate old phenomena in a different way. Philosophers of space and time have had to evaluate this suggestion, and indeed have accepted the superiority of the new language in many contexts.

Problems answered experimentally or accurately become non-philosophical. Hansson (2008a, 476–477) also reviews the claim that many philosophical topics move into a dedicated field of science once clear answers are at hand. He gives the example of psychology, which parted from philosophy after the introduction of experiments.

The term 'parted' is a somewhat problematic term in this connection, since according to Hansson himself, cooperation between empirical science and philosophy is both possible and fruitful. To rephrase the point, it has become *possible* to pursue experimental psychology without relying on philosophical considerations, or better, without the need to develop new concepts.

Since accuracy in the rules for the application of a term is one desideratum of an explication, it is unsurprising that fulfillment of this desideratum often marks the end of philosophical work. Furthermore, empirical research involving auxiliary terms is only possible if there are meaning postulates to connect the auxiliary and basic terms. The split between philosophy and (some areas of) psychology can therefore be seen as the result of the development of a precise language to establish such connections. Of course, once experimental research is pursued, the language can still be modified on the basis of the experimental results. Since a language can be so precise that it allows for rigorous derivations, Hansson's point also applies, for example, to the parting of some areas of mathematics and symbolic logic from philosophy.[41]

The autonomy of applied philosophy. According to Hansson (2008a, §8), the philosophy of science is not an application of epistemology in the way that applied mathematics is an application of pure mathematics. Rather, philosophers of science develop their own theories which are related to—but not derivable from—epistemology.

Hansson's point becomes obvious when considering that epistemology is usually more general than philosophy of science, in that it aims to explicate terms such as 'belief', 'justification', etc. (and, more generally, form concepts) for as many contexts as possible. Philosophy of science, on the other hand, explicates concepts within the context of scientific theories and scientific practice. Given the different domains, it is to be expected that the explicata differ: Not all contexts that are relevant in epistemology are relevant in the philosophy of science, and some contexts that are very important in the philosophy of science only play a minor role in general epistemology.

Contrary to Hansson's suggestion, there is an analogy between the philosophy of science and applied mathematics, namely when new mathematical concepts are developed for a specific application. A famous example is Dirac's "δ-function", which in fact cannot be treated as a function and led to the development of the theory of distributions. The perception of a disanalogy between the philosophy of science and applied mathematics may rest on a failure to distinguish between

[41] This is not a historical point: Ancient Greek geometry was already precise enough to be pursued without the need for concept formation, but it was still at least to a great extent pursued in connection with philosophy.

rigorous derivations and concept formation. If some concepts apply to a great variety of contexts, then any derivations that involve only these concepts will apply to each of these contexts as well. On the other hand, there is no reason to assume that the concept most fruitful for a great variety of contexts is also the most fruitful for each specific one.

Philosophical truths are not eternal. Hansson (2008a, §6) suggests that many philosophers see their discipline as independent of empirical, synthetic results, which have no relevance in the philosopher's realm of eternal, analytic truths. But, Hansson contends, Quine (1963) has shown that there is no uncontroversial line between analytic and synthetic statements, and philosophers who ignore empirical results (e. g., relativity theory) to arrive at claims that are "analytically true" (e. g., about time) are often just "demonstrably wrong".

It is understandable that Hansson considers the analytic-synthetic distinction to be a problem for the connection between empirical science and philosophy, since analytically true sentences cannot be demonstrably wrong if such a demonstration would be empirical. This is because the very definition of an analytic truth (as discussed here) is that it has no empirical import. Thus the thorough critiques by Mates (1951), Martin (1952), Kemeny (1963), Creath (1991), Stein (1992), George (2000), and Loomis (2006) of Quine's attack on the analytic-synthetic distinction may seem to pose a problem for Hansson as well. However, the relevance of empirical results for philosophical work can be established without questioning the distinction. As my discussion of the analytic component of postulates in general has shown, meaning postulates are often chosen to be true because of empirical assumptions, and in the case of the Carnap sentence, empirical results are the final arbiter about the relevance of the meaning postulates (see §2.8.2). In other words, analytic truths cannot be demonstrably wrong, but they can be demonstrably irrelevant.

One might reinterpret Hansson's claim to state that a language chosen without reliance on empirical results is very unlikely to accommodate them better than a language that was chosen with these results in mind. This claim is almost trivially true and suggests that those philosophers who, for instance, do not consider relativity theory when explicating 'time' do not intend their explicatum to accommodate all the empirical results that relativity theory is meant to accommodate.

Philosophy *of* or *with* a discipline. With respect to philosophical endeavors related to the sciences, Hansson (2008a, §7) distinguishes between the philosophy *of* science and philosophy *with* science. He states that philosophers of economics, for example, use the "tools of philosophy" to investigate how economists reason, so that philosophers relate to economists very much like social scientists to their objects of study. Philosophy with economics, on the other hand, consists of research conducted in collaboration with economists (for example, on the development of new representations of human beliefs, preferences, and norms).

Clearly, philosophers and scientists can work together when developing the same concepts and rigorously deriving results involving them. This is Hansson's idea of "philosophy with the sciences", and it is to be expected in areas where the foundational concepts have not yet been developed fully, for then the construction of new representations is a major element of research, and rigorous derivations and empirical research cannot yet be pursued independently. Of course, it is also possible to suggest improvements of concepts that have already been explicated in the sciences and used with much success. Such improvements are more likely to occur when dealing with problems that are not at the center of scientific research and thus may not have carried much weight in previous explications.

There is some unclarity in Hansson's description of "philosophy of science", for he does not specify "the tools of philosophy". But, given his comparison of philosophy of science to social science, Hansson probably has the naturalizable, non-normative part of ordinary language philosophy in mind, as its application would lead to descriptions of the rules of language use in the sciences.

* * *

A closer look at the "tools" of artificial language philosophy, i. e. rigorous derivations and language choice, clarifies the possibilities for a "philosophy of science" in Hansson's terminology, and also reveals several relations between science and philosophy of science that are missing from Hansson's list. Within artificial language philosophy, Hansson's philosophy of science is probably best captured as the explication of concepts that are not explicated in the sciences themselves but still used in those contexts, including such general scientific concepts as *explanation* and *probability*. These concepts may not connect very well to others, and indeed may be confusing in certain contexts. For example, according to the fine-tuning problem in physics, on the commonly used scales, the range of values under which the universal constants of physics allow life to exist are small. Therefore the existence of life is very improbable, and thus life is in need of an explanation (cf. Ratzsch 2009, §4.1). However, the actual usage of 'explanation', 'probability', and 'scale' in the sciences probably does not allow for these inferences. The first step towards solving this problem is an explication of the three terms, and this explication would fall within the domain of philosophy.[42] If the explicata still do not allow for these inferences, the explication of the terms is also the last step.

An example of a discipline-specific scientific concept that, once introduced, has been explicated more extensively in philosophy than in the respective science itself is the notion of *gene* (cf. Waters 1994). In this case, however, Hansson's distinction between philosophy *of* and *with* science becomes very blurry indeed. The explication of such a concept is farther removed from Hansson's philosophy with science when the explication's goals are different from those in the scientific

[42] Note that it may be necessary to explicate the concepts differently for different scientific domains.

discipline, leading to different conditions of adequacy and evaluations of fruitfulness. One instance of this is the explication of an initially discipline-specific concept for simultaneous use in other disciplines. *Life*, for example, is comparably well-explicated within biology, but not for simultaneous use in robotics. A philosophical inquiry into the implications of artificial life may therefore have to develop its own explicatum. *Life* may also have to be explicated differently when used in ethical theories, and thus some scientific concepts may need to be explicated for simultaneous use in a non-scientific domain. Waters (2004, §6) discusses the conditions of adequacy for such interdisciplinary explications in depth. Interdisciplinary explications are also desirable if the same term is already used for two slightly different concepts in two different fields, which can lead to fruitful interactions in some circumstances, but also abject confusion in others. 'Information' is a paradigmatic example.

Finally, there are concepts that are not used in the sciences at all, but whose explications must take scientific results into account. *Personal identity* or *free will* may not occur (centrally) in the scientific literature, but for many contexts, their explications will have to take into account scientific results about, for example, the functioning of the brain and the predictability of individual behavior.

2.12 An application: Idealization and abstraction

The preceding discussion has been fairly general. In this section, I explicate the notions of 'idealization' and 'abstraction', which will also be applied in the remainder of the book.[43] Since the concept of *entailment* in predicate logic will occur in the definientia of the explicata, the language in which the explication takes place is the metalanguage of predicate logic. The cost of formalizing this metalanguage would far exceed the benefits for the example, and thus I will rely on an informal version of the formal semantics presented in §2.8. For instance, I will formalize neither the conditions of adequacy, nor the eventual definitions of 'idealization' and 'abstraction'.

2.12.1 Preliminaries

There are many suggested elucidations of 'idealization' and 'abstraction', but most have been developed solely with the goal of capturing the actual usage of these terms in scientific discourse (cf. Jones 2005). Thus few of the elucidations are explications in the classical sense. Even fewer of the elucidations have been developed to be inferentially relevant—that is, to distinguish between valid and invalid inferences—and thus to relate to questions of justification. Of course,

[43] Parts of this section have been presented under the title "Justifying idealization by abstraction" at the SOPHA 2009, Université de Genève, Switzerland, September 4, 2009. I thank the audience for helpful discussions.

matters of justification have been discussed in connection with abstraction and idealization, but only *using* the terms, not as a guide to explicating them. I will develop explications of 'idealization' and 'abstraction' that are connected to the concept of inference in a simple way. My starting points will be generally accepted uses of the two terms, but whenever individual uses are vague, incompatible, or unhelpful, I will choose those that are inferentially relevant.

I will assume that 'abstraction' and 'idealization' denote relations between descriptions, that is, sets of sentences in a vocabulary $\mathcal{V}$. Since descriptions do not have to be true, this allows me to be silent about questions of realism or antirealism. Being so noncommittal befits an explication because it should yield an explicatum that, like any meaning postulate, gives meanings to terms without being creative; in other words, the explicatum should not entail statements about any terms other than the ones being explicated (though it may *suggest* specific explications of other terms).

Idealizations are widely accepted to be special kinds of distortions, and abstractions are sometimes seen as special kinds of omissions. Therefore, I will first give uncontentious explications of 'distortion' and 'omission' and analyze their relationship. These explications then suggest explicata for 'idealization' and 'abstraction' that are applicable to the theory of measurement and questions about the role of mathematics in the natural sciences.

2.12.2 Distortion and omission

To *distort* a description Θ is to give a description Δ that is incompatible with Θ. Incompatibility is always relative to a system of inference. For example, 'This passport is red' is logically but not analytically compatible with 'This passport is blue', given the meanings of 'blue' and 'red'. 'This passport is red' is analytically compatible with 'This is a US passport', but incompatible given my knowledge that US passports are blue. In order to capture these different systems of inference, I will assume in the following that an inference is a logical entailment ('$\vDash$') given some fixed set of background assumptions Λ. Analytic inference has been described by Carnap (1952) using meaning postulates as background assumptions. By the usual definition of incompatibility, a description Δ *distorts* a description Θ if and only if $\Theta \cup \Delta \cup \Lambda \vDash \bot$, where '$\bot$' is some inconsistent sentence. If $\Theta \cup \Lambda \vDash \varphi$ and $\Delta \cup \Lambda \vDash \neg\varphi$, I will call φ a consequence of Θ that is distorted by Δ given Λ.

An equally simple explication can be given for 'omission': A description Ω *omits* from a description Θ if and only if $\Theta \cup \Lambda \vDash \Omega$ and $\Omega \cup \Lambda \nvDash \Theta$. In the usual logical parlance, Θ is, given the background assumptions, stronger than Ω. If $\Theta \cup \Lambda \vDash \varphi$, $\Omega \cup \Lambda \nvDash \varphi$, and Ω omits from Θ, I will say that Ω *omits* φ from Θ. As long as $\Theta \cup \Lambda$ is consistent, no omission Ω is a distortion (and thus no distortion an omission), for otherwise $\Theta \cup \Lambda \vDash \Omega$ and $\Theta \cup \Lambda \vDash \Theta \cup \Lambda \cup \Omega \vDash \bot$.

While there are often pragmatic reasons against omitting anything from a description, there is also a reason for it: The omitting description Ω of a description

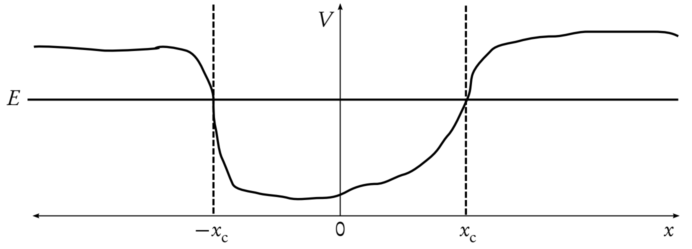

Figure 2.1: A particle with energy E in an irregular potential well.

Θ is more robust than the description Θ itself, because, given the background assumptions, there are more circumstances in which Ω is true than Θ. If, for example, Θ has consequences that may be or may become false, a description that omits these consequences from Θ is more likely to be and to stay true. But in cases where the possibly false consequences are important (e. g. they describe a phenomenon of interest), omitting these consequences is not an option.

The robustness that results from an omission can be used to justify a distortion. Not every consequence of a distortion Δ is incompatible with a description Θ, that is, there is some description Ω that can be inferred from Δ *and* from Θ. Since Δ and Θ are incompatible, neither can be inferred from Ω, and thus Ω omits from both. Now, if the initial description Θ is true, then its distortion Δ is false. And since Ω follows from Θ, Ω is true, and thus, is a true description that can be inferred from the false description Δ. If Θ's consequences of interest can be inferred from Ω, then Δ only distorts the unimportant consequences of Θ. In this sense, Ω justifies Δ.

This relation between distortion and omission is hardly surprising, but it deserves notice because it is a common technique of justification in both the sciences and philosophy. Take, as an example, the four principles of biomedical ethics suggested by Beauchamp and Childress (2008): respect for autonomy, nonmaleficence, beneficence, and justice. The authors infer these positions from two different ethical systems, deontological ethics and utilitarianism. Having established these principles, they infer derivative rules from them (veracity, privacy, confidentiality, and fidelity). The four principles omit from each of the underlying ethical systems, and hence are more robust than both. Clearly, utilitarianism is a distortion of deontological ethics, and vice versa. Assuming Beauchamp and Childress's inferences are correct, the four principles form a common omitting description. The arguments for the derivative rules can now proceed from the four principles, and those consequences of deontological ethics and utilitarianism that the four principles omit are unimportant for these arguments.

A more precise example can be found in physics, where it is common to employ distortions to simplify calculations. Take a quantum mechanical particle with

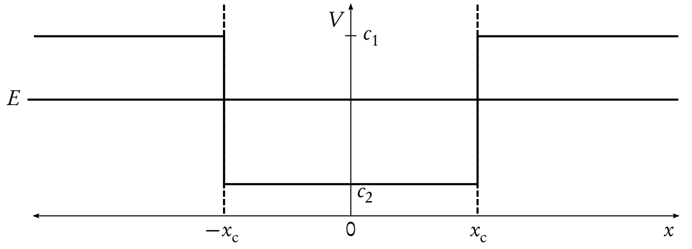

Figure 2.2: A particle with energy E in a rectangular potential well.

energy E in a one-dimensional potential well ($V(x) > E$ for all $|x| > x_c$, $V(x) \leq E$ for all $|x| \leq x_c$). The well may have an irregular shape, making calculations of the wave function difficult (figure 2.1). A rectangular distortion of the well ($V'(x) = c_1 > E$ for all $|x| > x_c$, $V'(x) = c_2 < E$ for all $|x| \leq x_c$, figure 2.2) simplifies the calculations greatly, but it is not obvious what the calculation actually shows—after all, the real well is irregular, not rectangular. In this situation, the best strategy is to find a description that omits both from the description of the irregular well *and* from the description of the rectangular well. Both descriptions are of a well, that is, a system in which only one interval ($[-x_c, x_c]$) is classically accessible to the particle. And this is already enough to use the stationary part of the Schrödinger equation, $\psi''(x)/\psi(x) = 2m[V(x) - E]/\hbar^2$, which in this case is assumed to be in the background assumptions Λ, to infer that the wave function bends towards the x-axis (i. e., is convex) in the classically accessible interval and away from it (i. e., is concave) in the classically inaccessible areas (figure 2.3). From this, one can infer that the wave function never oscillates in the classically forbidden areas, but can oscillate in the allowed interval (Schwabl 1995, §3.6). This pattern of the wave function can be correctly inferred from either description of the well, because both the initial description and its distortion have a common omitting description. If one is interested in the pattern, the omitting description justifies the distortion because it allows the intended inference, that of the description of the pattern.

It is also of note that the description of the pattern can be inferred without using a limit procedure. Batterman (2002) has argued that distortions reached by limit procedures play a central role in inferring descriptions of general, robust patterns in phenomena. However, the quantum well gives an example of an extremely general, robust pattern whose description can be inferred from an omitting description without such a limit procedure. The above analysis suggests that distortions, with or without a limit procedure, do not justify their consequences—what matters is the existence of a description that omits the distorted consequences.

Of course, limit procedures may help in finding an omitting description, and this also holds for distortions in general (cf. Redhead 1980, §5.i). Furthermore,

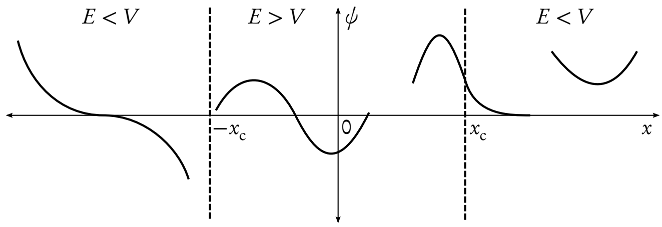

Figure 2.3: Possible shapes of the wave function for a particle with energy E in *any* potential well.

distortions may be used because the relevant omission is often unknown, so that distortions will be made in the *hopes* of not distorting any of the description's important consequences. If a variety of different distortions have a common omission, this can be a good inductive reason to believe the omission to be true (Redhead 1980, §5.iii). Therefore, distortions are an often expedient or even *practically* indispensable method for describing phenomena.

2.12.3 Idealization and abstraction

Since an idealization is commonly accepted to be a kind of distortion (Nowak 2000; O'Neill 1988; Batterman 2002; M. Jones 2005; N. Jones 2008), it is useful to determine *what* kind of distortion an idealization should be. Unfortunately, the notion of abstraction is far less clear, and many of the individual explications are incompatible (for an overview, see M. Jones 2005, introduction). However, it, too, can be usefully related to distortion and omission.

To start with, there are at least three pre-analytic notions of abstraction. N. Jones (2008), for example, discusses the claim that abstractions are nothing but idealizations. The claim may seem apt when looking at the functional dependencies between physical quantities: If, say, the frequency $\nu = f(l, w)$ of a pendulum depends on the length of its string l and the amount of friction w, then intuitively, it seems that abstracting from the friction amounts to making the idealization $\nu = f'(l)$. In what follows, however, I will not identify abstraction with idealization because there are many cases in which they are distinct, and more importantly, because there are alternative explications of the terms that distinguish between them and are inferentially relevant. Carnap (1939, §24) uses a second notion of abstraction when he states:

> We find among the concepts of physics [...] differences of abstractness. Some are more elementary than others, in the sense that we can apply them in concrete cases on the basis of observations in a more direct way than others.

Here, the farther removed a concept is from observation, the more abstract it is; statements are more abstract, the more abstract concepts they use.[44] This notion of abstraction is also not the one I wish to explicate because it is not clear how this distinction can be inferentially relevant.

In the same monograph, Carnap (1939, §2) uses another notion of abstraction when he discusses the components of a theory of language:

> [Pragmatics] take[s] into considerations the first component, whether it be alone or in combination with the other components. [Semantic] inquiries are made in abstraction from the speaker and deal only with the expressions of the language and their relations to their designata. [...] [Syntax] abstract[s] even from the designata and restrict[s] the investigation to formal properties [...] of the expressions and relations among them.

In this notion, an abstraction is a kind of omission, as is also assumed by Nowak (2000), O'Neill (1988), and Jones (2005). Carnap's use of the term further suggests the kind of omission. Syntax abstracts by omitting all and only those consequences of a description that are not about the expressions of a language. Analogously, semantics omits all and only those consequences that are not about expressions or designata. This is the notion of abstraction that I will explicate, for it will turn out that it can lead to an inferentially relevant explicatum.

First, let those consequences of a description Θ whose terms are from a subset $\mathscr{S} \subseteq \mathscr{V}$ of the vocabulary be the *$\mathscr{S}$-consequences* $\Theta|_{\mathscr{S}}$ of Θ ($\Theta|_{\mathscr{S}} := \{\sigma \mid \Theta \vDash \sigma$ and σ is a $\mathscr{S}$-sentence$\}$). Then call a description A an *abstraction* of a description Θ *in terms of* $\mathscr{S}$ (*in $\mathscr{S}$-terms*) if and only if A omits all and only those sentences that cannot be inferred from Θ's $\mathscr{S}$-consequences, that is, if and only if $A \cup \Lambda \mathrel{\vDash\!\dashv} \Theta|_{\mathscr{S}} \cup \Lambda$. It is sometimes more convenient to specify the terms that are *not* in $\mathscr{S}$, and in these cases, one can speak of *abstracting from* $\complement\mathscr{S} := \mathscr{V} - \mathscr{S}$.

Since inferences and omissions are relative to background knowledge, an abstraction can contain or consist entirely of terms that do not occur in the original description, if these terms occur in the background assumptions. In the quantum well example, this change of terminology occurs when the well is described in terms of classically accessible and inaccessible areas. The abstraction allows one to infer the description of the wave function that is of interest, namely that it is convex in accessible areas and concave in inaccessible ones.

If abstractions are supposed to justify idealizations in the way that omissions justify distortions, the explication of 'idealization' is now straightforward, because a description and its idealization should have a common abstraction. This suggests the following: If a description I *idealizes* a description Θ *in the vocabulary* $\complement\mathscr{S}$, then it distorts only those consequences of Θ that contain terms of $\complement\mathscr{S}$. In other

[44] Arguably, this meaning of 'abstract' is a comparative version of the epistemological interpretation of abstract objects (cf. Rosen 2009).

words, $\Theta \cup I \cup \Lambda \vDash \bot$ and $I|_{\mathscr{S}} \cup \Lambda \mathrel{\Dashv\!\vDash} \Theta|_{\mathscr{S}} \cup \Lambda$.

I do not suggest that this necessary criterion for idealization be taken as sufficient, because idealizations are typically assumed to be better than the original descriptions (e. g. by being mathematically more tractable). But this concession to the current use of the term does not hamper the use of abstraction in justifying idealization, because any idealization in the vocabulary $\complement\mathscr{S}$ must leave all $\mathscr{S}$-consequences unchanged, and thus any abstraction from $\complement\mathscr{S}$ omits the distorted consequences.

Abstractions can play a more important role in justifying idealizations than omissions can play in justifying distortions, because the relation between omission and distortion can easily be trivialized. Any two descriptions with the same (possibly tautological) consequence have that consequence as a common omitting description. In the interesting cases, there is a stronger common omitting description (e. g. the quantum well) that allows the inference of the description of interest (e. g. the oscillations of the wave function). The relation between idealization and abstraction is not in as much danger of trivialization because an idealization in $\complement\mathscr{S}$ does not distort *any* consequence of the original description in $\mathscr{S}$. When the description of interest can be given in $\mathscr{S}$ (which must therefore be non-empty), this ensures that the idealization is justified, because it has a non-tautological abstraction in $\mathscr{S}$-terms from which the description can be inferred.

Since inferences and therefore omissions are relative to background knowledge, an abstraction can contain or be entirely made up of terms that do not occur in the original description. Of course, these new terms do occur in the background assumptions, for example in definitions of the terms of the original description. This change of terminology occurs in the four principles of Beauchamp and Childress (2008). The utilitarian description of a situation contains a description of the situation in non-moral terms and in terms of its utility. It does not contain a reference to autonomy, nonmaleficence, beneficence, or justice. The description of the same situation with the help of the four principles does not contain a reference to utility, because it abstracts from it. Of course, the four principles have consequences in terms of utility, but these consequences are omissions from utilitarianism.

* * *

There are many methods of omitting from a description beyond abstracting from some subvocabulary. An especially important one consists, to put the matter very vaguely, in adding ‘approximately’ in front of a set of sentences Θ. This can be formalized semantically by moving from the actual structure $\mathfrak{M} \vDash \Theta$ to a vagueness set $\mathbf{M}$ with $\mathfrak{M} \in \mathbf{M}$, so that $\mathfrak{M}$ is a precisification of $\mathbf{M}$. This means making the interpretation of the terms in Θ vague by moving to multiple models instead of one and employing the subvaluationist account of truth.

Syntactically, an approximation means moving from Θ to $\Theta' \approx \Theta$, where the latter expression is empty until '$\approx$' is defined. There are many ways to do so; in probably the simplest case, Θ contains a quantified identity statement $\forall x\, f\, x = g\, x$, and Θ' contains in its stead an approximation statement as known from mathematics, say, that $|f(x) - g(x)| < \varepsilon$ for all or almost all x in some interval and for some positive ε. Of course, this approximation statement would have to be appropriately formalized. I discuss another possibility for arriving at a definition in 11.4.

2.12.4 Connections to other accounts

$\mathscr{S}$ is typically not chosen arbitrarily, because it should be possible to describe the important consequences of the original description in $\mathscr{S}$-terms. In that sense, the terms in $\complement\mathscr{S}$ are unimportant, and the $\mathscr{S}$-terms may refer to what Nowak (2000, §2) calls "essential factors". It is also plausible that an "abstractions", as described by Cartwright (1989, ch. 5), is (in the terminology suggested here) an idealization in which the $\mathscr{S}$-terms refer to the causal factors of a situation. The explications suggested here may therefore help to clarify these two earlier approaches to idealization and abstraction.

The explicata for 'abstraction' and 'idealization' are easily related to the formalism of Ramsey sentences. $\mathsf{R}_{\mathscr{S}}(\Theta)$ entails just those consequences of Θ that contain only terms from $\mathscr{S}$. Thus, if Λ and the descriptions Θ and Θ' are finite, and $\{\mathsf{R}_{\mathscr{S}}(\Theta \cup \Lambda)\} \cup \Lambda \mathrel{\vDash\!\!\!\dashv} \{\mathsf{R}_{\mathscr{S}}(\Theta' \cup \Lambda)\} \cup \Lambda$, then Θ' is an idealization of Θ if and only if the two descriptions are incompatible. Furthermore, $\mathsf{R}_{\mathscr{S}}(\Theta \cup \Lambda)$ is an abstraction of Θ in terms of $\mathscr{S}$. If $\mathscr{S}$ contains all and only basic terms, $\mathsf{R}_{\mathscr{S}}(\Theta \cup \Lambda)$ can be taken as the empirical content of $\Theta \cup \Lambda$; this shows very clearly that the explicatum for 'abstract' suggested here does not explicate the notion of being far removed from observation, and thus can connect usefully to the theory of measurement.

Measurement and meaningful sentences

According to Suppes (1959, 131),

> [a]n empirical hypothesis, or any statement in fact, which uses numerical quantities is empirically meaningful only if its truth value is invariant under the appropriate transformations of the numerical quantities involved.

A numerical quantity (measurement) is represented as a function from physical objects to numbers; an appropriate transformation leads only from one adequate function to another (Suppes 1959, 132). An adequate function must fulfill conditions specific to the measurement it represents. Suppes (1959, 135) states the conditions for a function m representing mass measurement as

$$\begin{aligned}\Lambda := \{&\forall x \forall y\, [(Px \wedge Py) \rightarrow (x \precsim y \leftrightarrow mx \leq my)],\\ &\forall x \forall y\, [(Px \wedge Py) \rightarrow (m(x * y) = mx + my)]\}\,,\end{aligned}$$

where '$\precsim$' stands for 'is at most as heavy as', '$*$' stands for physical combination, and P represents physical objects. Suppes (1959, 135) notes that "the functional composition of any similarity transformation φ with the function m yields a function $\varphi \circ m$ which also satisfies" Λ. Therefore, Suppes (1959, 138) suggests that a formula S is empirically meaningful "if and only if S is satisfied in a model $\mathfrak{M}$ [...] when and only when it is satisfied in every model [...] related to $\mathfrak{M}$ by a similarity transformation." This definition relies on similarity conditions only because they lead from one model of Λ to another such model, so more generally, any transformation mapping all the functions that adequately represent the measurement in question to other such functions should be appropriate (cf. Przełęcki 1974a).

To see the connection to abstraction and idealization, note that appropriate transformations do not change the interpretations of any statements involving only empirical terms (in this case $\{\precsim, *\} =: \mathscr{S}$, but generalizations are straightforward). Therefore the statements containing only these terms are abstractions from statements containing non-empirical terms as well (here 'm', '$\leq$', and '$+$'). Any set Θ of sentences that is true in some model $\mathfrak{M}$ of Λ can be taken to make up a description. Any appropriate transformation yields another model $\mathfrak{M}'$ of Λ, and any set Θ' of sentences that is true in $\mathfrak{M}'$ but distorts Θ is an idealization of Θ. This is because appropriate transformations only change the interpretation of the functions that represent measurements (here 'm'), so that $\Theta \cup \Lambda | \mathscr{S} \mathrel{\rightleftharpoons} \Theta' \cup \Lambda | \mathscr{S}$. In effect, this means that all and only meaningful measurement sentences can be captured in the more abstract language of empirical terms, since meaningful sentences are robust under idealization. More precisely, from any set of empirical sentences that, together with Λ, determines its models up to isomorphism, the truth or falsity of each meaningful sentence can be inferred (see §6.5). To take an example from Suppes (1959, 134), $m(o_1) = 4$ is not meaningful, but $\varphi \mathrel{\rightleftharpoons} m(o_2) \leq m(o_1)$ is meaningful because $\{o_2 \precsim o_1\} \cup \Lambda \models \varphi$.

Abstraction and explanation

In defense of the theoretical indispensability of mathematics, Pincock (2007, 263) uses Euler's solution to the Königsberg bridge problem to argue that

> when a scientist accepts a mathematical statement like "$C(S) = 40$" ["The temperature of S is 40°C"] or "The bridge system forms a non-Eulerian graph" there is an implicit adjudication between the mathematical properties that the scientist believes are appropriate to

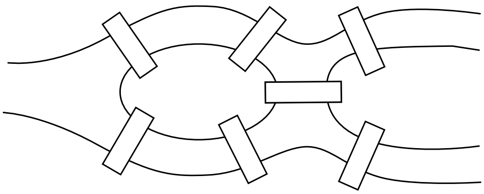

Figure 2.4: The bridges of Königsberg in 1735 (adapted from Pincock 2007, figure 1).

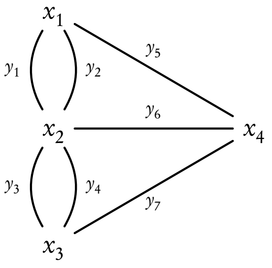

Figure 2.5: The graph for Königsberg. Labels are added for convenient comparison with the variables in formula (2.39).

> ascribe to the physical system and those that are deemed inappropriate.

From this he concludes that "mathematics allows us to make claims about higher-order or large-scale features of physical systems while remaining neutral about the basic or micro-scale features of such systems" (Pincock 2007, 255). In effect, then, Pincock argues that a mathematical statement accepted to be a description of a physical system entails some sentences that are considered to be true of the system, and some sentences that may be false. The latter sentences are silently ignored, and in this way, mathematics allows one to remain neutral with respect to some aspects of the physical system.

I will argue that in Pincock's example of the bridges of Königsberg, the possibility of remaining neutral about some aspects of a physical system has nothing to do with mathematics—and everything to do with abstraction. This conclusion also elucidates an otherwise puzzling aspect of the first quotation: The statement '$C(S) = 40$' seems to be *less*, not more, neutral than a non-mathematical description of temperature measurement, which should also be compatible with the Kelvin scale, for instance.

In 1735, Leonhard Euler proved that it is impossible to cross every bridge of Königsberg (see figure 2.4) exactly once and end up at the starting point of the walk. He identified the banks and islands of Königsberg with vertices and the bridges with the edges of an undirected graph (figure 2.5) and showed that this graph cannot be traced in one line that starts and ends at the same vertex. Pincock claims that this explanation of the impossibility (the explanandum) exemplifies the way in which mathematics allows us to stay neutral about the microstructures of physical systems, because the proof does not rely on the microstructure of Königsberg, and so does not "fail if the microphysics of the bridges [is] altered"

(Pincock 2007, 260).

One assumption in Pincock's argument is that the inference of a sentence φ fails to explain whatever φ describes if its premises are false, another is that the example shows that mathematics can be used to arrive at premises that stay true under microphysical changes (i. e., are robust). To identify the robust premises in this case, it is important to note that graphs can be represented in different ways. For example, one can start with a directed graph, an ordered pair $\langle A, f : I \to \{\langle a, b\rangle \mid a, b \in A\}\rangle$ with $I \cap A = \varnothing$, where A is a set of vertices and f a family of directed edges. An undirected graph can then be defined as a directed graph in which for each directed edge, there is another directed edge in the opposite direction. Another option is to define the edges of an undirected graph as a multiset of two-element subsets of A or as a family of two-element subsets of A: $G = \langle A, f : I \to \{\{a, b\} \mid a, b \in A\}\rangle$ with $I \cap A = \varnothing$.[45]

The above representations are reductive explications of 'vertex' and 'edge', in that they define the two terms in set theoretic terms (Benacerraf 1965, III.B). To abstract from the set theoretic terms that occur in the reductive explication of D, define

$$\forall y[Ey \leftrightarrow y \in I] \tag{2.31}$$

$$\forall x[Vx \leftrightarrow x \in A] \tag{2.32}$$

$$\forall yxx'[Cyxx' \leftrightarrow f(y) = \{x, x'\}]\ . \tag{2.33}$$

Since edges and vertices are not individuated, there are no further terms in the vocabulary. Then the abstraction of the definition of G in terms of $\mathscr{S} = \{V, E, C\}$ is as follows:

$$\forall x[Ex \to \neg Vx] \tag{2.34}$$

$$\forall yx_1x_2(Cyx_1x_2 \leftrightarrow Cyx_2x_1) \tag{2.35}$$

$$\forall yx_1x_2x_3x_4[Cyx_1x_2 \wedge Cyx_3x_4 \to (x_1 = x_3 \wedge x_2 = x_4) \vee (x_2 = x_3 \wedge x_1 = x_4)] \tag{2.36}$$

$$\forall yx_1x_2(Cyx_1x_2 \to Ey \wedge Vx_1 \wedge Vx_2)\ . \tag{2.37}$$

These formulas express that f is a function (2.36) mapping edges to sets (2.35) of vertices (2.37), which are distinct from edges (2.34).

The description of Königsberg in the language of graph theory relies on the identification of the parts of Königsberg with vertices and edges:

$$\{\forall x(Vx \leftrightarrow \mathit{Bank}\,x \vee \mathit{Island}\,x), \forall x(Ex \leftrightarrow \mathit{Bridge}\,x), \\ \forall y \forall x_1 \forall x_2(Cyx_1x_2 \leftrightarrow \mathit{connects}\,yx_1x_2)\}\ . \tag{2.38}$$

[45]Some vertices of the graph (figure 2.5) are connected by multiple edges (i. e. the graph is not simple). So, *pace* Pincock (2007, 258), the edges cannot be represented by a set of two-element sets of vertices.

Here, $\ulcorner \mathit{Bank}\, x \urcorner$, $\ulcorner \mathit{Island}\, x \urcorner$, $\ulcorner \mathit{Bridge}\, x \urcorner$, and $\ulcorner \mathit{connects}\, y x_1 x_2 \urcorner$ stand for $\ulcorner x$ is a bank$\urcorner$, $\ulcorner x$ is an island$\urcorner$, $\ulcorner x$ is a bridge$\urcorner$, and $\ulcorner y$ connects x_1 and $x_2 \urcorner$, respectively. With these definitions, the axioms (2.34)–(2.37) of graph theory are true of Königsberg and can be included in Λ, and a full description of Königsberg entails

$$\begin{aligned} \kappa \mathrel{\rightleftharpoons} \exists x_1 \ldots x_4 \exists y_1 \ldots y_7 \Big[& \bigwedge_{1\le i<j\le 4} x_i \neq x_j \land \forall z \Big(V z \rightarrow \bigvee_{i=1}^{4} z = x_i \Big) \land \\ & \bigwedge_{1\le i<j\le 7} y_i \neq y_j \land \forall z \Big(E z \rightarrow \bigvee_{i=1}^{7} z = y_i \Big) \land \\ & C y_1 x_1 x_2 \land C y_2 x_1 x_2 \land C y_3 x_2 x_3 \land C y_4 x_2 x_3 \land \\ & C y_5 x_1 x_4 \land C y_6 x_2 x_4 \land C y_7 x_3 x_4 \Big] . \end{aligned} \tag{2.39}$$

κ is a *true* description of Königsberg that, together with the axioms of graph theory, is also complete.[46] As Pincock emphasizes, κ is also very robust because, as long as no bridge is torn down or newly built and the river does not change its course too much, it will remain true of Königsberg. However, there is no reason to assume that this has anything to do with mathematics, since any description of Königsberg in terms of banks, islands, bridges, and connections can also be true and robust as long as it abstracts from enough detail—witness "There are two banks and two islands, with two bridges connecting the first bank with the first island, two bridges connecting the first island with the second bank…".[47] On the other hand, a geometrical description of Königsberg that includes the shape of the riverbed is not particularly robust at all (under this aspect, figure 2.4 and Pincock's version of it have always been false).

It now holds that $\Lambda \cup \{\kappa\} \vDash \varrho$ with

$$\varrho \mathrel{\rightleftharpoons} \neg \exists x_1 \ldots \exists x_7 \exists y_1 \ldots \exists y_7 \Big(\bigwedge_{1\le i<j\le 7} y_i \neq y_j \land \bigwedge_{i=1}^{6} C y_i x_i x_{i+1} \land C y_7 x_7 x_1 \Big) ,$$

where ϱ is the translation of the explanandum with the help of (2.38). The explanation uses mathematics only if predicate logic is considered mathematics.

It can now be seen that Pincock's use of the bridges of Königsberg to argue that the statement $C(S) = 40$ allows neutrality with respect to the microstructure of the physical system confuses two distinct situations. κ is an abstraction from a description of Königsberg that includes river banks, bridges, etc., and it is true and robust without any implicit adjudication to ignore some sentences. In the case

[46] Pincock's mapping account of the application of mathematics is therefore easily applicable.

[47] Note that this description is, within the vagaries of ordinary language, even complete in its vocabulary.

of measuring temperature, however, there is an implicit adjudication, as in the case of mass measurement. In Suppes's terminology, only those statements that are invariant under the appropriate transformations are empirically meaningful and hence accepted. And they are meaningful not because they are given in mathematical terms, but because they can also be given in empirical terms, that is, they are entailed by a description that *abstracts from* the mathematical terms.[48] As for $m(o_1) = 4$, this does not hold for $C(S) = 40$. This is why "the bridge system and this particular graph [in figure 2.5] seem much more intimately connected than the system with a temperature and the number 40" (Pincock 2007, 259).

2.13 Some conclusions

The explicata for 'abstraction' and 'idealization' suggested here fulfill the standard requirements for a successful explication (Kuipers 2007, §1). They are similar to the explicanda, precise, and fairly simple. Most importantly, they are fruitful, in that they relate in clear ways to each other, as well as to concepts developed in the theory of measurement and the formalism of Ramsey sentences. Furthermore, they clarify the role of mathematics in the sciences. This is why my explications can be considered (to some extent) successful—not because they capture specific intuitions.

This independence from intuitions and previous use has, as argued, the consequence that explications are searches for conventions. This does not mean that they are arbitrary, but, quite to the contrary, that they are chosen according to expedience,[49] on their own or in relation to other concepts or empirical results. This influence of empirical results on concepts suggests that the distinction between *a priori* and *a posteriori* truths may not be very helpful, since the meaning postulates, while not forced upon us by the world, are still chosen because of the way the world works.

The relevance of *a posteriori* knowledge in the choice of concepts and the demand for fruitfulness also suggests an important role for the empirical investigation of science, both within the history of science and within the sociology or experimental philosophy of science. If the motivation for methodological naturalism is correct and philosophical concept formation would become more effective by learning from scientific concept formation, it is important to understand how science goes about forming concepts. And this, of course, requires empirical research.

However, this defense of naturalism in philosophy only goes so far. It does

[48] Pincock's mapping account is again easily applicable to this abstraction, even if '$C(S) = 40$' is an idealization (cf. Pincock 2007, 271–272).

[49] No doubt like many conventions; ISO 216 is a convention about paper formats (A4, A3 etc.), but there is no question that living without is to the severe detriment of one's quality of life. I cannot stress this strongly enough.

not extend to ontological naturalism, for instance. To the contrary, if the sciences indeed engage in artificial language methodology as I have argued, and if, *pace* Lewis, not all scientific terms can be explicitly defined in previously known terms, science is free to introduce new objects by convention.[50] But this means that there is no fixed set of ontological objects on which philosophical concepts and philosophical ontology can rest. What is more, if both philosophy and the sciences rely on artificial language methodology and in fact often interact in their development of concepts, the introduction of new objects by philosophical theories (as suggested, for example, by Church in the opening quote of §2.10) does not differ from the introduction of new objects by scientific theories, and may even lead to the acceptance of these new objects in the sciences. Then ontological naturalism becomes an ill-defined position.

[50] For a discussion of such "meaning postulates with $\mathscr{A}$-terms controlled by an existential quantifier", see Stopes-Roe (1958), Kokoszyńska (1964), Przełęcki (1969, §8.II), and Winnie (1967).

Chapter 3

The Received View in the philosophy of science

The Received View on scientific theories as developed by Rudolf Carnap, Carl Gustav Hempel, and Herbert Feigl, among others, was arguably "the epistemic heart" of logical empiricism (Suppe 2000).[1] In this view a scientific theory is formalized as a set of sentences (called *theoretical sentences*) of predicate logic that contain only logical or mathematical terms and the terms of the theory (*theoretical terms*). The theoretical terms are connected to terms that refer to observable properties (*observational terms*) through sets of *correspondence rules*, sentences that contain both theoretical and observational terms. The observational terms are given a semantic interpretation, which, through the correspondence rules and theoretical sentences, restricts the possible semantic interpretations of the theoretical terms (Feigl 1970, 5–6; Carnap 1939, §24). Thus, in the terminology of artificial language methodology, the observational terms are basic terms and the theoretical terms are auxiliary terms.

The formalization of scientific theories and the analysis of formalized theories have been very fruitful in the past and promise to stay fruitful in the future (see,

[1]Parts of this chapter have been presented under the title "Carnap's unchanging correspondence rules" at the *Second SIFA Graduate Conference* at the Cogito Research Centre, Bologna, Italy, on October 29, in 2009 and under the title "Two constants in Carnap's philosophy of science" at the *HOPOS 2010* conference at the Central European University, Budapest, Hungary, on June the 25, in 2010. Parts have been presented under the title "On a straw man in the philosophy of science—A defense of the Received View" at the *Salzburgiense Concilium Omnibus Philosophis Analyticis* at the Universität Salzburg, Austria, September 10, 2010 and at the *British Society for the Philosophy of Science Annual Meeting* at University College, Dublin, Ireland, on July 9, 2010. I thank the audiences for helpful discussions. A paper under the latter title is forthcoming in *HOPOS: The Journal of the International Society for the History of Philosophy of Science* (Lutz 2012). I thank two anonymous referees for their comments.

for example, Suppes 1968, Betti and de Jong 2008, Betti et al. 2009, Leitgeb 2009). However, the Received View (the name is due to Putnam 1962, 240), also called the Syntactic Approach (van Fraassen 1970), the Syntactic View (Wessels 1974, 215), the Standard Conception (Hempel 1970), or the Orthodox View (Feigl 1970), is currently not an accepted framework for formalization (cf. Suppe 2000, §1).[2] Accordingly, any analysis or argument that presupposes this view faces strong opposition. If the Received View is indeed as unmitigated a disaster as it seems to many, this state of affairs is just as it ought to be. On the other hand, if the Received View is a fruitful framework for the formalization of scientific theories, this state of affairs is unfortunate, not only because it deprives philosophers of science of a means for analyzing theories, but also because there are analyses that presuppose the Received View, and that, if correct, would provide deep and helpful insights into scientific theories and epistemological questions.

One such analysis is given by Hempel and Oppenheim (1948) and subsequent authors, who developed increasingly sophisticated and general explications of 'explanation' on the basis of the Received View. Their explications have now been succeeded by a host of mutually incompatible accounts, which (arguably) have not always reached the precision and breadth of the original explications. Similarly, the explication of 'reduction' by Nagel (1949, 1951) has not been improved upon so much as abandoned, and in many contemporary discussions, the term is not explicated at all but rather used in an intuitive way as a primitive concept. The criteria of empirical significance helpfully summarized and then rejected by Hempel (1965c) have never recovered their status, and the same holds for the explications of empirical content developed by Craig (1953) and Ramsey (1929). Building partly on Ramsey's results, Carnap (1952, 1958) explicated the notion of analyticity, a concept that is now often considered impossible to explicate, useless, or empty. Since it allows a clear analytic-synthetic distinction, it is especially this last explication that supports Suppe's stance on the importance of the Received View for logical empiricism. Therefore, if "being a logical empiricist really is not a live option for a twenty-first-century philosopher" (Richardson 2007), the perceived failure of the Received View certainly contributed to this state of affairs.

Such explications (let alone whole philosophical research programs) are generally not given up without good reason, and indeed there have been many severe criticisms of the Received View and of the whole project of logical empiricism. As Carus (2007, 6–7, footnote removed) diagnoses,

> most analytic philosophers have striven to distance themselves from logical empiricism. [...] In the wider intellectual world, meanwhile, the reaction against 'logical positivism' is even more pronounced.

[2]It is thus somewhat unfortunate that the name 'Received View' has stuck. On the other hand, it is by now such an obvious misnomer that it is much more easily recognized as a proper name than 'Syntactic View', which, given the importance of a semantic interpretation of the observational terms, is a misnomer as well.

> Despite recent historical interest in the movement, it is still regarded with almost universal disdain. It functions in the humanities and social sciences as a kind of 'other', against which almost anyone's own position may be defined or identified. The baleful influence of 'logical positivism' was felt so widely that it is now a recognised term of abuse in almost every field outside natural science. [...] This widespread rejection in turn exacerbates the embarrassment felt by analytic philosophers and redoubles their hurry to disavow any residual connection with the barbaric past.

Few have tried to defend logical empiricism. One, of course, is Carus (2007, 7) himself, who argues that many criticisms of logical empiricism "are seriously misguided. The 'logical empiricism' they reject was never propounded by any of its leading figures, whose actual doctrines have been largely ignored". However, like most other recent defenses (e. g. Maher 2007, Lutz 2009, 2011a, Justus 2011), his defense pertains only to "Carnap's larger programme of explication and pluralistic language engineering" (Carus 2007, 39). Most individual projects within logical empiricism, and specifically the Received View, remain undefended.[3] This seems to leave current projects with a close affinity to the Received View in a lurch. It can be argued that, for example, the partial structures approach (see §4.3) and most criteria of empirical significance (see part II) can be phrased within the Received View, approximative explanations are deductive-nomological (Pearce and Rantala 1985), asymptotic explanations provide "a friendly amendment to the deductive-nomological account" (Belot 2005, 144, n. 34), intertheoretic reductions are Nagelian (Dizadji-Bahmani et al. 2010), and Carnap has provided a useful explication of analyticity (Papineau 2009b). Thus, if the Received View is a failure, so much the worse for partial structures, criteria of empirical significance, asymptotic and approximative explanation, intertheoretic reduction, and analyticity.

In this chapter, I will attempt to rehabilitate the Received View as it emerges from the later writings of Carnap, Hempel, and (to a smaller extent) Feigl, and defend it against some of the major criticisms. My strategy will be that of Carus (2007, 7): Rather than discuss the criticisms' conceptual merits, I will focus on their presumptions about the Received View. Many of these, I argue, are contentious or wrong, making the criticisms attacks on a straw man. Systematic defenses of different aspects of the Received View follow in the next chapter.

It should be noted that some misconceptions of the Received View could be easily avoided by reference to its by now canonical exposition[4] by Suppe (1974a). For instance, Burgos (2007, 157), probably misled by the name 'Syntactic View', claims that according to the Received View, "[n]othing non-syntactic constitutes

[3] Carus (2007, 39, n. 61) lists some exceptions.

[4] Canonical within the philosophy of science (see, for example, Hughes 1989, 80; Ereshefsky 1991, 62; Morrison and Morgan 1999, 2, n. 1; Muller 2010, §2, n. 1), and sometimes even the history of philosophy of science (Mormann 1991, 64; Stern 2007, 307, n. 2; Uebel 2008, §3).

any theory. In particular, whatever is meant or referred to by a term or statement (i. e., the semantic aspect), it is not constitutive of theories". In contradistinction, Suppe (1974a, §II.E) correctly emphasizes that the Received View includes a semantic interpretation (cf. Carnap 1939, §24).

But some misconceptions are not forestalled by Suppe's exposition. Of these, I will take up the Received View's alleged demand for exhaustive axiomatizations (§3.4) and its perceived dismissive attitude towards models (§3.7). And even Suppe's generally careful exposition contains mistakes, of which I will discuss the claims that the Received View demands axiomatizations in first order logic (§3.3); that the bipartition of the vocabulary into theoretical and observational terms is unduly vague (§3.6.2) and leads to overly complex concepts (§3.6.3); that the correspondence rules connecting theoretical and observational vocabulary are considered part of the theory (§3.6.4), cannot take other theories into account (§3.6.5), and are overly simple (§3.6.6); and that the Received View is intended to make the meaning of the term 'theory' more precise (§3.8). The discussion of the last of these misconceptions in particular will suggest that the Received View did not fail, but was abandoned before it had been properly developed or, in fact, received.

3.1 The decline and fall of the Received View

The following is a historical introduction into the most striking feature of the Received View, the connection of theories with observations by way of correspondence rules between theoretical and observational terms. This history, beginning with *Der logische Aufbau der Welt* (Carnap 1928a, *Aufbau* from now on), is fairly well known, although I will emphasize aspects that are important in the sequel and have not been discussed much in the literature.[5]

With respect to the correspondence rules between theoretical and observational terms, Carnap (1963a, §9) himself describes the development of logical empiricism as a gradual liberalization in his "Intellectual Autobiography" ("Autobiography" from now on). Initially, every kind of knowledge "was supposed to be firmly supported" by the experiences as described by Wittgenstein's principle of verifiability, "which says that it is in principle possible to obtain either a definite verification or a definite refutation for any meaningful sentence" ("Autobiography", 57). In the *Aufbau*, §35, this principle is expressed in the concepts of reducibility and construction. Every meaningful sentence is supposed to be translatable into a sentence about experiences, and this means that "the concepts of science are explicitly definable on the basis of observation concepts" ("Autobiography", 59), which are thus the referents of the observational terms.

In "Testability and Meaning" (Carnap 1936, 1937, "Testability" from now on),

[5] A concordance between the short titles of the works and the references in the bibliography is given at the beginning of the bibliography on page 413.

Carnap relaxes this claim because he has come to the opinion that it is impossible to define disposition terms explicitly in non-dispositional observational terms. Instead, he suggests that new terms should be introduced by reduction sentences, that is, necessary and sufficient conditions (see §7.3).

In a short contribution to the *Unity of Science Forum*, Carnap summarizes "Testability" (Carnap 1938, fn. 1) and points to the *Foundations of Logic and Mathematics* (Carnap 1939, *L&M* from now on) for an elaboration of two methods of constructing a scientific language. One method starts with the observation terms as primitive and successively introduces theoretical terms through reduction sentences as in "Testability". The second method starts with theoretical terms, which are already related to each other through the postulates of a theory. These theoretical terms are taken as primitive and further theoretical terms are successively introduced to finally arrive at observational terms. In the second method, Carnap suggests, it may be possible to explicitly define all terms. In both methods, only the observational terms are interpreted. Carnap (1938, 34) claims:

> The first way is interesting from the point of view of empiricism because it allows a closer check-up with respect to the empirical character of the language of science. By beginning our construction at the bottom, we see more easily whether and how each term proposed for introduction is connected with possible observations.

With its reliance on reduction sentences, the first method is supposed to relate theoretical terms more easily to observational terms because, according to Carnap (1936, 447, thm. 7), it ensures that the theoretical terms can be reduced to observational terms.[6] It is easy to see that the conditions for theoretical terms given by the second method can be very complicated. For not all definitions of observational terms in theoretical terms lead to necessary or sufficient conditions for the theoretical terms. As an example, consider the definition of an observational term O by four theoretical terms T_1, T_2, T_3, T_4 in

$$\forall x \big(Ox \leftrightarrow [(T_1 x \wedge T_2 x) \vee (T_3 x \wedge T_4 x)]\big) . \tag{3.1}$$

The applicability of O to any object is neither necessary nor sufficient for the applicability of any of the four theoretical terms. Also, the second method does not demand that all theoretical terms occur non-trivially in the definition of an observational term. Some theoretical terms may only be related to other theoretical terms through the postulates of the theory, which are not further restricted.

In *L&M* itself, Carnap elaborates on the distinction between the two methods

[6] Considering the welter of definitions in "Testability", it is unsurprising that Carnap did not notice that this is essentially an empty claim. In §7.5, I show that (chain-)reducible terms can be completely unrestricted in their interpretation.

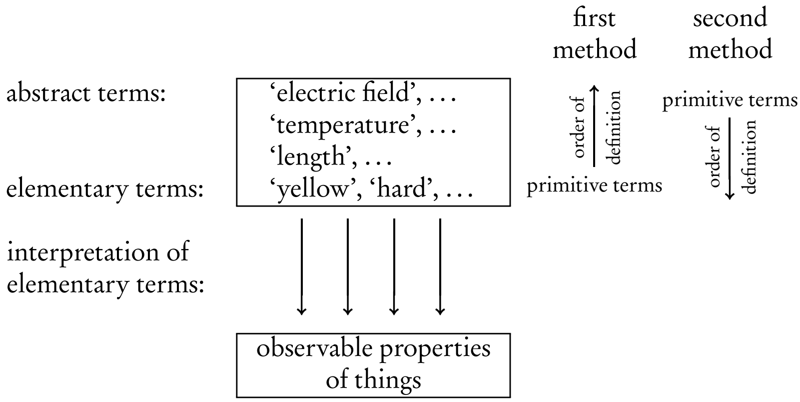

Figure 3.1: Carnap's diagram of two methods of giving an empirical interpretation to theoretical terms (Carnap 1939, 205).

for relating theoretical terms (he calls them "abstract") and observational ("elementary") terms (figure 3.1). Carnap (1939, 206) again notes that in the second method,

> so it seems at present, explicit definitions will do. More special laws, containing less abstract terms, are to be proved on the basis of the axioms. At least this is the direction in which physicists have been striving with remarkable success, especially in the past few decades. But at the present time, the method cannot yet be carried through in the pure form indicated. For many less abstract terms no definition on the basis of abstract terms alone is as yet known; hence those terms must also be taken as primitive. And many more laws, especially in biological fields, cannot yet be proved on the basis of laws in abstract terms only; hence those laws must also be taken as primitive.

As is clear from figure 3.1, *L&M* also marks a clear commitment to a semantic interpretation of the observational terms. The theoretical terms are interpreted solely through the correspondence rules as given by one of the two methods.

While the first method describes the observational import of abstract terms very clearly, scientists "are inclined to choose the second method" (Carnap 1939, 206, §24, emphasis removed). Restricting the first method to explicit definitions (which are special reduction sentences) would lead to "a science in sensationalistic form which Goethe in his polemic against Newton, as well as some positivists, seems to have had in mind", since every sentence could be translated into an observation sentence (see claim 7.1). The restriction is not feasible, however, because "it turns out—and this is an empirical fact, not a logical necessity—that it is not possible to arrive in this way at a powerful and efficacious system of laws", so that science always uses terms that are not explicitly definable in observational terms.

Carnap (1962, §5) gives one example in the *Logical Foundations of Probability* (*Probability* from now on): 'higher temperature than', as an explication of 'warmer than', should be such that it can be used instead of 'warmer than' itself. Carnap stresses that this requirement has to be fulfilled in most but not all cases, and adds:

> The converse of the requirement [...] would be this: the concept Temperature is to be such that, if x is not warmer than y (in the prescientific sense), then the temperature of x is not higher than that of y. It is important to realize that this is not required, not even "in most cases". When the difference between the temperatures of x and y is small, then, as a rule, we notice no difference in our heat sensations.

Clearly, then, 'temperature' cannot be defined in terms of 'warmer than'.

In two later papers, Carnap does not even mention the first method of interpreting theoretical terms, probably because he has come to the conclusion that this method is insufficient for the theoretical terms of science. In "The Methodological Character of Theoretical Concepts", Carnap (1956b, 53, "Theoretical concepts" from now on) writes:

> At the time of ["Testability"], I still believed that all scientific terms could be introduced as disposition terms on the basis of observation terms either by explicit definitions or by so-called reduction sentences. Today I think, in agreement with most empiricists, that the connection between the observation terms and the terms of theoretical science is much more weak than it was conceived [...] in my earlier formulations [...].

Accordingly, he goes on to develop a weaker criterion of empirical significance for terms and sentences that I will discuss in §7.4.2. In "Beobachtungssprache und theoretische Sprache", Carnap (1958, "Beobachtungssprache" from now on) develops a general method to distinguish between the analytic and the synthetic content of a theory. In "Theoretical concepts", 47, §V, Carnap states that the correspondence rules (C-rules)

> specify the relation R which [...] relates to an observable space-time region u, e. g., an observable event or thing, a class u' of coordinate quadruples which may be specified by intervals around the coordinate values x, y, z, t.
>
> On the basis of these C-rules for space-time designations, other C-rules are given for [theoretical terms]. These rules [...] hold for any space-time location. They will usually connect only very special kinds of value distributions of the theoretical magnitudes in question with an observable event. For example, a rule might refer to two material

> bodies [. . .] observable at locations u and v [. . .]. Another rule may connect the theoretical term "temperature" with the observable predicate "warmer than" in this way: "If u is warmer than v, then the temperature of u' is higher than that of v'."

Carnap here states that observable space-time regions can be assigned space-time coordinates, and that observational terms are assigned specific values of theoretical magnitudes. He also emphasizes the asymmetry between observational terms and theoretical magnitudes when noting that some, but not all, distributions of theoretical magnitudes are assigned an observable event.[7]

In "Beobachtungssprache", 243, Carnap repeats this construction and notes that there is an asymmetry not only between theoretical magnitudes and observational predicates, but also between space-time coordinates and space-time regions:

> We can [. . .] introduce a system of space-time coordinates by assigning an ordered quadruple of real numbers to each small body at any time. Through generalization, every such quadruple is then taken as a representative of a space-time point (that is, an unobservable, theoretical object). Then physical magnitudes (e. g. mass-density) are introduced, which have a value for each space-time point, for example a real number. In our system, a function of this kind is reconstructed by a function F of quadruples of numbers.[8]

This passage is a paraphrase of the one from "Theoretical concepts" above. The space-time region u is given by a small body, and the theoretical magnitudes are the physical magnitudes. In this passage, however, Carnap is even clearer about the impossibility of translating every theoretical statement into an observational statement. In line with his remarks about Goethe in *L&M*, Carnap (1958, 237) considers every term explicitly definable in observational terms also to be part of the observational vocabulary.

Carnap's introductory text *Philosophical Foundations of Physics: An Introduction to the Philosophy of Science* (Carnap 1966, *Introduction* from now on), based on a 1958 seminar but updated to include, for example, the results of "Beobachtungssprache", serves as the stopping point of my history of the decline of logical

[7]In this quote, Carnap blatantly flouts his own point that the relation between 'temperature' and 'warmer than' does not hold in all cases. This is excusable, though, because this correspondence rule only serves as an illustration for pedagogical purposes.

[8]"Wir können [. . .] ein raum-zeitliches Koordinatensystem einführen, in dem zunächst jedem kleinen Körper zu irgend einem Zeitpunkt ein geordnetes Quadrupel reeller Zahlen zugeordnet wird. Dann wird durch Verallgemeinerung jede solche Quadrupel als Vertreter eines Raum-Zeit-Punktes (also eines unbeobachtbaren, theoretischen Gegenstandes) genommen. Dann werden auch physikalische Grössen eingeführt (z. B. Massendichte), die für jeden Raum-Zeit-Punkt einen Wert haben, etwa eine reelle Zahl. Eine Funktion dieser Art wird in unserem System durch eine Funktion F von Quadrupeln reeller Zahlen nachkonstruiert."

empiricism. Carnap (1966, v) was urged by friends to write the book "because not many of [his] views on problems in the philosophy of science had been published". The book, which Carnap (1966, vi) believed "may also serve as a general introduction to the philosophy of science", provides an interesting glimpse of what he considered core issues in the philosophy of science.

The shorter parts of the *Introduction* concern laws, explanation, and probability, causality and determinism, and statistical laws. The two longest parts are about the structure of space and concept formation in the sciences. In both parts, Carnap stresses the conventional elements in, for example, the conceptualizations of time and length (Carnap 1966, ch. 8, 9) and the conceptualization of the structure of space (Carnap 1966, ch. 15). The third longest part deals with the structure of scientific theories, in which Carnap relies on the bipartition of the vocabulary into an observational and a theoretical part and the connection between the two via correspondence rules. There, he repeats the claim that when introducing observational concepts based on theoretical concepts (his second method of relating theories to observations), it may be possible to rely on definitions. When starting from observational concepts and introducing theoretical ones (the first method), on the other hand, a restriction to definitions is impossible (Carnap 1966, 234).

Carnap's second method for interpreting theoretical terms is a major point in Hempel's contribution to the *Encyclopedia of Unified Science*, his monograph *Fundamentals of Concept Formation in Empirical Sciences* (*Concept Formation* from now on), which is referenced in the *Introduction*. There, Hempel (1952, 684) stresses repeatedly the importance of

> *theoretical constructs*, i. e., the often highly abstract terms used in the advanced stages of scientific theory formation, such as 'mass, 'mass point' [...], 'volume', 'Carnot process' [...], 'proton', 'ψ-function', etc [...]. Terms of this kind are not introduced by definitions or reduction chains based on observables; in fact, they are not introduced by any piecemeal process of assigning meaning to them individually. Rather, the constructs used in a theory are introduced jointly, as it were, by setting up a theoretical system formulated in terms of them and by giving this system an experiential interpretation.

One way to give an experiential interpretation to a theoretical system consists in defining further concepts with the help of theoretical constructs and interpreting those concepts directly. This is Carnap's second method of giving empirical meaning to abstract terms described in *L&M* (figure 3.1); Hempel's treatment of 'mass' as such a theoretical term (Hempel 1952, §12) was even suggested by Carnap (Hempel 1952, 738, n. 72). Hempel (1952, §§6–7) considers the use of theoretical constructs another step in the liberalization of empiricism: The claim that all terms are explicitly definable in observational terms is the narrower thesis of empiricism (Hempel 1952, 676), while the claim that all terms are reducible to observational terms in the sense of "Testability" is the liberalized thesis of empiricism (Hempel 1952, 683). The need for reduction sentences shows that the narrower thesis is false, and the need for theoretical constructs shows that the liberalized thesis is false.

While Carnap saw value in the concept of correspondence rules until the end, Hempel (1970, §6) extended his critique of explicit definitions and reduction sentences to theoretical constructs and dismissed the whole concept of correspondence rules as misguided. Further criticisms of correspondence rules (Suppe 1972, §II) and the use of axiomatizations in general (Hempel 1970, §3, Suppe 1974a, §IV.f) contributed to the downfall of the Received View. Hempel (1970, 1974), for example, abandoned it completely. In its stead, the Semantic View on scientific theories has become an important framework for the reconstruction of theories, possibly the dominant one (Suppe 1989, 3, French and Ladyman 1999, 103). In this view, theories are not formalized as sets of postulates in predicate logic, but rather via a set theoretical predicate given in the metalanguage, as suggested by Suppes (1992, §2). The set theoretical predicate directly determines the model theoretic structures (Suppe 1989, 4, van Fraassen 1980, 44) (see also §4.1). Instead of relying on correspondence rules, theories are connected to observations through model theoretic notions such as embeddings (see §4.2.1).

3.2 Three early works of Carnap

Towards a historical defense of the Received View, I will discuss a somewhat neglected part of the history of the Received View, found in three of Carnap's works leading up to the *Aufbau*: his article "Über die Aufgabe der Physik und die Anwendung des Grundsatzes der Einfachstheit"[9] (Carnap 1923, "Aufgabe" from now on), his article "Dreidimensionalität des Raumes und Kausalität. Eine Untersuchung über den logischen Zusammenhang zweier Fiktionen"[10] (Carnap 1924, "Dreidimensionalität"), and the monograph *Physikalische Begriffsbildung*[11] (Carnap 1926, *Begriffsbildung*).

Both Suppe (1974a, 12) and Feigl (1970, 3) cite "Aufgabe" as Carnap's first exposition of the logical empiricist's view of scientific theories (Feigl also mentions Campbell 1920). Mormann (2007, 159, n. 13) criticizes this classification:

> Feigl once went so far to trace back the essentials of the Logical Empiricist account of empirical theories to an early (pre-Vienna) paper of Carnap that may well be classified as belonging to his neo-Kantian period [. . .]. This stance betrays, to put it mildly, that Feigl did not pay too much attention to the amendments that had taken place since then.

[9] "On the task of physics and the application of the principle of maximal simplicity."

[10] "Three-dimensionality of space and causality. An investigation of the logical connection between two fictions."

[11] *Physical Concept Formation.*

However, there are a number of reasons to consider "Aufgabe" a contribution to logical empiricism's view of theories. For one, Mormann's claim rests on the text's being written during Carnap's neo-Kantian period, but even if Carnap wrote the article while holding neo-Kantian views, this does not mean that they are manifest or relevant in the article. Carnap himself, for example, does not seem to think so. When discussing the influence of Kant's views on his own work in his autobiography, Carnap (1963a, 12) mentions his doctoral dissertation *Der Raum. Ein Beitrag zur Wissenschaftslehre*[12] (Carnap 1922, *Der Raum* from now on), specifically the chapter on intuitive space (Carnap 1963a, 4), but not "Aufgabe". As influences for "Aufgabe", Carnap rather lists Poincaré and Hugo Dingler, and at another point Hilbert, Poincaré, and Duhem (Carnap 1963a, §13 and 77–78).

Furthermore, Carnap (1963a, 15) considers "Aufgabe" to belong to the same "period" as *Begriffsbildung*. *Begriffsbildung* was written during his time in Vienna, and its main points were taken up by Hempel in *Concept Formation* and Carnap in the *Introduction*.

Perhaps most importantly, Carnap begins §13 of his autobiography, entitled "The Theoretical Language", with "Aufgabe". The section continues with *Begriffsbildung* and ends with "Theoretical concepts". This is relevant because §13 occurs in Part II of Carnap's autobiography, entitled "Philosophical Problems", where "[i]n each section, a certain problem or complex of problems [is being] dealt with". So Carnap himself considers "Aufgabe" a starting point of the development that led to one of his core articles on scientific theories. In the "Autobiography", 15, Carnap also summarizes "Aufgabe" in his later terminology:

> I imagined the ideal system of physics as consisting of three volumes: The first was to contain the basic physical laws, represented as a formal axiom system; the second to contain the phenomenal-physical dictionary, that is to say, the rules of correspondence between observable qualities and physical magnitudes; the third to contain descriptions of the physical state of the universe for two arbitrary time points. From these descriptions, together with the laws contained in the first volume, the state of the world for any other time-point would be deducible (Laplace's form of determinism), and from this result, with the help of the rules of correspondence, the qualities could be derived which are observable at any position in space and time. The distinction between the laws represented as formal axioms and the correlations to observables was resumed and further developed many years later in connection with the theoretical language.

Note that here, too, Carnap points out the continuity between "Aufgabe" and his later discussions of theoretical terms.

In "Aufgabe" itself, Carnap (1923, §I, all emphases removed) also introduces

[12] *Space. A Contribution to the Theory of Science.*

his main point with a reference to Poincaré and Dingler, but not Kant:

> The main thesis of conventionalism, marshaled by Poincaré and continued by Dingler, says that the construction of physics needs stipulations that are subject to our choice. [. . .] But the choice of these stipulations shall not be arbitrary, but rather be governed by specific principles, where in the end the principle of maximal simplicity decides.[13]

The laws of physics in the first volume can be chosen according to the principle of maximal simplicity, Carnap (1923, §I) states, because they are not determined by experience. Whether, for example, the world has a Euclidean or non-Euclidean geometry depends on the objects that are chosen to be rigid bodies, an example he repeats in the *Introduction* when discussing the conventional elements in the concept of length. Therefore, Carnap (1923, 97) concludes, the laws of physics are

> synthetic sentences a priori, but not exactly in the Kantian transcendental-critical sense. Because that would mean that they express necessary conditions for the object of experience, themselves determined by the forms of intuition and of thought. [. . .] But the content [of this volume] is actually often a matter of choice. [. . .]
>
> The choice is only to be made according to methodical principles, especially that of maximal simplicity. As a label for the first volume the term "hypothetico-deductive system" is therefore preferable to the Kantian term "a priori" [. . .].[14]

This is also the only passage in "Aufgabe" that contains an explicit reference to Kant, and it is far from an endorsement of the Kantian doctrine of the synthetic a priori as adhered to in *Der Raum*, but rather, in keeping with the "Autobiography", an endorsement of Poincaré's conventionalism (see also Carus 2007, 122).

The connections between "Aufgabe" and Carnap's later view on scientific theories are obvious. He makes a distinction that "cannot be emphasized too strongly"[15] between the domain of perception and the domain of physical theories

[13] "Die Hauptthese des von Poincaré aufgestellten und von Dingler weitergeführten Konventionalismus besagt, daß zum Aufbau der Physik Festsetzungen getroffen werden müssen, die unserer freien Wahl unterliegen. [. . .] Die Wahl dieser Festsetzungen soll aber nicht etwa willkürlich geschehen, sondern nach bestimmten methodischen Grundsätzen, wobei letzten Endes der Grundsatz der Einfachstheit die Entscheidung zu treffen hat."

[14] "synthetische Sätze a priori, allerdings nicht genau im Kantischen transzendental-kritischen Sinne. Denn das würde bedeuten, daß sie notwendige Bedingungen des Gegenstandes der Erfahrung ausdrücken, selbst bedingt durch die Formen der Anschauung und des Denkens. [. . .] In Wirklichkeit ist aber [der] Aufbau [des ersten Bandes] vielfach unserer Wahl anheimgestellt. [. . .]

Die Wahl ist nur nach methodischen Grundsätzen zu treffen, insbesondere dem der Einfachstheit. Zur Kennzeichnung des ersten Bandes is demnach dem Kantischen Apriori-Begriffe vorzuziehen der Begriff ‚hypothetisch-deduktives System' [. . .]."

[15] "kann gar nicht scharf genug betont werden"

("Aufgabe", 99). The connection between the two domains is given "in a way through a kind of dictionary that specifies which objects (elements, complexes, processes) in the second domain correspond to the ones of the first"[16]. This dictionary is just a metaphor for correspondence rules. Here are Carnap's examples of correspondence rules in "Aufgabe", 99–100:

> "To such a blue (designated, for example, according to the Ostwald color-system) corresponds a specific periodic movement of electrons (designated by the frequency)"[.] [. . .] "To such a pungent smell (chlorine smell, system of designation missing) corresponds a batch of electron complexes of a specific structure (Cl-atoms)"; "To such a heat experience (system of designation missing) corresponds a specific average kinetic energy of some amount of electron-complexes (atoms or molecules)."[17]

Compare this to two examples for correspondence rules given in the *Introduction*, 233:

> An example for such a rule is: "If there is an electromagnetic oscillation of a specified frequency, then there is a visible greenish-blue color of a certain hue". Here something observable is connected with a nonobservable microprocess.
>
> Another example is: "The temperature (measured by a thermometer, and therefore, an observable in the wider sense explained earlier) of a gas is proportional to the mean kinetic energy of its molecules". This rule connects a nonobservable in molecular theory, the kinetic energy of molecules, with an observable, the temperature of a gas.

In the *Introduction*, the Ostwald color system is substituted by the hue of the colors, and the missing system of designation for heat experiences is circumvented by using an observational term in the wider sense, the temperature according to a thermometer. The connection to heat experiences is given through the further correspondence rule specified in "Theoretical concepts", 48: The observational relation 'warmer than' is interpreted through temperature.

Sometimes, Carnap notes in "Aufgabe", 103, only the physical laws are considered part of the theory, so that correspondence rules and descriptions have to be considered part of the "physiological psychology of the senses"[18] and "a

[16] "gewissermaßen durch eine Art Wörterbuch, das angibt, welche Gegenstände (Elemente, Komplexe, Vorgänge) im zweiten Gebiet den einzelnen des ersten entsprechen"

[17] "‚Einem solchen Blau (bezeichnet z. B. nach dem Ostwaldschen Farbsystem) entspricht eine gewissen periodische Elektronenbewegung (bezeichnet durch die Schwingungszahl)'[.] [. . .] ‚Einem solchen stechenden Geruch (Chlorgeruch, Bezeichnungssystem fehlt) entspricht ein Gemenge von Elektronenkomplexen bestimmter Struktur (Cl-Atome)'; ‚Einer solchen Wärmeempfindung (Bezeichungssystem fehlt) entspricht eine gewisse durchschnittliche kinetische Energie einer Menge von Elektronenkomplexen (Atomen oder Molekülen).'"

[18] "physiologische Psychologie der Sinne"

(non-existing) descriptive complete science"[19]. The idea that there is no complete science to determine the correspondence rules anticipates Carnap's claim in *L&M* that many terms of the non-physical sciences cannot yet be defined in physical terms.

Carnap gives another example of a correspondence rule in *Begriffsbildung*, 60, that is similar to the correspondence rule connecting a pungent smell with an electron configuration: Specific electron configurations are assigned specific atoms or crystals, say, chloride and sodium or sodium chloride. These configurations are then assigned the qualities 'white' and 'salty'. Compare this with the example from *L&M*, 207:

> [L]et us imagine a calculus of physics is constructed, according to the second method [cf. figure 3.1], on the basis of primitive specific signs like 'electromagnetic field', 'gravitational field', 'electron', 'proton', etc. The system of definitions will then lead to elementary terms, e. g. to 'Fe', defined as a class of regions in which the configuration of particles fulfils certain conditions, and 'Na-yellow' as a class of space-time regions in which the temporal distribution of the electromagnetic field fulfils certain conditions. The semantical rules are laid down stating that 'Fe' designates iron and 'Na-yellow' designates a specific yellow color. (If 'iron' is not accepted as sufficiently elementary, the rules can be stated for more elementary terms.)

Begriffsbildung has strong connections to "Aufgabe". The principle of maximal simplicity from "Aufgabe" is a recurring theme: The introduction of correction factors for measurement devices (Carnap 1926, 30), the choice of a measurement scale for temperature (Carnap 1926, 36–37) and measurement scales (Carnap 1926, 23) in general, and the choice between Euclidean and non-Euclidean geometry are all justified by the simplicity of the resulting laws. These questions of simplicity occur in connection with the formation of physical concepts, the topic of the monograph. Consider the concept of temperature—different objects, when in contact with our skin, cause different heat sensations. While it would be possible to declare two objects to have the same temperature when they feel equally warm, this choice would "show itself to be inexpedient considering the fact of 'heat exchange'"[20]. This is because one body might cool down noticeably when in contact with another, while the other does not noticeably warm. Then, "for the sake of later laws of nature that occur in connection with the concept of specific heat"[21], the other body is ascribed a change in temperature that does not relate to a noticeable warming (Carnap 1926, 17–18). Carnap (1926, 57) also states that "the events in future and past are determined if the state is fixed for one point of

[19] "einer (nicht bestehenden) deskriptiven Gesamtwissenschaft"

[20] "würde sich jedoch als unzweckmäßig erweisen angesichts der Tatsache des ‚Wärmeausgleichs'"

[21] "späteren Naturgesetzen zuliebe, die beim Begriff der spezifischen Wärme auftreten"

time"[22] (cf. "Aufgabe", 101).

Begriffsbildung also has strong connections to much of Carnap's and Hempel's later work: There is the claim that counting is dependent on time (Carnap 1926, 15), which can be found again in the *Introduction*, 60, the discussion of rigid bodies (*Begriffsbildung*, 25, and *Introduction*, 91), the distinction between strong and weak periodicity (*Begriffsbildung*, 39–40, and *Introduction*, 80–81), and the five rules for physical magnitudes that appear in both *Begriffsbildung*, 22–23, and the *Introduction*, 63–65. The strong claim that all measurement is counting in *Begriffsbildung*, 15, also occurs in Hempel's *Concept Formation*, 719, albeit weakened to hold for measurements that rely on a concatenation operation for physical objects. With respect to measurement, Hempel (1952, 737, n. 64) refers directly to *Begriffsbildung*. Most importantly, Hempel's core claim that concept formation goes hand in hand with theory formation and is influenced by empirical results can already be found in *Begriffsbildung* and in "Aufgabe", since Carnap argues that the aim of concept formation is to arrive at theories that are simple given one's experiences.

Begriffsbildung also relates to "Dreidimensionalität", 107–108, in which Carnap relies, as in "Aufgabe", on the distinction between the domain of perception and the domain of physical theories. He speaks of experiences of the first level, which are the phenomena, and experiences of the second level, which can be the experiences of physics or, since this level is subject to conventional choices, also the ordinary ones. Ordinary experiences are those involving everyday concepts, where objects have qualities like color, hardness, etc. and concepts like *substance* and *causality* apply. The contents of the experiences of the first level are the "primary world", and those of the second level the "secondary world", where the secondary world can be the "physical" or the "ordinary world" ("Dreidimensionalität", 108). The connections between the secondary and primary worlds is, with a reference to "Aufgabe", given through a "correspondence relation"[23] ("Dreidimensionalität", 108, n. 1).

The *Aufbau*, finally, puts the discussions of "Aufgabe", "Dreidimensionalität", and *Begriffsbildung* in a wider perspective. Here, Carnap (1928a, §§1–2, §35, §156) develops a constructional system in which all statements are translatable into statements that contain only one primitive relation (although more are possible). The lower levels of the system constitute the autopsychological objects out of the primitive relation, the intermediate levels the physical objects out of the autopsychological objects, and the upper levels the heteropsychological and cultural objects, constituted from the physical objects. The correspondence rules of "Aufgabe" and *Begriffsbildung* are found in the intermediate levels, where the physical objects are constituted. But more than just the correspondence rules are

[22] "[D]as Geschehen in Zukunft und Vergangenheit is bestimmt, wenn der Zustand in einem Zeitpunkt festliegt."

[23] "Zuordnungsbeziehung"

needed to constitute the physical objects from the autopsychological objects.

The autopsychological objects only go up to the visual field, the colors, and a preliminary time order. The time order is preliminary because it is not always complete (*Aufbau*, §120). From these objects, colored world points and world lines are constituted by demanding that a certain set of desiderata be "satisfied as far as possible"(*Aufbau*, §§125–127), by using similarity assumptions (*Aufbau*, §135) and, once the upper levels have been constituted, by using observations of other people (*Aufbau*, §144). When world lines stay in close proximity to each other over a period of time, the class of the corresponding world points are visual things (*Aufbau*, §128), which, when assigned tactile and other qualities, lead to perceptual things (*Aufbau*, §134). Finally, "the entire space-time world, with the assignment of sense qualities to the individual world points, we call the perceptual world" (*Aufbau*, §133). Note that while perceptual qualities are assigned to things, the perceptual world is the resulting assignment of perceptual qualities to space-time points. The visual things seem to belong to the "ordinary world" of "Dreidimensionalität", while the perceptual world also allows the construction of the other secondary world—the "physical world"—which is constructed in §136 of the *Aufbau*.[24] When constructing the physical world, Carnap (1928a, §136) refers to "Aufgabe", "Dreidimensionalität", and *Begriffsbildung*, and describes the role of the perceptual world in more detail. He repeats the point that the perceptual world does not contain laws of the kind that hold in the physical world, but laws that are "of a much more complicated nature".[25]

As a final observation, note that in the construction of the world as described in the *Aufbau*, the method of interpreting theoretical terms that is described in "Aufgabe" is not involved in the assignment of sensory qualities to space-time points, but covers only the second step, the correlation of numbers as values of state magnitudes to these sensory qualities. The Kantian notions of *Der Raum* would occur somewhere in the construction leading to the assignment of sensory qualities to space-time points. So whatever Kantian notions about synthetic a priori statements Carnap had at the time of writing, they would not have been relevant for the construction developed in "Aufgabe". This is the final reason why Mormann's claim that "Aufgabe" cannot be considered a contribution to the logical empiricist's theory of science is incorrect.

3.3 The exclusive use of first order logic

According to Suppe (1974a, 16, 50), proponents of the Received View always assumed an axiomatization of scientific theories in first order predicate logic or,

[24]In the translation, the paragraph has the title "The World of Physics". The original uses the same phrase ("Die physikalische Welt") as "Dreidimensionalität".

[25]"von viel verwickelterer Gestalt". Note the similarity of this choice of words to "Dreidimensionalität", where the laws are "verwickelter Art" (of a complicated kind).

in Carnap's later writings, first order logic augmented by modal operators. In contemporary summaries of the Received View, the option of modal operators is typically not mentioned, unlike the demand for first order logic (to list just a tiny fraction: Stegmüller 1978, 40; Beatty 1980, 419; Ereshefsky 1991, 61; Morrison and Morgan 1999, 2; van Fraassen 2000; Kitcher 2001, 152; Burgos 2007, 157; Frigg and Hartmann 2008, §4.1; Muller 2010, §2).

Suppes (1967, 56–57) also speaks of the "standard sketch" of scientific theories, by which he clearly means the Received View, and later discusses "standard formalizations", axiomatizations in first order logic (Suppes 1967, 58), without explicitly identifying the two (some authors who cite the text do so, however, for example Suppe (1974a, 114, n. 241), Beatty (1980, appendix 1), and Ereshefsky (1991, 62)). In his discussion, Suppes (1967, 58) puts forth what has become one of the standard criticisms of the view's supposed reliance on first order logic: "Theories [...] like quantum mechanics [...] need to use [...] many results concerning the real numbers. Formalization of such theories in first order logic is utterly impractical". The impracticality of first order axiomatization is, next to the perceived need for an exhaustive axiomatization discussed below, one pillar of Stegmüller's criticism of the Received View (Stegmüller 1978, 39–40).

Beyond being cumbersome, first order logic has a deficiency that even in principle cannot be avoided: It cannot determine structures with an infinite domain up to isomorphism. As Suppe (2000, §3) puts it, a "general problem [with the Received View] was that the Löwenheim-Skolem theorem implied that [the theory's] models must include both intended and wildly unintended models"; the Löwenheim-Skolem theorem holds only in first order logic. The Semantic View does not have this problem, because the models of the theory are not defined to be those of a set of sentences in the object language, but are rather defined directly in the metalanguage. The Semantic View is superior to the Received View in this respect since structures can be determined up to isomorphism, as van Fraassen (1989, §8.6) also stresses: "[W]hen a theory is presented by defining the class of its models, that class of structures cannot generally be identified with an elementary class of models of any first-order language". French and Ladyman (1999, 116–117) also make this point.[26]

Whether this criticism is valid—that is, whether something essential about scientific theories is missed when they are described in first order predicate logic—is beyond the scope of this book (but see the remarks in §4.1.3 and by Pearce and Rantala 1985, 135). Rather, I want to note that the Received View is not restricted to first order logic in any of the sources that Suppe (1974a, §II.E, n. 107) cites for the final version of the Received View (Carnap 1956b, 1958, 1963c, 1966, Hempel 1965g, 1963). Indeed there is clear evidence that Hempel and Feigl allowed infinite

[26]There are other criticisms of the Received View that stem from its perceived reliance on first order logic. For instance, Klein (2011) argues that because it does not allow quantification over predicates, the Received View does not allow the reduction of multiply realized properties.

type theory, and overwhelming evidence that Carnap, from the beginning of the Received View to its final version, assumed it.

For one, even though Carnap's "Aufgabe" was published after a talk by Hilbert in which he isolated first order logic as a distinct subsystem of logic, "Aufgabe" was published before this result appeared in print (cf. Anellis 1996). Arguably, the distinction between first order logic and type theory was not completely clarified until 1935 (Hodges 2001, 1–4). Hence it is unsurprising that Carnap does not mention first order logic or any restriction of the mathematical formalism that might point towards the restriction to first order logic, except for a reference to the ideal physics as an "axiomatic system of pure deduction"[27] (Carnap 1923, 103). Syntactic deduction is complete for first order logic, but not for type theory (Gödel 1930, 1931), and relying only on syntactic deduction essentially reduces the possible inferences of type theory to those of first order logic (Henkin 1950). But since these results were unknown when Carnap wrote "Aufgabe", this should not be read as an exclusive endorsement of first order logic by Carnap. Furthermore, Carnap thought that type theory was complete (Reck 2007). Indeed, given his focus on type theory in his research on axiomatics (cf. Reck 2007), it would be rather surprising if he had demanded first order axiomatizations.

A restriction to first order logic is furthermore incompatible with the logical pluralism of Carnap's Principle of Tolerance (Carnap 1934a, §17), which grew out of the protocol sentence debate in philosophy of science (Carus 2007, 252–256). Carnap (1956b, §II) sometimes considers different restrictions on the *observational* part of a language (since this part is considered to be completely interpreted by observation), but the theoretical language (L_T) is unrestricted (Carnap 1956b, 46):

> For L_T we do not claim to have a complete interpretation, but only the indirect and partial interpretation given by the correspondence rules. Therefore, we should feel free to choose the logical structure of this language as it best fits our needs for the purpose for which the language is constructed.

The choice of a logic for the description of theories is a practical one. If Carnap had considered first order logic to be the most expedient for all theories, of course, then he might have developed the Received View based on first order logic, simply because he would not have expected type theory to be relevant for science.

But Carnap is in clear agreement with Suppes (1967, 58) that first order logic is too cumbersome, as he explicitly suggests infinite type theory for reconstructions. Already in the *Aufbau*, Carnap (1928a, §30) writes that

> Russell has applied his theory [of types] only to formal-logical structures, not to a system of concrete concepts (more precisely: only to variables and logical constants, not to nonlogical constants). Our object spheres are Russell's "types" applied to extralogical concepts.

[27] "axiomatisches System reiner Deduktion"

Since the concept of object spheres is used throughout Carnap's reconstruction (e. g., §§109–118), there is no question about his stance on its expedience.

It is clear that Carnap's later writings on the Received View allow for type theory because each of the works on which Suppe's exposition is based either explicitly uses infinite type theory or refers to a work that does. In "Theoretical concepts", 43, for instance, he conjectures that "acceptance of the following three conventions C1–C3 is sufficient to make sure that L_T includes all of mathematics that is needed in science":

> Conventions about the domain D of entities admitted as values of variables in L_T.
>
> C1. D includes a denumerable subdomain I of entities.
>
> C2. Any ordered n-tuple of entities in D (for any finite n) belongs also to D.
>
> C3. Any class of entities in D belongs also to D.

Convention C1 includes denumerably many objects in the domain, C2 includes relations of any of the domain's elements in the domain, and C3 includes classes of any of the domain's elements in the domain. The result is that any predicate or relation in the domain can be in the scope of another predicate or relation, and this is just infinite type theory. Given the incompleteness of type theory, Carnap (1956b, 51, 62) relies on semantic entailment rather than syntactic deduction for his rules of inference.

The article that includes this construction was motivated by Hempel's unpublished (at the time) discussion of Carnap's philosophy of science (Hempel 1963).[28] Accordingly, Carnap (1963c, §24.A, n. 38) refers to this article in his direct response to Hempel. He points out some differences between their presentations, but none with respect to the use of type theory. Carnap (1963c, §24.C, n. 41), also refers to another elaboration of his position (Carnap 1958, 237), in which he is even more explicit about his logical assumptions:

> Let the structure of L_T be such that it contains a *type theory* with an infinite series of domains D^0, D^1, D^2, etc.; D^n is called the n^{th}-level domain. Each variable and each constant belongs to a specific type. Each variable of type n has D^n as its domain and each constant of type n refers to an element of D^n.[29]

In his popular introduction to the philosophy of science, Carnap (1966, 253, n. 2)

[28] Carnap had read Hempel's discussion in 1954 (Carnap 1963c, §24.B).

[29] "Die Struktur von L_T sei so, dass sie eine *Stufenlogik* mit einer unendlichen Folge von Bereichen D^0, D^1, D^2, usw. enthält; D^n heisst der Bereich n-ter Stufe. Jede Variable und jede Konstante gehört zu einer bestimmten Stufe. Jede Variable n-ter Stufe hat D^n als Wertbereich, und jede Konstante n-ter Stufe bezeichnet ein Element von D^n."

refers to this text as an elaboration of his position and gives a simplified version of the conventions C1–C3 from "Theoretical concepts".

Accordingly, Carnap (1954, §44, §48) uses higher order logic extensively in axiomatizations of mathematical concepts like the natural numbers, as well as physical notions like space-time topology. He also relies on higher order logic for more straightforwardly philosophical investigations, such as discussions of philosophical method (Carnap 1934a, §§26–27) and the definition of scientific concepts in phenomenalistic terms as in the *Aufbau* (Carnap 1928a, §§107–122).

Hempel is not as vocal as Carnap about the use of type theory, but he also does not dismiss it. Specifically, he does not distance himself from Carnap's use of type theory, even when he refers to Carnap's expositions of the Received View in type theory (Hempel 1965g, 194–197; 1963, §I, n. 2).

In some cases, he attaches considerable weight to analyzes of first order logic. For example, in his discussion of the eliminability of theoretical terms, Hempel (1965g, 211–212) states that Craig's theorem applies "provided that [the theory] satisfies certain extremely liberal and unconfining conditions". Elsewhere, Hempel (1963, 698) states that Craig's theorem applies "in a very comprehensive class of cases". But, as Hempel himself notes, Craig's theorem assumes first order logic.

On the other hand, Hempel (1973, 264–265) states later, after having abandoned the Received View, that

> the precisely characterized languages by reference to which certain philosophical problems have been studied are often distinctly simpler than those required for the purposes of science. For example, Carnap's theory of reduction and confirmability, and his vast system of inductive logic are limited to languages with first-order logic, which certainly does not suffice for the formulation of contemporary physical theories. The same remark applies to various studies by Carnap and other empiricists that deal with the structure and function of scientific theories [...].

It is not clear to which of Carnap's studies on the structure of scientific theories Hempel refers. Without being seriously mistaken, he cannot mean Carnap's expositions of the Received View, quoted above. The last conception of reduction and confirmability[30] in which Carnap (1937) relies on first order logic was quickly followed by a generalization that allows correspondence rules of any logical kind for reduction (Carnap 1938, 1939; see also Carnap 1963b, 424).

This misrepresentation of Carnap's views notwithstanding, the passage shows that at least in 1973, Hempel was of the opinion that a restriction to first order logic puts more than just an "extremely liberal and unconfining" condition on scientific theories. So either Hempel had changed his mind at that point, or he thought that

[30]See Gemes (1998b, §1.4) for the curious relation between the two via the concept of empirical significance.

even without coverage of physical theories, first order logic would suffice for a "very comprehensive class" of theories. Or, in a more dubious interpretation, he may have thought that a sentence S is not part of the observational content of a theory T if T entails S, but S is not deducible from T. Then the observational content of a theory described in type theory could be captured by transcribing it into first order logic (Henkin 1950, cf. Leivant 1994, §5.5) and then applying Craig's theorem.

All of the preceding claims by Hempel about first order logic are descriptive, that is, he states its conditions on the axiomatizations of scientific theories. So while he may have been mistaken about the expedience of type theory, at no point does he *demand* that it be excluded from axiomatizations. Instead, Hempel (1965g, 201–202) uses type theory himself to arrive at explicit definitions for real-valued measurement results in observational terms. He notes that using type theory "will be considered too high [a price] by nominalists", but adds that "it would no doubt be generally considered a worthwhile advance in clarification if for a set of theoretical scientific expressions explicit definitions in terms of observables can be constructed at all". Just like Carnap, Hempel is of the opinion that a properly observational language must not have a logical apparatus that is too strong: He cautions that "the definiens will normally be teeming with symbols of quantification over individuals and over classes and relations of various types and will be far from providing finite observational criteria of application". However, he proceeds to give the definition anyway, showing that he does not consider exclusive reliance on first order logic a necessity.

Besides using type theory for philosophical analyses, Hempel also considers higher order axiomatizations of scientific theories to be consistent with the demands of the Received View. Like Feigl (1970, 8), Hempel (1970, §3) lists the axiomatizations given by Suppes as acceptable within the Received View, and he explicitly notes that they use "set theory and mathematical analysis" which are "more powerful" than first order logic. This may be questioned, since there are first order axiomatizations of set theory and thus of mathematics, insofar as set theory can be a foundation of mathematics in general. But Hempel's note does show once again that he does not consider exclusive reliance on first order logic necessary. The clearest example, however, is Woodger's axiomatization of parts of cell biology (Woodger 1939). Both Feigl (1970, 8) and Hempel (1952, 687) accept this axiomatization as a possible reconstruction, and Woodger (1939, §III) explicitly uses type theory. Finally, Hempel (1970, §2, n. 4) also gives a list of expositions of the Received View; among them are some of the ones by Carnap (1956b, 1966) described above, and another one in which Carnap (1939, §14) explicitly discusses type theory. There is thus no indication in Hempel's work that he thinks that axiomatizations have to rely on first order logic only, and there are multiple passages in which he either uses type theory himself, endorses reconstructions that do, or counts works that endorse type theory as expositions of the Received

View.

Of course, there is an abundance of formulas in first order logic and simple first order theories in Hempel's and Carnap's writings. In discussing Carnap's studies of probability theory, which rely solely on first order logic, Hempel (1973, 265) describes and seems to agree with Carnap's view of these analyses:

> [...] Carnap often stressed that these studies are intended only as the first stage in the development of more comprehensive theories, and that the solutions they offer may well permit of extension to more complex situations.

In other words, the studies focus on cases in which the problem seems solvable, in the hope that the results can be generalized to more complicated situations. A nice example of such a generalization is described by Psillos (2000, §1, n. 7), who argues that Carnap rediscovered the Ramsey sentence approach while trying to generalize Craig's theorem to type theory. That Hempel shared Carnap's view on individual analyses as a starting point for generalizations is suggested by his discussion of the problems of confirmation. After noting that a criterion of confirmation should be applicable to hypotheses of any logical form, Hempel (1965f, §5) suggests that a criterion that is only applicable to laws of the form 'All ravens are black' might "still be considered as stating a particularly obvious and important sufficient condition of confirmation".

First order axiomatizations also often appear as examples in Hempel's writings. On the use of Ramsey sentences, Hempel (1965g, §9, my emphasis) states that their

> logical apparatus is more extravagant than that required by [the original theories]. *In our illustration, for example*, [the original theories] contain variables and quantifiers only with respect to individuals (physical objects), whereas the Ramsey-sentence [...] contains variables and quantifiers also for properties of individuals.

When discussing correspondence rules, Hempel (1963, 692, my emphases) again writes: "*For example*, the logical framework *might* be that of the first-order functional calculus with identity". He adds, again ignoring all of the later publications by Carnap (1938, 1939, 1956b, etc.) on the topic: "This is, in fact, one of the principal cases with which Carnap's theory of reduction is concerned".[31]

Hempel's inaccurate remarks about Carnap's reliance on first order logic may help to explain how this misconception became part of philosophical folklore. Suppes's switch from the discussion of the "standard sketch", which is the Received View, to the discussion of "standard formalizations", which are not demanded in the Received View, clearly has also contributed to the confusion, at least in the

[31] Of course, Hempel mentions first order logic as *one* of the principal cases. But if type theory is meant to be the other principal case, his remark is wholly mystifying.

cases of Beatty (1980), Ereshefsky (1991), and Suppe (1974a). Suppe popularized this misconception through his canonical exposition. Another possible explanation stems from Carnap's exclusive reliance on first order logic in his explication of probability (Carnap 1950b). This part of Carnap's work may have been influential enough to overshadow his use of type theory in other works. Finally, Quine's stance that higher order logic is not proper logic but rather mathematics (Quine 1970) probably also contributed to the idea that proponents of the Received View always referred to first order logic when speaking of logic.

The likely influence of Quine's stance and the logicians who followed him in this respect (cf. Leivant 1994, §5.2) finds a peculiar analogy in a critique of the Received View on the grounds that it relies on a distinction between theoretical and observational vocabulary. In this critique, van Fraassen (1980, §3.6) remarks that "logicians attached importance to restricted vocabularies, and that was seemingly enough for philosophers to think them important too", and concludes that the Received View "focussed attention on philosophically irrelevant technical questions". This may very well be true in the case of first order logic, except that the philosophers in question are not the proponents of the Received View, but their critics. Now, second order logic is already enough "to capture *directly* most all mathematical practice" (Leivant 1994, §5.2, §7; cf. Väänänen 2001, 515), and capturing mathematical practice only gets easier for higher orders. Thus the slogan "Mathematics for the philosophy of science, not meta-mathematics", emphatically endorsed by critics of the Received View (Muller 2010, §2; cf. van Fraassen 1972, 309; 1989, 221–222),[32] makes sense as a criticism of "standard formalizations" only under the identification of first order logic with meta-mathematics and higher order logic with mathematics. But the slogan "Higher order logic for the philosophy of science, not first order logic" could just as well have been Carnap's.

3.4 Exhaustive axiomatization

Although the Received View does not demand first order logic, it may still turn out to be overly cumbersome if it demands the exhaustive axiomatization of scientific theories and all of the mathematics they contain. In reference to van Fraassen (1980), who explicitly argues against the Received View, Suppes (1992, 207–208) criticizes the philosophers of science who rely on "standard formalizations":

> Suppose we want to give a standard formalization of elementary probability theory. [Only] after stating a group of axioms on sets, and another group on the real numbers, [are we] in a position to state the axioms that belong just to probability theory as it is usually

[32]The slogan is attributed to Patrick Suppes, but the closest published phrase I could find is: "[T]he basic methods appropriate for axiomatic studies in the empirical sciences are not metamathematical (and thus syntactical and semantical), but set-theoretical" (Suppes 1954, 244).

> conceived. In this welter of axioms, those special to probability can easily be lost sight of.
>
> More important, it is senseless and uninteresting continually to repeat these general axioms on sets and on numbers whenever we consider formalizing a scientific theory. No one does it, and for good reason.

In Suppes's sense, a standard formalization of a theory is thus not only restricted to first order logic, but also must specify all of the axioms on every mathematical concept that occurs in the theory.

Suppes suggests instead a reconstruction of theories in the Semantic View, which does not require specifying all of the mathematical axioms, and accordingly leads to much simpler formalizations. Speaking in favor of the Semantic View, Stegmüller (1979, §1) argues that Carnap's approach to scientific theories could only be executed by philosophers with technical abilities far beyond even those of Montague (1962), who gave an exhaustive axiomatization for Newtonian mechanics. Muller (2010, §7), after suggesting improvements to both the Received View and the Semantic View that bring them quite close together, concludes that

> the gap remains sufficiently wide to prefer the [improved Semantic View] over the [improved Received View], because the starting point for [the semantic] construal remains a set-theoretical predicate [...], and because the formal theories [...] need never be spelled out (the formalising labour that is mandatory in the [Received View] need not be performed).

Muller explicitly assumes that the welter of axioms Suppes ascribes to standard formalizations also occurs in the Received View. Van Fraassen (1972, 306) makes this assumption as well.

It seems, indeed, like an unacceptable burden on philosophical analysis to demand such "formalizing labor"—the exhaustive axiomatization of *every* concept employed by the theory under consideration. Fortunately, contrary to what its critics claim, the Received View does not demand this. It is true that both Hempel (1952, 733, n. 24) and Feigl (1970, 8) mention Woodger's exhaustive axiomatization of parts of biology as in keeping with the Received View, but this indicates only that exhaustive axiomatizations are allowed, not that they are required. Besides Woodger's axiomatization, Hempel also mentions the axiomatization of the theory of relativity by Reichenbach and the axiomatization of game theory by von Neumann and Morgenstern, while Feigl refers to the axiomatizations of Reichenbach and of Suppes himself, none of which are exhaustive. Hempel (1970, 149–150) later calls Reichenbach's treatment "axiomatically oriented, though not strictly formalized", but still considers Suppes's treatments to be full-fledged axiomatizations. Hempel (1974, 247, my emphasis) also calls Suppes's axiomatization "technically much *more* elegant and rigorous than Reichenbach's", suggesting that the exhaustiveness of axiomatizations comes in degrees.

Relatedly, Carnap (1939, §15) notes that it would be "practically impossible to give each deduction which occurs the form of a complete derivation in the logical calculus [. . .]. But it is essential that this dissolution is theoretically possible and practically possible for any small part of the process". In other words, obvious or uncontentious steps in a derivation may be skipped, as well as the axioms on which they rely if the axioms are not needed at other points in the derivation. Carnap (1939, §16) gives a general description of this method, which seems to describe Suppes's strategy for avoiding the "welter of axioms" quite well:

> Each of the nonlogical calculi [that are applied in science] consists, strictly speaking, of two parts: a logical *basic calculus* and a *specific calculus* added to it. The basic calculus could be approximately the same for all those calculi; it could consist of the sentential calculus and a smaller or greater part of the functional calculus as previously outlined. The specific partial calculus does not usually contain additional rules of inference but only additional primitive sentences, called *axioms*. As the basic calculus is essentially the same for all the different specific calculi, it is customary not to mention it at all but to describe only the specific part of the calculus.

Since "the functional calculus as previously outlined" contains type theory and mathematics (Carnap 1939, §14), this allows a large number of axioms to go unmentioned.

The restriction of the exposition to a specific calculus is quite common according to Carnap (1939, §23), as he notes that "the customary formulation of a physical calculus is such that it presupposes a logico-mathematical calculus as its basis". When Carnap (1939, §16) gives a simple theory of thermal expansion as an example of a calculus, he accordingly uses—but does not axiomatize—its logical and mathematical concepts. In a derivation within this calculus, mathematical inferences take up one step and are simply marked "Proved mathem. theorem". Hempel (1958, §3) goes one step further when analyzing the role of theoretical terms in a simple theory of buoyancy, by both using mathematics without giving all axioms (postulate 3.4) and describing the logical relations in natural language (postulate 3.3).

Carnap's description of the *additional* primitive sentences in a nonlogical calculus as 'axioms' may have led to the misconception that the Received View demands an exhaustive *logical* calculus, rather than (at most) an exhaustive specific one. This distinction may not have been made clear enough in many of the works of the Received View, although Carnap (1950b, 15–16), for example, refers to his earlier distinction between the basic and the specific calculus (Carnap 1939, §16) when discussing formalizations of probability theory, and states:

> The formalization (or axiomatization) of a theory or of the concepts

> of a theory is here understood in the sense of the construction of a formal system, an *axiom system* (or postulate system) for that theory. [...]
>
> In the discussions of this book we are [...] thinking of those semiformal, semi-interpreted systems which are constructed by contemporary authors, especially mathematicians, under the title of axiom systems (or postulate systems). In a system of this kind the axiomatic terms (for instance, in Hilbert's axiom system of geometry the terms 'point', 'line', 'incidence', 'between', and others) remain uninterpreted, while for all or some of the logical terms occurring [...] and sometimes for certain arithmetical terms [...] their customary interpretation is—in most cases tacitly—presupposed.

This passage is a general description of how the tacit presupposition of axioms is to be understood, and it could well serve as a conclusion to the quote by Suppes (1992, 207–208) at the beginning of this section (page 108). It also accurately describes Carnap's own methodology for analyzing probability (Carnap 1950b), in that he uses mathematics without axiomatizing mathematical concepts. And this is not a singular case: Already in the *Aufbau*, Carnap (1928a, §112) defines 'quality class' by using, but not axiomatizing, arithmetic division (cf. Carnap 1928a, §97).

Therefore, the Received View cannot be criticized for demanding an unwieldy and overly difficult formalization from all philosophers of science. As their own practice shows, Hempel and Feigl consider axiomatizations exhaustive enough even when they do not list each and every axiom used in a derivation. Carnap not only presents and uses axiomatizations that leave out all of mathematics, but he also describes in general how a theory can be axiomatized with a specific calculus alone, while the basic calculus is presupposed.

Of course, one may still criticize the Received View on the grounds that *any* kind of axiomatization, whether exhaustive or not, is fruitless.[33] Such a criticism, however, could not be made from within the Semantic View of Suppes or Stegmüller, since it relies on non-exhaustive axiomatizations. Such an argument would also need to establish that, for example, Suppes (1968), Betti and de Jong (2008), Betti et al. (2009), and Leitgeb (2009) are mistaken in their defense of formal approaches. And finally, as Halvorson (2012, 203) puts it:

> [W]hat would "informal philosophy of science" look like? Should the informal philosopher of science eschew all use of mathematical notation or concepts? But how then should the informal philosopher of science discuss quantum mechanics or general relativity or string theory?

It is plausible that not all scientific theories can be fruitfully axiomatized, but it also seems clear that some can. In §3.8.2, I will argue that this comparably weak

[33] I thank an anonymous referee for *HOPOS* for this point.

claim is enough to justify the Received View.

3.5 The formalization of theories

Not all formalizations of a scientific theory are equally good. While the Received View does not demand exhaustive axiomatizations in first order logic, it may nevertheless assume formalizations that are fruitless or useless. To counter this charge, I will show that throughout Carnap's philosophical career, he relied on formalizations of scientific theories in phase space, which is also van Fraassen's preferred formalization.

"Aufgabe", 97, only contains a very abstract reference to physical laws as axiom systems and the claim that the physical world is free from perceptual qualities. This is fleshed out in "Dreidimensionalität", 107, where the physical world is said to contain

> only space and time magnitudes and certain non-perceptual state magnitudes. In their purest form, these three kinds of magnitudes also have a character not in the slightest comparable to space, time, or perceptual qualities, but are bare assignments of numbers, that is, relational terms. For intuitiveness, the terms space, time, processes, changes, etc. will still be used.[34]

In "Aufgabe", the third volume of the ideal physics describes the world by giving the value of each magnitude at each point in space for two different times. In "Dreidimensionalität", 120, Carnap notes that one can equivalently give the value and the derivative of each magnitude at each point in space for just one time. In "Aufgabe", 101, Carnap refers to Russell and Mongré-Hausdorff for the proof that derivatives cannot be considered instantaneous magnitudes, so that a single moment in time cannot describe the physical world completely; the equivalence however allows one to use the more expedient description at a single moment. In contemporary vocabulary, Carnap is in "Dreidimensionalität" describing a phase space with position and velocity as generalized coordinates.

The laws in this pure physical world are free from the concept of causation as it occurs in the ordinary world ("Dreidimensionalität", 108):

> The processes of the physical world do *not act* upon each other, but for them a dependency holds that must be considered a pure mathematical-functional relation [. . .].[35]

[34] "in der es nur Raum- und Zeitgrößen und gewisse nichtsinnliche Zustandsgrößen gibt. In der reinsten Form haben auch diese drei Größenarten keinerlei mit Räumlichkeit, Zeitlichkeit oder Sinnesqualität vergleichbaren Charakter, sondern sind bloße Zahlbestimmungen, d. h. Relationsterme. Aus Gründen der Anschaulichkeit werden trotzdem die Bezeichnungen Raum, Zeit, Vorgänge, Veränderungen usw. beibehalten."

[35] "Die Vorgänge der physikalischen Welt *wirken nicht* aufeinander, sondern es gilt für sie eine

In "Dreidimensionalität", 118, Carnap clearly describes the laws of theories as restrictions on phase space:

> If any element of a class depends on other elements in such a way that it is uniquely determined whenever some subclass of the others is determined, we call the dependency-relation a "deterministic law" and the class "deterministic". [. . .]
>
> Laws of dependency that do not result in a unique determination for some element even when all others are determined, but still restrict the possibilities for this element, we call "restricting laws".[36]

The conception of a phase space is elaborated in chapter III of *Begriffsbildung*, entitled "Abstract level: The four-dimensional world events"[37]. Here, Carnap first suggests identifying space-time points by four-tuples so that a physical description of the world consists in assigning the values of the basic physical magnitudes to each point in four-dimensional space-time. In a second suggestion, he goes beyond this geometrical description and lets descriptions of the world consist of sets of tuples with the first four values being the space-time points, the rest being the values of the physical magnitudes. This then is the purely numerical phase space description of "Dreidimensionalität". As in "Dreidimensionalität", laws of nature are in both cases restrictions on the possible descriptions—restrictions on the possible assignments of values to space-time points in the geometrical case, and restrictions on the possible sets of tuples in the numerical case (*Begriffsbildung*, 58). The full correspondence rules that lead to the qualities of 'white' and 'salty' begin, at the most abstract level, with purely mathematical descriptions in tuples. These tuples are assigned to electron configurations, which are assigned to specific crystals, which are finally assigned to the phenomenal qualities (*Begriffsbildung*, 60).

Carnap (1963a, 15–16) himself considers this aspect of *Begriffsbildung* to be closely connected to his later work, writing:

> I described the world of physics as an abstract system of ordered quadruples of real numbers to which values of certain functions are co-ordinated; the quadruples represent space-time points, and the functions represent the state-magnitudes of physics. This abstract conception of the system of physics was later elaborated in my work on the theoretical language.

Abhängigkeit, die als reine mathematisch-funktionale Beziehung aufzufassen ist [. . .]."

[36] "Wenn irgendein Element einer Klasse derart von andern Elementen abhängt, daß es eindeutig bestimmt ist, sobald eine gewisse Teilklasse der übrigen festliegt, so nennen wir die Abhängigkeitsbeziehung ein „determinierendes Gesetz" und die Klasse „determiniert". [. . .]

Abhängigkeitsgesetze, die zwar für irgendein Element, selbst wenn alle übrigen bestimmt sind, nicht eindeutige Bestimmtheit ergeben, aber doch die Möglichkeit für dieses Element einschränken, nennen wir „beschränkende Gesetze"."

[37] "Abstrakte Stufe: das vierdimensionale Weltgeschehen"

Of course, the points of *Begriffsbildung* to which Carnap here refers essentially recapitulate the same points of “Dreidimensionalität”. Carnap relies on both works in the *Aufbau*, §136, where the physical world that he aims to construct is a “(purely numerical) structure”. The formalism suggested in these three texts is also used in “Theoretical concepts” and “Beobachtungssprache” (see §3.1), with the same stress on the purely mathematical character of the space-time tuples and the assignment of other numbers to those tuples as physical magnitudes at those space-time points. The discussion in the two early works makes clear the role of physical theories as restrictions on phase space.

Now compare Carnap’s account of the formalization of scientific theories to that of van Fraassen. In an outline of his version of the Semantic View (endorsed, for example, by Suppe 1974a, 222–223), van Fraassen (1970, 328–329) focuses on

> the formal structure of *nonrelativistic* theories in physics [. . .]. A physical theory then typically uses a mathematical model to represent the behavior of a certain kind of physical system. A physical system is conceived of as capable of a certain set of *states*, and these states are represented by elements of a certain mathematical space, the *state-space*. [. . .] To give the simplest example, a classical particle['s] [. . .] state-space can be taken to be Euclidean 6-space, whose points are the 6-tuples of real numbers $(q_x, q_y, q_z, p_x, p_y, p_z)$.

Van Fraassen (1970, 330) then distinguishes between laws of coexistence, laws of succession, and laws of interaction. In the non-statistical case (van Fraassen 1970, §5.1), laws of coexistence select the physically possible subset of the state-space (van Fraassen 1970, 330), laws of succession select, in the instantaneous state picture, the physically possible trajectories in the state-space (van Fraassen 1970, 331), and laws of interaction at least in principle reduce to the above (van Fraassen 1970, 332).

It is striking how close van Fraassen’s schema for the formalization of scientific theories resembles Carnap’s in “Aufgabe”, “Dreidimensionalität”, *Begriffsbildung*, “Theoretical concepts”, and “Beobachtungssprache”. Both accounts describe a classical particle as located in four-dimensional space-time, being assigned the relevant physical magnitudes (momentum in van Fraassen’s example). In both accounts, the phase space[38] is restricted by the scientific theory.

The difference between the accounts lies mainly in their use of semantics: In Carnap’s account, the phase space description is given in the object language and is related to observations via correspondence rules in the object language. With the adoption of Tarski’s semantics in *L&M*, a semantic interpretation is given to the observations in the metalanguage. In van Fraassen’s account, the interpretation of physical theories relies on “elementary statements”:

[38] ‘Phase space’ seems to be a more common term for this description of physical systems than ‘state-space’.

> Besides the state-space, the theory uses a certain set of *measurable physical magnitudes* to characterize the physical system. This yields the set of *elementary statements* about the system (of the theory): each elementary statement U formulates a proposition to the effect that a certain such physical magnitude m has a certain value r at a certain time t. (Thus we write $U = U(m, r, t)$ [...].)
>
> [...] For each elementary statement U there is a region $h(U)$ of the state-space H such that U is true if and only if the system's actual state is represented by an element of $h(U)$. [...] The mapping h (the *satisfaction function*) is the third characteristic feature of the theory. [...] The exact relation between $U(m, r, t)$ and the outcome of an actual experiment is the subject of an auxiliary theory of measurement, of which the notion of "correspondence rule" gives only the shallowest characterization.

The swipe about correspondence rules references an article by Suppes (1967). I will show in §3.6.6 that both Suppes and van Fraassen greatly underestimate the complexity allowed by the correspondence rules as suggested by Carnap and Hempel. In van Fraassen's account, the phase space description is given in the metalanguage and semantically related to the elementary statements in the object language, which are then related to measurement outcomes via correspondence rules, possibly in the object language. Van Fraassen is silent about the semantic interpretation of the measurement outcomes.

At the theory level, that is, the restriction of the phase space by the theory, the only difference between Carnap and van Fraassen is that the former considers the description to be part of the object language, the latter considers it part of the metalanguage. But since both Carnap and van Fraassen take their respective languages to include all of mathematics and allow as much or as little explicit formalization as expedient (see §3.4), this difference is merely verbal.

Later versions of van Fraassen's the Semantic View left out elementary statements, relying only on the restriction of the phase space, and a relation of the theory to observations in the metalanguage (for example, van Fraassen 1989, 365, n. 34; French and Ladyman 1999). But without an object language, there cannot be a metalanguage (see n. 32), and so this account is in fact very close to that of Carnap in his early works, where a theory is described in phase space (with the help of *some* language, of course) and connected within the language to a distinguished set of (observation) sentences, without mention of their semantic interpretation. The main difference seems to be that Carnap assumes that the language is or can be completely formal, while van Fraassen only assumes that it is precise enough to describe the notions of model theory (see §4.2.1) and of the phase space formalism. The alleged difference between the object language and the metalanguage then is rather the difference between a well-defined language and the semi-regulated language of mathematics as it is found in textbooks and research papers.

3.6 Connecting theoretical and observational terms by correspondence rules

Many criticisms of the Received View essentially maintain that it is misguided to assume a bipartition of the language into observational and theoretical terms, and a connection between those terms via correspondence rules. Starting from a history of the nature of correspondence rules, I will argue that Carnap's examples for observational and theoretical terms were correct within the framework of the Received View, and that the bipartition of a language does not have to be unduly complex. Furthermore, correspondence rules for the theoretical terms of one theory can come from other theories, and neither have to be (nor have been) overly simple.

3.6.1 The structure of correspondence rules

The typical view on Carnap's position on correspondence rules (see §3.1) is that in the *Aufbau*, Carnap assumed the explicit definability of all theoretical terms in observational terms, but that later he progressively weakened his stance until he assumed that (i) theoretical terms are not definable in observational terms and (ii) observational terms may be definable in theoretical terms. I now want to show that in fact Carnap always assumed (i), and that (ii) is, if anything, a strengthening of his original position.[39]

In "Aufgabe", as in the *Introduction*, the correspondence rules are such that any statement in observational terms can be described in theoretical terms, but not vice versa. Carnap (1923, 100) writes about the correspondence rules (the "dictionary"):

> The dictionary can be used in both directions: It serves both for the translation of a phenomenal fact into the physical and conversely. It has to be noted, however, that the correlation is unique only in the second case; while, for two different reasons, a specific phenomenal content generally corresponds to not only one specific physical fact but an infinite set of them.[40]

Carnap thus claims that the dictionary associates with each physical state exactly

[39] I thank Christopher French for helpful discussions about this section.

[40] "Das Wörterbuch ist in beiden Richtungen benutzbar: es dient sowohl zur Übersetzung eines phänomenalen Tatbestandes in den entsprechenden physikalischen, wie auch umgekehrt. Es ist jedoch zu bemerken, daß nur im zweiten Falle die Zuordnung eindeutig ist; während aus zwei verschiedenen Gründen einem bestimmten Empfindungsinhalt i. A. nicht nur ein bestimmter physikalischer Tatbestand entspricht, sondern eine unendliche Menge von solchen."

one phenomenal state, but with each phenomenal state an infinite set of physical states.

Carnap's first reason for the lack of uniqueness in the latter case is the existence of multiple microscopic realizations of physical macrostates in, for example, thermodynamics. His second reason is the perception threshold (Carnap 1923, 100). Later, in *Probability*, Carnap argues that because of this perception threshold, it cannot be required that there is a difference in temperatures only if there is a difference in heat sensations (see §3.1). Carnap adds two more reasons in "Dreidimensionalität", 126: First, a sensation does not uniquely identify the location of its physical source; second, a sensation (e. g., of a color) corresponds to a multitude of physical states (e. g., a multitude of frequency distributions of electromagnetic waves).

In modern terminology, this feature of the correspondence rules can be phrased syntactically or semantically. (Of course, in 1923 Carnap could not have made that distinction easily, since he first learned of it from Tarski, whom he did not meet until 1930 (Carnap 1963a, 60, 30).) If the phenomenal and physical facts are given by sentences (that is, syntactically), Carnap claims that for every observational sentence ω, the correspondence rules entail that ω is true if and only if one of infinitely many theoretical sentences is true. A special case of this situation occurs when the theory entails that the infinitely many theoretical sentences are true if and only if *one* other theoretical sentence is true. For example, a physical theory may entail that one of infinitely many descriptions of physical microstates is true if and only if some description τ of a physical macrostate is true. In this special case, the correspondence rules together with the theory entail that ω is equivalent to τ, that is, the observational sentence can be translated into a theoretical sentence.[41] If this holds for *all formulas* of the observational language, it is known from the theory of definition that every observational term can be explicitly defined in theoretical terms. Thus if Carnap assumed the translatability of all observational formulas, his earliest view is equivalent to (ii), that all observational terms are definable in theoretical terms (although without the qualification that this *may* be the case). Otherwise, his later view was *stronger* than his earlier view.

Conversely, Carnap assumed that for every theoretical sentence τ, the correspondence rules entail that if τ is true, then exactly one observational sentence is true, but the converse does not hold. Thus Carnap's earliest view was that no theoretical sentence is equivalent to an observational sentence, or in other words, that no theoretical sentence is translatable into an observational one. It is known from definition theory that then no theoretical term is definable in observational terms. Hence Carnap's earliest view entails (i).

If the phenomenal and physical facts are given by model theoretic structures (that is, semantically), Carnap claims that a structure for the theoretical vocabulary can be expanded in at most one way to a model of the correspondence rules, since

[41] Note that Carnap does not use 'translation' in the modern sense.

there is exactly one observational structure for each theoretical one. Thus the correspondence rules entail an explicit definition for each observational term (according to Beth's theorem) for correspondence rules in first order logic and (according to theorem 3 by Tarski 1935) for a finite number of correspondence rules in finite type theory (cf. Leivant 1994, §5.1). On the other hand, each structure for the observational vocabulary can be expanded to infinitely many models of the correspondence rules. This means that the correspondence rules do not entail an explicit definition for each theoretical term (according to Padoa's theorem).

In other words, already in his very first publication on the structure of theories, Carnap claims that (i) theoretical terms cannot be defined in observational terms, and he might also claim that (ii) observational terms can be defined in theoretical terms. Later, however, Carnap only went so far as to state that given the contemporary state of science it *seems* that in the future, all observational terms can be explicitly defined—see §3.1). Of course, when he wrote "Aufgabe", definition theory and formal semantics were not yet developed enough to phrase this consequence so clearly, but this is what his position entails. Furthermore, Carnap (1936, 168) referred to Tarski's article (Tarski 1935) very soon thereafter, and so could have seen the relation then.

In his "Autobiography", Carnap (1963a, 15–16) notes the close connection of his later work to the phase space approach in *Begriffsbildung*, but he does remark upon the similarly close connection to the relation of physical magnitudes and experiences described in *Begriffsbildung*: As in "Aufgabe", the relation is given through a translation (*Begriffsbildung*, 60):

> The retranslation of the pure statements about numbers of the abstract physics into statements about qualities is possible because a specific distribution of values is unambiguously assigned to specific physical qualities, and finally specific perceptual qualities.[42]

An example would be the translation of the quadruples of numbers into average kinetic energies as described in "Beobachtungssprache", from there into temperature as described in the *Introduction*, and finally into heat experience as described in *Probability*.

Since the assignments are unique, *Begriffsbildung* in effect repeats the claim that observational states are uniquely determined by physical states. Carnap does not explicitly state that the reverse is not the case, but repeats his claim from "Aufgabe" and "Dreidimensionalität" that there is a perception threshold (Carnap 1926, 17):

> At first, we could try [...] to assign a higher temperature to one

[42] "Die Rückübersetzung der reinen Zahlenaussagen der abstrakten Physik in Qualitätsaussagen ist möglich, weil einer bestimmten Werteverteilung bestimmter Zustandsgrößen stets eindeutig bestimmte physikalische Qualitäten, schließlich bestimmte Sinnesqualitäten zugeordnet sind."

> body than another when it effects a stronger perception of warmth [...]. However, this assignment procedure would prove itself to be inexpedient in light of the fact of "heat exchange". This is because, when two bodies come into contact [and] only one changes perceptibly, then (for the sake of later laws of nature in connection with the concept of specific heat) the other one is assigned a converse change of imperceptible magnitude. [43]

This illustration of the perception threshold led Carnap in the earlier two articles to conclude that the theoretical states are not uniquely determined by the observational states. In *Begriffsbildung*, Carnap does nothing to counter this conclusion. He thus does not give a perfectly clear endorsement of the thesis about the indefinability of theoretical terms that he would later hold; this clear endorsement is given in the *Aufbau*, of all places—even though it is typically seen as the most explicit endorsement of the explicit definability of all theoretical terms.

The "physicoqualitative correlation" that Carnap describes in the *Aufbau*, §136, consists, first, of a one-to-one correspondence between the world points of physics (the space-time points of *Begriffsbildung*) and the world points of the perceptual world. Second, the many-one relation of "Aufgabe" and *Begriffsbildung* holds between the physical magnitudes and the perceptual qualities. While each physical state can be assigned a perceptual state,

> in the opposite direction, the correlation is not unique; the assignment of a quality to a world point in the perceptual world does not determine which structure of state magnitudes is to be assigned to the neighborhood of the corresponding physical world point of the world of physics; the assignment merely determines a class to which this structure must belong.

In his autobiography, Carnap (1963a, 19, §2) notes the relation of this part of the *Aufbau* to his later work:

> For the construction of the world of physics on the basis of the temporal sequence of sensory qualities, I used the following method. A system of ordered quadruples of real numbers serves as the system of co-ordinates of space-time points. To these quadruples, sensory qualities, e. g., colors, are assigned first, and then numbers as values of physical state magnitudes. [...] In general, I introduced concepts by

[43] "Zunächst könnten wir versuchen, [...] einem Körper dann eine höhere Temperatur zuzuschreiben als einem anderen, wenn er eine stärkere Wärmeempfindung [...] hervorruft. Dieses Zuschreibungsverfahren würde sich jedoch als unzweckmäßig erweisen angesichts der Tatsache des „Wärmeausgleichs". Wenn nämlich zwei Körper in Kontakt treten [und] nur bei dem einen eine Änderung wahrnehmbar [ist], so wird, (späteren Naturgesetzen zuliebe, die bei dem Begriff der spezifischen Wärme auftreten) dem andern eine entgegengesetzte Änderung von nicht wahrnehmbarer Größe zugeschrieben."

> explicit definitions, but here the physical concepts were introduced instead on the basis of general principles of correspondence, simplicity, and analogy. It seems to me that the procedure which is used in the construction of the physical world, anticipates the method which I recognized explicitly much later, namely the method of introducing theoretical terms by postulates and rules of correspondence.

Typically, the discussion of Carnap's attempt in the *Aufbau* to define all theoretical terms in observational ones has followed Quine (1969b), and focused on the first step at which Carnap's procedure fails: The construction of the phenomenal world through simplicity and analogy. Carnap (1963a) himself treats the step to the phenomenal world together with the step to the physical world, and thereby brushes over the point I have made here: Even if Carnap had succeeded in explicitly defining the concepts of the perceptual world, the physical magnitudes and anything that relies on physical magnitudes for its construction are not explicitly definable, according to his own position.

Contrary to this result, Friedman (1992) claims that Carnap rather gives a method for arriving at explicit definitions in spite of the one-to-many relation between phenomenal and physical facts. Friedman (1992, 21) wonders how it is possible, "as Carnap claims in §179 [of the *Aufbau*], to translate all statements of science into 'statements about the basic objects, namely, about relations between elementary experiences'". And Friedman (1992, 21–22) suggests the following answer:

> Section 136 of the *Aufbau* refers us to "Aufgabe" for more details on the physical-qualitative coordination. Although Carnap repeats the claim that the coordination between "phenomenal facts" and corresponding state-magnitudes is only unique [. . .] in the direction from the latter to the former, he there outlines a procedure for nonetheless approximating to a unique assignment of physical state-magnitudes by focusing on a small neighborhood of a given phenomenally characterized space-time point and working back and forth using the laws of physics (1923, pp. 102–03). The crucial point is that the laws of physics, together with an unambiguous determination of phenomenal qualities from physical state-magnitudes, provide a methodological procedure for narrowing down the ambiguity in the assignment of physical state-magnitudes: in principle, a unique assignment is thereby constructed after all. It appears, then, that in §136 of the *Aufbau* Carnap intends to achieve an unambiguous constitution of the world of physics by just such a methodological procedure.

I do not think that this is Carnap's claim. The method to which Friedman refers is described in a passage in which Carnap relaxes the idealizing assumptions about the third volume of an ideal physics, the complete knowledge of the state of the world. "Then the task is to calculate from the observed state of a bounded area, namely our environment in space-time, the state of a different space-time area."[44] As a technical problem of this task, Carnap notes that to calculate even the state of an arbitrarily small area just for one second would demand knowledge of the state of the world in a 300 000 km radius. The bigger problem is that in principle the physical state of an area cannot be uniquely determined from observations, because the dictionary contains only one-many relations.

The method to come to predictions is given in the following passage, which I quote almost in full:

> For the following reason, a physics that is very far away even from this more cautious fiction can make predictions based on observations at all: To a specific observation report belong indeed an infinite set of physical states of the area, and therefore a set of states of the same cardinality for the future state under investigation [. . .]. But in many cases, retranslating this infinite set of physical states into perceptual contents results in a comparably small set of perceptual contents, which in advantageous cases form a continuous region of qualities (i. e., a domain of similar colors). First, the aspiration is now to make observations that do not result in several unconnected regions of qualities for the future point in time, and then to narrow down the range of the one region of qualities as much as possible. [. . .] In special cases and if the time interval is not too big, [the deficiencies of the prediction] can be completely eliminated, that is, a unique prediction can be reached. [. . .] This holds for the prediction of *perceptual contents*, which is the only one demanded in practice. In contradistinction, science always stays infinitely far away from the unique prediction of physical states even for arbitrarily small time intervals.[45]

[44] "Dann lautet die Aufgabe: aus dem beobachteten Zustande eines beschränkten Bereiches, nämlich unserer raumzeitlichen Umgebung, den Zustand eines anderen Raum-Zeitbereiches zu berechnen."

[45] "Daß trotzdem eine Physik, die auch von dieser vorsichtigeren Fiktion noch sehr weit entfernt ist, überhaupt Vorausberechnungen auf Grund von Beobachtungen anstellen kann, hat folgenden Grund: Zu einem bestimmten Beobachtungsbefund gehört allerdings eine unendliche Menge physikalischer Zustände des Bereiches, und damit auch eine gleichmächtige Menge solcher Zustände für den zu erkundenden zukünftigen Augenblick [. . .]. Aber bei Rückübersetzung dieser unendlichen Menge physikalischer Zustände in Empfindungsinhalte ergibt sich in vielen Fällen eine verhältnismäßig kleine Menge von Empfindungsinhalten, die in günstigen Fällen ein stetiges Qualitätsbereich bilden (d. h. ein Bereich ähnlicher Farbtöne). Das Bestreben ist nun zunächst darauf gerichtet, die Beobachtungen so anzustellen, daß sich nicht mehrere unzusammenhängende Qualitätsbereiche für den zukünftigen Zeitpunkt ergeben, und dann darauf, die Grenzen des einen Qualitätsbereiches möglichst zu verengern. [. . .] In besonderen Fällen für nicht zu lange Zeitabstände können [die Mängel der Voraussage] völlig beseitigt, also Eindeutigkeit der Voraussage, erreicht werden. [. . .] Das gilt für die praktisch allein verlangte Voraussage von *Wahrnehmungsinhalten*. Von der eindeutigen Voraussage physikalischer Zustände dagegen bleibt die Wissenschaft auch bei noch so kleinen Zeitabständen immer unendlich weit entfernt."

So, contrary to Friedman's claim, Carnap does not give a method for explicitly defining the terms of physics. Rather, he points out that for some cases, exact prediction of a future observation in a small region of space-time, but not of a future physical state, is possible. His argument against the prediction of physical states rests on the one-many relation between observations and physical states, the assumptions that physical states can be determined only through observations (this is implicit), and that for each physical state, there is exactly one physical state in the future (a set of physical states evolves over time into a set of physical states of the same cardinality). Since at any point in time, one can only determine an infinite set of physical states, any prediction can therefore also only determine an infinite set of physical states. (Carnap does not consider the possibility of using observations at more than one point in time to narrow down the set of physical states.)

Carnap states that, for practical purposes, it is only necessary to predict future perceptual contents, not future physical states. And this, he argues, is possible even uniquely if the future point in time is not too far away and the circumstances are also favorable in other (not further determined) respects. Otherwise, uniqueness has to be given up, so that the goal is then to come to a prediction of one continuous and as small as feasible range of possible perceptions.

In "Dreidimensionalität", 123–124, Carnap denies not only the possibility of unique predictions of phenomenal contents, but even of any restrictions on the possible elements of experience at all. Nonetheless, probabilistic predictions can be given:

> While [. . .] neither deterministic nor restricting laws hold, there are *frequency functions* both for the spatial distribution of the simultaneous elements and the chronological series. The possibility of predictions is based on such frequency functions of a complicated kind.[46]

3.6.2 Carnap's bipartition of the language

I have already mentioned in the introduction of this chapter that some defenses of the Received View can be put forth based on Suppe's exposition. Suppe himself also defends the Received View against some criticisms, most notably those by Putnam (1962). Putnam argues, first, that there is no acceptable conceptualization of 'partial interpretation of theoretical terms' and, second, that the bipartition of the vocabulary of science into observational and theoretical terms is not possible. The second charge rests on the claim that in actual scientific theories, the theoretical terms of the Received View are not the terms that are introduced by theories

[46] "Es gelten [. . .] zwar weder determinierende noch beschränkende Gesetze, aber doch *Häufigkeitsfunktionen* sowohl für die räumliche Verteilung der gleichzeitigen Elemente, als auch für die zeitliche Reihe. Auf solchen Häufigkeitsfunktionen verwickelter Art beruht die Möglichkeit von Voraussagen."

and that observation reports do not only contain observational terms.

Suppe incisively criticizes both claims, the first one by devising a partial interpretation (Suppe 1971),[47] the second one by noting that Putnam's criticism of the bipartition of the vocabulary relies on ordinary language, while the Received View assumes a rational reconstruction, which could allow the *introduction* of a bipartition (Suppe 1972). In other words, Putnam's second charge rests on a confusion of the rational reconstruction of theories with the description of scientific theory and practice.[48] With respect to observation reports, Carnap (1931a, 437–438) himself already notes that

> With the "primary" protocol I mean the one we would get if we were to keep writing and processing of the protocol strictly separate, that is, include no sentences in the protocol that were indirectly attained. [...] A primary protocol would be very cumbersome. It is hence practically expedient that the formulations of the protocol already use inferred designations.[49]

Thus Carnap already assumed that, in the ordinary language sense, observation reports do not only contain observation terms.

According to Suppe (1974a, 83), it is hard to say what a bipartition introduced in a rational reconstruction would look like, however, because "the ways advocates of the Received View have attempted to specify the distinction fail to specify it precisely or in such a way that their paradigm examples of observation terms and theoretical terms clearly do qualify as such". So first, it is not always clear which terms are observable and which are theoretical, and second, those terms that are allegedly clearly observable or clearly theoretical are not.

An example of the imprecision of the attempts is, according to Suppe, Carnap's explication of 'observable' as "properties directly perceived by the senses" (Carnap 1966, 255), because it is not clear whether the property has to be always directly perceivable or just sometimes. In the former case, Carnap's examples of 'blue' and 'warmer' as observable terms (Carnap 1966, 259) are incorrect, Suppe (1972, 6) argues, because sufficiently small blue objects are not observable, and an object of −250°C is warmer than one of −273°C. Therefore, a property has to be observable if it is sometimes directly perceivable. Then 'being a gas' and 'being electrically charged' are observable because sometimes, it is possible to smell a gas or receive a shock. One could doubt this conclusion arguing that one does not smell that

[47] Note, however, that Suppe adds the possibility of interpreting theoretical terms directly. This is not necessary for his defense, and also historically inaccurate, as I spell out in §3.7.

[48] In still other words, Putnam criticizes an application of artificial language philosophy for being an incorrect application of ordinary language philosophy.

[49] "Mit dem „ursprünglichen" Protokoll ist dasjenige gemeint, das wir erhalten würden, wenn wir Protokollaufnahme und Verarbeitung der Protokollsätze im wissenschaftlichen Verfahren scharf voneinander trennen würden, also in das Protokoll keine indirekt gewonnenen Sätze aufnehmen würden. [...] Ein ursprüngliches Protokoll würde sehr umständlich sein. Daher ist es für die Praxis zweckmäßig, daß die Formulierung des Protokolls schon abgeleitete Bestimmungen verwendet."

something is a gas, but rather only perceives certain manifestations of being a gas. But this argument cannot be correct, because analogously one does not directly perceive that an object is hard, but only certain manifestations of it being hard (Suppe 1972, 7, n. 7).

Suppe's argument fails, and is in fact already anticipated and countered by Carnap himself. For Suppe uses the term 'warmer than' simply as a synonym for 'higher temperature than'. But this is not what Carnap has in mind, since he points out in *Begriffsbildung*, 35–36, that because of the perception threshold, there are cases of higher temperature that do not feel warmer, and that indeed *temperature* is an abstract concept that corresponds to no single physical measurement device (and thus *a fortiori* not to human perception) (see §3.6.6). In "Aufgabe", Carnap is very explicit that whether concepts like *warmer* or *blue* apply is a matter of perception. The contents of perception do not occur at all in theoretical physics, which is not obvious because terms like 'pressure' and 'heat' are used in both domains ("Aufgabe", 99). This confusion of physical and observational concepts is probably easiest to see in Suppe's response to the criticism that one does not smell that something is a gas, but only a specific gas smell. He claims that then, one would also not feel that an object is hard, but only one manifestation. Obviously, he uses 'hard' in the physical sense, while the manifestation of hardness then would be exactly the meaning of 'hard' that Carnap is after.

That Carnap fails to specify precisely the distinction between observational and theoretical terms is unsurprising, for one because what one person can perceive may be imperceptible for another. In "Testability", Carnap (1935a, §16) discusses "sufficient reduction bases", sets of terms to which all other can be reduced with the help of reduction sentences. According to Carnap, the visual sense provides a sufficient reduction basis, and so does the visual sense restricted to shades of gray or even only black and white, since colors can be determined with the help of a spectrometer with a gray or black-and-white display. Since a spectroscope may use tangible scale-marks and pointers and thus be used by a blind and deaf person, the sense of touch provides a sufficient reduction basis as well. Thus different persons may have different observation terms based on their abilities. What is more, according to Carnap (1932, 224) each person can *choose* the observation terms depending on the context:

> *Every concrete sentence of the physicalistic system-language can serve as protocol-sentence in some situation.* Let G be a law (i. e., general sentence of the system-language). To check G, one has to derive concrete sentences that relate to specific space-time points [. . .]. From these concrete sentences one has to derive further concrete sentences using other laws and logico-mathematical rules of derivation, until one reaches sentences that one wants to accept in the specific case. And it is a matter of choice which sentences one intends to use as such endpoints of the reduction, that is, as protocol-sentences. When-

> ever one wants to—for instance, if there are doubts or one wants to consolidate the scientific hypotheses more securely—one can reduce those sentences previously accepted as endpoints again to other ones and choose those to be endpoints. In any case, this reduction for the purpose of checking *G* has to stop at some point. But in no case does this reduction have to stop at a specific point. From any sentence, one can go further back; *there are no absolute primary sentences* for the construction of science.[50]

Carnap thus explicitly considers it a matter of choice which sentences are taken to be primitively interpreted (cf. Oberdan 1990). Furthermore, Carnap's partition of the set of concepts into different levels in the *Aufbau* already suggests that 'observability' is not a categorical property but a comparative relation, so that concepts can be more or less observable. In *Begriffsbildung*, Carnap (1926, ch. III) accordingly speaks of abstract distributions of values first being assigned "physical qualities, and finally perceptual qualities". Similarly, in *L&M*, Carnap (1939) describes an "order of definitions", leading in successive steps from observational to theoretical terms and vice versa (cf. figure 3.1). When giving examples of correspondence rules, he notes that "if 'iron' is not accepted as sufficiently [observational], the rules can be stated for *more* [observational] terms" (Carnap 1939, 207, my emphasis).

3.6.3 Suppe's bipartition of the language

Suppe's solution to the problems of determining which terms are observable is to classify not properties, but occurrences of properties as observable or unobservable. With this bipartition, Suppe (1974a, 84) suggests,

> one could employ separate terms terms of *L* to refer to observable entities or attribute occurrences and others to refer to nonobservable ones—for example, one might employ 'red_o' to refer to observable occurrences of the property *red* and 'red_t' to refer to nonobservables.

[50] "*Jeder konkrete Satz der physikalistischen Systemsprache kann unter Umständen als Protokollsatz dienen*. *G* sei ein Gesetz (d. h. allgemeiner Satz der Systemsprache). Zum Zweck der Nachprüfung sind aus *G* zunächst konkrete, auf bestimmte Raum-Zeit-Stellen bezogene Sätze abzuleiten [. . .]. Aus diesen konkreten Sätzen sind mit Hilfe anderer Gesetze und logisch-mathematischer Schlußregeln weitere konkrete Sätze abzuleiten, bis man zu Sätzen kommt, die man im gerade vorliegenden Fall anerkennen will. Dabei ist es Sache des Entschlusses, welche Sätze man jeweils als derartige Endpunkte der Zurückführung, also als Protokollsätze verwenden will. Sobald man will, – etwa wenn Zweifel auftreten oder wenn man die wissenschaftlichen Thesen sicherer zu fundieren wünscht, – kann man die zunächst als Endpunkte genommenen Sätze ihrerseits wieder auf andere zurückführen und jetzt diese durch Beschluß zu Endpunkten erklären. In jedem Fall muß man mit der Zurückführung zum Zweck der Nachprüfung irgendwo haltmachen. In keinem Fall aber ist man gezwungen, an einer bestimmten Stelle haltzumachen. Man kann von jedem Satz aus noch weiter zurückgehen; *es gibt keine absoluten Anfangssätze* für den Aufbau der Wissenschaft."

All observational terms that also apply to unobservable occurrences in the original theory would therefore be split up into two, one term that applies only to observable, one term that applies only to unobservable instances.[51] Note that this comparably simple introduction of new terms becomes vastly more complicated when more-place predicates and functions are considered. It is indeed not obvious that for all many-place predicates that can apply to both observable and unobservable objects at once, two new predicates can be found so that one applies only to unobservable objects, the other only to observable objects. This is also what Suppe (1972, 8) notes:

> Some provision will have to be made to allow comparisons between nonobservable and observable occurrences of properties, and also for comparative relations whose applications straddle the observable, nonobservable border. It is not clear whether this proposal is workable; but it is clear that it will be rather complicated if it is.

Even if the proposal is workable, however, Suppe (1974a, 84) argues that it might be a Pyrrhic victory:

> Such an artificial drawing of the observational–theoretical distinction certainly is going to make the Received View reconstruction of a theory very complex, introducing a degree and kind of complexity not found in theories as they are employed in actual scientific practice.

And this added complexity is not worth the cost, because the observational–theoretical bipartition does not mark anything of epistemic interest: Suppe (1974a, 85) concurs with Putnam (1962) that the distinction between theoretical terms, those that come from a scientific theory, and non-theoretical terms is important, but that the distinction between observational and non-observational terms does not capture this. And observation reports, which are also epistemically significant, often employ non-observational terms. Therefore, the bipartition of a language into observational and theoretical terms is not useful.

The conventionality of the distinction between observational and theoretical terms discussed in §3.6.2 already counters the argument that the observational–theoretical distinction does not mark anything of interest. A clear response to Suppe's claim that the bipartition of the vocabulary must lead to an overly complex language can be derived from the fact that Carnap considered it possible to define observational terms in theoretical terms, but not *vice versa*.

Giving an explicit definition of phenomenal terms in observational terms does not overstep the restrictions on correspondence rules that Carnap set in

[51] Putnam (1962, 241, n. 2) dismisses this option on the grounds that "'[b]eing an observable thing' is, in a sense, highly theoretical and yet applies only to observables!" This objection fails because first, not all senses of 'theoretical' are relevant to the Received View, and second, 'being an observable thing' does not occur as a predicate in Suppe's reconstructed observational language.

"Aufgabe". And this allows a reconstruction of scientific language in which a term that is applied both to observable and unobservable objects can be replaced by one observable and one theoretical term, but in a far less problematic way than Suppe suggests. Suppe assumes that 'red' becomes 'red_o' and 'red_t', where the extensions of the two new terms are disjoint:

$$\forall x[Rx \leftrightarrow (R_o x \vee R_t x)] \wedge \forall x(R_o x \rightarrow \neg R_t x)\,, \tag{3.2}$$

where 'R_o' stands for 'red_o' etc. But instead, 'red' could be considered a theoretical term (R) that applies to anything that reflects light of a certain spectrum, even if too small to see (which is Suppe's implicit assumption). 'red_o' could then be the restriction of this term to those objects that are actually visible (V) to the naked eye as determined by the whole of science, which includes physiology:[52]

$$\forall x[R_o x \leftrightarrow (Rx \wedge Vx)]\,. \tag{3.3}$$

There are now at least two ways in which one can define 'perceptibly warmer than' (W) if one again follows Suppe and assumes that being warmer is the same as having a higher temperature (T). One way is to simply restrict all comparisons to those objects that can still be touched (S) because they are neither too hot nor too cold, that is

$$\forall x \forall y[Wxy \leftrightarrow (Sx \wedge Sy \wedge Tx > Ty)]\,. \tag{3.4}$$

Even ignoring the perception threshold, this definition might simply be physiologically incorrect, however, because objects that have too high or too low a temperature probably still feel very hot or very cold, only that there is no difference felt between any of the objects that are too hot or those that are too cold. So a better definition of 'perceptibly warmer than' (including a perception threshold of t) would be

$$\forall x \forall y\big(Wxy \leftrightarrow [\exists z(Sz \wedge Tx \geq Tz \wedge Tz \geq Ty) \wedge Tx > Ty + t]\big)\,. \tag{3.5}$$

All the objects that have too high a temperature to be touched could then be defined as of equal perceptible warmth, and would be warmer than any of the objects that can be touched or have too low a temperature to be touched. Analogously for temperatures that are too low. Of course, the physiologically correct formula is likely to be still more complicated (with, for instance, t being dependent on temperature and individual), but not as an artifact of the Received View's formalism, but because of the complexity of the relation.

The "unusual consequences" of an artificial introduction of a bipartition into natural language that Suppe (1972, 7) identifies then become less unusual as well.

[52]Note that V is not an observational term and thus Putnam's criticism (see n. 51) does not apply.

> If I take an *O*-red [red_o] object of minimal area and smash it to pieces, the pieces will not be *O*-red, but rather *N*-red [red_t]. And [i]f I combine together a number of *N*-red blood specks I will obtain an *O*-red blood patch.

With the definition (3.3) suggested above, this would be true, but missing the important point: The red_o object of minimal area already *is* red_t, and its parts stay that way when smashed. But smashing the object results in objects that are not longer observable, and therefore are not red_o. Similarly, a red_o blood patch is also red_t. These considerations show that Suppe's method for devising a bipartition of the vocabulary rests (like Putnam's criticism) on the incorrect assumption that theoretical terms cannot apply to observable objects.

Finally, Suppe's method rests on the assumption that the vocabulary of a rational reconstruction must take as its starting point natural language. But this does not have to be the case, as Bohnert (1963, §II) and Carnap (1963c, 938–939) argue (see §2.4).

3.6.4 Correspondence rules as components of theories

Closely connected to the bipartition of the language is the role of correspondence rules, the rules that connect the observational terms with the theoretical terms. After an exposition of criticisms by Schaffner (1969) and Suppes (1962, 1967), Suppe (1974a, 109, footnote omitted) concludes that

> the Received View's treatment of correspondence rules is inadequate in three important respects: first, it mistakenly views them all as components of theories, rather than as auxiliary hypotheses; second, the account ignores the fact that correspondence rules often constitute explanatory causal chains which employ other theories as auxiliary hypotheses; third, insofar as correspondence rules characterize the experimental connections between phenomena and theory, the account is oversimple and epistemologically misleading.

Suppe's three criticisms of the role of correspondence rules are echoed, for example, by Thompson (1988, §3).

To incorporate correspondence rules in the theories is mistaken, Suppe argues, because if correspondence rules are part of the theory, then any change in the correspondence rules is a change in the theory. Since correspondence rules incorporate experimental methods, any new procedure to apply the theory to phenomena leads to a new theory. But this, he states, is at best misleading, since the theory stays the same, even if it can be applied in a new way.

This first of Suppe's criticisms is probably best compared to Carnap's note in "Aufgabe", 103–104, that it is a matter of linguistic convention whether correspondence rules are considered part of the theories or not (see page 97). At worst, Suppe's criticism could then be seen as an argument why it would be inexpedient to consider correspondence rules parts of the theories. Furthermore, Carnap treated the correspondence rules as separate from the theories most of the time simply because he assumed that correspondence rules could incorporate other theories, as I will argue next.

3.6.5 Correspondence rules and other theories

According to Suppe, the application of a theory typically relies on theoretical hypotheses, that is, hypotheses that derive from other theories. To apply a theory that describes movement in a vacuum to real experiments, for example, one needs theoretical hypotheses that describe the effect of air resistance on the predictions of the theory (Suppe 1972, 15). That the Received View does not take this into account is Suppe's second criticism.

But there is no reason to believe that in the Received View, correspondence rules cannot contain laws from other theories, as Suppe claims. Quite to the contrary, Carnap (1923, 103) specifically notes in "Aufgabe" that the connection from the axioms of a theory all the way down to the observations would need the whole of science which, at this point, does not exist (see §3.2). In the *Aufbau*, §106, Carnap is similarly explicit: All the definitions he suggests in the "constructional system" are based on the results of empirical sciences:

> As concerns the content of our constructional system, let us emphasize that it is only a tentative example. The content depends upon the material findings of the empirical sciences; for the lower levels in particular, upon the findings of the phenomenology of perception, and psychology. The results of these sciences are themselves subject to debate; since a constructional system is merely the translation of of such findings, its complete material correctness cannot be guaranteed.

In *Begriffsbildung*, Carnap even spells out specific laws that are to be part of the definitions of specific concepts, for example the law of thermal expansion that is taken into account for the definition of length (see §3.6.6). Finally, the claim that correspondence rules cannot incorporate other theories flies in the face of the idea of the unity of science, a core notion of logical empiricism (cf. Cat 2011, §4).

3.6.6 The complexity of correspondence rules

In his third criticism, Suppe charges that the role of correspondence rules in the Received View is overly simple and misleading because it does not take into account the hierarchy of theories that, according to Suppes (1962, 260), mediates between a theory and the phenomena:

> One of the besetting sins of philosophers of science is to overly simplify the structure of science. Philosophers who write about the

> representation of scientific theories as logical calculi then go on to say that a theory is given empirical meaning by providing interpretations or coordinating definitions for some of the primitive or defined terms of the calculus. What I have attempted to argue is that a whole hierarchy of models stands between the model of the basic theory and the complete experimental experience. Moreover, for each level of the hierarchy there is a theory in its own right. Theory at one level is given empirical meaning by making formal connections with theory at a lower level.

'Model' here refers to the model of a theory, so that for each level including the level of data, there is a theory, and through formal connections of these theories, the model of the basic theory is eventually connected to the model of the data. Many of the connections between the levels are statistical.

Statistical connections are indeed absent in most of the discussions of correspondence rules, and I will come back to this point in §3.8.2. Overall, however, it is rather surprising that Suppes and Suppe charge the Received View with oversimplifying the relation between theory and observation. For in "Aufgabe", Carnap does not even consider science to be developed enough to give the full correspondence rules, and in the *Aufbau*, Carnap states that the correspondence rules are only formalizing empirical findings, so that the correspondence rules are as complicated as the sciences determine them to be. In *Begriffsbildung*, Carnap spells out in some detail what is needed to develop quantitative concepts from qualitative ones and how to incorporate correction factors into the formation of new concepts like *length*. In other words, he gives a description of how a theory containing the concept *length* connects to quantitative or qualitative observations, correction factors and all.

The hierarchical structure of the connection between theory and observations is already described in *Begriffsbildung* (see §3.5) and is a core motif in the *Aufbau*, with its succession of steps to define the terms of the physical world. It also occurs in Carnap's later writings (e. g., Carnap 1939) (cf. figure 3.1).

But even leaving out the question of correction factors and levels of theories, Carnap's and Hempel's view on the relation between observational and measurement terms is in line with contemporary views. Already in *Begriffsbildung*, 22–23, Carnap discusses the relations and functions that have to hold for the objects of a domain so that one can assign a physical magnitude to them. These are

> 1. The topological definition of a magnitude
>   Precondition and reason for the introduction of a magnitude is an empirical finding of the sort that two relations hold between the objects (bodies, processes) of some domain: one transitive, symmetrical and one transitive, asymmetrical. [. . .] One then stipulates that numbers have to be assigned to the objects in such a way that

a) objects in the transitive, symmetric relation are assigned the same number
b) an object in the transitive, asymmetric relation to another is assigned a lower number than the other object.

2. The metrical definition of the magnitude through three stipulations [...]
   a) One has to choose a *scale*, that is, stipulate when two scale-distances, that is, the difference between two values of the respective magnitude, should be considered equal.
   b) One has to choose a *zero point* of the scale, that is, stipulate when an object is to be assigned a value of zero.
   c) One has to choose a *unit*, that is, stipulate when an object is to be assigned the value of one.[53]

Compare this with Carnap's discussion of the introduction of numerical values in the *Introduction*, 63–65:

> Rule 1 [...] states that, if the relation E_M holds between objects a and b, the two objects will have equal values of the magnitude M. In symbolic form:
>
> $$\text{If } E_M(a,b), \text{ then } M(a)=M(b)\,. \tag{3.6}$$
>
> Rule 2 [...] says that, if the relation L_M holds between a and b, the value of the magnitude M will be smaller for a than for b. In symbolic

53

"1. Die topologische Definition einer physikalischen Größe
Voraussetzung und Anlaß für die Einführung einer Größenart ist ein erfahrungsmäßiger Befund von der Art, daß zwischen den Objekten (Körpern, Vorgängen) irgend eines Bereiches zwei Beziehungen bestehen: eine transitive, symmetrische und eine transitive, asymmetrische. [...] Es wird dann bestimmt, daß die Zuschreibung der Zahlen zu den Objekten des Bereiches so geschehen soll, daß
   a) den Objekten, zwischen denen die transitive, symmetrische Beziehung besteht, gleiche Zahlen zugeschrieben werden,
   b) einem Objekt, daß in der transitiven, asymmetrischen Beziehung zu einem anderen steht, eine niedere Zahl als dem andern zugeschrieben wird.
2. Die metrische Definition der Größe durch drei Festsetzungen [...]
   a) Es ist eine *Skalenform* zu wählen, d. h. eine Festsetzung darüber zu treffen, wann zwei Skalenstrecken, also zwei Größendifferenzen der betreffenden Größe, als gleich gelten sollen.
   b) Es ist ein *Nullpunkt* der Skala zu wählen, d. h. eine Festsetzung darüber zu treffen, wann einem Objekt der Größenwert Null zugeschrieben werden soll.
   c) Es ist eine *Einheit* zu wählen, d. h. eine Festsetzung darüber zu treffen, wann einem Objekt der Größenwert Eins zugeschrieben werden soll."

> form:
>
> $$\text{If } L_M(a,b), \text{ then } M(a) < M(b)\,. \qquad (3.7)$$
>
> [...] Rule 3 tells us when to assign a selected numerical value, usually zero, to the magnitude we are attempting to measure. [...]
>
> Rule 4, usually called the rule of the unit, assigns a second selected value of the magnitude to an object by specifying another easily recognized, easily reproducible state of that object. This second value is usually 1, but it may be any number different from the number specified by Rule 3. [...]
>
> Rule 5 [...] specifies the empirical conditions ED_M, under which we shall say that [...] the difference between any two values of the magnitudes for a and for b is the same as the difference between two other values, say, for c and for d. This fifth rule has the following symbolic form:
>
> $$\text{If } ED_M(a,b,c,d), \text{ then } M(a) - M(b) = M(c) - M(d)\,. \qquad (3.8)$$

Since Carnap (1966, 54–55) assumes E_M to be transitive and symmetric and assumes L_M to be transitive and asymmetric, Rule 1 corresponds to condition 1a) from *Begriffsbildung*, Rule 2 to condition 1b), Rule 3 to 2b), Rule 4 to 2c), and Rule 5 to 2a). As in *Begriffsbildung*, 23, Carnap (1966, 65) considers Rule 5 to be the most important one of the last three rules (conditions 2a–c).

In the *Introduction*, 54–57, Carnap points out that E_M and L_M must be more than just transitive and symmetric or asymmetric to allow the introduction of a numerical magnitude. For a thorough discussion, he refers to Hempel's *Concept Formation*, §§10–11. Hempel's discussion is based on an article by Suppes (1951) on axioms for extensive quantities, which is a contribution to standard measurement theory (cf. Krantz et al. 1971, §§3.1, 3.5.1). Carnap's discussion, both its formal assumptions and its methodological considerations were at the core of a course at the George Washington University Department of Continuing Education on calibration laboratory management and a series of seminars given in the 1970s at the National Bureau of Standards (Simpson 1981, 291). In an article based on this course, Simpson (1981, abstract) develops Carnap's theory of measurement and "its applications to metrology [...] as an aid to program planning and evaluation", e. g. quality control. Clearly, then, the Received View on correspondence rules is not far from actual science, at least as far as measurement is concerned.

In the case of axiomatizations, it turned out that Carnap had often relied on the same kind of formalizations that would later be used in the Semantic View. Analogously, the actual correspondence rules that Carnap and Hempel considered are in line with those of the Semantic View (as developed by Suppes) and with measurement theory in general. Therefore Carnap and Hempel did not rely on notions of the relation between observational terms and theoretical (measurement) terms that are oversimplified or naïve compared to those of the Semantic View. (Note also how nicely the distinction between the observational terms E_M and L_M and similar terms (cf. Krantz et al. 1971, §§1.1–1.2) on the one hand, and the theoretical term M (cf. Krantz et al. 1971, §§1.1–1.2) on the other fits into the Received View's bipartition of the language.)

Of course, the correspondence rules are not discussed in such depth in all publications; "Aufgabe", for example, contains almost no information about correspondence rules, and those that it does contain are indeed very simple. But that does not mean that correspondence rules *must* be sketchy, or even that they were generally sketchy in the Received View.

I now want to go further and argue that both Carnap and Hempel had accurate ideas about the *development* of measurement concepts. In *Begriffsbildung*, 35–36, Carnap describes the introduction of a scale on the basis of a conception of 'warmer than'. This is done by choosing a value of zero for the thawing point of water, a value of 100 for the boiling point of water, and choosing the change of temperature to be proportional to the change of volume of mercury in this range. Carnap points out that, since the concept of volume depends on the concept of length, and the concept of length based on rigid bodies has a correction factor for thermal expansion and thus involves temperature, this may seem circular. But this circularity can be avoided in two ways: First, simply by ensuring that length is defined relative to a fixed temperature of the measuring rod. Second, by taking into account the different levels of precision: The initial concept of length can be determined without a correction factor, then be used to determine the concept of temperature, which can then be used again to determine a refined concept of length, etc. Hempel (1952, 739, n. 77) similarly refers to the "method of successive approximations" described by physicist Victor F. Lenzen (1938, §13), which is Carnap's second method, as an example of the interplay between concept and theory formation. Materials other than mercury lead to different scales, and the physical concept of "thermodynamic" temperature finally was chosen to be different from all these scales, so that for each material, a set of correction factors has to be used. This choice has the advantage that "the laws of thermodynamics, which have gained a fundamental relevance in the newer physics, have the simplest form"[54].

The important point now is that Carnap's (and Lentzen's) conceptual analysis has historical examples: In the already mentioned recent study of the development of the concept of *temperature*, for example, Chang gives examples in the sections "The Iterative Improvement of Standards: Constructive Ascent" (Chang 2004, 44–48) and "Accuracy through Iteration" (Chang 2004, 212–217). In conclusion, Carnap and Hempel not only had a realistic idea of correspondence rules, but

[54] "die Gesetze der Thermodynamik, die in der neueren Physik eine fundamentale Bedeutung gewonnen haben, die einfachste Form annehmen."

also correctly analyzed or (in Hempel's case) relied on the iterative nature of the formation of numerical concepts.

3.7 Models

Clearly, the Received View does not demand overly simple exhaustive axiomatizations in first order logic, but it does demand axiomatizations. This emphasis on axiomatization has been another source of criticism, since it suggests a dismissive attitude towards scientific models. The attitude is summarized well by Frigg and Hartmann (2008, §4.1):

> Within [the syntactic view], the term model is used in a wider and in a narrower sense. In the wider sense, a model is just a system of semantic rules that interpret the abstract calculus [. . .]. In the narrower sense, a model is an alternative interpretation of a certain calculus [. . .]. Proponents of the syntactic view believe such models to be irrelevant to science. Models, they hold, are superfluous additions that are at best of pedagogical, aesthetical or psychological value [. . .].

A model in the wider sense is just any (Tarskian) model theoretic model of the calculus. Frigg and Hartmann do not claim that the Received View does not allow such models, for obvious reasons: Carnap (1939, II.10; 1963c, §10.II) made explicit use of model theoretic models, for example by relying on semantic entailment for the rules of inference of scientific theories (Carnap 1956b, 51, 61). Hempel similarly embraced model theoretic models (Kim 1999, 6) and relied on them, for example, in his discussion of vague terms (Hempel 1939, §2).

Frigg and Hartmann leave implicit that a model in the narrower sense is an alternative to some literal interpretation. The literal interpretation of electrodynamics, for example, refers to waves in an electric field, while an alternative interpretation may refer to waves in an elastic solid. Hence, technically, any model in the wider sense *except* the literal interpretation is a model in the narrower sense. According to Frigg and Hartmann, then, proponents of the Received View consider it superfluous to distinguish any specific non-literal interpretation within the class of models in the wider sense.

Frigg and Hartmann cite Carnap (1939) and Hempel (1965b) as proponents of this dismissive attitude towards models in the narrower sense. Given the central role of models in science and current philosophy of science, the attitude has provoked a lot of criticism, and has generally been seen as a disadvantage of the Received View when compared to the Semantic View, which is considered to be more hospitable to scientific models (see, for example, da Costa and French 1990; Morrison and Morgan 1999; Suppe 2000, Bailer-Jones 2003, §2; Muller 2010, §2). This is again neatly summarized by Frigg and Hartmann (2008, §4.1):

> The semantic view of theories [. . .] reverses this standpoint [of the Received View towards models] and declares that we should dispense with a formal calculus altogether and view a theory as a family of models. Although different version[s] of the semantic view assume a different notion of model [. . . ,] they all agree that models are the central unit of scientific theorizing.

Note that the last sentence just means that all versions of the Semantic View agree that the central unit of scientific theorizing should be called 'models', but disagree on what the central unit is. Unless the disagreement is over specific kinds of models in the narrower sense of Frigg and Hartmann, some versions of the Semantic View therefore assume that models in the narrower sense are *not* the central unit of scientific theorizing. In this respect, the Received View could thus be in agreement with those versions.

In the following, I will discuss Carnap's and Hempel's stance towards two kinds of models: I will argue that Carnap and Hempel in fact allow and use models, understood as possibly idealized theories with limited scope. Furthermore, they only claim that *one specific use* of models in the narrower sense is not essential to scientific theories, and are joined in this attitude by a major proponent of (one version of) the Semantic View, Patrick Suppes.

First, however, I note the failure of one specific defense of the Received View with respect to its stance on models. Suppe (1974a, 90) claims that "Carnap and Hempel do make it clear at various places that independent nonobservational semantic interpretations [of theories] are permissible". While it "must be admitted [. . .] that they tended to do so begrudgingly and also to belittle the importance of giving such interpretations" (Suppe 1974a, 90, n. 191), Suppe (1974a, 91) argues that in the Received View, theoretical terms can be

> interpreted as referring to electrons, electron emissions, and so on, where 'electron', 'electron emission', and so forth, have their normal meaning in scientific language. If we look at theoretical terms such as 'electron', we find that [. . .] much of the meaning concerns extra-observational associations—for example, for electrons there might include various features of the billiard-ball model, various classical intuitions about macroscopic point-masses, and so on.

Thus, alternative interpretations of the calculus of a theory can play a role in the Received View. However, one of the sources that Suppe cites to support this defense is a statement by Hempel about a new approach to scientific theories (Achinstein et al. 1974, 260), which he proposed after having given up the Received View. In describing his new approach, Hempel (1974, IV) explicitly states that the "extensive theoretical use of antecedent terms appears to me to throw into question the [Received View's] conception of the internal principles of a theory as an axiomatized system whose postulates provide 'implicit definitions' for its extralogical terms". Hempel thus does not assert that direct interpretations of the calculus of the theory (through the "antecedent terms" used in the theory) are permissible in the Received View. Quite the opposite, he considers the need for (or use of) direct interpretations of theoretical terms a reason to *give up* the Received View.

Hempel's interpretation of the Received View seems correct in that Suppe's two other sources for his defense are very explicit about the impossibility of such non-observational interpretations. Carnap (1939, 204) notes the possibility of assigning meanings to the theoretical terms in the metalanguage, but adds that if there is someone "who does not know physics but has normal senses and understands a language in which observable properties of things can be described", and we want to enable him to "apply [the theory] to his observations in order to arrive at explanations and predictions", this strategy is useless. Therefore "we have to give semantical rules for elementary terms only, connecting them with observable properties of things". This is not even a begrudging endorsement of the use of non-observational interpretations.

In Suppe's third source, Hempel (1963, 696) argues with reference to Carnap (1939) that interpreting theoretical terms in a metalanguage offers "little help towards an understanding of those expressions. For the criteria will be intelligible only to those who understand the metalanguage in which they are expressed". The argument for this conclusion is detailed by Rozeboom (1970, 204–205).

Carnap (1956b, 47, §V) is indeed very clear that there is no interpretation of the theoretical language L_T independent of the correspondence rules: "There is no independent interpretation of L_T. [...] The [theoretical terms] obtain only an indirect and incomplete interpretation by the fact that some of them are connected by the [correspondence rules] with observational terms". Hempel's and Carnap's reluctance to assume a direct interpretation of theoretical terms is unsurprising, since they held that such a direct interpretation is impossible. This is the thesis of semantic empiricism (cf. Rozeboom 1962), a core assumption of logical empiricism, and the basis of, for example, Carnap's use of the Ramsey sentence and his solution to the problem of analyticity (Psillos 2000).

A good starting point for a defense of the Received View on models is rather the discussion of two different kinds of models given by Hempel (1965a, §6). He calls instances of the one kind 'theoretical models', noting that they are also known as "mathematical models" (Hempel 1965a, 445–446). "Broadly speaking, and disregarding many differences in detail", he continues, "a theoretical model of this kind has the character of a theory with more or less limited scope of application". As examples, Hempel lists models of learning, conflict behavior, and other social, political, and economic phenomena. Models of this kind, Hempel notes, are often idealizations in that they disregard factors relevant for the phenomenon under study, oversimplify the relations of their parameters, and may be applicable only under very specific conditions.

Clearly, theoretical models are models neither in the narrower sense nor in the wider sense of Frigg and Hartmann. In contemporary philosophy of science, however, the term ‘model’ may often refer to theoretical models. A case in point is a recent discussion by Weisberg (2007) of a predator-prey model. Vito Volterra, in an effort to analyze the populations of Adriatic fish, stipulated certain properties of a predator-prey relationship between the populations to arrive at two coupled differential equations. These equations allowed him to predict the qualitative effects of fishing on the different populations. Weisberg (2007, §2.1) notes that Volterra himself recognized that “his model was extremely simple and highly idealized with respect to any real world phenomenon”—in short, that it was a theoretical model as defined by Hempel. Hempel’s notion of a theoretical model is also at least one way to make sense of the existence of incompatible models (cf. Frigg and Hartmann 2008, §5.1), for one because two theoretical models with restricted domains may be incompatible when their domains are taken to be unrestricted. More importantly, two theoretical models with overlapping domains may disregard different relevant factors and oversimplify the relations of their parameters in different ways.[55]

Thus, an important meaning of ‘model’ in science and philosophy of science may be the theoretical model in Hempel’s sense. Since theoretical models are theories, there is no question that they play an important role in the Received View. In fact, Carnap’s simple theory of thermal expansion and Hempel’s simple theory of buoyancy are theoretical models, and so is Carnap’s axiomatization of space-time topology, as it idealizes physical particles as points (Carnap 1954, 197–198). Those versions of the Semantic View that consider theoretical models to be the central unit of scientific theorizing are therefore, in this respect, in agreement with the Received View.

Hempel calls the other kind of model that he considers ‘analogical’. He first defines two sets L_1 and L_2 of sentences to be *syntactically isomorphic* if L_1 can be obtained from L_2 by renaming the nonlogical constants that appear in its sentences. A system S_1 is then an analogical model of a system S_2 with respect to the sets of laws L_1 and L_2 if L_1 is true of S_1, L_2 is true of S_2, and L_1 and L_2 are syntactically isomorphic. It is easy to see that, by renaming the nonlogical constants, the interpretation of L_1 by S_1 can be turned into an interpretation of L_2 and vice versa. Therefore, an analogical model of this kind is an alternative interpretation of a calculus, and thus for Frigg and Hartmann a model “in the narrower sense”.

Hempel (1965a, 440–441) lists three ways in which analogical models can be useful. First, an analogical model “may make for ‘intellectual economy’” because “all the logical consequences of the [one system’s laws] can be transferred to the new domain by simply replacing all extra-logical terms by their counterparts”. A set of mutually analogical models allows for the development of one “general

[55] I thank Christopher Belanger for impressing upon me the importance of incompatible models.

mathematical theory" that describes all systems at once, "without distinguishing between the different subject matters to which the resulting theory can be applied". The intellectual economy of this is straightforward: Investigate the logical or mathematical features of the laws once, and apply the results to all systems with syntactically isomorphic laws.

Second, analogical models can "facilitate one's grasp" of the laws in a new domain by "exhibiting a parallel with [laws] for a more familiar domain" (Hempel 1965a, 441). The wave equations of electrodynamics, for example, can be qualitatively analyzed by simply thinking about waves in visible things, because those are sometimes governed by syntactically isomorphic equations.

"More important", Hempel continues, "well-chosen analogies or models may prove useful 'in the context of discovery', i. e., they may provide heuristic guidance in the search for new explanatory principles". This is the third use of analogical models. Hempel (1965a, 445) elucidates:

> Considering the great heuristic value of structural analogies, it is natural that a scientist attempting to frame a new theory should let himself be guided by concepts and laws that have proved fruitful in previously explored areas. But if these should fail, he will have to resort to ideas that depart more and more from the familiar ones.

As an example, Hempel adduces the development of quantum mechanics, which started out close to classical mechanics but became considerably less analogical, thereby gaining in scope.

The first two uses of analogical models are, in a sense, of psychological value, just as Frigg and Hartmann contend. The third, heuristic use of analogical models is not so, nor is its main value pedagogical or aesthetic. That the heuristic use does not play a major role in the Received View on scientific theories is unsurprising, because the Received View was not developed for analyses in the context of discovery, but rather for analyses in the context of justification (cf. Feigl 1970, 3–4, 13–14).

Hempel (1965a, 438–439) only dismisses the use of analogical models as essential for explanation, and his argument is straightforward: Assume that some feature of a system is to be explained. To find an analogical model of the system, the system's laws have to be established. These laws are all that is needed to give an explanation of the feature, and so no analogical model is required. This argument rests on the D-N-schema of explanation, and it is not obvious how this translates to other explications of explanation. But whether it translates or not, Hempel does establish that an analogical model (in his sense) is not needed to derive statements about a system from the system's laws.

Carnap's dismissal of models is similarly confined. Discussing 'understanding' in physics (Carnap 1939, 209–210), he notes:

> When abstract, nonintuitive formulas, as, e. g., Maxwell's equations

> of electromagnetism, were proposed as new axioms, physicists endeavored to make them "intuitive" by constructing a "model", i. e., a way of representing electromagnetic microprocesses by an analogy to known macroprocesses, e. g., movements of visible things.

Not only did these attempts fail, Carnap states, but they are also unnecessary:

> It is important to realize that the discovery of a model has no more than an aesthetic or didactic or at best a heuristic value, but is not at all essential for a successful application of the physical theory.

Like Hempel, Carnap does not dismiss theoretical models or question the usefulness of analogical models in the context of discovery (which he does not even mention here). He only objects to the demand that every formalism be supplied with an analogical model in terms of macroprocesses; that is, he objects to analogical models as *necessary* for the theory's application. Assuming that the theory is successfully applied by deriving true statements from it, this objection is justified by Hempel's argument against the need for analogical models in explanations.

When Carnap (1966, 232–233) does mention the context of discovery, he accepts the prominent role of analogical models:

> [I]magine that we are [. . .] preparing to state for the first time some theoretical laws about molecules in a gas. [. . .] We do not know the exact shape of molecules, so let us suppose that they are tiny spheres. How do spheres collide? There are laws about colliding spheres, but they concern large bodies. Since we cannot directly observe molecules, we assume their collisions are analogous to those of large bodies [. . .]. These are, of course, only assumptions; guesses suggested by analogies with known macrolaws.

Carnap (1966, 174–175) only cautions that visual models (which, arguably, are the most important analogical models) are neither necessary nor sufficient for applying theories, but can be helpful. And these models can turn out to be true:

> A physicist must always guard against taking a visual model as more than a pedagogical device or makeshift help. At the same time, he must also be alert to the possibility that a visual model can, and sometimes does, turn out to be literally accurate. Nature sometimes springs such surprises. [. . .]
>
> A theory may move away from models that can be visualized; then, in a later phase, when more is known, it may move back again to visual models that were previously doubted.

Thus Carnap, like Hempel, considers analogical models important, but not necessary for the application of a theory to the system.

Let me compare this stance with the one taken by Suppes (1960) in one of the articles that cemented the Semantic View's position as an account of scientific theories that is especially hospitable to models (see Muller 2010, §2). Suppes (1960, 289) argues that

> the concept of model in the sense of Tarski may be used without distortion and as a fundamental concept in [physics, social science, and mathematical statistics]. In this sense I would assert that the meaning of the concept of model is the same in mathematics and the empirical sciences. The difference to be found in these disciplines is to be found in their use of the concept.

Suppes's claim is thus that in all three disciplines 'model' means what Frigg and Hartmann call 'model in a wider sense'. Therefore, Suppes's argument cannot possibly establish that the Semantic View is more hospitable to analogical models than the Received View. Suppes (1960, 289) also explicitly excludes Hempel's theoretical models from his discussion by cautioning against

> one very common tendency, namely, to confuse or amalgamate what logicians would call the model and the theory of the model. It is very widespread practice in mathematical statistics and in the behavioral sciences to use the word 'model' to mean the set of quantitative assumptions of the theory, that is, the set of sentences which in a precise treatment would be taken as axioms [...].

Since theoretical models are theories, not models of theories, they are excluded from Suppes's discussion like analogical models. Note that Suppes here in effect claims that many of the so-called models used in the sciences are actually theories, and thus play an important role in the Received View.

Suppes (1960, 290) does acknowledge that "many physicists want to think of a model of the orbital theory of the atom as being more than a certain set theoretical entity. They envisage it as a very concrete physical thing built on the analogy of the solar system". It is not exactly clear whether Suppes has Hempel's analogical models in mind here, or simply considers the "concrete physical thing" the thing with the actual nucleons and electrons (i. e., the literal interpretation of the formalism). If the latter, this is not a model in either the wide or the narrow sense. But Suppes (1960, 291) clearly refers to an analogical model in his discussion of

> Kelvin's and Maxwell's efforts to find a mechanical model of electromagnetic phenomena. Without doubt they both thought of possible models in a literal physical sense, but it is not difficult to recast their published memoirs on this topic into a search for set-theoretical models of the theory of continuum mechanics which will account for

> observed electromagnetic phenomena. Moreover, it is really the formal part of their memoirs which has had permanent value. Ultimately it is the mathematical theory of Maxwell which has proved important, not the physical image of an ether behaving like an elastic solid.

According to Suppes, Kelvin and Maxwell aimed to find the regularities that govern electromagnetic phenomena by using an analogy with mechanical phenomena. Therefore, they used analogical models in the context of discovery, in the sense discussed by Hempel and Carnap. But Suppes's interest lies in the possibility of describing electrodynamics with a mathematical model, that is, of giving a model in the wider sense. And the permanent value of Maxwell's theory is, I take it, its applicability, for which Suppes considers the analogical model irrelevant—as Hempel establishes in his argument against their role in theoretical explanations.

Overall, Suppes's attitude seems at most as hospitable towards scientific models as that of Hempel and Carnap. To conclude that the Semantic View is directly related to scientific models because it uses model theory is but a fallacy of equivocation, and whether scientific models are better formalized in predicate logic or model theory is an open question whose answer will probably depend on the kind of model under consideration and what is meant by 'better'.

3.8 The concept and object of explication

Suppe (1974a, 58; 2000, S104) develops a direct argument against the Received View that is closely related to the question of the axiomatizability of scientific theories. As the first premise, he "take[s] it as being reasonably clear from Carnap's and Hempel's writings that they intend their analysis to provide an *explication* of the concept of a scientific theory". Carnap (1950b, §§2–6) gives a detailed discussion of the idea behind and the structure of an explication. The explication of a vague or otherwise unclear notion, the explicandum, consists in the development of a new concept, the explicatum[56], that is to take the explicandum's place in some analyses. The explicatum has to fulfill four requirements: Some similarity to the explicatum, precision in the rules for its use, fruitfulness in the development of theories, and simplicity (to the extent allowed by the previous three requirements).[57]

The requirements, Suppe (1974a, 59) notes, are "rather vague as to the relationship in which the explicatum should stand to the explicandum", and he therefore introduces as his second premise an "adequacy criterion" given by Chomsky (1957, §2.1), which "seems to be in accord with Carnap's position" (Suppe 1974a, 59). According to this criterion, an explication is adequate only if "the explicatum denote[s] all the clear-cut instances, and none of the clear-cut noninstances" of the

[56]Reichenbach (1951, 49) prefers 'explicans' on linguistic grounds, Stein (1992, 281) defends Carnap's choice of terminology.

[57]See page 19.

explicandum. In other words, the explicatum has to concur with the explicandum within the boundaries of the explicandum's vagueness, so that explication is a kind of precisification.[58] This is also how, for example, Bishop (1992, 268) and Gaines (2010, 168) interpret Carnap's account.[59]

This criterion of adequacy allows the Received View to be tested, because if it "can be demonstrated that there are clear-cut examples of scientific theories which do not admit of the required canonical formulation [of the Received View], or else show[n] that certain clear-cut examples of nonscientific theories fit their analysis", then this shows "the inadequacy of the Received View" (Suppe 1974a, 60). It is not difficult for Suppe (1974a, 65) to find examples to establish his third premise, that many scientific theories cannot be axiomatized. These include

> Darwin's theory of evolution, Hoyle's theory on the origin of the universe, [and] Freud's psychology [...]. Furthermore, it is manifest that most theories in cultural anthropology, most sociological theories about the family; theories about the origin of the American Indian [...] are all such at present that any attempts at axiomatization would be premature and fruitless [...].

Since "[s]ome theories do admit of fruitful axiomatization, however", Suppe (1974a, 63) concludes that "the Received View is plausible for some but not all scientific theories". Similarly, Beatty (1980, appendix 1) considers the impossibility of axiomatizing "evolutionary theory, Freudian psychology, theories of the origin of the universe, and many others" to be a major problem of the Received View (and curiously traces this impossibility to the Received View's perceived restriction to first order logic).

First, note that Suppe's conclusion (that the Received View is plausible for some, but not all theories), does not follow from his premises; the correct conclusion is that the Received View is false. Second, $\ulcorner\{\forall x(Px \lor Qx), Oc \leftrightarrow Pc\}\urcorner$ is axiomatizable and fulfills all further requirements of the Received View if P and Q are uninterpreted theoretical predicates, O is an interpreted observational predicate, and c is an interpreted observational constant. Some such sets of sentences are clearly not called 'scientific theory' in ordinary language, for example if $\ulcorner Px\urcorner$ stands for $\ulcorner x$ is foo$\urcorner$, $\ulcorner Qx\urcorner$ for $\ulcorner x$ is bar$\urcorner$, c for $\ulcorner$the cat$\urcorner$, and $\ulcorner Ox\urcorner$ for $\ulcorner x$ is on the mat$\urcorner$. Thus the Received View is false independently of Suppe's third premise. Given the ease of this apparent disproof of the Received View, it should come as no surprise that Suppe's first and second premises are false.

[58] In the terminology of §2.8.3, every element of the vagueness set of a vocabulary is a possible precisification of the vocabulary's interpretation.

[59] Carus (2007, 285–287) lays out how the even stronger requirement that explicatum and explicandum must have *the same* meaning leads to a criticism of explication itself.

3.8.1 The concept of explication

Carus (2007, 257–259) describes how an earlier incarnation of explication, Carnap's notion of translation (Carnap 1934a, §§74–79), was indeed meant to preserve the meaning of the original term (which was not yet called 'explicandum') as closely as possible. But for explication itself, Suppe's sole source for the first premise is Chomsky (1957, §2.1), who references neither Carnap nor Hempel for his criterion of adequacy for explication. He does mention Goodman (1951, 5–6), but Goodman (1951, §I,1) lists precisification only as one kind of explication (he speaks of 'constructional definition'). Contrary to Suppe's first premise, Goodman (1951, 5) notes that for explications in general, scientists and philosophers often "trim and patch the use of ordinary terms to suit their special needs, deviating from popular usage even where it is quite unambiguous".

In his discussion of Goodman's view on explication, Carnap (1963c, §21) does not take issue with the possible deviation of the explicatum from the explicandum. This is unsurprising given that Carnap (1950b, 5–6) himself allows for such deviation, and even uses the explication of 'fish' as an example, just like Goodman:

> [O]ne might perhaps think that the explicatum should be as close to or as similar with the explicandum as the latter's vagueness permits. However, it is easily seen that this requirement would be too strong, that the actual procedure of scientists is often not in agreement with it, and for good reasons. [...] In the construction of a systematic language of zoölogy, the concept Fish designated by this term has been replaced by a scientific concept designated by the same term 'fish'; let us use for the latter concept the term 'piscis' in order to avoid confusion. When we compare the explicandum Fish with the explicatum Piscis, we see that they do not even approximately coincide. The latter is much narrower than the former; many kinds of animals which were subsumed under the concept Fish, for instance, whales and seals, are excluded from the concept Piscis.

This settles the question of Carnap's (and Goodman's, for that matter) stance on Chomsky's criterion of adequacy.

Hempel's position, on the other hand, is more ambiguous. Often he seems to be in accord with Carnap: In a monograph on concept formation, Hempel (1952, 663) refers to Carnap's exposition without taking exception to the possible deviation of the explicatum from the explicandum, and states:

> Explications, having the nature of proposals, cannot be qualified as being either true or false. Yet [...] they have to satisfy two major requirements: First, the explicative reinterpretation of a term or—as is often the case—of a set of related terms must permit us to reformulate [...] at least a large part of what is customarily expressed by means

> of the terms under consideration. Second, it should be possible to develop, in terms of the reconstructed concepts, a comprehensive, rigorous and sound theoretical system.

Note that Hempel does not demand that the explicatum coincide with the explicandum to the extent that the latter's vagueness permits. He only demands that what can be expressed with the explicandum can also be expressed with the explicatum (or the explicata), and even this only in a large portion of cases, but not necessarily in all. Most importantly, according to Hempel explications cannot be false, which they could be if there was a condition of adequacy for the relation of explicandum and explicatum.

Another passage by Hempel also suggests his agreement with Carnap. In a review of Goodman's exposition of explication, Hempel (1953, 113–114) states:

> It seems to me important to note [. . .] that the stage of rigorous construction in philosophy [. . .] presupposes a preconstructional clarification of the explicanda under investigation. [. . .] But in the pursuit of its objective, analysis cannot be content with a purely descriptive account of linguistic behavior patterns: it has to point out the pitfalls inherent in the various modes of usage [. . .]. And from here on, it is only a short step [to] explicitly proposing certain modifications of existing usage which will enhance clarity and which promise to be theoretically fruitful. Once this last step has been taken, the stage is set for the development of a constructional system for the readjusted explicanda[.]

It is of no relevance in this context that Hempel here, in a quite puzzling shift of terminology, allows the explicanda to deviate from actual usage, rather than allowing the explicata to deviate from the clear cases of the explicanda. The important point is that Hempel allows the explicata to deviate from clear cases of actual usage.

Perhaps most importantly, Hempel, like Carnap, allows the explicatum to have a logical structure different from the explicandum. In fact, Hempel (1952, §10) argues that the explication of a classificatory explicandum by a comparative explicatum is often a sign of an investigation's maturity (see also Hempel and Oppenheim 1936), as the explication of 'warm' by 'warmer than' illustrates (cf. Carnap 1950b, §4). And if the logical structure of the explicatum is different from that of the explicandum, it is not even clear what it would mean for the explicatum to denote "all the clear cut instances and none of the clear cut non-instances" of the explicandum.[60]

However, Hempel (1950, §6) also answers the question of "how [. . .] to judge

[60]That is why Laudan (1986, 120) misrepresents Carnap's conception of explication when he criticizes that "as far as Carnap was concerned, one of the necessary tests for determining whether the philosopher had done a proper job of explicating the methodological terminology of science

the adequacy of a proposed explication, as expressed in some specific criterion of cognitive meaning" with two criteria, the first of which relies on the fact that

> there exists a large class of sentences which are rather generally recognized as making intelligible assertions, and another large class of which this is more or less generally denied. We shall have to demand of an adequate explication that it take into account these spheres of common usage; hence an explication which, let us say, denies cognitive import to descriptions of past events or to generalizations expressed in terms of observables has to be rejected as inadequate.

Unfortunately, Hempel is not explicit about whether an explicatum always has to coincide with the explicandum's clear cases, but speaks of an explication "taking into account" common usage, which is determined by "rather general" agreement on the positive instances and "more or less general" agreement on the negative ones.

In a defense of his and Oppenheim's explication of scientific explanation, Hempel (1965a, §11) similarly states:

> Like any other explication, the construal here put forward has to be justified by appropriate arguments. In our case, these have to show that the proposed construal does justice to such accounts as are generally agreed to be instances of scientific explanation, and that it affords a basis for a systematically fruitful logical and methodological analysis of the explanatory procedures used in empirical science.

While the phrase 'do justice to such accounts' is not as clear as one might wish, this is quite possibly the passage in which Hempel comes closest to demanding that the explicatum include all clear positive instances of the explicandum (note that he does not demand agreement on negative instances). It was written in the years of 1963 and 1964 at the Center for Advanced Study (Hempel 1965e), about 18 years after his monograph on concept formation was researched and perhaps written (Hempel 1952, 731, n. 1), and a year before he gave a lecture that would form the core of his rejection of the Received View (Hempel 1970, 142, n. 1). Maybe Hempel had simply changed his mind. At the Center, he also met Thomas Kuhn for the first time, whose ideas "certainly contributed to [his] shift from an antinaturalistic stance to a naturalistic one" in the later years (Hempel 1993). Maybe the passage marks the first tiny step in this direction.

On the other hand, the passage sounds suspiciously similar to the preceding one published in 1950, which precedes the publication of Hempel's monograph

---

involved ascertaining whether one's proffered explication (the 'explicans') could be substituted without alteration of truth-value into all or most of those contexts where the explicandum occurred!" Whenever the logical structure of the explicatum is different from that of the explicandum, a substitution is not possible at all, as it would not result in a syntactically well-formed sentence.

on concept formation (though not the research that led to it). The dissonance may therefore just be the result of loose language, which leaves—barely—enough room to conform to Carnap on the relation of explication and precisification. Hempel's position on explication may also simply have been inconsistent. When it comes to explicating 'cognitive significance', for instance, he states that "cognitive significance in a system is a matter of degree", and sees this as a reason for disposing of the concept altogether. Instead of "dichotomizing this array [of systems] into significant and non-significant systems", he states, one should compare systems of sentences by their precision, systematicity, simplicity, and level of confirmation (Hempel 1951, 74; cf. Hempel 1965c, 117). But the explication of a classificatory concept by a comparative one is exactly what he, at about the same time, claims to be an indicator of an investigation's maturity.

In summary, Carnap clearly did not conceive of explication as a specific kind of precisification, and unless Hempel was very confused about Carnap's stance, neither did Hempel. Considering that Hempel's position was unclear at some points, I want to note that explication is a core concept of what Rorty (1967a) has called 'ideal language philosophy' (Carnap 1963c, §19; Maxwell and Feigl 1961, 488; Lutz 2009, §2). In a defense of ideal against ordinary language philosophy, Maxwell and Feigl (1961, 491) are very explicit about the possible deviations of an explicatum from its explicandum:

> [W]e see absolutely no reason to believe that examination of ordinary use in the "paradigm", normal cases can provide us with definitive rules for "proper" use in the unusual and novel cases. [...]
>
> Furthermore—and this is of crucial importance—consideration of atypical cases often points up possible inadequacies and may suggest improvements in our conceptualization of the "normal" cases.

Because Maxwell and Feigl consider philosophical problems to be often linked to the unusual and novel cases, their first point is an explicit rejection of the viability of ordinary language philosophy: In order to tackle philosophical problems, ordinary language might have to be precisified, but ordinary language itself gives no clue as to which precisification is right. The second point extends this rejection to the areas where ordinary language is unambiguous, for even a precisified language might be inadequate, and thus language use in the clear cases might require modification. The rendering of explication as precisification means giving up the core of ideal language philosophy.[61]

3.8.2 The object of explication

Feigl certainly considered the Received View to be about explication, not just precisification: Maxwell and Feigl (1961, 489–490) call explication 'rational reconstruction', and in a defense of the Received View, Feigl (1970, 13) speaks of

[61] See Carus (2007, especially ch. 10–11) for a much more thorough discussion of this point.

the Received View as being concerned with "the rational reconstruction of theories". But this does not mean that Feigl agrees with Suppe's second premise, according to which the Received View is meant as an explication of the term 'theory'. For Feigl does not speak of the rational reconstruction of 'theory', or the rational reconstruction of the concept (or notion) of a theory, but the rational reconstruction of *theories*. In the same text, he does speak of the "analysis of the *notion of* evidential support" and the "*concept of* probability" (Feigl 1970, 9, my emphases), so it is improbable that Feigl would simply ignore the use-mention distinction when speaking about theories. Indeed, in an earlier passage of the same text he remarks that "logicians of science [. . .] analyze *a given theory* in regard to its logical structure" (Feigl 1970, 8, emphasis changed), that is, a specific theory, not the term denoting all of them. My thesis is therefore that the Received View is not meant to explicate 'theory', but rather to provide a framework in which specific theories like the general theory of relativity or evolutionary theory can be explicated. Again in the words of Feigl (1970, 13): "[T]he 'orthodox' view of scientific theories can help in clarifying their logico-mathematical structure, as well as their empirical confirmation (or disconfirmation)".

One piece of evidence for this view is that there is, to my knowledge, no text by Hempel, Carnap, or Feigl in which either mentions a successful explication of the term 'theory'. Quite to the contrary, when Hempel (1983, §6) discusses "Carnap's views on the analytic elaboration of methodological concepts and principles [to which he] refers [. . .] as *explication*", he notes that among those philosophical issues that Carnap so elaborated are "standards for a rational appraisal of the credibility of empirical hypotheses", but he does not include the concept of an empirical hypothesis or theory itself. He writes:

> Explication plays an important role in analytic philosophy, where it has often been referred to as logical analysis or rational reconstruction. All the accounts proposed by analytic empiricists for such notions as verification, falsification, confirmation, inductive reasoning, types of explanation, theoretical reduction, and the like are instances of explication.

This is a rather comprehensive list of the concepts at the core of logical empiricism's philosophy of science. If 'theory' had been explicated as well, it would be very surprising for Hempel to not have mentioned such a central concept at all.[62]

After having given up the Received View in favor of his competing account, Hempel was actually asked whether his new view was meant as a description of the actual use of 'theory' (in a discussion published in the conference proceedings edited by Suppe (1974b) himself). Sylvain Bromberger asks (Achinstein et al. 1974, 261):

[62] As an analogy, consider the claim "All of Einstein's enduring contributions to physics such as the explanations of Brownian motion and critical opalescence, the prediction of stimulated emission and the Bose-Einstein condensate, the explanation of the photo-effect, and the like were made before 1930."

> Exactly what is the status of this analysis of a theory? [. . .] As an analysis of the concept of a theory, that is, the concept that is embodied in our use of the word 'theory' in English or equivalent ones in other languages, I think it is demonstratively false. [. . .] The view that a theory consists of an uninterpreted calculus and rules of interpretation [. . .] was a program which, if successful, would have made explicit the rules that govern or that ought to govern acceptance and rejection of theories. [. . .] Now do you envisage your new [analysis] as entailing a program that will show how the theories that we in fact have might ultimately be analyzed?

Bromberger here claims that Hempel's new account does not correctly describe the actual use of the term 'theory' in ordinary language, and then asks whether it is meant to achieve the goal of the Received View. He takes this goal to be the successful explication of the rules for accepting and rejecting theories.

Hempel's reply is telling:

> Professor Bromberger is right in stressing the programmatic side of the standard conception. One of its objectives was to explicate, and appraise from the point of view of an analytic-philosophical conscience, the principles governing concept formation in scientific theories. Another objective was similarly to exhibit and appraise the principles governing the testing of scientific theories. [. . .]
>
> My paper was intended principally as a criticism of the basic assumptions by means of which the standard construal tackles its task; I did not put forward a properly developed alternative.

Again, Hempel does not mention the explication of the term 'theory' as a goal of the Received View. Instead, he agrees with Bromberger that one goal was the explication of the rules of acceptance and rejection, that is, for the testing of scientific theories. He adds that another goal was the explication of the rules of concept formation. Furthermore, Hempel seems to agree with Bromberger that his new account fails to correctly describe the actual use of the term 'theory' and that his new account might nonetheless achieve the goals of the Received View. Therefore, he cannot be of the opinion that the goal of the Received View is to describe the use of the term 'theory'.

In a critical discussion of the Received View, Hempel (1970, 148) explicitly addresses the relation between the Received View and explication: "[T]he standard construal [. . .] was intended [. . .] as a schematic explication that would clearly exhibit certain logical and epistemological characteristics of scientific theories". Since the Received View is not an explication but an explication schema, it is incomplete until it is applied to something. The explication of a scientific theory would then be an instantiation of this schema. In other words, the Received View provides a framework for explicating specific scientific theories.

Beyond the general impression of Carnap's and Hempel's writings, Suppe (1974a, §IV, A, n. 126) cites an early article by Carnap (1931a, §§5, 7) as evidence for the claim that a theory that "does not admit of a canonical reformulation meeting the conditions [...] of the Received View [...] is not a genuine scientific theory". This article, he states, contains a "very explicit version of the claim for the initial version of the Received View", where the "initial version" is that from the time of the Vienna Circle. That this claim still holds for the final version is, according to Suppe, borne out by the addendum to the article's English translation, in which Carnap (1963b) "reaffirms this claim in its essential form".

In the two sections to which Suppe refers, Carnap (1931a, 453) argues for physicalism, the thesis that "every scientific sentence can be translated into the physicalistic language"[63] and that through this, "the whole of science becomes physics"[64] (Carnap 1931a, 463, emphasis removed). His argument for these conclusions relies on the explicit definability of all scientific terms in observational terms, and the explicit definability of all observational terms in physical terms. Because explicit definability is transitive, the explicit definability of all scientific terms in physical terms follows immediately from these two claims. In the addendum, Carnap (1963b) notes that the explicit definability of all scientific terms in observational terms has to be given up and substituted by "reducibility through a kind of conditional definitions" (cf. Carnap 1936, 1937) or relations "still more flexible" (cf. Carnap 1956b). For then-current presentations of physicalism, Carnap refers to two discussions by Feigl (1958, 1963) and a work of his own (Carnap 1963c, §7). Both of Feigl's works discuss the reducibility of mental states to physical states.

It is surprising to me that Suppe sees in these discussions a justification of his second premise. Clearly, the texts are about the terms that can be reduced to physical language, and according to the thesis of physicalism held by Carnap, any meaningful term can be reduced to physical terms. But to infer from this position that the Received View explicates 'theory' would at least require additional premises, for example, the premises that all physical theories can be formalized according to the Received View, that all and only sets of sentences reducible to a theory are themselves theories, and that all and only scientific theories can be reduced to physics with the help of the reduction statements. But the latter premise was explicitly denied by Feigl (1963, 241–245) and Carnap (1963c, 883) at the time of the final version of the Received View, and also in the very article Suppe uses to support his second premise. There, Carnap (1931a, 449) states with respect to the reducibility of biology to physics that "the thesis of the universality of the physical language [...] is not about the reducibility of the biological *laws* to the physical, but the reducibility of the biological *terms* to the physical [...]. And this reducibility can, in contradistinction to the former, be easily shown".[65] In

[63] "Unsere Überlegungen [...] führen somit zu dem Ergebnis, daß jeder wissenschaftliche Satz in die physikalische Sprache übersetzbar ist."

[64] "Dadurch [...] wird die gesamte Wissenschaft zu Physik."

[65] "[Bei der] These von der Universalität der physikalischen Sprache [...] handelt es sich nicht um

conclusion, Suppe's textual evidence fails to make his case for the second premise.

Since the Received View is a framework intended to help in the explication of theories, there is no problem if it also allows one to explicate things that are not scientific or not theories. It is more problematic if the Received View relies on assumptions that are not fulfilled by all theories, so that it cannot aid in their explications. Like Suppe, Carnap (1939, 202) was clearly aware that some scientific theories could not, at that point, be fruitfully formalized, and hence a fortiori could not be fruitfully formalized according to the Received View. He writes:

> Any physical theory, and likewise the whole of physics, can [...] be presented in the form of an interpreted system, consisting of a specific calculus (axiom system) and a system of semantical rules for its interpretation [...]. It is, of course, logically possible to apply the same method to any other branch of science as well. But practically the situation is such that most of them seem at the present time to be not yet developed to a degree which would suggest this strict form of presentation.

So Carnap held the view that in principle, all theories can be fruitfully formalized. He plausibly also held the view that the better developed a theory is, the closer it comes to being so formalizable. But Carnap was also of the opinion that Suppe expresses in his third premise, that many theories cannot be fruitfully formalized "at present".

Clearly, then, Carnap was not of the opinion that all theories can be explicated in the Received View. I think it is most plausible that Carnap's attitude to the Received View was like his attitude to first order axiomatization as communicated by Hempel: It is the first stage in the development of more comprehensive frameworks for explication, and it may permit generalizations to deal with theories that cannot be axiomatized in (modal) predicate logic. The explications and analyses based on the Received View (for example, of theory testing or concept formation), therefore have the form of conditionals: If a theory can be reconstructed according to the Received View, then the respective explication or analysis is applicable.

Carnap (1923) in fact expresses this view in his earliest paper on the Received View. He introduces an ideal physics, consisting of a completely axiomatized theory, a set of correspondence rules, and a complete description of the physical world at two points in time. Carnap (1923, 96) describes the value of such a fiction thusly:

> To determine the direction that physics should take at any stage, the fiction of a completed construction of physics can be of great help, as

die Zurückführbarkeit der biologischen *Gesetze* auf die physikalischen, sondern um die Zurückführbarkeit der biologischen *Begriffe* [...] auf die physikalischen. Und diese Zurückführbarkeit kann, im Unterschied zu der ersteren, leicht erwiesen werden."

> it were, as a target at infinite distance.[66]

Later in the text, Carnap actually loosens one of his assumptions about the completed construction of physics, the assumption that at two points in time, the state of all physical magnitudes is known at all points in space. By considering what the restriction to observable phenomena means for the possible values of the physical magnitudes, Carnap tries to arrive at a more general theory about scientific theories starting from a special case, just as in his remark about first order logic. Note that this restriction to statements about observable states has the effect that the theoretical terms are interpreted only through the observational terms and the correspondence rules, that is, it leads to semantic empiricism. Thus this generalization becomes a core feature of the later Received View.

Shimony describes this search for generalizations as fundamental to Carnap's way of working. In an homage after Carnap's death, he writes that Carnap "took particular delight in technical advances which permitted him to widen the scope of his investigations without loss of precision" (Feigl et al. 1970, XXVI). And indeed Carnap (1956b, 49) considers another generalization of the Received View. After giving examples of correspondence rules (C-Rules), he states:

> In the above examples, the C-rules have the form of universal postulates. A more general form would be that of statistical laws involving the concept of statistical probability [. . .]. A postulate of this kind might say, for example, that, if a region has a certain state specified in theoretical terms, then there is a probability of 0.8 that a certain observable event occurs [. . .]. Or it might, conversely, state the probability for the theoretical property, with respect to the observable event. Statistical correspondence rules have so far been studied very little.

First, note that with the inclusion of statistical relations, correspondence rules in the Received View can be in every respect as complex as the "hierarchy of models" that Suppes (1962, 260) would describe six year later, lamenting the "besetting sin of philosophers of science [. . .] to oversimplify the structure of science."[67] Second, if the Received View was meant as a definitive framework for theory explication, such a generalization would not make sense. Hence I think it is clear that Carnap considered the Received View's reliance on formalization in predicate logic a restriction that could be shed in further generalizations. Accordingly, Carnap spent the rest of his life studying probability.[68]

[66] "Für die Feststellung der Richtung, in der die Physik auf irgendeiner Stufe weiterschreiten soll, kann die Fiktion eines vollendeten Aufbaues der Physik, gewissermaßen als Zielpunkt im Unendlichen, gute Dienste leisten."

[67] It is doubtful, however, that probabilistic correspondence rules can lead generally to explicit definitions of all observational terms in theoretical terms, as Carnap hoped for his second method of interpreting theoretical terms (see figure 3.1).

[68] Similarly, Reichenbach (1951, 48) supposes that Carnap's theory of probability "is intended

Thus, the Received View was meant to give a framework in which one can explicate individual theories that are formulated precisely enough to be axiomatized, and was meant to allow the incorporation of new developments in logic and mathematics to widen its applicability. Criticizing the Received View for its inability to capture a specific kind of non-formal reasoning in the sciences (say, causal inferences) thus simply amounts to a request to develop inference systems that can capture this kind of reasoning (say, default logic).

3.9 Aftermath

Suppe (2000) locates the end of the Received View as a philosophical program in Hempel's presentation at a conference in Urbana on the structure of scientific theories, held in March 1969:

> The *Received View on Theories* was the epistemic heart of Logical Positivism. Twelve hundred persons were in the audience the night it died. [. . .] The Received View had been under sustained attack for a decade and a critical mass of main protagonists had been assembled to fight it out. Carl Hempel [. . .] was expected to present the Received View's latest revision. Instead he told us why he was abandoning both the Received View *and* reliance on syntactic axiomatizations (Hempel 1974). Suddenly we knew the war had been won, and the Symposium became an energized exploration of where to go now.

I will come back to the description of a philosophical investigation as a siege in §3.10. My main interest here lies in the key role that Hempel's critique of the Received View seems to have played. Hempel's presentation was to a large extent based on a presentation given at a conference on the problems of correspondence rules at the Minnesota Center for Philosophy of Science held in May 1966 in Minneapolis. If, in keeping with Suppe's metaphor, the Received View indeed died with Hempel's disavowal, it did not die in March 1969—this is only when its body turned up. And while the Urbana conference did not have any proponent of the Received View, the Minneapolis conference saw two spirited defenses, one by Feigl (1970) and one by Rozeboom (1970).

In his presentation "The crisis of philosophical semantics", Rozeboom (1970, 202) focuses on the interpretation of theoretical terms and explicitly notes the possibility of theoretical terms being closer or less close to observational terms, as Hempel and Carnap before him. The goal here is to "set up a notion of the observation language [that allows] the possibility of later agreeing with somebody who says that the observation terms also contain theoretical meanings" (Achinstein

to overcome the shortcomings of his reduction chains, which are applicable only in cases where probability relations can be practically replaced by logical implications, and which thus are too primitive instruments for the construction of scientific language".

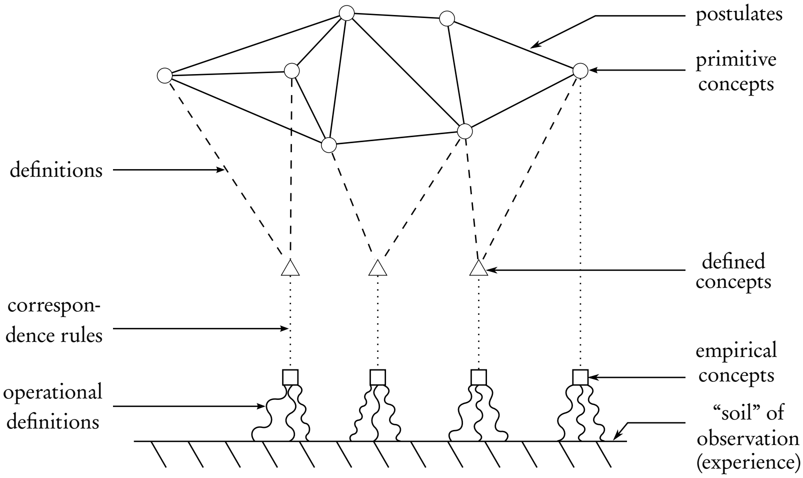

Figure 3.2: The relation of theoretical and observational terms according to Feigl (1970, 6). Feigl's figure differs from mine in that it does not include the labels on the left side and uses dashed lines both for definitions and correspondences rules.

et al. 1970, 242) by defining the bipartition of the vocabulary relative to one language, so that one language's theoretical terms can be another's observation terms. Feigl (1970) gives a more general defense of the Received View in his presentation "The 'orthodox' view of theories: remarks in defense as well as critique". The probably best known part of his discussion is the diagram of the logical relations between the theoretical terms and the observations (figure 3.2).

According to Feigl (1970, 5–6), the primitive theoretical terms are interrelated through the postulates, and used to define other concepts. These concepts are then linked to concepts that refer to "items of observation" like mass and temperature. "These empirical concepts", Feigl states, "are in turn 'operationally defined', i. e., by a specification of the rules of observation, measurement, experimentation, or statistical design which determine and delimit their applicability and application".

Note that the correspondence rules determine a "one-to-one correspondence" (Feigl 1956, 21) between the defined terms and the empirical terms—that is, for each observational term, there is a coextensive theoretical term. In Feigl's diagram, the operational definitions could seem to be semantic interpretations, but as his description makes clear, they embody many non-interpretative elements like rules of measurement and statistical design. Furthermore, operational definitions had already been discussed in depth by Hempel as best understood as reduction sentences, that is, syntactical notions. In Carnap's two methods to interpret postulate systems, on the other hand, the distinction between syntax and semantics is clear: The observable terms have a semantic interpretation, as shown in his diagram (figure 3.1), and whatever interpretation the theoretical terms have comes through their connection to the observational terms by the correspondence rules. In the first method, these correspondence rules take the form of reduction sentences from observational to increasingly theoretical terms (figure 3.3), in the second method, the correspondence rules consist of explicit definitions of increasingly observational terms by more theoretical ones or, since "Theoretical concepts", conditional probabilities between sentences involving observational and theoretical terms (figure 3.4).

Thus Feigl's account does not describe very well Carnap's account of the Received View, in which there is no distinction between explicit definitions (or reduction sentences) and correspondence rules, and the observational terms are not related to the observations by operational definitions but by semantic interpretations. Furthermore, Carnap allows also probabilistic correspondence rules, which is not captured in Feigl's account. Hempel argues that real valued functions like mass and temperature cannot be specified by operational definitions or reduction sentences in a purely observational language, so that Feigl's categorization of mass and temperature as empirical concepts is too inclusive. These misrepresentations are unrelated to Feigl's defense of the Received View, however.

The all-out critique by Hempel (1970) at the conference, "On the 'standard conception' of scientific theories", makes for a puzzling read: Much of what Hempel criticizes in the Received View has not been proposed by either Carnap or himself. His counterproposal is actually closer to the Received View than the view he criticizes. Even taking into account Hempel's misrepresentation of Carnap's work with respect to first order logic (see §3.3), this may sound preposterous, and I have been grappling with this puzzle without finding a good explanation, but let me first lay out the evidence, and then suggest a possible explanation of Hempel's misrepresentation of Carnap's and his very own earlier views.

Hempel's new account relies on a theory-relative distinction of the vocabulary into a theoretical vocabulary, containing terms newly introduced by the theory T, and a pre-theoretical or antecedent vocabulary. T is then presented as a pair of sets of sentences, $T = \langle B, I \rangle$, the internal principles I and the bridge principles B. Sentences in I contain only theoretical terms, while there are no restrictions on bridge principles. Alternatively, T can be represented as the set of the logical consequences of $B \cup I$ (Hempel 1970, 142–144). The Received View, Hempel states, represents a theory T as a pair of and uninterpreted calculus C and a set of correspondence rules R (Hempel 1970, 146–147). The correspondence rules "give empirical import or applicability to the calculus by interpreting some of its formulas in empirical terms—namely in terms of the vocabulary that serves to describe the phenomena which the theory is to explain". While in his new account, the bridge laws are part of the theory, in the Received View, the status of the correspondence rules is unclear. "One plausible construal" puts them in the metalanguage, stipulating the truth of certain sentences in the language of the

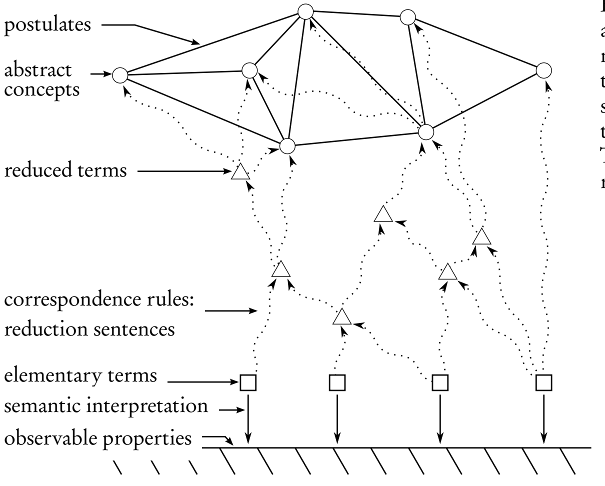

Figure 3.3: Feigl's diagram adapted to Carnap's first method of interpreting theoretical terms for one specific choice of the observational/theoretical bipartition. The dotted arrows point from reducing to reduced terms.

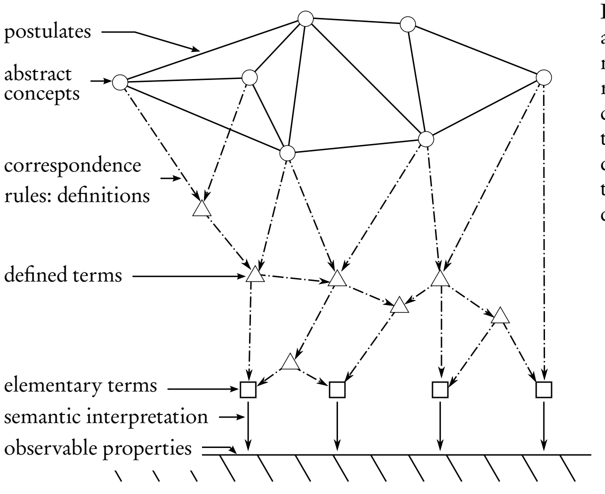

Figure 3.4: Feigl's diagram adapted to Carnap's second method of interpreting theoretical terms for one specific choice of the observational/theoretical bipartition. The dot-dash arrows point from terms of the definiens to the definienda.

theory "containing both theoretical and pre-theoretical terms".

This construal of the correspondence rules is central for one of Hempel's criticisms. Questioning the advantages of the Received View's assumption of axiomatization, he notes that axiomatization yields a "general criterion determining, for any sentence S, whether S is asserted by the theory". But this is not a defense of the Received View, he states, because "the standard construal assumes axiomatization only for the formulas of the uninterpreted calculus C rather than for all the sentences asserted by T" (Hempel 1970, 148–149, §3). This is very, very puzzling. First, because Carnap clearly assumes the axiomatization of the correspondence rules in his publications, starting from "Aufgabe". This is unsurprising, since correspondence rules derive from other, also axiomatized theories. Second, and more importantly, because this statement renders Hempel's article inconsistent. In a footnote to the above quotation, Hempel states that "by contrast[,] in the investigations by Ramsey (in 'Theories') and by Craig concerning the avoidability of theoretical terms in favor of pre-theoretical ones, axiomatization of the entire theory is presupposed." (Hempel 1970, n. 8). But in the *Introduction*, Carnap uses Ramsey's method to avoid theoretical terms, and Hempel (1970, 146, n. 4) cites the *Introduction* as one exposition of the Received View. Furthermore, he states that he has "himself relied on the standard construal" in "Dilemma", where he uses Craig's and Ramsey's methods himself. Thus Hempel's conceptualization of bridge principles as axiomatized is that of the Received View's conceptualization of correspondence rules.

Another criticism of Hempel's relies on an elaboration of the Received View's concept of empirical terms (Hempel 1970, 153, §4):

> the assumption of an axiomatized uninterpreted calculus [. . .] suggests that the basic principles of a theory [. . .] are formulated exclusively by means of a 'new' theoretical vocabulary [. . .]. Actually, however, the internal principles of most scientific theories employ not only 'new' theoretical concepts but also 'old', or pre-theoretical, ones [. . .]."

This seems to confuse both the explication with the description of theories and Hempel's own new account with the Received View: What vocabulary most scientific theories do, in fact, employ, is irrelevant for a rational reconstruction. As Suppe has argued at length, it is only relevant whether the theories can be reconstructed with a specific vocabulary. And as pointed out in connection with the bipartition of the language (see §3.6.2), how theoretical terms are distinguished from observational terms is a matter of convention. Hempel's new account is committed to the distinction between antecedent vocabulary and vocabulary newly introduced by one specific theory, but this is not the distinction of the Received View.

Both of Hempel's criticisms can be presented in relation to his new account: In the first case, he criticizes the Received View on the grounds that, unlike his new account, correspondence rules are given in the metalanguage, while in fact the Received View does not differ in this respect from his new account. In the second case, Hempel attributes a position of his new account (that theoretical terms are those introduced by the theory under investigation) to the Received View, while in fact only his new account relies on this assumption. This confusion about basic assumptions may be the result of something like a Gestalt-switch in Hempel's thinking: Friedman (2003) points out how Hempel was very close to Neurath's naturalism during the time of the Vienna Circle, and how he followed Carnap in focusing on explication while in close contact with him. Also under the influence of Kuhn, who joined him in Princeton, Hempel swayed back to a naturalistic description of science. It seems to me that after this paradigm shift, Hempel had, as Kuhn conjectured for scientists in general, difficulties appreciating the basic assumptions of the previous paradigm, the Received View.

While Kuhn contributed to Hempel's disavowal of the Received View (Hempel 1993) and thus moved him farther away from Carnap, there is also a clear influence of Carnap on Kuhn via Hempel with respect to meaning holism. In what way Carnap was even of the *same* opinion as Kuhn in this respect has been an object of discussion, for example between Earman (1993) and Irzik and Grünberg (1995). Their perceived disagreement, however, stems from Irzik and Grünberg misunderstanding Earman's point. "By semantic holism", Irzik and Grünberg, 289 "mean the doctrine that the theoretical postulates of a theory contribute to he meaning of theoretical terms occurring in them and that a change in the meaning of theoretical postulates results in a change in meaning", and they state that, contrary to Earman, Carnap was a meaning holist. Their point is straightforward: According to Carnap, theoretical terms are partially interpreted via correspondence rules and the postulates of the theory, and therefore a theory change leads to a change of the interpretation of theoretical terms. Earman, however, considers and rejects the possibility of Carnap being a meaning holist *by way of* a denial of the analytic/synthetic distinction Earman (1993, 11–12). Given Carnap's steadfast adherence to the distinction, such an attempt must fail. Earman (1993, 12) does allow for Carnap being a meaning holist in other ways, and indeed explicitly refers to a discussion by Friedman (1987) for an argument to that effect.

There *is* a connection between Carnap and Kuhn with respect to the lack of an analytic/synthetic distinction and meaning holism, but it is genealogical, not conceptual. Kuhn (1993, 312) phrases this kind of meaning holism thusly: "When the process [of learning scientific terms] is complete, the language or concept learner has acquired not only meanings, but also, inseparably, generalizations about nature". That is, concept formation in the sciences is "inseparably" connected to empirical knowledge. Kuhn (1993, 312) finds this position in Hempel's *Concept Formation* and states: "Though it was many years before I saw its full relevance to my emerging position, it fascinated me from the start, and its role in my intellectual development must have been considerable". I have already noted that Hempel's monograph borrows heavily from Carnap's *Begriffsbildung*, and even arrives at its core point, the close connection between concept and theory formation, based on considerations from Carnap's "Aufgabe" and *Begriffsbildung*. Hempel points this out in connection with reduction sentences both in *Concept Formation*, 680, and in his discussion of Carnap's philosophy of science (Hempel 1963, 691), where he adds that

> the introduction of fruitful new concepts in science is always intimately bound up with the establishment of new laws, as is shown quite clearly already in Carnap's early little work, *Physikalische Begriffsbildung*, which presents a lucid elementary analysis of the operational and the logical aspects of concept formation in physics.

Hempel wrote the first version of this article in 1954, two years after the publication of *Concept Formation*. There is little room for doubting that he held this opinion already when writing the latter. This very straightforward genealogical relation between Carnap's and Kuhn's meaning holism must not be mistaken for a conceptual relation: Carnap never made the step from the factual interrelation of concept and theory formation to the "inseparable" connection of meaning to factual statements. Conceptually, Carnap's insistence on the analytic/synthetic distinction blocked this path, as Earman notes. On the other hand, Carnap's concept of explication, or his earlier principle of maximal simplicity, clearly states the enormous role of empirical facts in the development of scientific concepts. It is only that a theory is not all empirical fact, so that concept and theory formation go hand in hand without making the analytic/synthetic distinction impossible.

Another genealogical influence of Carnap on purported adversaries goes through Poland. The formal semantics that I have presented in §2.8 are developed in a little and little known book on "the logical syntax and semantics of the language of empirical theories" by Marian Przełęcki (1969, 1). His analysis is based on "Testability", *L&M*, "Theoretical concepts", *Concept Formation*, "Dilemma", and other works in the tradition of the Received View. Like these, he assumes a distinction between observational and theoretical terms and allows a direct interpretation only for observational terms. Przełęcki (1975, 284) thought of himself as "positivistically-minded" and of the monograph as an introduction to the Received View (Przełęcki 1974b, 402).

The major invention of Przełęcki's account is the inclusion of vague terms (Przełęcki 1969, §3.II) in his discussion. This also allows him to consider vague observational terms, and here he parts ways with Carnap's account: While in Carnap's treatment, observational terms do not apply to unobservable objects (in that, for example, an object too small to see is determinately not red), Przełęcki construes observational terms as completely vague for unobservable objects (in that, for example, for an object too small to see it is undetermined whether it is red or not red) (Przełęcki 1969, 39–41). Note that Przełęcki's semantics is thus exactly that of Andreas (2010) as far as true theories are concerned (see §2.10.2). Note also that this construction still blocks Suppe's claim that parts of a red object that are too small to see are also red. Przełęcki (1969, ch. 10) further suggests a conception of theoretical terms as relative to a theory, along the lines of Rozeboom (1970). This he does under the heading "Towards a more realistic account", echoing both Carnap in "Aufgabe" and Rozeboom in their position that the Received View is an idealization with the hope of being generalizable.

Przełęcki's work is interesting for one because it has both been claimed to be in the tradition of the Received View (Pearce 1981, 3) and to be a precursor or even an elaboration of the Semantic View (da Costa and French 1990, 249; Volpe 1995, 566) by the respective view's proponents. Van Fraassen (1980, 64, n. 22; 1989, 227) even traces the inspiration for his concept of empirical adequacy, the core concept of his account, to Przełęcki's work. The Semantic View was initially developed as competitor to the Received View (see Suppe 1974a, §V.C), and so it is surprising that there is such a direct connection between the two.

As I have pointed out above, both Feigl and Hempel considered the semantic axiomatizations by Patrick Suppes to be compatible with the Received View, and at the Urbana conference, that is, after having abandoned the Received View, Hempel stated that "[t]his procedure [semantic axiomatization] has certain logical attractions, but these do not invalidate any of the reservations that I expressed concerning some of Professor Suppes's arguments for the desirability of axiomatization in science, for those arguments were essentially independent of the particular mode of axiomatic formalization" (Achinstein et al. 1974, 257). Suppe's account of the death of the Received View is therefore blatantly misleading: Hempel did abandon syntactic axiomatizations, but not because they are syntactic, but because they are axiomatizations. At no point in his discussion does he take issue with the reliance on syntax, and his criticisms are even directed explicitly at a defense of semantic axiomatizations by Suppes (1968).

3.10 Conclusion

> [Heraclitus's] words, like those of all the philosophers before Plato, are only known through quotations, largely made by Plato or Aristotle for the sake of refutation. When one thinks what would become of any modern philosopher if he were only known through the polemics of his rivals, one can see how admirable the pre-Socratics must have been, since even through the mist of malice spread by their enemies they still appear great.
>
> (Russell 1961, 64)

I do not think that it needs malice to misrepresent a position, but honest opponents (and even neutral expositors, like Russell arguably was) may very well make mistakes in their exposition, especially when the topic is emotionally charged. And Suppe (2000, §1) speaks of the Received View as having been "under sustained attack", with the "war [having] been won" because the Received View "died". Van Fraassen (1989, 365–366) speaks about the Received View as a "tragedy", and uses

(van Fraassen 1989, 365–366, n. 8)

> the word deliberately: It was a tragedy for philosophers of science to go off on these logico-linguistic tangles, which contributed nothing to the understanding of either science or logic or language. It is still unfortunately necessary to speak polemically about this, because so much philosophy of science is still couched in terminology based on a mistake.

Rozeboom (1970, 196) prefaces his discussion of the Received View with a regret:

> [B]y the late 1950's the empiricist analysis of scientific theory had pushed to the brink of what could have been—and might still become—a revolutionary breakthrough in the philosophy of cognition. However, the dominating style of philosophical argument, persuasive and holistically critical rather than discovery oriented, has severely impeded realization of this prospect. By "persuasive and holistically critical", I mean dialectic which seeks primarily to recruit allegiance to some favored doctrine while treating any flaw of discomfiture in prima facie competing doctrines as sufficient ground for their total dismissal.

Thus I think it is plausible that the discussion about the Received View has been polemical enough to explain why even crystal clear philosophical positions have been misrepresented.

One such crystal clear position is the Received View's stance on type theory (§3.3). Neither Carnap, Hempel, nor Feigl ever restrict reconstructions of theories to first order logic. Carnap assumes type theory in all of his expositions of the Received View, and uses it in his philosophical discussions (e. g., his solution to the problem of empirical content and analyticity and his discussion of philosophical method) as well as in his analyses of specific mathematical and scientific theories. Hempel refers to Carnap's expositions of the Received View without ever taking issue with the use of type theory. And he himself uses type theory in his exposition of the relation between observational and measurement terms, and, like Feigl, lists explications of theories that use type theory as compatible with the Received View. That first order logic is cumbersome to use and inadequate to describe specific mathematical structures up to isomorphism cannot, therefore, form a viable argument against the Received View.

The proponents of the Received View also did not demand exhaustive axiomatizations of all of the mathematics that appears in a scientific theory (§3.4). Carnap describes how mathematical and logical constants with a standard interpretation can be used without mentioning their axiomatizations, and Hempel and Feigl consider reconstructions of theories that are not exhaustive to be compatible with the Received View. Furthermore, both Carnap and Hempel relied on non-exhaustive axiomatizations in their analyses of, for example, probability, scientific inference, and the role of theoretical terms. Hence, the difficulty of arriving at exhaustive axiomatizations cannot count against the Received View. What is more, the formalization of scientific theories usually chosen by Carnap is the same phase space approach as that of van Fraassen (§3.5).

The bipartition of the language is neither as implausible nor as complicated as claimed. Arguments to the contrary essentially rest on a misunderstanding of the structure of the correspondence rules and the unjustified assumption that theoretical terms never apply to observational objects (§3.6). Relatedly, the correspondence rules in the Received View are not nearly as misguided or naïvely simple as claimed by critics. In fact, they allow both hierarchies and statistical relations as demanded by Suppes. And both Carnap and Hempel assumed correspondence relations between observational terms and numerical terms that are in line with Suppes' and other's work on measurement theory. What is more, both Carnap and Hempel had historically accurate ideas about the formation of new concepts.

Also, of the many meanings and uses of the term 'model', Hempel and Carnap doubt only the indispensability of one use of models under one meaning (§3.7): Neither considers it *necessary* that the laws of a theory be given an alternative, visualizable interpretation to determine the consequences of the theory. Nonetheless, Hempel describes a host of conveniences that come with such analogical models, and both Carnap and Hempel see much value in analogical models in the context of discovery. And both Carnap and Hempel themselves employ theoretical models, one of the many other possible kinds of models. Because of this and since the Received View is, in fact, at least as hospitable to analogical models as the Semantic View described by Suppes, there is no obvious reason to dismiss the Received View or prefer the Semantic View because of the relevance of scientific models.

Additionally, not all explications are precisifications (§3.8.1), so it is prima facie not a problem when the explicatum of a term does not conform to the explicandum in all clear cases. Carnap explicitly takes this stance, and Hempel relies on Carnap's account of explication. To demand that explications be precisification would furthermore undermine the basic tenet of ideal language philosophy and thus of the Received View. Finally, the Received View is not meant as an explication of the term 'theory' (§3.8.2). Rather, it is meant as a framework in which specific scientific theories can be explicated. Furthermore, it is meant as a precise framework that allows one to explicate some theories, while the explication of other theories would require either their further development or a generalization of the Received View. Because of the last two points, it is not a failure of the Received View that it does not make the use of 'theory' in ordinary language more precise.

Thus I end this chapter on a gloomy note, for I think it has by now become clear that a viable philosophical positions was not so much analyzed and criticized as caricatured and smeared. The following chapters will be more positive, however, as they are, to use Suppe's words, an "exploration of where to go now".

Chapter 4

Defending the Received View

In the previous chapter, I have given a historical defense of the Received View on scientific theories and models by showing that many of the criticisms of the Received View rely on incorrect presumptions about the Received View. Often, I have also pointed to a host of discussion from its proponents that already expound what its critics claim is incompatible with the Received View. In this chapter, I will give a systematic defense of the Received View by showing that it indeed allows features that, according to its critics, are impossible within the Received View. Specifically, I will argue that it allows describing van Fraassen's notion of empirical adequacy (thus capturing van Fraassen's ideas about the relation between theory and observation) and partial structures (thus formalizing the lack of knowledge of a domain). First, however, I will compare syntactic and semantic formalizations of theories more generally.

4.1 Syntactic and semantic formalizations

Suppes (1968, 654–656) lists a variety of rewards that come with formalizing a theory: explicitness, standardization, abstraction from non-essential aspects, objectivity, and the possibility of identifying self-contained, minimal assumptions. In particular, he suggests that these rewards can be reaped by using formalizations in set theory or first order predicate logic (Suppes 1968, 653). Presumably, he would also argue that formalizations in higher order logic and model theory can lead to the same rewards.[1]

[1] A very early version of this section has been presented at Herman Philipse's *Dutch Research Seminar in Analytic Philosophy* at Utrecht University. Parts have been presented under the title "What's right with a syntactic approach to theories and models?" at the *EPSA 09* at the Vrije Universiteit, Amsterdam, The Netherlands, on October 23, 2009 and at the workshop *Perspectives on Structuralism* at the Center for Advanced Studies/Munich Center for Mathematical Philosophy, Ludwig-Maximilians-

Let me for now call the reliance on formalizations in predicate logic of first or higher order 'syntactic', and the reliance on formalizations in set or model theory 'semantic'.[2] In this terminology, the Received View is a specific syntactic approach that additionally assumes a bipartition of the vocabulary and allows a direct interpretation only of the basic terms. In the following, I will argue that, contrary to common opinion, a number of problems that syntactic approaches allegedly face are solvable if they are solvable in semantic approaches. In discussing the relative merits of syntactic approaches, I will ignore the ontological question of whether theories can be identified with either kind of description (or, for that matter, with platonic objects, sets of propositions, thoughts, actions, connection weights in brains, combinations thereof…). Given that scientific theories are typically not formalized according to either kind of approach, they are probably ontologically different from both kinds of descriptions.[3]

There are good reasons to suspect that the use of predicate logic has fallen prey to the "persuasive and holistically critical" style of philosophical argument against the Received View that Rozeboom lamented; that is, it is often dismissed because of other, logically independent aspects of the Received View. For example, in the introduction to the proceedings of the Urbana conference, Suppe (1974a, 114) concludes that

> it is amply clear from the discussion of the observational-theoretical distinction and correspondence rules above that many of the epistemically relevant distinctions concerning theories cannot be drawn syntactically, and thus that the Received View's insistence on axiomatic canonical reformulation is untenable. Hence, if formalization is desirable in a philosophical analysis of theories, it must be of a semantic sort.

In the discussion of the theoretical-observational distinction, however, Suppe (1974a, II.B) only discusses its relation to the analytic-synthetic distinction and the problems of making the theoretical-observational distinction based on a partition of the vocabulary. Suppe's criticism of correspondence rules (Suppe 1974a, II.E) is based on his presentation of Suppes's hierarchy leading from observations to theories, which Suppe considers incompatible with the Received View. His description, however, is itself phrased in syntactic terms (cf. Suppe 1974a, 108, n. 225), and therefore clearly not an example of principled restrictions on syntactic

Universität München, Germany, on February 17, 2012. I thank the participants for helpful discussions. This section has also profited a lot from a reading group at Tilburg University with Reinhard Muskens and Stefan Wintein.

[2] Although this is standard terminology, it is somewhat incongruous: Set theory and predicate logic are foundational theories in that they can each be used as a sort of basic language in which to formalize other theories. Model theory, on the other hand, is one of those theories that can be so formalized.

[3] Ontologically different *given the majority's way of talking about theories*, that is. Relying on an artificial language methodology also in the metalanguage, I assume with Church (see §2.10) that the semantics of a theory is also a matter of choice. I will not, however, defend this assumption here.

approaches. None of his criticism rests on the Received View's use of predicate logic.

A similar pattern can be found in the already mentioned conclusion about the relation between theory and observation by van Fraassen (1980, 56):

> The syntactically defined relationships are simply the wrong ones. Perhaps the worst consequence of the syntactic approach was the way it focused attention on philosophically irrelevant technical questions. It is hard not to conclude that those discussions of axiomatizability in restricted vocabularies, 'theoretical terms', Craig's theorem, 'reduction sentences', 'empirical languages', Ramsey and Carnap sentences, were one and all off the mark—solutions to purely self-generated problems, and philosophically irrelevant. The main lesson of twentieth-century philosophy of science may well be this: no concept which is essentially language-dependent has any philosophical importance at all.

There are two possibilities to interpret this quote. On the one hand, the conclusion is drawn from arguments against an observational-theoretical distinction based on a bipartition of the vocabulary (van Fraassen 1980, §3.6),[4] and refers to Ramsey and Carnap sentences and theoretical vocabularies. This suggests that van Fraassen's critique is directed at the bipartition of a theory's vocabulary into an observational and a theoretical part. On the other hand, the talk of syntactically defined relationships, language-dependence, and Craig's theorem *suggests* that there is something wrong with the use of any syntactic concept, whether depending on a bipartition of the vocabulary or not.

My contention is that some philosophers of science have inadvertently developed a motte-and-bailey doctrine as described by Shackel (2005, 298–299). In a motte-and-bailey castle, the area of a bailey is determined by the weak defense at its perimeter, typically a palisade and a moat, the motte is a hill inside the bailey, sometimes surmounted by a keep (see figure 4.1). An attack on the bailey is far more likely to succeed than an attack on the motte, but if the bailey is taken, the defendants can simply retreat to the motte and wait until the attackers have given up. Then they can reclaim the bailey, a much nicer place than the motte. More spacious, for one thing. Analogously, a motte-and-bailey doctrine relies on an ambiguity between two positions: One easy to defend but comparably uninteresting (the motte), and one hard to defend but very interesting (the bailey). The rhetorical trick consists in retreating to the motte position when defending, and moving back to the bailey position when free to develop an argument, helped by the ambiguity.

The motte in this case is the claim that the Received View as typically conceived is untenable. This position is easy to defend, because beyond the use of predicate

[4] Van Fraassen (1980, 54, 220, n. 12) restricts the application of Craig's theorem to those cases as well.

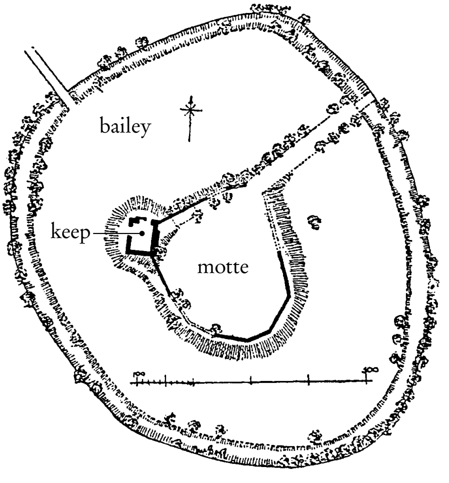

Figure 4.1: Plan of Duffus Castle, a motte-and-bailey castle near Elgin, Morray, Scotland (MacGibbon and Ross 1887, 279, labels added).

logic, the Received View is incorrectly taken to rely on exhaustive axiomatizations in first order logic, on a dismissal of scientific models, and correctly taken to rely on the mentioned bipartition of the scientific vocabulary, on the restriction of the interpretation to observational terms, and on the description of the relation between the axioms of a theory and observation statements by correspondence rules. Any of these additional assumptions can be questioned, and all of them have been. The bailey is the claim that any syntactical approach to scientific theories and models is untenable. This is a sweeping claim, which in van Fraassen's case, for instance, entails that predicate logic self-generates problems and does not allow developing concepts with any philosophical importance at all. The ambiguity that allows this motte-and-bailey doctrine is explicit in the synonymous use of the terms 'Received View' and 'syntactic view', and implicit, for example, in the above passages by Suppe and van Fraassen.

The previous chapter criticized a host of aspects of the motte position. In this section, I will consider the bailey position, the claim that syntactic approaches to theories and models are untenable.[5] I will try to defend syntactic approaches *relative to* semantic approaches, in which theories and models are formalized in formal semantic models or, and this will be repeatedly a point of discussion, in set theoretical structures.

Semantic approaches are widely thought to avoid a number of perceived shortcomings of syntactic approaches: (i) Syntactic approaches often have unintended models, unlike semantic approaches, (ii) syntactic approaches require and have

[5]It is already evidence for the irrelevance of the distinction between syntactic and semantic approaches in the philosophy of science that a rather thorough defense of the Received View is possible without ever countering the more fundamental bailey position.

failed to provide an account of the relation between language and the world, (iii) semantic approaches are language independent, (iv) the relation between theory and observation is misleading or wrong in syntactic approaches, (v) the description of scientific theories in syntactic approaches is cumbersome, (vi) with their focus on models, semantic approaches are closer to actual scientific practice. I will argue that if the first three of these problems can be solved for semantic approaches, they can be solved for syntactic ones as well. That the same holds for the last three problems then follows easily. Since I have already argued historically in §3 that the last three alleged problems do not occur in the Received View of Carnap, Hempel, and Feigl, this amounts to the systematic claim that their view is consistent.

Given its thesis, this section can probably count as a contribution to the "endless silly, largely unpublished debates over what semantic approaches can do that syntactical or statement approaches intrinsically cannot" which Suppe (2000, S103) laments. There are still justifications for its existence. First, beyond the intrinsic strength of axiomatizations in predicate logic, my discussion also covers the connection of a theory's description to the world, the relation between formalizations in set theoretical structures and formalizations in models, and the language independence of the approaches. Hence at most some of it is silly. Second, even if the debates are silly, many philosophers of science do hold the view that an analysis of science better use a semantic approach, and sometimes make it sound like semantic approaches are intrinsically superior. After all, Suppe (1974b, 114) himself claims early in the history of the discussion that "if formalization is desirable in a philosophical analysis of theories, it must be of a semantic sort". And looking back at the discussion since he made this claim, Suppe (2000, S110, my emphasis) concludes that "by construing theories in terms of families of models, semantic analyses—*and they alone*—have real potential for parlaying such new philosophical wisdom [gained by focusing on models] into enhanced understanding of theories". These claims suggest, at least on the surface, that there is an intrinsic advantage of semantic approaches. Third, if the debates are silly, it might be a good idea to finally put them to rest. This, incidentally, is what I would like to contribute to. As a first step, I may note that this contribution to the debate is published.

4.1.1 Translating between sentences and models

According to van Fraassen (1980, 44),

> [t]he syntactic picture of a theory identifies it with a body of theorems, stated in one particular language chosen for the expression of that theory. This should be contrasted with the alternative of presenting a theory in the first instance by identifying a class of structures as its models. In this second, semantic, approach the language used to express the theory is neither basic nor unique; the same class of

structures could well be described in radically different ways, each with its own limitations.

In the semantic approach, a theory is thus formalized by a class of structures, in the syntactic approach by a set of sentences. It seems plausible that every syntactic description of a scientific theory can be captured by a semantic one, because any set Σ of sentences determines the set $\mathbf{S}$ of its models through the mapping

$$\mathbf{\Phi} : \Sigma \mapsto \mathbf{S} := \{\mathfrak{S} \mid \mathfrak{S} \models \Sigma\} \,. \tag{4.1}$$

Σ has a fixed vocabulary $\mathscr{V}$, containing m_i-place predicates P_i, n_j-place functions F_j, constants c_k, and in higher order logic their respective types. $\mathscr{V}$ does not disappear by the mapping $\mathbf{\Phi}$, since every structure $\mathfrak{S} \in \mathbf{S}$ contains a mapping from $\mathscr{V}$ to a set of extensions with the corresponding arity and type. $\mathscr{V}$, sometimes called the 'signature of of $\mathfrak{S}$' can thus be read off uniquely from $\mathfrak{S}$ (Hodges 1993, 4).

However, $\mathbf{\Phi}$ does lose some information because it cannot distinguish between equivalent sets of sentences, that is, if $\Sigma \models\!\!\!\!\dashv \Theta$, then $\mathbf{\Phi}(\Sigma) = \mathbf{\Phi}(\Theta)$. This can pose problems, for example when modifying a theory: One formulation of a theory can be vastly superior to an equivalent one when it needs to be generalized or adjusted, as van Fraassen (1980, §3.5) has pointed out. Relatedly, the formulation is also relevant when it comes to the inductive support of parts of the theory: If the data support one postulate but not another, a formulation that keeps the two postulates separate is arguably better than one that contains a single postulate equivalent to their conjunction. Thus, if irrelevant conjunctions indeed pose a problem for an explicatum of 'confirmation' (cf. Fitelson 2002), the possibility of reformulating a theory allows hiding the conjunction, and thus make the irrelevant conjunct harder to detect. And if the problem of irrelevant conjunctions can *only* be solved by distinguishing between equivalent formulations, then semantic approaches are incapable of a solution.[6]

The loss of distinction between equivalent sets of sentences does not pose a problem, however, if the results of an analysis of a scientific theory are invariant under the theory's equivalent reformulation; and outside of questions of induction, many interesting analyses of scientific theories are so invariant. Conversely, it is often considered a problem if an analysis of the theory is not (witness, for example, Hempel 1965f, §§4–5; Carnap 1956b, 56; Winnie 1970, 294–295).[7]

[6]This, of course, assumes semantic approaches that indeed do not incorporate a theory's specific formulation.

[7]This also makes it easy to see that one well-known criticism of the Received View rests on a non-sequitur: Van Fraassen (1980, 55) states that in the Received View, a theory *TN*(0) that postulates absolute space cannot be empirically equivalent to a theory *TNE* that does not postulate absolute space, because even if absolute space is considered a theoretical term (and therefore existentially quantified over in a Ramsey sentence), *TN*(0) still postulates the existence of something that *TNE* does not postulate. But as the Ramsey sentence (2.15a) of the toy theory of pain (2.14) on page 48 shows,

Conversely, any set of structures $\mathbf{S}$ yields a set of sentences Σ through the mapping

$$\mathbf{\Psi} : \mathbf{S} \mapsto \Sigma := \{\varphi \mid \mathfrak{S} \models \varphi \text{ for all } \mathfrak{S} \in \mathbf{S}\} . \tag{4.2}$$

$\mathbf{\Psi}(\mathbf{S})$ is thus the set of sentences true in $\mathbf{S}$ (see §2.8.1). For many logics, $\mathbf{\Psi}$ is not the inverse of $\mathbf{\Phi}$. For instance, $\mathbf{S}$ may not be closed under elementary equivalence, while in first order logic, $\mathbf{\Phi}(\mathbf{\Psi}(\mathbf{S}))$ is. Of course, the set $\mathbf{S}' := \mathbf{\Phi}(\mathbf{\Psi}(\mathbf{S}))$ is mapped onto itself by $\mathbf{\Phi} \circ \mathbf{\Psi}$. Conversely, any set Σ of sentences that is closed under entailment is mapped onto itself by $\mathbf{\Psi} \circ \mathbf{\Phi}$. For such sets of sentences and sets of structures, $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are therefore duals in first order logic. Call the class of models of a single set of sentence *Δ-elementary*. It is known that there are classes of relational first order structures that are not Δ-elementary, but are complements of a Δ-elementary classes, or are the union of Δ-elementary classes. All and only classes that are unions of Δ-elementary classes are closed under elementary equivalence (Bell and Slomson 1974, 141–144).

Higher order logic can often distinguish between elementary equivalent but non-isomorphic structures, and the use of logics that allow formulas with infinitely many quantifiers, conjuncts, or disjuncts increases the number of classes for which $\mathbf{\Psi}$ is the dual of $\mathbf{\Phi}$. For simplicity, I will assume in the following that the classes of structures under discussion do not need to be described as the unions of different classes of models of sets of sentences.[8] But as a matter of principle (that is, independently of the logic used), $\mathbf{\Psi}$ always loses information because it cannot distinguish between two non-identical sets of element-wise isomorphic structures, that is, if $\mathfrak{S} \simeq \mathfrak{T}$ for all $\mathfrak{S} \in \mathbf{S}, \mathfrak{T} \in \mathbf{T}$, then $\mathbf{\Psi}(\mathbf{S}) = \mathbf{\Psi}(\mathbf{T})$ even if $\mathbf{S} \neq \mathbf{T}$. However, if the results of an analysis of a scientific theory are invariant under isomorphic transformations of theories (a natural demand that is usually fulfilled),[9] this does not pose a problem.

It is sometimes argued that as a matter of principle, more information is lost in the mapping $\mathbf{\Psi}$ than simply the distinction between isomorphic structures. Suppe (2000, S104), for example, argues that syntactic approaches are in principle unable to capture some analyses of scientific theories, because for syntactic descriptions,

> [a] general problem [is] that the Löwenheim–Skolem theorem implie[s] that [...] models must include both intended and wildly unintended models. Unintended models provide potential counterexamples.
>
> Blocking them more concerns eliminating syntactical-approach

equivalent reformulation can sometimes eliminate a perceived existential commitment. A toy theory for van Fraassen's example that is syntactically isomorphic to the toy theory for pain would be 'All objects are in an absolute space and every object in an absolute space moves straight unless acted upon'. I will discuss this problem more generally in connection with structural realism in §11.6.

[8] In other words, I will assume that the classes can be captured as the models of a single set of sentences or cannot be captured as models of sentences at all.

[9] See, for example, van Fraassen's notion of a theory (§4.2.1).

> artifacts than dealing with substantive analysis. [. . .] For example, Kitcher's (1989) unification explanation account has a very simple idea. But he develops it syntactically spending most of the paper trying to block unintended consequences that are artifacts of his formalism. [. . .]
>
> This is the correct sense of [the] claim symbolic logic is an inappropriate formalism.

First and foremost, Suppe's criticism is not directed at syntactic approaches (or symbolic logic) in general, because the Löwenheim-Skolem theorem does not hold in higher order logic. And the standard examples of structures that cannot be captured in first order logic, e. g. the natural numbers and the reals, can be described up to isomorphism in higher order logic if a full semantics is assumed (Enderton 2009, §2).[10] There are structures that cannot be described up to isomorphism, not even in predicate logic of transfinite order (Enderton 2009, §2, §4). However, set theory itself can be described up to isomorphism by (and only by) the addition of axioms about the existence of specific inaccessible cardinals (Väänänen 2001, 516). Since the proof theory of higher order logic is not complete for full semantics (Väänänen 2001, 505), entailment needs to be defined semantically in terms of structures. However, there is no good reason to disallow the use of structures in syntactic approaches, since they presume only that scientific theories can be analyzed by way of their description in predicate logic. They do not presume that the analysis itself must proceed wholly in the object language. As I have noted in §3.7, the use of formal models to define entailment was even part of the Received View.

Second, for predicate logic of any order, some structures *can* be characterized in that language up to isomorphism. If the theory has a finite domain, for example, all structures that are syntactically equivalent in a first order language are isomorphic (Hodges 1993, §2.2, ex. 5), so that there are no unintended models.

Third, whether the existence of unintended models poses a problem depends on the kind of analysis sought after. The answers to questions that can be phrased in the object language, for example, do not depend on isomorphism, since otherwise they would provide a means of distinguishing between non-isomorphic models. And an analysis that requires isomorphism in only a finite subdomain of the theory's domain (for example in the domain of observations) is immune to the problem even in a first order language. In general, it is doubtful that the observational content of a theory will be changed by the mapping $\boldsymbol{\Psi}$, as the number of observations will always stay well below any inaccessible cardinal.

Finally, Suppe's criticism rests on an equivocation of 'unintended model' and 'non-standard model', the latter referring to a model that is syntactically equivalent but not isomorphic to a standard model. Even though Kitcher spends a lot of

[10] Leivant (1994, §3.1, §5.4) and Väänänen (2001, 504–505) discuss the difference between full semantics and Henkin semantics. I will come back to this in §4.1.4.

work blocking unintended consequences, the unintended consequences are not syntactically equivalent to intended ones and can therefore be blocked by syntactic means. Kitcher's account is a good example of how difficult it can be to develop a formalization of an idea, but not of a failure of a syntactic approach because of the Löwenheim-Skolem theorem. Put crudely, if the unintended consequences could be blocked by using a predicate logic of higher order, Kitcher would probably not have shied away from it for lack of a complete proof theory.

4.1.2 Language independence

According to the quote by van Fraassen (1980, 44) above, "the same class of structures could well be described in radically different ways, each with its own limitations", and in the words of Suppe (1989, 4), the semantic approach

> construes theories as what their formulations refer to when the formulations are given a (formal) semantic interpretation. Thus, 'semantic' is used here in the sense of formal semantics or model theory in mathematical logic.

French and Ladyman (1999, 114–115) similarly assume that the structures used in semantic approaches do not contain a vocabulary when they discuss a criticism of semantic approaches they attribute to Mauricio Suárez: If a semantic approach uses models as they are defined in model theory, it is still dependent on a language, since a model "is a structure and an interpretation of a formal language in terms of that structure (that is, a map from the symbols of the syntax to elements of the structure)". If models are taken to involve such a mapping, French and Ladyman (1999, 114) write,

> it is clear that the celebrated claim of the linguistic independence of considering models (and not first-order formalizations of theories), stressed by adherents of the semantic approach as giving it a clear advantage over the syntactic view, is simply not true.

They quote a concurring passage by van Fraassen (1989, 366):

> The impact of Suppes's innovation [switching to models] is lost if models are defined, as in many standard logic texts, to be partially linguistic entities, each yoked to a particular syntax. In my terminology here the models are mathematical structures, called models of a given theory only by virtue of belonging to the class defined to be the models of that theory.

"Thus," French and Ladyman (1999, 115) conclude, "van Fraassen should be interpreted as talking about *structures* by those who wish to understand model in the sense of the 'standard logic texts'".

However, it is not exactly clear what French and Ladyman mean by 'structure', except that it contains the extensions of symbols, and not the symbols itself. They do claim that van Fraassen's "emphasis on structure is compatible with this definition of model theory from a contemporary textbook", according to which model theory "is the study of the construction and classification of structures within specified classes of structures". Of course, everything in this quote from Hodges (1993, ix) depends on his definition of 'structure'. A few pages further on (Hodges 1993, 1), there is evidence that his definition might not be what French and Ladyman think it is:

> Model theorists are forever talking about symbols, names and labels. A group theorist will happily write the same Abelian group multiplicatively or additively, whichever is more convenient for the matter in hand. Not so the model theorist: for him or her the group with '$\cdot$' is one structure and the group with '$+$' is a different structure. Change the name and you change the structure.

One of the reasons that Hodges (1993, 2) gives for this focus on symbols is that

> we shall often want to compare two structures and study the homomorphisms from one to the other. What is a homomorphism? [...] [A] homomorphism from structure A to structure B is a map which carries each operation of A to *the operation with the same name in B*.

But, of course, it is the *definition* of a structure that shows whether symbols play a role in model theory. Here is, for example, the part of his definition that deals with relations (Hodges 1993, 2, my notation):

> For each positive integer n [a structure contains] a set of n-ary relations on $|\mathfrak{A}|$ (i. e. subsets of $|\mathfrak{A}|^n$), each of which is named by one or more n-ary **relation symbols**. If R is a relation symbol, we write $R^{\mathfrak{A}}$ for the relation named by R.

It is clear that the symbols play an important role in a structure: They identify the extensions by naming them.

So where does the confusion stem from? Why would one assume that a structure contains not, for example, pairs of relation symbols and relations $\langle R, R^{\mathfrak{A}}\rangle$, but only the relations $R^{\mathfrak{A}}$? Again, some hints come from Hodges's text. For one, many model theorists take a liberal stance on what counts as a name. Hodges (1993, 3), for example, puts "no restrictions at all on what can serve as a name. For example any ordinal can be a name, and any mathematical object can serve as a name of itself". If an object can serve as its own name, it is understandable that the two might be confused. This possibility for misunderstanding is increased by a liberal stance on notation. Again Hodges (1993, 4, my notation):

> Some writers define $\mathfrak{A}$ to be an ordered pair $\langle|\mathfrak{A}|, \mathscr{I}\rangle$ where $\mathscr{I}$ is a function taking each symbol S to the corresponding item $S^{\mathfrak{A}}$. The important thing is to know what the symbols and the ingredients are, and this can be indicated in any reasonable way.
>
> For example a model theorist may refer to the structure
>
> $$\langle \mathbb{R}, +, -, \cdot, 0, 1, \leq \rangle .$$
>
> With some common sense the reader can guess that this means the structure whose domain is the set of real numbers, with constants 0 and 1 naming the numbers 0 and 1, a 2-ary relation symbol $\leq$ naming the relation $\leq$, 2-ary function symbols $+$ and $\cdot$ naming addition and multiplication respectively, and a 1-ary function symbol naming minus.

Seeing structures written as tuples might easily lead to the belief that structures are tuples. Of course, that this is a misunderstanding can be seen unequivocally from Hodges's definition of 'structure'. His is a standard definition also used by, for example, Chang and Keisler (1990, §1.3), who even explicitly define $\mathfrak{A}$ to be an ordered pair $\langle|\mathfrak{A}|, \mathscr{I}\rangle$. Smith (2008) calls structures as defined by Hodges '*labeled structures*', but I will stick with 'structures'. Like syntactic approaches, semantic approaches using structures therefore have one specific object language.

French and Ladyman would have found a definition better suited to their position in a less contemporary textbook (Bell and Slomson 1974, §3.2) in which

> a *relational structure* is an ordered pair
>
> $$\mathfrak{A} = \langle A, \{R_n \mid n \in \omega\}\rangle ,$$
>
> where [...] for $n \in \omega$ R_n is a finitary, say $\lambda(n)$-ary relation on A. [...] The relational structure $\mathfrak{A}$ will count as an interpretation of the language L $[= \{P_n : n \in \omega\}]$ if the degrees of the relations R_n correspond to the degrees $[\delta(n)]$ of the predicate letters P_n. That is, for $n \in \omega$, $\delta(n) = \lambda(n)$. In this case we say that the relational structure $\mathfrak{A}$ is a *realization* of the language L and that L is *appropriate* for the structure $\mathfrak{A}$.

This definition of what one could call '*pure structures*' (Smith 2008) seems to be used by many proponents of the semantic approach (e. g., da Costa and French 1990, French and Ladyman 1999). It is free of any specific vocabulary up to the vocabulary's similarity type, that is, any appropriate vocabulary can be used with the structure.

A structure $\mathfrak{A} = \langle A, \mathscr{I}\rangle$ contains an interpretation $\mathscr{I}$ mapping, say, the relations $\{P_i\}_{i \in I}$, from a vocabulary $\mathscr{V}$ to their extensions, $\{P_i^{\mathfrak{A}} \mid i \in I\}$. A pure structure, on the other hand, contains instead an indexed set $\{P_i^{\mathfrak{A}}\}_{i\in I}$ of extensions.[11] But this introduces a vocabulary through the back door: The mapping from I to the set $\{P_i^{\mathfrak{A}} \mid i \in I\}$ that is needed to define such an indexed set is is the same as an interpretation with the vocabulary I. Any claim that I is not a vocabulary but an index set has to rely on commitments (or rather declarations) outside of the formalism.

In the set theoretic semantic approach by Suppes and Stegmüller, pure structures are written as tuples with the domain as first element, which can avoid this hidden dependence on a specific vocabulary: Tuples may, for example, be introduced as primitives, with axioms like $\langle a_1,\dots a_n\rangle = \langle b_1,\dots,b_n\rangle \Leftrightarrow a_1 = b_1,\dots,a_n = b_n$. They may also be defined through sets, as in Kuratowski's definition. Of course, one can still introduce a vocabulary by assigning, for example, natural numbers to the positions of the tuple, but this assignment is not unique. In fact, any mapping from a well ordered index set (an ordered index set with a smallest element of each subset) with the right cardinality would do. The common method of representing a tuple $\langle a_1,\dots,a_n\rangle$ as a set of pairs $\{\langle 1,a_1\rangle,\dots,\langle n,a_n\rangle\}$ again introduces a vocabulary, the natural numbers $\{1,\dots,n\}$. This, incidentally, is the vocabulary that Carnap (1958, 242) chooses to name physical objects.

Since structures can have infinite vocabularies, but tuples are finite, pure structures cannot be tuples if they are to be able to express anything that structures can express. In line with the possibility of assigning some finite well-ordered index set to tuples, I therefore suggest to let a pure structure $\hat{\mathfrak{A}}$ contain a domain A and a mapping from an arbitrary well-ordered index set to a set of extensions on the domain. The arbitrariness of the index set can be captured by identifying any two mappings that only differ in their index sets:

Definition 4.1. A *representative of a pure structure* $\hat{\mathfrak{A}}$ is a triple $\langle A,a,\prec\rangle$, where $a : I \longrightarrow \mathscr{I}(\mathscr{V})$ is a mapping from the index set I to the image of an interpretation $\mathscr{I}$, and $\prec$ is a well-ordering of I. Two triples $\langle A,a,\prec\rangle$ and $\langle A,a',\prec'\rangle$ *represent the same pure structure*, $[\langle A,a,\prec\rangle] = [\langle A,a',\prec'\rangle]$, if and only if there is an order isomorphism $f : I \longrightarrow I'$ and $a \circ f = a'$.

The use of an interpretation $\mathscr{I}$ is simply a way to ensure that a maps only to set theoretical objects that can be extensions of predicate, function, or constant symbols. The definition determines a tuple if and only if the index sets of the representatives are finite. In both the finite and infinite case, the order of the objects in $\mathscr{I}(\mathscr{V})$ is preserved as in an ordered set, and additionally, elements of $\mathscr{I}(\mathscr{V})$ can occur repeatedly.

[11] It is clear from their use of ω as an index set that Bell and Slomson (1974) intend '$\{R_n \mid n \in \omega\}$' to be an indexed set, in my notation '$\{R_n\}_{n\in\omega}$' (rather then the set of the indexed set's elements). For if $R_2 = R_3$, then $\{R_i\}_{i\in\{1,2,3\}} \neq \{R_i\}_{i\in\{1,2\}}$, while $\{R_i \mid i \in \{1,2,3\}\} = \{R_i \mid i \in \{1,2\}\}$; Bell and Slomson thus need indexed sets to allow for different names with identical extension.

Definition 4.1 does not introduce a specific vocabulary, and is as language independent as the use of tuples: Even if not represented as a mapping from natural numbers to extensions, a tuple *can* be assigned a well-ordered index set, and this definition of pure structures only makes this possibility explicit. To stress the point: The infinity of vocabularies in definition 4.1 is not artificially introduced. Both indexed sets and tuples assume some set or multiple sets of entities that provide a vocabulary. Index sets can be used as vocabularies immediately, since they are already mappings from some set to a set of extensions. Tuples are either defined as a mapping from a range of natural numbers to a set of extensions or allow the introduction of such a mapping without any further assumptions except about the elements of the index set. In definition 4.1, the identification of structures with different index sets avoids this one further assumption.

Muller (2010) has suggested a similar conception for the use in semantic approaches, but he considers his conception an extension of semantic approaches, and justifies it with the need to connect theories to the world with its help. I will shortly revisit Muller's suggestion for connecting theories to the world below; at this point I only want to stress (again) that definition 4.1 is not an extension of the concept of a structure used in semantic approaches, but rather an acknowledgment of the infinity of vocabularies implicitly contained in it.

In the definition of 'structure' given by Bell and Slomson (1974), the index set I plays the role of the set of relation symbols $\{P_i \mid i \in I\}$ used by Chang and Keisler (1990) and Hodges (1993). For examples relevant in the following, see the definitions of 'reduct', 'isomorphism', and 'substructure' by Chang and Keisler (1990, 20–23) and by Bell and Slomson (1974, 153, 73), respectively. Definition 4.1 requires a somewhat more elaborate modification of standard definitions, of which I will only give the modification for the notion of embedding. Since pure structures are given by classes of mappings with well-ordered index sets, embeddings between pure structures can be defined via a bijection between the index sets of their representatives and a function from one pure structure's domain to the other's, just like for labeled structures. In this definition (which I give only for pure first order structures), the position in the ordering of the index sets plays the role of the element of the vocabulary.

Definition 4.2. A *pure first order structure* $\hat{\mathfrak{A}}$ can be *embedded* in a pure first order structure $\hat{\mathfrak{B}}$ if and only if for any two representatives $\hat{\mathfrak{A}} = [\langle A, a, \prec\rangle]$ with $a : I \longrightarrow \mathscr{I}(\mathscr{V})$ and $\prec$ an ordering on I, and $\hat{\mathfrak{B}} = [\langle B, b, \prec'\rangle]$ with $b : J \longrightarrow \mathscr{I}(\mathscr{V}')$ and $\prec'$ an ordering on J, there is an order isomorphism $g : I \longrightarrow J$ and an injective mapping $h : A \longrightarrow B$ such that

1. for all $c \in I$ mapped to constants by a, $h(a(c)) = b(g(c))$,
2. for all $F \in I$ mapped to n-ary functions by a and all $x_1, \ldots, x_n \in A$, $h\big(a(F)(x_1, \ldots, x_n)\big) = b\big(g(F)\big)(h(x_1), \ldots, h(x_n))$, and

$$\begin{tikzcd} I \arrow[r, "a"] \arrow[d, "g"] & \mathscr{I}(\mathcal{V}) & A \arrow[d, "h"'] \\ J \arrow[r, "b"] & \mathscr{I}(\mathcal{V}') & B \end{tikzcd}$$

Figure 4.2: Functions between arbitrary representatives of two structures $\hat{\mathfrak{A}} = [\langle A, a, \prec \rangle]$, $\hat{\mathfrak{B}} = [\langle B, b, \prec' \rangle]$ so that $\hat{\mathfrak{A}}$ can be embedded in $\hat{\mathfrak{B}}$.

3. for all $P \in I$ mapped to n-ary relations by a and all $x_1, \dots, x_n \in A$, $\langle x_1, \dots, x_n \rangle \in a(P) \Leftrightarrow \langle h(x_1), \dots, h(x_n) \rangle \in b(g(P))$,

h is called an *embedding* of $\hat{\mathfrak{A}}$ in $\hat{\mathfrak{B}}$. If h is surjective, $\hat{\mathfrak{A}}$ and $\hat{\mathfrak{B}}$ are called *isomorphic*.

Embeddability of pure structures only has to be shown for one representative of each pure structure:

Lemma 4.1. *Pure first order structure $\hat{\mathfrak{A}}$ can be embedded in pure first order structure $\hat{\mathfrak{B}}$ if and only if there are two representatives $\hat{\mathfrak{A}} = [\langle A, a, \prec \rangle]$ with $a : I \longrightarrow \mathscr{I}(\mathcal{V})$ and $\prec$ an ordering on I, and $\hat{\mathfrak{B}} = [\langle B, b, \prec' \rangle]$ with $b : J \longrightarrow \mathscr{I}(\mathcal{V}')$ and $\prec'$ an ordering on J, an order isomorphism $g : I \longrightarrow J$, and an injective mapping $h : A \longrightarrow B$ such that conditions 1–3 of definition 4.2 are fulfilled.*

Proof. The proof from left to right is immediate. For the other direction, assume $[\langle A, c, \prec^* \rangle] = \hat{\mathfrak{A}}$ with $c : K \longrightarrow \mathscr{I}(\mathcal{V})$ and $[\langle B, d, \prec'^* \rangle] = \hat{\mathfrak{B}}$ with $d : L \longrightarrow \mathscr{I}(\mathcal{V}')$. By definition 4.1, there are order isomorphisms $i : K \longrightarrow I$ and $j : J \longrightarrow L$ (see figure 4.3). Then there is an order isomorphism $k : K \longrightarrow L$ and a one-to-one mapping $h : A \longrightarrow B$ such that

1. for all $m \in K$ mapped to constants by c, $h(c(m)) = h \circ a \circ i(m) = b \circ g \circ i(m) = b \circ j^{-1} \circ j \circ g \circ i(m) = d(j \circ g \circ i(m)) = d(k(m))$,

2. for all $F \in K$ mapped to n-ary functions by c and all $x_1, \dots, x_n \in A$, $h\big(c(F)(x_1, \dots, x_n)\big) = h\big(a \circ i(F))(x_1, \dots, x_n)\big) = b \circ g \circ i(F)(h(x_1), \dots, h(x_n)) = b \circ j^{-1} \circ j \circ g \circ i(F)(h(x_1), \dots, h(x_n)) = d(k(F))(h(x_1), \dots, h(x_n))$, and

3. for all $P \in I$ mapped to n-ary relations by c and all $x_1, \dots, x_n \in A$, $\langle x_1, \dots, x_n \rangle \in c(P) \Leftrightarrow \langle x_1, \dots, x_n \rangle \in a \circ i(P) \Leftrightarrow \langle x_1, \dots, x_n \rangle \in a \circ i(P) \Leftrightarrow \langle h(x_1), \dots, h(x_n) \rangle \in b \circ g \circ i(P) \Leftrightarrow \langle h(x_1), \dots, h(x_n) \rangle \in b \circ j^{-1} \circ j \circ g \circ i(P) \Leftrightarrow \langle h(x_1), \dots, h(x_n) \rangle \in d(k(P))$

$\square$

The definition of embedding for pure structures respects the standard definition of embedding for structures (Hodges 1993, 5) in that there is a connection between structures and pure structures, and under this connection, the two definitions of embedding are interchangeable. More precisely, any structure $\mathfrak{A}$ gives

$$\begin{tikzcd}
K \arrow[d, "i"] \arrow[dr, "c = a \circ i"] \arrow[ddd, bend right, "k = j \circ g \circ i"'] & & \\
I \arrow[r, "a"] \arrow[d, "g"] & \mathscr{I}(\mathscr{V}) & A \arrow[d, "h"] \\
J \arrow[r, "b"] \arrow[d, "j"] & \mathscr{J}(\mathscr{V}') & B \\
L \arrow[ur, "d = b \circ j^{-1}"'] & &
\end{tikzcd}$$

Figure 4.3: Functions between four representatives of two structures $\hat{\mathfrak{A}} = [\langle A, a, \prec\rangle] = [\langle A, c, \prec^*\rangle]$, $\hat{\mathfrak{B}} = [\langle B, b, \prec'\rangle] = [\langle B, d, \prec'^*\rangle]$.

rise to a representative of a pure structure $\hat{\mathfrak{A}}$ by a well-ordering of its vocabulary, and any pure structure $\hat{\mathfrak{A}}$ contains a structure $\mathfrak{A}$ among its representatives. From the definition of embedding for pure structures, it follows that a pure structure $\hat{\mathfrak{A}}$ can be embedded in a pure structure $\hat{\mathfrak{B}}$ if and only if they are represented by some structures $\mathfrak{A}$ and $\mathfrak{B}$ with the same vocabulary and the same ordering of the vocabulary, and $\mathfrak{A}$ can be embedded in $\mathfrak{B}$:

Claim 4.2. *Pure first order structure $\hat{\mathfrak{A}}$ can be embedded in pure first order $\hat{\mathfrak{B}}$ if and only if there are first order structures $\mathfrak{A} = \langle A, \mathscr{I}\rangle$ and $\mathfrak{B} = \langle B, \mathscr{J}\rangle$ and a well-ordering $\prec$ such that $\hat{\mathfrak{A}} = [\langle A, \mathscr{I}, \prec\rangle]$, $\hat{\mathfrak{B}} = [\langle B, \mathscr{J}, \prec\rangle]$, and $\mathfrak{A}$ can be embedded in $\mathfrak{B}$.*

Proof. For the proof from left to right, choose any representatives $\langle A, a, \prec\rangle$ of $\hat{\mathfrak{A}}$ and $\langle B, b, \prec'\rangle$ of $\hat{\mathfrak{B}}$. Now choose for $\mathscr{V}$ the index set of a, $\mathscr{I} = a$, and $\mathscr{J} = b \circ g$, where g is the order isomorphism from I to J. The claim follows from definition 4.2 because for any $Q \in \mathscr{V}$, $h(Q^{\mathfrak{A}}) = h \circ a(Q) = b \circ g(Q) = \mathscr{J}(Q) = Q^{\mathfrak{B}}$ in the domain of $\hat{\mathfrak{A}}$, which is also the domain of $\mathfrak{A}$.

For the proof from right to left, choose $g = \mathrm{id}$. The claim follows with lemma 4.1 because for any $Q \in \mathscr{V}$, $h(\mathscr{I}(Q)) = h(Q^{\mathfrak{A}}) = Q^{\mathfrak{B}} = \mathscr{J}(g(Q))$. □

Putting it slightly differently, $\hat{\mathfrak{A}}$ can be embedded in $\hat{\mathfrak{B}}$ if and only if for *any* two structures $\mathfrak{A}$ and $\mathfrak{B}$ that represent, with a common well-ordering of their vocabulary, $\hat{\mathfrak{A}}$ and $\hat{\mathfrak{B}}$, $\mathfrak{A}$ can be embedded in $\mathfrak{B}$:

Corollary 4.3. *Pure first order structure $\hat{\mathfrak{A}}$ can be embedded in $\hat{\mathfrak{B}}$ if and only if for any structures $\mathfrak{A} = \langle A, \mathscr{I}\rangle$ and $\mathfrak{B} = \langle B, \mathscr{J}\rangle$ and any well-ordering $\prec$ such that $\hat{\mathfrak{A}} = [\langle A, \mathscr{I}, \prec\rangle]$ and $\hat{\mathfrak{B}} = [\langle B, \mathscr{J}, \prec\rangle]$, $\mathfrak{A}$ can be embedded in $\mathfrak{B}$.*

Proof. Immediately from claim 4.2 and lemma 4.1. □

Conversely, a structure $\mathfrak{A}$ can be embedded in a structure $\mathfrak{B}$ if and only if, under some well-ordering of their vocabulary, they represent pure structures $\hat{\mathfrak{A}}$ and $\hat{\mathfrak{B}}$, respectively, such that $\hat{\mathfrak{A}}$ can be embedded in $\hat{\mathfrak{B}}$. The proof assumes the axiom of choice.

Corollary 4.4. *First order structure* $\mathfrak{A} = \langle A, \mathscr{I} \rangle$ *can be embedded in* $\mathfrak{B} = \langle B, \mathscr{J} \rangle$ *if and only if* $[\langle A, \mathscr{I}, \prec \rangle]$ *can be embedded in* $[\langle B, \mathscr{J}, \prec \rangle]$ *for some well-ordering* $\prec$.

Proof. Note that $\mathscr{I}$ and $\mathscr{J}$ must have the same vocabulary. The proof from right to left follows immediately from theorem 4.2. For the proof from left to right, note that any vocabulary $\mathscr{V}$ can be well-ordered (assuming the axiom of choice). The claim again follows immediately from theorem 4.2. □

Again, putting this slightly differently, $\mathfrak{A}$ can be embedded in $\mathfrak{B}$ if and only if under *any* well-ordering, they represent pure structures $\hat{\mathfrak{A}}$ and $\hat{\mathfrak{B}}$, respectively, such that $\hat{\mathfrak{A}}$ can be embedded in $\hat{\mathfrak{B}}$:

Corollary 4.5. *First order structure* $\mathfrak{A} = \langle A, \mathscr{I} \rangle$ *can be embedded in structure* $\mathfrak{B} = \langle B, \mathscr{J} \rangle$ *if and only if* $[\langle A, \mathscr{I}, \prec \rangle]$ *can be embedded in* $[\langle B, \mathscr{J}, \prec \rangle]$ *for all well-orderings* $\prec$.

Proof. Immediately from corollary 4.4 and lemma 4.1. □

These results show that language independent pure structures and typical model theoretic operations between them are well-defined. It also shows that nothing is lost by discussing embeddings only for structures or only for pure structures. Language independence can be achieved at any point by introducing an ordering for the vocabulary, turning the structures into representatives of pure structures. Conversely, the ordering can be eliminated at any point in the discussion by choosing, for all pure structures at once, one of the possible vocabularies. It is clear that other model theoretic notions besides embedding have similar analogues for pure structures.

The independence from a specific vocabulary which French and Ladyman (1999) call 'much celebrated' seems indeed important, since language is inherently conventional, or at least relative to a group of speakers. Understanding 'propositions' "in the medieval sense of the term", that is, as "interpreted sentences of some particular language" Suppe (1974a, 204–205), argues that therefore theories cannot be the propositions in which they are formulated:

> Suppose a theory is first formulated in English, and then is translated into French. The English formulation and the French formulations constitute different collection of propositions; if theories were collections of propositions, then the translation of the theory into French would produce a new theory; but, of course, it does not—it is the same theory reformulated in French.

Suppe's point is clear: The ideal gas law does not change only because one uses the names 'pression', 'volume', 'constante de Boltzmann', and 'températur' instead of 'pressure', 'volume', 'Boltzmann constant', and 'temperature'. The use of structures as defined above avoids the use of any specific vocabulary, and thus supports the claim by van Fraassen (1989, 222) that "in discussions of the structure of theories [language] can largely be ignored".

The freedom the use of pure structures brings with it is limited, of course, because an extension's position in the pure structure takes over the role of its name in a structure. This is important when it comes to questions of embeddings between structures, since, for example, the pure structure corresponding to the tuple $\langle\{1,2,3\},\{1,2\},\{3\}\rangle$ is isomorphic to (and therefore embeddable in) $\langle\{a,b,c\},\{a,b\},\{c\}\rangle$, but not in $\langle\{a,b,c\},\{c\},\{a,b\}\rangle$.

The *simultaneous* switch of all structures to pure structures indeed leads to a freedom to choose between an infinity of vocabularies, and thus the freedom from any specific one. So assume that two models $\mathfrak{A}$ and $\mathfrak{A}'$ of two sets of sentences Θ in $\mathscr{V}$ and Θ' in $\mathscr{V}'$, under some order of their vocabularies, represent the same pure structure $\hat{\mathfrak{A}}$. Then the models are reducts of some structure $\mathfrak{B}$ in $\mathscr{V} \cup \mathscr{V}'$ which is a model of both Θ and Θ'. But $\mathfrak{B}$ is also a model of identity claims between any two elements of $\mathscr{V}$ and $\mathscr{V}'$ that have the same position in the pure structure. In the vocabulary $\mathscr{V} \cup \mathscr{V}'$, each symbol of $\mathscr{V}$ can thus be used as the definiens of the symbol of the same position in $\mathscr{V}'$ and vice versa. These identifications therefore allow the definitional extension of either theory to the other. Since explicit definitions are non-creative, neither theory gains content by the addition of the definitions, and thus it is possible to go from, say, Θ to Θ' by first extending Θ through the definitions and then eliminating the $\mathscr{V}$-symbols from the resulting theory. The effect, of course, is a simple renaming of the symbols, since up to equivalent reformulation, the two theories are syntactically isomorphic.

But in its neutrality with respect to the vocabulary, the semantic view is too liberal, for not every renaming makes sense. In the ideal gas law for example, the translation should allow the renaming of 'températur' into 'temperature', but not into 'pression'. For syntactic approaches, this restriction on renamings can be expressed by introducing analytic sentences. Any renaming not entailed by analytic sentences is not allowed. Then, 'temperature' can be renamed as 'températur', but not as 'pression'.

There is another problem with arbitrary renaming, for it is not only too liberal in some respects, it is also too restrictive in others. It does not, for example, avoid the language dependence referred to by Suppe (1974a, 204–205), who notes, relying on the medieval notion of 'proposition', that

> quantum theory can be formulated equivalently as wave mechanics or as matrix mechanics; whichever way it is formulated, it is the same theory, though its formulations as wave mechanics will constitute a collection of propositions which is different from the collection of

propositions resulting from its formulations as matrix mechanics.

Obviously, the difference between matrix mechanics and wave mechanics goes beyond a mere renaming of the predicate-, function-, and object-symbols. But this is a problem for the semantic view as well, because if more than the names of the extensions changes, the description in terms of pure structures changes as well. Hence when Hendry and Psillos (2007, 137, my emphasis) in connection with matrix and wave mechanics state that "a (semantic) model of one *could be turned* into a model of the other", they effectively point out that the pure structures are not the same.

The use of explicit definitions can again provide a solution to this kind of language and structure dependence. So far, only the simplest kind of definition was used, a simple identification of symbols. That way, it is clear that a theory Θ and a theory Θ' are equivalent given the definitions. But this comparison of two theories can be generalized to include any kind of explicit definitions. Two theories Θ, Θ' are definitionally equivalent if and only if both can be extended by explicit definitions such that they become equivalent. Then their models can be turned into each other by a procedure analogous to the one described above: First expand the model of one theory to include the defined symbols of the definitions. This expansion is unique. Then reduce the resulting interpretation to the vocabulary of the other theory (cf. Hodges 1993, 61). This allows the identification of theories that differ not only in the vocabulary they use, but also in the structures of their models. The procedure therefore also allows for structural differences in the description of theories, and thus can not only be used to extend the syntactic approaches, but also the semantic ones.[12]

Unlike the demand for the identification of inferentially equivalent sets of sentences, the demand for the identification of definitionally equivalent sets of sentences clearly must be restricted. The first reason is that not every symbol of one language can be translated into any symbol of another, as the example of 'pression' and 'températur' shows. The second reason is that when it comes to induction, Goodman (1965) argues that there are good and bad names, or, as he puts it, projectible and non-projectible terms. A theory containing the names 'green' and 'blue' and some term 't' for a specific time is definitionally equivalent to a theory that instead contains the names 'grue' and 'bleen' and 't' when, first, 'grue' is defined as 'green before t, blue at and after t', and *mutatis mutandis* for 'bleen', and, second, 'green' is defined as 'grue before t, bleen at and after t' and *mutatis mutandis* for 'blue'. But, assuming that t lies in the future, a theory like 'All emeralds are green' leads to correct predictions based on induction, while 'All emeralds are grue', leads to false predictions. For 'green', but

[12] There is no obvious reason why the identification of theories could not be relaxed even further. One could, for example, identify theories Θ in $\mathscr{T}$ and Λ in $\mathscr{L}$ if and only if Θ can be definitionally extended to entail Λ and vice versa. Once this step is taken, one could consider two theories identical when they are mutually reducible in one or another sense of reducibility (see §11).

not 'grue' is projectible, as far as emeralds are concerned.[13] The relativization of the projectibility of terms to the domain of emeralds is crucial, however. As Davidson (1966) argues, 'grue' is a perfectly projectible predicate as far as 'emeraters' are concerned, where an emerater is an emerald before t, and a large body of water thereafter; 'green', on the other hand, fails to be projectible relative to 'emerater'. It may thus be that there are no bad names, only bad combinations of names: 'Grue' and 'emerater' is good, as is 'green' and 'emerald', while 'green' and 'emerater' is as bad as 'grue' and 'emerald'. I will not pursue these considerations further, and will here just express the hope that it may be possible to describe good combinations of names via sets of systematic redefinitions of *all* terms in a vocabulary.

4.1.3 Sentences, structures, and the world

A set Σ of sentences of predicate logic is not enough for applying a theory to the world, for if $\mathfrak{A} \models \Sigma$, any set of the same cardinality as A can be made into a structure of Σ as well:

Claim 4.6. *Let* $|\mathfrak{A}| = A$ *and* $\mathrm{card}(A) = \mathrm{card}(B)$. *Then there is a* $\mathfrak{B} \simeq \mathfrak{A}$ *with* $|\mathfrak{B}| = B$.

Proof. If $\mathrm{card}(A) = \mathrm{card}(B)$, then there exists a bijection $g : A \longrightarrow B$. For each relation $P_i^{\mathfrak{A}}$ in $\mathfrak{A}$, define $P_i^{\mathfrak{B}} := g(P_i^{\mathfrak{A}})$, for each function $f_j^{\mathfrak{A}}$ and each tuple $\bar{b} \in B^{n_j}$ (or arguments for functions of higher order), define $f_j^{\mathfrak{B}}(\bar{b}) = g(f_j^{\mathfrak{A}}(g^{-1}\bar{b}))$, and for each constant $c_k^{\mathfrak{A}}$, define $c_k^{\mathfrak{B}} := g(c_k^{\mathfrak{A}})$. It is straightforward to show that $\mathfrak{B} \simeq \mathfrak{A}$. □

Corollary 4.7. *Let* $\mathfrak{A} \models \Sigma$ *and* $\mathrm{card}(A) = \mathrm{card}(B)$. *Then there is a* $\mathfrak{B}$ *with* $|\mathfrak{B}| = B$ *and* $\mathfrak{B} \models \Sigma$.

Proof. From claim 4.6, because sets of sentences can determine structures at most up to isomorphism. □

Considering that any bijection between a set and the domain of $\mathfrak{A}$ will do for the proof of claim 4.6, it is clear that any element of A can be exchanged for any other element. In light of claim 4.7, it is clear that a set of sentences can be connected to the world (beyond statements about the number of its objects) only if there is a way to distinguish between isomorphic structures. For this reason, Przełęcki introduces the the set of intended structures **N** (see page 43) and I have introduced the set of possible structures **M** (definition 2.2). **M** allows Σ to make more than cardinality claims about the world, and **N** allows checking whether Σ is true in the world (rather than just makes a correct cardinality claim about

[13] This consideration is arguably a generalizations of the considerations of the importance of differences in equivalent axiomatizations discussed in §4.1.1, which also revolve around non-deductive inferences.

the world). The definitions of **M** and **N** allow the assumption that the vocabulary $\mathscr{V}$ of Σ is bipartitioned into a directly interpreted set $\mathscr{B}$ of terms and a set $\mathscr{A}$ of terms interpreted only via the interpretation of $\mathscr{B}$. But the bipartition can be trivial, specifically, the definitions do not assume that $\mathscr{A} \neq \varnothing$. Therefore one can choose $\mathscr{B} = \mathscr{V}$ and interpret *all* terms in $\mathscr{V}$ directly.

It is important to see that the requirement for possible structures does not trivialize syntactic approaches by allowing all semantic tools: As postulated in §2.8.1 and definition 2.2, the set **M** of possible structures only distinguishes between isomorphic structures, even if $\mathscr{B} = \mathscr{V}$; hence the isomorphic closure of **M** is always the class of all $\mathscr{V}$-structures. Therefore, the isomorphic closure of the possible models of Σ is always the class of all models of Σ (see the proof of claim 2.1)—in other words, in syntactic approaches, non-isomorphic structures have to be distinguished by sentences of the object language.

In the Received View, an interpretation maps the basic ("observational") vocabulary to extensions that contains objects of the world. This gives a general way of connecting syntactic descriptions to the world: Their interpretations map at least one symbol to an extension that contains an object of the world—a *worldly extension*, as I will call it from now on. Interpretations with a worldly extension are accordingly *worldly interpretations*, and the analogous holds for models and structures.

If there are worldly structures, the connection between syntactic descriptions and the world can be determined by worldly possible structures. These can be given by actually showing (pointing to) members of worldly sets and relying on the psychological fact of intersubjective agreements about similarity between experiences (Przełęcki 1969, 35–38).[14] More commonly, the worldly sets are described in a metalanguage (thereby leaving out the question of how the terms of the metalanguage are connected to the world; see §3.7). As van Fraassen (1989, 222) puts it: "Any effective communication proceeds by language, except in those rare cases in which information can be conveyed by the immediate display of an object or happening". Either approach assumes that it is possible to build sets out of worldly objects, which is a somewhat problematic assumption. Since an interpretation maps to such sets, it is also assumed that it is possible to have a function from a vocabulary to worldly extensions. Whether these assumptions are plausible is a matter of discussion, but reaches too far into the philosophy of mathematics to be discussed here. It will turn out that their status is irrelevant for the comparison of semantic and syntactic approaches with respect to their

[14]The reliance of logical empiricists on the ostensive interpretation of terms gave rise to the opinion that in effect, logical empiricism reduces theories to pointing and grunting. For a critique that is another example of the "holistically critical" attitude, see Angelides (2004, 402–403), who takes one consequence of this reliance to be "the current, popular idea that the natural sciences need not be governed by moral concerns; or if they do, such moral governance only need be justified by appeals to governmental, technological, or, happily in this most learned of eras, divine authority". For a constructive discussion of ostensive interpretations see Eberle (1990).

connection to the world.

The connection to the world has sometimes been claimed to be a strong point of semantic approaches, because they escape the problem of connecting linguistic entities to the world (Chakravartty 2001, 327). However, the connection between pure structures and the world is not completely straightforward either:

Corollary 4.8. *Let* $|\hat{\mathfrak{A}}| = A$ *and* $\mathrm{card}(A) = \mathrm{card}(B)$*. Then there is a* $\hat{\mathfrak{B}} \simeq \hat{\mathfrak{A}}$ *with* $|\hat{\mathfrak{B}}| = B$*.*

Proof. From claims 4.6 and 4.3. □

Thus just as syntactic approaches must have a means to distinguish between isomorphic structures, semantic approaches must have a means to distinguish between isomorphic pure structures. This suggests a more precise distinction between syntactic and semantic approaches:

Definition 4.3. *Syntactic approaches* describe theories with sentences in the object language and with structures. Non-isomorphic structures are only distinguished by sentences in the object language.

Definition 4.4. *Semantic approaches* describe theories with pure structures.

Semantic approaches thus distinguish between isomorphic and non-isomorphic pure structures by a (set theoretic) description of the pure structures.

How the connection between pure structures and the world is to be envisaged in semantic approaches depends on whether the pure structures are meant to be worldly or non-worldly. Suppe's statement quoted in section 4.1.2 that semantic approaches construe theories as the formal referents of the theories' formulations (see page 171) suggests that the pure structures used in semantic approaches contain worldly extensions, that is, are worldly themselves. Da Costa and French (2000, fn. 2) also seem to assume that the pure structures are worldly when they state that "the set-theoretic models are constructed in set theories with Urelemente (individual[s], systems, portions of the universe, real things,...)."

A pure structure has as one of its representatives a structure (with a vocabulary), and accordingly the existence of a worldly pure structure entails the existence of a worldly structure. This structure then provides the connection of the linguistic entities with the world. Therefore, if a semantic approach successfully connects theories to the world by using pure worldly structures, syntactic descriptions can be connected to the world as well.[15]

[15] For pure structures taken as indexed sets along the lines of Bell and Slomson (rather than equivalence classes of indexed sets), the modification of the discussion is straightforward: The index set I can be used directly as a vocabulary, providing the interpretation $a : I \longrightarrow a(I)$. If a specific vocabulary $\mathscr{V}$ is desirable, any isomorphism $g : \mathscr{V} \longrightarrow I$ leads to the interpretation $\mathscr{I} = a \circ g : \mathscr{V} \longrightarrow a(I)$. The discussions below can modified analogously.

Mostly, however, the pure structures discussed in semantic approaches are not supposed to be worldly (see, e. g., French and Ladyman 1999). They are simply abstract set theoretic entities—no dogs, observations, or electrons are members of the sets. Theses non-worldly pure structures then have to be connected to the world somehow. I will discuss four ways to do so.

In the first way, the theory described by pure structures is supposed to be isomorphic to some worldly pure structure.[16] But because of corollary 4.6, if the theory is about more than just the cardinality of its domain, there has to be one isomorphism (with an order isomorphism g between the index sets) or a set of distinguished isomorphisms that connect each non-worldly pure structure $\hat{\mathfrak{A}} = [\langle A, a, \prec \rangle]$ with those worldly pure structures that the theory structure is supposed to refer to. The result is again a worldly pure structure, since for each g and worldly pure structure $\hat{\mathfrak{A}}$, $[\langle A, a \circ g, g^{-1}(\prec)\rangle]$ is again a worldly pure structure, and thus the previous discussion applies.

A second way to connect a non-worldly pure structure to the world is chosen by Muller (2010). He suggests taking a pure structure (which he takes to be a tuple), and allowing it to be assigned any compatible vocabulary. The appropriate vocabulary can then be chosen according to expedience. The connection to the world is then given through an interpretation of the vocabulary. First, it is clear that this approach to connect semantic descriptions to the world presumes that it is possible to connect syntactic descriptions to the world. Furthermore, what Muller describes is basically a pure structure, since the index set of the structure can be assigned any vocabulary, that is, any other index set.

A third way to connect non-worldly pure structures to the world is the similarity of the pure structure to the system it is meant to represent. French and Ladyman (1999) argue successfully that this relation, left largely undefined, is much too vague to be of much use. Thus even if, within the boundaries of its vagueness, *similarity* connects theories to the world, the success of this approach has been achieved only by significantly lowering the precision of the analysis. Assuming that syntactic approaches aim for more precision, this approach therefore at best achieves a different goal.

French and Ladyman (1999, 115, cf. 113) argue instead that the connection to the world is simply not a problem to be solved by semantic approaches:

> The theoretical models [the pure structures of the theory] are held to relate to models of the phenomena and these are just other structures. That these represent real events and processes cannot be determined by the content of the theory, but is a pragmatic fact about our language [...] and it is unreasonable to demand that the semantic view explains the nature of representation in general.

Note that this clearly falls short of the idea that semantic approaches are *easier*

[16] In semantic approaches, isomorphisms are often used as analogues to Tarski's definition of truth.

to connect to the world than syntactic approaches. Without an account of the relation between theories and the world, French and Ladyman's stance completes a terminological confusion. Because of their reliance on possible structures, syntactic approaches are semantic, and now, stripped of their connection to the world and relying on pure structures, semantic approaches stay on the level of the language of set theory, and hence are purely syntactic.

In French and Ladyman's approach, there is simply one set of sentences about structures (those describing models of the phenomena) that are distinguished as representing real events. It is of historical interest to note that the idea to connect the sentences of a language to the world by determining their relations to a distinguished subset of the whole language goes back to Neurath (1932), who postulated the translatability of all scientific sentences into protocol sentences (cf. Carnap 1932). If neither predicate logic nor set theory are given a formal connection to the world, they are simply languages that are assumed to describe the world in some not further specified way. Then, for example, the question of non-standard models does not even arise (neither for first order nor higher order logic, nor set theory) unless an additional metalanguage is artificially introduced (Väänänen 2001). If there is a metalanguage, however, set theory and higher order logic are equally expressive (Väänänen 2001, 506–507). In a sense, then, the discussion in §4.1.1 about the possibilities to capture structures up to isomorphism stacks the deck against syntactic approaches: It rests on the assumption that set theory is not a language that describes the world, but *is* the world (specifically, the world is a worldly structure), and the task of syntactic approaches is to describe the set theoretic world. If one instead treats, with French and Ladyman, set theory just as another language, it cannot describe the world more precisely than higher order logic. Väänänen (2001) further argues that neither set theory nor higher order logic can describe the world more precisely than first order logic.

The set theoretical notions used in semantic approaches can be axiomatized within first or higher order predicate logic (Väänänen 2001). In this case, the one plausible advantage of semantic approaches stripped of their direct relation to the world becomes that they unify a major part of the mathematical component of scientific theories: Many mathematical concepts can be defined through the membership relation, which becomes the only non-logical mathematical constant. All axioms involving the defined constants then become theorems of the definitional extension of the axioms of set theory. This, of course, does not touch the problem of non-mathematical constants; some of these may still not be definable through set-membership, although their relations may be *axiomatizable* by sentences involving set-membership.[17] The difference between syntactic and semantic approaches *excluding* their respective connection of theories to the world then amounts to the restriction to set theory on the side of semantic approaches, and no such restriction on the side of syntactic approaches. If a theory is formalized

[17] This may also be the case with some mathematical concepts.

in a semantic approach, switching to a syntactic approach would thus allow the abstraction from all terms that do not originally occur in the formalized theory, including the membership relation. A semantic approach, on the other hand, would cease to be a semantic approach if one abstracted from the membership relation.

In conclusion, then, the connection to the world is equally problematic for syntactic and semantic approaches. And a solution for one approach, for example the pragmatic one suggested by van Fraassen (2006), also provides a solution for the other.[18]

4.1.4 On the motte position

I have so far argued directly against the bailey position, the position that syntactic approaches are untenable. I now want to show that the criticism of syntactic approaches is indeed a motte-and-bailey doctrine, that is, I want to discuss briefly why not all syntactic approaches have to rely on first order logic or exhaustive axiomatization, use partial interpretations, or ignore the role of scientific models.

The use of higher order predicate logic

As I have argued at length in §3.3, the restriction to first order logic is already historically inaccurate with respect to the Received View, which should be a hint that it is an inappropriate restriction for syntactic approaches in general. A possible justification of the restriction may be the absence of a complete proof theory in higher order logic, which leads to a loss of the nice features of first order logic described by Rantala (1978), and also makes it necessary to use model theory and thus structures when determining which statements are entailed by a set of sentences. However, there is no good reason to disallow the use of structures in a syntactic approach. As argued in section 4.1.3, syntactic approaches need possible worldly structures to makes more than cardinality claims about the world. But the use of a set of worldly structures that is not closed under isomorphism is, if anything, more problematic than the use a set of non-worldly structures closed under isomorphism. Thus, given that model theory only needs such non-worldly structures, there seems to be no further problem for using entailment as a means of inference. As pointed out in §3.7, Carnap himself relied on semantic entailment.

The impracticality of syntactic axiomatizations

Sometimes the difference between syntactic and semantic approaches seem to consist mainly in the difference between exhaustive and non-exhaustive axiomati-

[18] I have argued that solutions for semantic approaches are also solutions for syntactic approaches. The argument for opposite direction can rely on the fact that every structure represents a pure structure and, if needed, corollary 4.5.

zations (cf. Stegmüller 1979, Suppes 1992). But of course, both an axiomatization in a syntactic approach and an axiomatization in a semantic approach can be non-exhaustive. On the other hand, a scientific theory may well be exhaustively axiomatized as, for example, by Woodger (1939), and the description of structures may be given within an exhaustive axiomatization of, say, ZFC. As discussed in §3.4, Carnap notes and argues for the necessity of non-exhaustive axiomatizations, and also gives guidelines on which axioms to include and which steps of an inference to spell out in detail.

Carnap does insist that in principle, well-developed theories can be exhaustively axiomatized. And while there is no reason that syntactic approaches have to insist even in principle on exhaustive axiomatization, some authors seem to suggest that (i) this is what makes an approach syntactical, and (ii) this is also what leads to problems. I have just argued against (i), but I also want to look shortly at (ii). French and Ladyman (1999, 116), in the article in which they argue that the connection to the world is not a problem that has to be tackled by semantic approaches, object to such an exhaustive axiomatization of set theory based on two points. First, they note that "there are many set-theories these days". This either means that there are different equivalent formalizations or that there are different non-equivalent formalizations. In the first case, some set theoretic axioms are simply not mentioned in a non-exhaustive axiomatization, in the second case, this means that the theory is incompletely described by a non-exhaustive axiomatization until one of the different formalizations is chosen. The first case is reasonable, as stated above, but also available to syntactic approaches. The second case is unproblematic if the different axiomatizations agree on those statements that are being analyzed. If they disagree, then the analysis cannot proceed until the relevant axioms are given (see §2.3.2).

The second point is that, "when *set* and the *membership* relation are characterized implicitly by means of an axiom-system, then the Löwenheim-Skolem theorem tells us of the existence of non-standard models of set theory" (French and Ladyman 1999, 116). One thus has the choice of not mentioning the axioms or accepting non-standard models. But as Väänänen (2001) argues convincingly, the idea of non-standard models is meaningless unless one assumes a metalanguage in which to describe mathematical objects, and an object language in which one can try and describe those same mathematical objects with the help of some interpretation. The metalanguage itself has no non-standard models *until* a meta-metalanguage is introduced.

Furthermore, even assuming that the set theory that French and Ladyman (1999) are envisioning has a metalanguage, it is not clear that the *existence* of non-standard models is the result of axiomatization, and not rather their *discovery*. Finally, since French and Ladyman (1999) exclude the connection to the world from the problems the semantic view has to solve, there is no clear need for a model theory, and hence neither for non-standard models; it would not be clear what these models would stand for—since they are not connected to the world, they are at best some interesting formal feature of the axiomatization.

So I conclude that there is no obvious reason why exhaustive axiomatization leads to *less* precision, to wit, the inability to exclude non-standard models, but even if it did, this is not more of a problem for syntactic approaches than for semantic approaches. For neither approach needs to rely on exhaustive axiomatizations.

Models

I have ended my discussion of models in the Received View with the conclusion that arguments for the advantage of semantic approaches over syntactic approaches with respect to scientific models rest on an equivocation of scientific and model theoretic model. Here, I want to point out that independently of which meaning of 'model' is chosen, syntactic approaches are on a par with semantic approaches.

If 'model' is understood model theoretically (that is, synonymously to 'structure'), then the possibility of using higher order logic allows capturing structures up to isomorphism (*modulo* an assumption about inaccessible cardinals), and the use of possible structures with $\mathscr{V} = \mathscr{B}$ allows capturing individual structures. Furthermore, semantic approaches relying on pure structures involve structures only indirectly, as representatives of pure structures (semantic approaches relying on indexed structures involve structures directly, with the index set as the vocabulary). If, on the other hand, 'model' refers to scientific models, then the question is whether semantic approaches can describe the world (or anything that is not a pure structure, for that matter) better than syntactic approaches. I have answered this question in the negative in §4.1.1 and §4.1.3.

Partial interpretations and correspondence rules

The partition of $\mathscr{V}$ into basic and an auxiliary terms, the direct interpretation of only the basic terms, and the interpretation of the auxiliary terms through the interpretation of the basic terms and the correspondence rules are specific to the Received View. They are also obviously independent from the decision between syntactic and semantic approaches. That the use of a subvocabulary is not restricted to syntactic approaches can be seen from the model theoretic concept of a reduct of a structure, in which all those extensions not belonging to the terms of a subvocabulary are eliminated (Hodges 1993, 9). The definition of a reduct for a structure can be extended to also cover pure structures by using the set of positions in a structure rather then the set of terms in a vocabulary. In fact, in part II I will present a number of criteria of empirical significance that have been suggested within semantic approaches and that rely on a bipartition of the vocabulary.

Conversely, there is no reason why, for example, van Fraassen's concept of embedding, which relies on substructures rather than reducts, cannot be captured in syntactic terms. One attempt to this effect has been undertaken by Turney (1990), who introduces the concept of implanting as the syntactic counterpart of van Fraassen's *embedding*, that is, isomorphism to empirical substructures of a theory. I doubt that Turney's definition of 'implanting' does justice to van Fraassen's notion of embedding, but do think that embedding has a syntactical counterpart. The important point is that, while embedding is defined semantically in terms of structures or pure structures, there is no *prima facie* reason that the notion cannot also be captured syntactically. As already mentioned, Suppe's description (Suppe 1974a, II.E) of how a theory relates to the data according to Suppes (1962) is given in syntactic terms, and thus already shows that it is possible to capture this conception syntactically.

There is also no reason why a syntactic approach cannot rely on possible structures with $\mathscr{V} = \mathscr{B}$. As I have argued in §4.1.3, this does not trivialize the distinction between syntactic and semantic approaches. Historically, Feigl (1950, "personal postscript"), for example, did not endorse the view of a partial interpretation of the vocabulary of theories, even though he clearly was a proponent of syntactic approaches, and even something very close to the Received View.

There is, however, the possibility that the relation of a theory to the observations cannot be captured at all in any formalism, for example because the relation is achieved through completely implicit, contextual knowledge of the scientists. While a discussion of this possibility would lead to far afield, I might just note that this assumes that the implicit knowledge or the context dependence cannot be made explicit. That this is so certainly needs some kind of proof and would, because of the preceding results, either hold for both semantic and syntactic approaches, or for neither.

4.2 Empirical adequacy in the Received View

I have argued that the standard account of measurement and Suppes' account of a hierarchy between observations and theories can be captured in the Received View. I have also argued that van Fraassen's account of the formalization of scientific theories is that of Carnap. I have also argued that *if* van Fraassen relies on a single language, then his account is in fact that of the early Carnap, who relies on protocol sentences in the same language as the theory to connect theory and observations.

I now want to show that van Fraassen's account of the relation between theory and observation, empirical adequacy, can even be captured in the Received View with its bipartition of the language, and even assuming that the pure structures in van Fraassen's account have to be described in the object language of the Received View. Turney (1990) argues for a weaker result, namely that empirical adequacy can be captured syntactically, but his formalization of empirical adequacy rests on a misunderstanding of van Fraassen's position (see §4.2.1).[19]

4.2.1 Definitions

Within constructive empiricism, van Fraassen (1980, 64) states,

> [t]o present a theory is to specify a family of structures, its *models*; and secondly, to specify certain parts of those models (the *empirical substructures*) as candidates for the direct representation of observable phenomena.

Furthermore the models of the theory "are describable only up to structural isomorphism" (van Fraassen 2008, 238; cf. 2002, 22). More formally, this can be phrased as follows:

Definition 4.5. A *theory* is a family $\{\mathfrak{T}_n\}_{n\in N}$ of structures (the *models of the theory*) such that each of its members $\mathfrak{T}_n = \langle T_n, P_i^{\mathfrak{T}_n}, F_j^{\mathfrak{T}_n}, c_k^{\mathfrak{T}_n}\rangle_{i\in I_n, j\in J_n, k\in K_n}$ has a set $\mathbf{E}_n$ of *empirical substructures*, such that for each $\mathfrak{E} \in \mathbf{E}_n$, $\mathfrak{E} \subseteq \mathfrak{T}_n$. With each model, a theory also contains every isomorphic structure and its corresponding[20] empirical substructures.

For convenience in the sequel, I am not using pure structures in the definition. Given corollary 4.5, this is only a notational choice. Strictly distinguishing between the set O of observable objects and the unobservable objects, van Fraassen (1980, 64) suggests to describe observable phenomena by structures as well: "The structures which can be described in experimental and measurement reports we can call *appearances*" (van Fraassen 2008, 286). This suggests

Definition 4.6. *Appearances* are given by a set $\mathbf{P}$ of structures such that the domain of each $\mathfrak{P} \in \mathbf{P}$ is a subset of O. A structure $\mathfrak{P} \in \mathbf{P}$ is an appearance.

Note that the set of appearances does not have to be closed under isomorphism. Van Fraassen (1980, 64) then defines a theory to be "empirically adequate if it has some model such that all appearances are isomorphic to empirical substructures of that model" (cf. van Fraassen 1991, 12):

Definition 4.7. A theory $\{\mathfrak{T}_n\}_{n\in N}$ is *empirically adequate* for appearances $\mathbf{P}$ if and only if there is some $n \in N$ such that for every $\mathfrak{P} \in \mathbf{P}$, there is an $\mathfrak{E} \in \mathbf{E}_n$ with $\mathfrak{E} \simeq \mathfrak{P}$.

[19] An early version of this section has been presented under the title "Syntactic characteristics of empirical substructures" at the *PhDs in Logic* conference at Gent University, Belgium, February 19, 2009. I thank the audience for helpful discussion.

[20] To be precise: If $f : T_m \longrightarrow T_n$ is an isomorphism between $\mathfrak{T}_m$ and $\mathfrak{T}_n$, then the set $\mathbf{E}_n$ of empirical substructures that corresponds to $\mathbf{E}_m$ contains all and only those structures $\mathfrak{E}_n$ for which there is an $\mathfrak{E}_m \in \mathbf{E}_n$ such that f is an isomorphism between $\mathfrak{E}_m$ and $\mathfrak{E}_n$.

Definition 4.7 defines the empirical adequacy of a theory *relative* to a set of appearances. In contradistinction, empirical adequacy simpliciter is defined as empirical adequacy for the set of *all* appearances (cf. Monton and Mohler 2008, §1.5). Therefore any set **P** of appearances that does not contain all of them may allow the deductive inference that some theory is not empirically adequate; but the inference that a theory *is* empirically adequate will have to be in some way ampliative. And it is for empirical adequacy simpliciter that van Fraassen (1980, 12, emphasis removed) claims that "acceptance of a theory involves as belief only that it is empirically adequate". Thus it is possible to determine by deduction that a theory should not be accepted (assuming that one tries to avoid false beliefs), but not that a theory *should* be accepted.

A note on terminology: Van Fraassen (1980, 66) and others (e. g., Turney 1990, 431; Suárez 2005, §4.1; Monton and Mohler 2008, §§1.5–1.6) occasionally speak of the empirical adequacy of a theory as the embeddability of the appearances into a model of the theory. But the two are not equivalent: $\mathfrak{P} \in \mathbf{P}$ can be embedded in $\mathfrak{T}_n$ if and only if $\mathfrak{P}$ is isomorphic to *any* substructure of $\mathfrak{T}_n$ (Hodges 1993, 6). The substructure does not have to be an *empirical* substructure. In the following, I will call an isomorphic mapping to an empirical substructure an *empirical embedding*.

In a more puzzling oversight, some exponents of empirical adequacy, (e. g. Suárez 2005, 39), rely on

Definition 4.8. A theory $\{\mathfrak{T}_n\}_{n\in N}$ is *idiosyncratically empirically adequate* for appearances **P** if and only if for every $\mathfrak{P} \in \mathbf{P}$, there are an $n \in N$ and an $\mathfrak{E} \in \mathbf{E}_n$ such that $\mathfrak{E} \simeq \mathfrak{P}$.

Definitions 4.7 and 4.8 are equivalent if there is only one appearance, $\mathbf{P} = \{\mathfrak{P}\}$ (as assumed by Turney 1990),[21] but not in general: Let the appearances be given by the set of the two structures $\{\langle\{1,2\},\{1,2\}\rangle, \langle\{3,4\},\{3\}\rangle\}$. Let the theory be given by the family with members $\mathfrak{T}_1 = \langle\{0,1,2,\},\{0,1,2\}\rangle$ and $\mathfrak{T}_2 = \langle\{3,4,5\},\{3,4\}\rangle$ as well as the singleton sets of empirical substructures $\mathbf{E}_1 = \{\langle\{1,2\},\{1,2\}\rangle\}$ and $\mathbf{E}_2 = \{\langle\{3,4\},\{3\}\rangle\}$. Let all other models of the theory be isomorphic to $\mathfrak{T}_1$ or $\mathfrak{T}_2$ and have the corresponding empirical substructures. Then the theory is idiosyncratically empirically adequate by virtue of the identity mapping on each of the appearances' domains, but it is not empirically adequate.

Since theories are closed under isomorphism, an appearance is empirically embeddable in a model of a theory if and only if it is an empirical substructure of a model of that theory (Hodges 1993, ex. 1.2.4b). Therefore a theory is idiosyncratically empirically adequate if and only if all appearances are empirical substructures of models of the theory (that is, in definition 4.8, $\mathfrak{E} \simeq \mathfrak{P}$ could be

[21] This is decidedly *not* what van Fraassen in general assumes (personal email from June 15, 2011), and it is also incompatible with his definitions of appearances and empirical adequacy quoted above: The "structure*s*" given by measurements are appearances, and in the case of empirical adequacy, "all appearance*s* are isomorphic to empirical substructure*s*" of a single model.

exchanged for $\mathfrak{E} = \mathfrak{P}$). This is not the case for empirical adequacy: Let the appearances be given by the set of the two structures $\{\langle\{a,b\},\{a,b\}\rangle, \langle\{c,d\},\{c\}\rangle\}$, where a,b,c, and d are distinct objects. Let the theory be given by the family with the member $\mathfrak{T}_1 = \langle\{1,2,3\},\{1,2\}\rangle$ and the set of empirical substructures $\mathbf{E}_1 = \{\langle\{1,2\},\{1,2\}\rangle, \langle\{2,3\},\{2\}\rangle\}$. Let all other models of the theory be isomorphic to $\mathfrak{T}_1$ and have the corresponding empirical substructures. Then the theory is empirically adequate, but every bijection from $\{1,2,3\}$—and thus every isomorphism for $\mathfrak{T}_1$—maps 2, the object shared by the empirical substructures, to a single object. Since the domains of the appearances do not share an element, the appearances therefore can never be empirical substructures of the same model of the theory.

4.2.2 Syntactic empirical adequacy

I first want to show that, as Turney tries to show, there is no important philosophical divide between syntactic and semantic approaches when it comes to empirical adequacy. For van Fraassen, a theory is given by a family $\{\mathfrak{T}_n\}_{n\in N}$ and a set of substructures $\mathbf{E}_n$ for each structure $\mathfrak{T}_n$. The individual elements of $\mathbf{E}_n$ are not distinguished among each other (apart from being different), and each element $\mathfrak{E} \in \mathbf{E}_n$ is determined by its domain E and $\mathfrak{T}_n$. Thus for each $\mathfrak{T}_n$, $\mathbf{E}_n$ is determined by the extension $\{E \mid \mathfrak{E} \in \mathbf{E}_n\}$ of a predicate S of the next order. Therefore each $\mathfrak{T}_n$ can be taken as the structure $\mathfrak{T}'_n = \langle T_n, P_i^{\mathfrak{T}_n}, F_j^{\mathfrak{T}_n}, c_k^{\mathfrak{T}_n}, \{E \mid \mathfrak{E} \in \mathbf{E}_n\}\rangle_{i\in I_n, j\in J_n, k\in K_n}$. The resulting family $\{\mathfrak{T}'_n\}_{n\in N}$ can then be captured by a set of (sets of) sentences Θ. I will again for simplicity assume that Θ is a single set of sentences.[22]

Each appearance $\mathfrak{P} \in \mathbf{P}$ can also be described by a set of sentences $\Phi_{\mathfrak{P}}$. But, of course, $\mathfrak{P}$ contains only one of a host of domains of appearances, so that the sentences in $\Phi_{\mathfrak{P}}$ must be taken as relativized to one of those domains. Giving the domain P of $\mathfrak{P}$ the name '$E_{\mathfrak{P}}$', this leads to a set $\Phi_{\mathfrak{P}}^{(E_{\mathfrak{P}})}$ of relativized sentences.[23] The set of all appearances can then be expressed as the set of sentences $\{SE_{\mathfrak{P}} \mid \mathfrak{P} \in \mathbf{P}\} \cup \{\Phi_{\mathfrak{P}}^{(E_{\mathfrak{P}})} \mid \mathfrak{P} \in \mathbf{P}\}$, containing for each predicate $E_{\mathfrak{P}}$ the higher order sentence $SE_{\mathfrak{P}}$ stating that $E_{\mathfrak{P}}$ is an empirical predicate, and the relativized descriptions of the appearances. Empirical adequacy then becomes the consistency of a theory with the appearances:

Claim 4.9. *Let* $\{\mathfrak{T}'_n\}_{n\in N}$ *be the family of structures* $\mathfrak{T}'_n = \langle T_n, P_i^{\mathfrak{T}_n}, F_j^{\mathfrak{T}_n}, c_k^{\mathfrak{T}_n}, S^{\mathfrak{T}'_n}\rangle_{i\in I_n, j\in J_n, k\in K_n}$ *with* $S^{\mathfrak{T}'_n} := \{E \mid \mathfrak{E} \in \mathbf{E}_n\}$, *and* Θ *be such that* $\{\mathfrak{T}'_n \mid n \in N\}$

[22] As noted in §4.1.1, this will often require strengthening the logic beyond first order.

[23] The relativization $\sigma^{(E_{\mathfrak{P}})}$ of a sentence σ consists of the restriction of all quantifiers in σ to $E_{\mathfrak{P}}$ (Hodges 1993, Theorem 5.1.1). For any set Σ of formulas, $\Sigma^{(E_{\mathfrak{P}})}$ is the set of the relativization of the elements of Σ.

is the set of models of Θ. Let each appearance $\mathfrak{P}$ be described up to isomorphism by $\Phi_{\mathfrak{P}}$. Then $\{\mathfrak{T}_n\}_{n\in N}$ is empirically adequate for $\mathbf{P}$ *if and only if*

$$\Theta \cup \{SE_{\mathfrak{P}} \mid \mathfrak{P} \in \mathbf{P}\} \cup \left\{\Phi_{\mathfrak{P}}^{(E_{\mathfrak{P}})} \,\middle|\, \mathfrak{P} \in \mathbf{P}\right\} \tag{4.3}$$

is consistent.

Proof. The set (4.3) is consistent iff it has a model. This is the case iff there is some model $\mathfrak{T}_n$ of Θ such that for all $\mathfrak{P} \in \mathbf{P}$, there is an $\mathfrak{E} \in \mathbf{E}_n$ with $\mathfrak{E} = \mathfrak{T}_n|E \vDash \Phi_{\mathfrak{P}}^{(E_{\mathfrak{P}})}$. The latter is is the case iff $\mathfrak{E} \vDash \Phi_{\mathfrak{P}}^{(E_{\mathfrak{P}})} \cup \{\forall x E_{\mathfrak{P}} x\} \mathrel{\vDash\!\!\!\dashv} \Phi_{\mathfrak{P}}$, that is, iff $\mathfrak{E}$ is isomorphic to $\mathfrak{P}$. □

There is another way to capture empirical adequacy syntactically that relies on the basic idea that the appearances $\mathbf{P}$ determine the intended structures of a theory $\{\mathfrak{T}_n\}_{n\in N}$. This is in line with the notion of intended structures, since they are meant to incorporate empirical information. Thus, for the appearances $\mathbf{P}$, define the intended structures $\mathbf{N_P}$ as the class of those structures $\mathfrak{N}$ that contain all $\mathfrak{P} \in \mathbf{P}$ as substructures, and additionally contain the higher order extension $S^{\mathfrak{N}} := \{|\mathfrak{P}| \mid \mathfrak{P} \in \mathbf{P}\}$. Then $\{\mathfrak{T}_n\}_{n\in N}$ is empirically adequate for $\mathbf{P}$ if and only if Θ is true in at least one intended structure $\mathfrak{N} \in \mathbf{N_P}$. If the observations are given syntactically as well, that is, that $\mathbf{N_P}$ is up to isomorphism described by $\Phi_{\mathbf{P}}$, then the theory is empirically adequate if and only if $\Theta \cup \Phi_{\mathbf{P}}$ is consistent.

4.2.3 Empirical adequacy in a bipartitioned language

Claim 4.9 captures empirical adequacy as a consistency condition, assuming that the phenomena and the theory rely on the same vocabulary. In the Received View, however, the observations are identified by their vocabulary $\mathscr{B}$, and theoretical claims in a disjoint vocabulary $\mathscr{A}$ are connected to the observations by correspondence rules.

One way to capture empirical adequacy in the Received View works in the special case that each $\mathbf{E}_n$ is a singleton set,[24] and the modified theory $\{\mathfrak{T}'_n\}_{n\in N}$ (see §4.2.2) can be given by a single set of first order sentences Θ. Then Θ determines the class $\{\mathfrak{A}|E \mid \mathfrak{A} \vDash \Theta\}$ of empirical substructures, and $\{\mathfrak{T}_n\}_{n\in N}$ is empirically adequate if and only if $\mathbf{P} = \{\mathfrak{P}\}$ with $\mathfrak{P} \in \{\mathfrak{A}|E \mid \mathfrak{A} \vDash \Theta\}$. In this case, one can use the model theoretic fact that any class of substructures $\{\mathfrak{A}|E \mid \mathfrak{A} \vDash \Theta\}$ of a theory Θ is a *pseudoelementary* class $\{\mathfrak{A}|_{\mathscr{B}} \mid \mathfrak{A} \vDash \Theta^*\}$ for some theory Θ^* with at most as many non-logical constants as Θ (Hodges 1993, theorem 5.2.1). Θ^* can be seen as a rational reconstruction of Θ that allows for the analyses given within the Received View. Thus Θ is empirically adequate in light of the appearances $\{\mathfrak{P}\}$ if

[24] Although this is not van Fraassen's assumption (see n. 21), it is often presumed in discussions of empirical adequacy.

and only if $\mathfrak{P}$ can be expanded to a model of Θ^*, which by lemma 2.4 is the case if and only if $\mathfrak{P} \vDash \mathsf{R}_{\mathscr{B}}(\Theta^*)$.

Given that the empirical adequacy of a theory can be phrased syntactically in one vocabulary, it is very simple to capture empirical adequacy in the bipartitioned vocabulary of the Received View more generally. For this, one can simply treat the original vocabulary as theoretical and introduce a second, observational term for each theoretical one. Thus, for each P_i one can introduce an observational term P_i^*, and analogously for function and constant symbols. The appearances can then be given by structures with the starred, observational vocabulary instead of the unstarred vocabulary, and the correspondence rules simply identify the starred with the respective unstarred terms. Either of the syntactical methods from §4.2.2 then allows capturing empirical adequacy in the Received View.

In keeping with Carnap's hope that all observational terms can be defined in theoretical terms, one can also restrict the identification of theoretical with observational terms to observable objects. This is perhaps the most faithful translation of empirical adequacy into the Received View. For each relation P_i, the correspondence rule then can have the form

$$\forall \bar{x}[P_i^* \bar{x} \leftrightarrow \exists Y(SY \wedge Yx_1 \wedge \cdots \wedge Yx_{m_i}) \wedge P_i \bar{x}] \,. \tag{4.4}$$

Thus the observational relations can be explicitly defined in theoretical terms, as Carnap had hoped.[25]

Things are not as straightforward for functions, since functions have to be defined over the whole domain. One could thus substitute an $n_j + 1$-ary observational relation Q_j^* for each n_j-ary theoretical function F_j with

$$\forall \bar{x}[Q_i^* x_1 \ldots x_{n_j+1} \leftrightarrow \exists Y(SY \wedge Yx_1 \wedge \cdots \wedge Yx_{n_j}) \wedge F_j x_1 \ldots x_{n_j} = x_{n_j+1}] \tag{4.5}$$

or disappoint Carnap's hope and rely on the conditional definition

$$\forall \bar{x}[\exists Y(SY \wedge Yx_1 \wedge \cdots \wedge Yx_{n_j}) \rightarrow (F_j^* \bar{x} = F_j \bar{x})] \,. \tag{4.6}$$

However, since the value of the observational functions is irrelevant outside the domains of the empirical substructures, the conditional definitions could be completed by conditionally defining that outside of the empirical subdomains, the observational functions all have the same, arbitrarily chosen value. The explicit definition of all constant symbols is discussed in §4.2.4.

Thus the Received View can capture not only only Suppes' hierarchy from observations to theories and typical measurement relations, but also van Fraassen's account of the relation of theory and appearances.

[25] Friedman (1982, 276–277) sketches a similar approach for capturing empirical adequacy syntactically.

* * *

Claim 4.9 throws doubt on van Fraassen's claim that science "*aims to give us theories that are empirically adequate*" (van Fraassen 1980, 12), at least if empirical adequacy is meant to be the *only* aim of science, or even a particularly interesting one. For empirical adequacy is cheap: A tautology is consistent with any set of sentences describing the appearances, and thus its set of models is empirically adequate. This is obvious given claim 4.9, but can also be seen in van Fraassen's original definition: Simply choose as theory the class of all structures and for each structure the set of all its substructures as empirical substructures.

Claim 4.9 and its use in a formalization of a theory according to the Received View also allows a comparison of empirical adequacy with accounts of the relation between theory and observations suggested in the Received View. In line with, for example, Carnap (1966, 240–241), Muller (2010, 90) describes the aim of scientific theories in the Received View to be *observational adequacy*, for which a theory has to entail the observations made so far. He adds (Muller 2010, 90, n. 8):

> This concept of observational adequacy (ObsAdeq) should not be confused with Van Fraassen's well-known concept of empirical adequacy (EmpAdeq). Three differences between ObsAdeq and EmpAdeq: (i) ObsAdeq relies on the distinction between theoretical and observational concepts whereas EmpAdeq does not rely on that; (ii) ObsAdeq depends on historical time whereas EmpAdeq is timeless (quantifies universally over historical times); (iii) EmpAdeq requires for its definition a conception of a scientific theory that results from the Model Revolution, because it cannot be defined in the [Received View].

I have shown that differences (i) and (iii) are easily overcome, that is, can be avoided by capturing van Fraassen's definition of a theory syntactically. I have already pointed out that without an ampliative inference, van Fraassen is restricted to empirical adequacy with respect to the observations made so far; conversely, there is no reason to assume that the logical empiricists were only interested in theories that entail observations made so far, but will not entail the observations made in the future. Thus difference (ii) is not one of principle, but at best one of primacy of definition: One could define 'preliminary empirical adequacy' as 'empirical adequacy given the appearances observed so far', and define 'eternal observational adequacy' as 'observational adequacy at the end of all time'.

Muller misses the real difference: A theory is empirically adequate if it is not shown to be wrong by observations, while a theory is observationally adequate if it asserts what is observed. For this reason, Sober's remark that "empirical adequacy consists in making predictions that are borne out in experience" (Sober 1990) is similarly misleading. For empirical adequacy consists in *not* making predictions that are *not* borne out in experience.[26] And one way to achieve this is not to make

[26]Sober's formulation could be interpreted to state that empirical adequacy consists in making *only*

predictions.

This raises the question what the predictions of a theory are when it is defined in van Fraassen's way, and again the syntactic reformulation guides the way. If a theory Θ entails the observations, every model $\mathbf{T}'_n$ (defined as in claim 4.9) of Θ is also a model of the observations. This is the case if and only if every empirical substructure $\mathfrak{E} \in \mathbf{E}_n$ of every structure in $\{\mathfrak{T}_n\}_{n \in N}$, a theory as defined in definition 4.5, is isomorphic to an appearance, an element of $\mathbf{P}$.

4.2.4 Generalizing empirical adequacy

Before discussing a shortcoming of van Fraassen's conception of empirical adequacy, I shortly want to defend it against an unjustified criticism: In his account of the hierarchy connecting measurements and theories, Suppes (1962) notes that measurement results are discontinuous sets of values, while many theories are given for continuous values. In their "cardinality objection", Bueno et al. (2002, 503) argue that this poses a problem for empirical adequacy, because the domains of the appearances "in general are finite". The implicit assumption of the cardinality objection is that empirical substructures always have infinite domains, but this is not necessarily so. For theories can have both infinite and finite empirical substructures, since for any model $\mathfrak{T}_n$ of a theory, the set of empirical substructures $\mathbf{E}_n$ can be closed under the substructure relation, that is, if $\mathfrak{E} \in \mathbf{E}_n$ and $\mathfrak{E}' \subseteq \mathfrak{E}$, then $\mathfrak{E}' \in \mathbf{E}_n$, so that with every empirical substructure $\mathfrak{E}$, all of $\mathfrak{E}$'s finite substructures are empirical substructures as well.

This response to the cardinality objections is hampered only by the problems stemming from the severe restrictions placed by definition 4.5 on the models $\{\mathfrak{T}_n\}_{n \in N}$ of a theory. For it follows from the definition of a substructure that every constant of a model $\mathfrak{T}_n$ has to be in the domain E of *each* of its substructures $\mathfrak{E} \in \mathbf{E}_n$. Furthermore, every function of the model $\mathfrak{E} \in \mathbf{E}_n$ must map all (tuples of) elements of E to elements of E (Hodges 1993, lemma 1.2.2).

If now a theory $\{\mathfrak{T}_n\}_{n \in N}$ is empirically adequate, every appearance is a substructure of some $\mathfrak{T}_n$, so that $\mathfrak{T}_n$'s domain T_n contains observable objects, all constants of $\mathfrak{T}_n$ are observable objects, and all functions of $\mathfrak{T}_n$ map observable objects to other observable objects. If now the theory is about, say, elementary particles, the observable objects are, for example, the results shown on the measurement instruments or (parts of) the measurement instruments themselves. These then have to satisfy those formulas that the theory ascribes to elementary particles, and all constants in the theory have to be results shown on measurement instruments or, respectively, (parts of) measurement instruments.[27]

predictions that are borne out in experience, allowing the demand to be vacuously fulfilled. In this case, it would be correct (but misleading).

[27]Because of claim 4.6, this can *formally* always be arranged by defining appropriate relations, functions, and constants for the observable objects (assuming the right cardinality of the empirical substructures' and the appearances' domains), but this would trivialize empirical adequacy (cf. van

One may object to these criticisms for two reasons. For one, in early works van Fraassen (1970, §3) relied on "elementary statements" and a "satisfaction function" to give the relation between a theory and observations, so that one could argue that the model theoretic formalization above does not capture van Fraassen's position. However, van Fraassen (1989, 365, n. 34) himself states that he soon "found it much more advantageous to concentrate on the propositions expressible by elementary statements, rather than on the statements themselves". Thus van Fraassen had abandoned the reliance on elementary statements and satisfaction functions by the time he defined empirical adequacy. More importantly, empirical adequacy is defined without reference to either of the two concepts, and thus an analysis of empirical adequacy does not have to take them into account either.

One may also object that the terms 'embedding', 'substructure', and 'isomorphism' are not meant literally, but refer to relations between theories and appearances given by either satisfaction functions or something completely different. Possible support comes from van Fraassen's standard example of embedding, the seven point geometry (1980, §3.1; 1989, §9.1), which is not an embedding in the sense of model theory (Turney 1990, 441–443). But this looks more like an oversight than a conscious decision. More generally, the terms are well-defined within, but not outside of model theory, where in general they also do not occur together. And the objection makes van Fraassen use these terms in a different, undefined way *without pointing this out*. It also renders downright nonsensical passages in which he uses technical results from model theory. For example, van Fraassen (1980, 43) discusses cases in which "every model of T_1 can be embedded in (identified with a substructure of) a model of T_2." The parenthetical equivalence claim relies on the model theoretic definition of 'embedding' and 'substructure', on the closure of the set of models of a theory under isomorphism, and the equivalence of embedding and the substructure relation for isomorphically closed classes of structures (Hodges 1993, ex. 1.2.4b). If the terms were not meant in the model theoretic sense, there would be no reason at all for this equivalence claim.

I now want to show that the problem resulting from van Fraassen's reliance on substructures can easily be solved in the Received View. In van Fraassen's definition 4.5 of a theory, the components of a theory that are connected to the appearances are given by empirical substructures. The Received View, on the other hand, allows theories to contain non-observational terms, which themselves do not have to be directly connected to the appearances. This suggests the following generalization of definition 4.5:

Definition 4.9. A *theory* is a family $\{\mathfrak{T}_n\}_{n\in N}$ of structures (the *models of the theory*) such that each of its members $\mathfrak{T}_n = \langle T_n, P_i^{\mathfrak{T}_n}, F_j^{\mathfrak{T}_n}, c_k^{\mathfrak{T}_n}\rangle_{i\in I_n, j\in J_n, k\in K_n}$ has a set $\mathbf{E}_n$ of *empirical relativized reducts*, such that for each $\mathfrak{E}\in\mathbf{E}_n$, there is a set $\mathscr{A}_{\mathfrak{E}}$ of terms such that $\mathfrak{E}\subseteq\mathfrak{T}_n|_{\mathscr{A}_{\mathfrak{E}}}$. With each model, a theory also contains every

Fraassen 2006); I am assuming here that the trivialization problem has been solved.

isomorphic model and its corresponding empirical relativized reducts.

This means that every empirical relativized reduct $\mathfrak{E} \in \mathbf{E}_n$ is a relativized reduct of $\mathfrak{T}_n$, $\mathfrak{E} = \mathfrak{T}_n | E_{\mathscr{A}_{\mathfrak{E}}}$, which is already implicitly used by Suárez (2005, 38) instead of the notion of a substructure in his discussion of empirical adequacy. The formalizations of empirical adequacy in the Received View given in the preceding section are easily generalized to this new definition.

The new definition solves the problems connected with the use of substructures: There can be unobservable constants, and functions from observable objects to unobservable ones, since the constants' and functions' symbols may not be in the vocabulary $\mathscr{A}_{\mathfrak{E}}$. The definition may also alleviate the problem that, to be empirically adequate, a theory has to describe observational relations between observable objects: A theory may, for instance, already contain or be extended to contain the term 'being part of' and terms for observable objects and relations. This would allow, say, elementary particles to be treated as parts of observable objects, rather than as observable objects themselves (cf. Przełęcki 1969, 86–87). Both the elementary particles and the everyday objects would then be in the (extended) theory's domain, but the observational vocabulary of the theory might then refer only to observable objects, not elementary particles.

Besides these technical problems that stem from the reliance on substructures, there are two ways in which empirical adequacy is too strict to be used in most contexts of scientific research because it relies on an isomorphism between the appearances and the empirical substructures. For one, most theories' implications for the appearances are not true up to arbitrary precision. But the theories are still empirically adequate up to some precision, that is, *approximately* empirically adequate. Approximately empirically adequate theories like the ray theory of light, quantum field theory, and general relativity are clearly useful although not empirically adequate according to definition 4.7. For this reason, van Fraassen (1989, 366, n. 5) himself suggests, but does not explicate, the notion of 'approximate embedding'.

Second, scientists are seldom if ever in the situation that they know what the exact appearances are. Rather, their knowledge only restricts what appearances are epistemically possible, for example because it stems from measurements with a certain amount of error. Definition 4.7 does not allow for such lack of knowledge, since it refers solely to the set of appearances. This makes sense, since a theory that is empirically adequate as far as we know is not always an empirically adequate theory. Still, definition 4.7 may define a concept that can never be applied in scientific research. As van Fraassen (1991, 12) puts it: "Empirical adequacy, like truth, admits of no degrees".

Both problems can be solved at once when the appearances are given by sets of sentences, for sets of sentences allow omissions as defined in §2.12. Whatever is not known about the appearances can be omitted from their description. To capture the notion of approximate empirical adequacy, one omits every sentence that, given the approximation, could be false.[28]

4.3 Partial structures in the Received View

I have so far discussed means to capture semantic approaches that rely on structures. The partial structures approach generalizes the previously discussed semantic approaches because *partial structure* generalizes *structure*. The partial structure approach is in the vanguard of semantic approaches (da Costa and French 2000; Le Bihan 2011, n. 3, §5) and one of the main reasons why the Received View is considered inferior to the semantic approaches (French and Ladyman 1999). I will show that the core notions of the partial structures approach, quasi-truth, partial homomorphism, and partial isomorphism, can be captured very naturally within the Received View in two different ways.

4.3.1 Definitions

The partial structures approach is motivated by a simple epistemological point: Most of the time, scientists do not have enough information about a domain to determine its structure with arbitrary precision. For most relations, it is at best known of *some* tuples of objects that they fall under the relation and known of *some* objects that they do not fall under it. For many if not most tuples this is unknown. Similarly, the value of a function is not know for all of its possible arguments. Partial structures are defined to take this lack of knowledge into account.

While most works on partial structures in the philosophy of science (e. g., da Costa and French 1990, Bueno 1997, da Costa and French 2000) do not consider functions, and the foundational paper by Mikenberg et al. (1986) does not consider constants, the respective definitions can be easily combined to give

Definition 4.10. $\tilde{\mathfrak{A}}$ is a *partial structure* for the symbols $\{P_i, F_j, c_k\}_{i\in I, j\in J, k\in K}$ if and only if

$$\tilde{\mathfrak{A}} = \left\langle A, P_i^{\tilde{\mathfrak{A}}}, F_j^{\tilde{\mathfrak{A}}}, c_k^{\tilde{\mathfrak{A}}} \right\rangle_{i\in I, j\in J, k\in K}, \tag{4.7}$$

where $A \neq \varnothing$, $P_i^{\tilde{\mathfrak{A}}} = \left\langle P_i^{\tilde{\mathfrak{A}},+}, P_i^{\tilde{\mathfrak{A}},-}, P_i^{\tilde{\mathfrak{A}},\circ} \right\rangle$ is a tripartition of A^{m_i} for each $i \in I$, $F_j^{\tilde{\mathfrak{A}}} : C_{\tilde{\mathfrak{A}},j} \longrightarrow A$ is a function with domain $C_{\tilde{\mathfrak{A}},j} \subseteq A^{n_j}$ for each $j \in J$, and $c_k^{\tilde{\mathfrak{A}}} \in A$ for each $k \in K$.

The definition of partial structures by Mikenberg et al. (1986, def. 1) is recovered for $K = \varnothing$, the definition by da Costa and French (1990, 255–256) and

[28] The two problems can also be solved by a direct generalization of van Fraassen's semantic notion of an appearance (Lutz 2011b,c).

da Costa et al. (1998, 605) for $J = \varnothing$.[29] Lack of knowledge is represented by non-empty sets $P_i^{\tilde{\mathfrak{A}},\circ}$ and sets $C_{\tilde{\mathfrak{A}},j} \subset A^{n_j}$.

Taking into account background knowledge, expressed by a set $\tilde{\Pi}$ of sentences, the *primary statements*, Mikenberg et al. (1986, def. 2.ii) and da Costa and French (1990, 256) give

Definition 4.11. Structure $\mathfrak{B}$ is $\tilde{\mathfrak{A}}$-*normal* for primary statements $\tilde{\Pi}$ if and only if $B = A$, $P_i^{\tilde{\mathfrak{A}},+} \subseteq P_i^{\mathfrak{B}} \subseteq A^{m_i} - P_i^{\tilde{\mathfrak{A}},-}$ for each $i \in I$, $F_j^{\mathfrak{B}}|C_j = F_j^{\tilde{\mathfrak{A}}}$ for each $j \in J$, $c_k^{\mathfrak{B}} = c_k^{\tilde{\mathfrak{A}}}$ for each $k \in K$, and $\mathfrak{B} \models \tilde{\Pi}$.

This allows the definition of quasi-truth, also called 'pragmatic truth' or 'partial truth':

Definition 4.12. Sentence φ is *quasi-true* in partial structure $\tilde{\mathfrak{A}}$ relative to primary statements $\tilde{\Pi}$ if and only if there is a structure $\mathfrak{B}$ that is $\tilde{\mathfrak{A}}$-normal for $\tilde{\Pi}$ and $\mathfrak{B} \models \varphi$.

One of the most important properties of quasi-truth is that incompatible sentences can be quasi-true without quasi-truth being trivial: Let $\tilde{\mathfrak{A}}$ be the partial structure $\langle A, \langle P_1^{\tilde{\mathfrak{A}},+}, P_1^{\tilde{\mathfrak{A}},-}, P_1^{\tilde{\mathfrak{A}},\circ} \rangle, c_1^{\tilde{\mathfrak{A}}} \rangle$ with $A = \{1,2,3\}$, $P_1^{\tilde{\mathfrak{A}},+} = \{1\}$, $P_1^{\tilde{\mathfrak{A}},-} = \{3\}$, $c_1^{\tilde{\mathfrak{A}}} = 2$, and $\tilde{\Pi} = \varnothing$. Then $P_1 c$ and $\neg P_1 c$ are both quasi-true, while $\neg \exists x P_1 x$ is not.

Bueno et al. (2002, 503–504) define partial homomorphisms between partial structures:

Definition 4.13. A *partial homomorphism* from partial structure $\tilde{\mathfrak{A}}$ to partial structure $\tilde{\mathfrak{B}}$ is a mapping $f : A \longrightarrow B$ for which the following holds: If $\bar{a} \in P_i^{\tilde{\mathfrak{A}},+}$ then $f(\bar{a}) \in P_i^{\tilde{\mathfrak{B}},+}$ and if $\bar{a} \in P_i^{\tilde{\mathfrak{A}},-}$ then $f(\bar{a}) \in P_i^{\tilde{\mathfrak{B}},-}$ for all $i \in I$, if $\bar{a} \in C_{\tilde{\mathfrak{A}},j}$ then $f(\bar{a}) \in C_{\tilde{\mathfrak{B}},j}$ and for all $\bar{a} \in C_{\tilde{\mathfrak{A}},j}$, $f\left(F_j^{\tilde{\mathfrak{A}}}(\bar{a})\right) = F_j^{\tilde{\mathfrak{B}}}\left(f(\bar{a})\right)$ for all $j \in J$, and $f\left(c_k^{\tilde{\mathfrak{A}}}\right) = c_k^{\tilde{\mathfrak{B}}}$ for all $k \in K$.

Bueno (1997, 596) introduces the notion of a partial isomorphism between partial structures containing only relations, which can be generalized as follows:

Definition 4.14. A *partial isomorphism* from partial structure $\tilde{\mathfrak{A}}$ to partial structure $\tilde{\mathfrak{B}}$ is a bijection $f : A \longrightarrow B$ for which the following holds: $\bar{a} \in P_i^{\tilde{\mathfrak{A}},+}$ if and only if $f(\bar{a}) \in P_i^{\tilde{\mathfrak{B}},+}$ and $\bar{a} \in P_i^{\tilde{\mathfrak{A}},-}$ if and only if $f(\bar{a}) \in P_i^{\tilde{\mathfrak{B}},-}$ for all $i \in I$, $\bar{a} \in C_{\tilde{\mathfrak{A}},j}$ if and only if $f(\bar{a}) \in C_{\tilde{\mathfrak{B}},j}$ and for all $\bar{a} \in C_{\tilde{\mathfrak{A}},j}$, $f\left(F_j^{\tilde{\mathfrak{A}}}(\bar{a})\right) = F_j^{\tilde{\mathfrak{B}}}\left(f(\bar{a})\right)$ for all $j \in J$, and $f\left(c_k^{\tilde{\mathfrak{A}}}\right) = c_k^{\tilde{\mathfrak{B}}}$ for all $k \in K$.

[29] While da Costa and French (1990, 255) and da Costa et al. (1998, 605) define partial structures *only* for relations, their further definition of $\tilde{\mathfrak{A}}$-normal structures presumes that partial structures can contain constants as well.

4.3.2 Partial structures generalized by vagueness sets of $\mathscr{B}$-structures

Partial structures express our empirical knowledge about some domain, expressed in vocabulary $\mathscr{V}$. In §2.8.1 and §2.8.2, such empirical knowledge is expressed by intended structures, a special case of which are vagueness sets. Every partial structure can be expressed with the help of a vagueness set (see definition 2.4 on page 45):

Definition 4.15. Let $\tilde{\mathfrak{A}} = \langle A, P_i^{\tilde{\mathfrak{A}}}, F_j^{\tilde{\mathfrak{A}}}, c_k^{\tilde{\mathfrak{A}}} \rangle_{i\in I, j\in J, k\in K}$ be a partial structure and $\tilde{\Pi}$ a set of primary statements. Then the vagueness set for $\{P_i^+, P_i^-, F_j^{+\circ}, c_k^{+\circ}\}_{i\in I, j\in J, k\in K}$ over A with penumbral connections $W(P_i, F_j, c_k)_{i\in I, j\in J, k\in K} = \tilde{\Pi}$ and $P_i^+ = P_i^{\tilde{\mathfrak{A}},+}$, $P_i^- = P_i^{\tilde{\mathfrak{A}},-}$, $F_j^{+\circ} = F_j^{\tilde{\mathfrak{A}}} \cup (A^{n_j} - C_{\tilde{\mathfrak{A}},j}) \times A$, and $c_k^{+\circ} = \{c_k^{\tilde{\mathfrak{A}}}\}$ for $i \in I, j \in J, k \in K$ *corresponds* to $\tilde{\mathfrak{A}}$ and $\tilde{\Pi}$.

Lemma 4.10. *For any partial structure $\tilde{\mathfrak{A}}$, the set of $\tilde{\mathfrak{A}}$-normal structures relative to $\tilde{\Pi}$ is the corresponding vagueness set.*

Proof. '$\Rightarrow$': Let $\mathfrak{B}$ be $\tilde{\mathfrak{A}}$-normal for $\tilde{\Pi}$. Then $B = A$ and $\mathfrak{B}$ fulfills $W(P_i, F_j, c_k)_{i\in I, j\in J, k\in K}$. Furthermore, $P_i^+ = P_i^{\tilde{\mathfrak{A}},+} \subseteq P_i^{\mathfrak{B}} \subseteq A^{m_i} - P_i^{\tilde{\mathfrak{A}},-} = A^{m_i} - P_i^-$ for each $i \in I$, $F_j^{\mathfrak{B}}|C_{\tilde{\mathfrak{A}},j} = F_j^{\tilde{\mathfrak{A}}}$ so that $F_j^{\mathfrak{B}} \subseteq F_j^{\tilde{\mathfrak{A}}} \cup (A^{n_j} - C_{\tilde{\mathfrak{A}},j}) \times A = F_j^{+\circ}$ for each $j \in J$, and $c_k^{\mathfrak{B}} = c_k^{\tilde{\mathfrak{A}}} \in \{c_k^{\tilde{\mathfrak{A}}}\} = c_k^{+\circ}$ for each $k \in K$. Thus $\mathfrak{B}$ is in the vagueness set corresponding to $\tilde{\mathfrak{A}}$ and $\tilde{\Pi}$.

'$\Leftarrow$': Let $\mathfrak{B}$ be in the vagueness set for $\{P_i^+, P_i^-, F_j^{+\circ}, c_k^{+\circ}\}_{i\in I, j\in J, k\in K}$ over A with penumbral connections $W(P_i, F_j, c_k)_{i\in I, j\in J, k\in K}$. Then $B = A$ and $\mathfrak{B} \models \tilde{\Pi}$. Furthermore, $P_i^{\tilde{\mathfrak{A}},+} = P_i^+ \subseteq P_i^{\mathfrak{B}} \subseteq A^{m_i} - P_i^+ = A^{m_i} - P_i^{\tilde{\mathfrak{A}},-}$ for each $i \in I$, $F_j^{\mathfrak{B}} \subseteq F_j^{+\circ} = F_j^{\tilde{\mathfrak{A}}} \cup (A^{n_j} - C_{\tilde{\mathfrak{A}},j}) \times A$ so that $F_j^{\mathfrak{B}}|C_{\tilde{\mathfrak{A}},j} = F_j^{\tilde{\mathfrak{A}}}$, and $c_k^{\mathfrak{B}} \in c_k^{+\circ} = \{c_k^{\tilde{\mathfrak{A}}}\}$ so that $c_k^{\mathfrak{B}} = c_k^{\tilde{\mathfrak{A}}}$. Therefore $\mathfrak{B}$ is $\tilde{\mathfrak{A}}$-normal for $\tilde{\Pi}$. □

Since not every vagueness set corresponds to a partial structure, the notion of a vagueness set is a proper generalization of the notion of a partial structure. Incidentally, this generalization solves two problems that stand in the way of many applications of partial structures in the analysis of scientific theories: While the interpretations of relation symbols in partial structures can capture fairly general cases of lack of knowledge, there can be no lack of knowledge whatsoever when it comes to constant symbols, because they are interpreted uniquely. And the interpretation of function symbols, practically important because many mathematized theories are teeming with functions, captures only a kind of lack of knowledge encountered very seldom in the sciences. For in a partial structure $\tilde{\mathfrak{A}}$, the values of a function $F_j^{\tilde{\mathfrak{A}}}$ are known with arbitrary precision over $C_{\tilde{\mathfrak{A}},j}$, but not at all over $A^{n_j} - C_{\tilde{\mathfrak{A}},j}$. In contradistinction, the measurement of, say, the time averaged intensity $\overline{\psi}$ of a light wave over some spatial interval $[x_1, x_2]$ will typically have a finite precision, giving a range of possible intensity values $[y_1, y_2] \subset \mathbb{R}^{\geq 0}$ for each point $x \in [x_1, x_2]$. While this can be neatly captured by vagueness sets, a partial structure can only capture measurements that, at any point x, give either a precise value $\overline{\psi}(x) = y_3 \in \mathbb{R}^{\geq 0}$, or no value at all.

Przełęcki (1976, 378–379) suggests to use vagueness sets to define approximate truth and call a set Θ of sentences *approximately true* in a vagueness set $\mathbf{N}$ if and only if there is at least one element of $\mathbf{N}$ in which all elements of Θ are true.

The relation between quasi-truth and approximate truth is given by

Claim 4.11. *Θ is quasi-true in partial structure $\tilde{\mathfrak{A}}$ for $\tilde{\Pi}$ if and only if Θ is approximately true in the corresponding vagueness set.*

Proof. Immediately from lemma 4.10. □

Therefore approximate truth is a generalization of quasi-truth. Note that if a partial structure corresponds to a singleton vagueness set, quasi-truth in the partial structure is equivalent to truth in the structure in the corresponding vagueness set.

The notions of a partial isomorphism and homomorphism can also be generalized with the help of vagueness sets:

Claim 4.12. *Let $\mathbf{A}$ and $\mathbf{B}$ be the vagueness sets corresponding to partial structures $\tilde{\mathfrak{A}}$ and $\tilde{\mathfrak{B}}$ with $\tilde{\Pi} = \varnothing$. Then f is a partial isomorphism between $\tilde{\mathfrak{A}}$ and $\tilde{\mathfrak{B}}$ if and only if for all $\mathfrak{A} \in \mathbf{A}$ there is a $\mathfrak{B} \in \mathbf{B}$ and for all $\mathfrak{B} \in \mathbf{B}$ there is an $\mathfrak{A} \in \mathbf{A}$ such that f is an isomorphism between $\mathfrak{A}$ and $\mathfrak{B}$.*

Proof. '$\Rightarrow$': Let $f : A \longrightarrow B$ be a partial isomorphism between $\tilde{\mathfrak{A}}$ and $\tilde{\mathfrak{B}}$ and let $\mathfrak{A}$ be in $\mathbf{A}$. By lemma 4.10, $\mathfrak{A}$ is $\tilde{\mathfrak{A}}$-normal for $\varnothing$. Since f is a bijection, it is an isomorphism from $\mathfrak{A}$ to $\mathfrak{B}$ with $P_i^{\mathfrak{B}} = f(P_i^{\mathfrak{A}})$ for each $i \in I$, $F_j^{\mathfrak{B}} x_1 \ldots x_{n_j} = f(F_j^{\mathfrak{A}} f^{-1}(x_1) \ldots f^{-1}(x_{m_j}))$ for each $x_1, \ldots, x_{n_j} \in B^{n_j}$, $j \in J$, and $c_k^{\mathfrak{B}} = f(c_k^{\mathfrak{A}})$ for each $k \in K$. Furthermore, $\mathfrak{B}$ is $\tilde{\mathfrak{B}}$-normal, because first, $P_i^{\tilde{\mathfrak{B}},+} = f(P_i^{\tilde{\mathfrak{A}},+}) \subseteq f(P_i^{\mathfrak{A}}) = P_i^{\mathfrak{B}} \subseteq f(P_i^{\tilde{\mathfrak{A}},-}) = P_i^{\tilde{\mathfrak{B}},-}$. Second, $C_{\tilde{\mathfrak{B}},j} = f(C_{\tilde{\mathfrak{A}},j})$ and hence $F_j^{\mathfrak{B}} x_1 \ldots x_{n_j} = f(F_j^{\mathfrak{A}} f^{-1}(x_1) \ldots f^{-1}(x_{n_j})) = f(F_j^{\tilde{\mathfrak{A}}} f^{-1}(x_1) \ldots f^{-1}(x_{n_j})) = F_j^{\tilde{\mathfrak{B}}} x_1 \ldots x_{n_j}$ for all $x_1, \ldots, x_{n_j} \in C_{\tilde{\mathfrak{B}}}$. Third, $c_k^{\mathfrak{B}} = f(c_k^{\mathfrak{A}}) = f(c_k^{\tilde{\mathfrak{A}}}) = c_k^{\tilde{\mathfrak{B}}}$. By lemma 4.10, $\mathfrak{B} \in \mathbf{B}$. By the same reasoning, if $\mathfrak{B} \in \mathbf{B}$, there is an $\mathfrak{A} \in \mathbf{A}$ such that f^{-1} and thus f is an isomorphism between $\mathfrak{A}$ and $\mathfrak{B}$.[30]

[30] This half of the proof generalizes the rough proof given by Bueno (2000, 279–280).

'$\Leftarrow$': Assume that $f : A \longrightarrow B$ is a bijection but not a partial isomorphism between $\tilde{\mathfrak{A}}$ and $\tilde{\mathfrak{B}}$. Then there are an $i \in I$ and some $x_1, \dots, x_{m_i} \in A$ such that (i) $P_i^{\mathfrak{A},+} x_1 \dots x_{m_i}$ and not $P_i^{\mathfrak{B},+} f(x_1) \dots f(x_{m_i})$ or (ii) not $P_i^{\mathfrak{A},+} x_1 \dots x_{m_i}$ and $P_i^{\mathfrak{B},+} f(x_1) \dots f(x_{m_i})$ or (iii) $P_i^{\mathfrak{A},-} x_1 \dots x_{m_i}$ and not $P_i^{\mathfrak{B},-} f(x_1) \dots f(x_{m_i})$ or (iv) not $P_i^{\mathfrak{A},-} x_1 \dots x_{m_i}$ and $P_i^{\mathfrak{B},-} f(x_1) \dots f(x_{m_i})$, or there is a $j \in J$ such that for some $x_1, \dots, x_{m_j} \in A$, (v) $C_{\tilde{\mathfrak{A}},j} x_1 \dots x_{m_j}$ and not $C_{\tilde{\mathfrak{B}},j} f(x_1) \dots f(x_{m_j})$ or (vi) not $C_{\tilde{\mathfrak{A}},j} x_1 \dots x_{m_j}$ and $C_{\tilde{\mathfrak{B}},j} f(x_1) \dots f(x_{m_j})$ or (vii) for some $(x_1, \dots, x_{m_j}) \in C_{\tilde{\mathfrak{A}},j}$, $f\left(F_j^{\tilde{\mathfrak{A}}} x_1 \dots x_{n_j}\right) \neq F_j^{\tilde{\mathfrak{B}}} f(x_1) \dots f x_{n_j}$, or (viii) there is a $k \in K$ such that $f c_k^{\tilde{\mathfrak{A}}} \neq c_k^{\tilde{\mathfrak{B}}}$. It is to be shown that $(*)$ there is an $\mathfrak{A} \in \mathbf{A}$ for which there is no $\mathfrak{B} \in \mathbf{B}$ or there is an $\mathfrak{B} \in \mathbf{A}$ for which there is no $\mathfrak{A} \in \mathbf{A}$ such that f is an isomorphism between $\mathfrak{A}$ and $\mathfrak{B}$.

If (i) holds for some $i \in I$ and $x_1, \dots, x_{m_i}$, then choose an $\tilde{\mathfrak{A}}$-normal structure with $P_i^{\mathfrak{A}} x_1 \dots x_{m_i}$, if (ii) holds for some $i \in I$ and $x_1, \dots, x_{m_i}$, then choose an $\tilde{\mathfrak{B}}$-normal structure with $P_i^{\mathfrak{B}} f(x_1) \dots f(x_{m_i})$, and analogously for (iii) and (iv). Then, if f is an isomorphism between $\mathfrak{A}$ and $\mathfrak{B}$, in case (i) $\mathfrak{B}$ is not $\tilde{\mathfrak{B}}$-normal because $P_i^{\mathfrak{B},+} f(x_1) \dots f(x_{m_i})$ does not hold, in case (ii) $\mathfrak{A}$ is not $\tilde{\mathfrak{A}}$-normal because $P_i^{\mathfrak{A},+} x_1 \dots x_{m_i}$ does not hold, and analogously for (iii) and (iv). If (v) holds for some $j \in J$ and $x_1, \dots, x_{n_j} \in A$, choose $F_j^{\mathfrak{B}} f(x_1), \dots, f(x_{n_j}) \neq f\left(F_j^{\tilde{\mathfrak{A}}} x_1 \dots x_{n_j}\right)$, and analogously for (vi). Then, if f is an isomorphism between $\mathfrak{A}$ and $\mathfrak{B}$, in case (v) $\mathfrak{A}$ is not $\tilde{\mathfrak{A}}$-normal because $F_j^{\mathfrak{A}} x_1 \dots x_{n_j} = f^{-1}\left(F_j^{\mathfrak{B}} f(x_1) \dots f(x_{n_j})\right) \neq F_j^{\tilde{\mathfrak{A}}} x_1 \dots x_{n_j}$, and analogously for (vii). For any $\tilde{\mathfrak{A}}$-normal structure $\mathfrak{A}$ and any $\tilde{\mathfrak{B}}$-normal structure $\mathfrak{B}$, if (vii) holds, $F_j^{\mathfrak{A}} x_1, \dots, x_{n_j} = F_j^{\tilde{\mathfrak{A}}} x_1 \dots x_{n_j} \neq F_j^{\tilde{\mathfrak{B}}} x_1, \dots, x_{n_j} = F_j^{\mathfrak{B}} x_1, \dots, x_{n_j}$ and if (viii) holds, $c_k^{\mathfrak{A}} = c_k^{\tilde{\mathfrak{A}}} \neq c_k^{\tilde{\mathfrak{B}}} = c_k^{\mathfrak{B}}$. $(*)$ follows by lemma 4.10. □

Claim 4.13. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be the vagueness sets corresponding to partial structures* $\tilde{\mathfrak{A}}$ *and* $\tilde{\mathfrak{B}}$ *with* $\tilde{\Pi} = \varnothing$. *Then* f *is a partial homomorphism between* $\tilde{\mathfrak{A}}$ *and* $\tilde{\mathfrak{B}}$ *if and only if for all* $\mathfrak{A} \in \mathbf{A}$ *there is a* $\mathfrak{B} \in \mathbf{B}$ *and for all* $\mathfrak{B} \in \mathbf{B}$ *there is an* $\mathfrak{A} \in \mathbf{A}$ *such that* f *is a homomorphism between* $\mathfrak{A}$ *and* $\mathfrak{B}$.

Proof. Similar to the proof of claim 4.12. □

4.3.3 Partial structures as $\mathscr{A}$-structures

When capturing partial structures via vagueness sets and identifying quasi-truth of a theory with approximate truth, one assumes that the theory is formulated in the same vocabulary in which our empirical knowledge (expressed by the partial structure) is formulated. But there is another way to capture partial structures that assumes a distinction between a basic vocabulary, in which our empirical knowledge is formulated, and an auxiliary vocabulary, in which our theories are formulated.

In a partial structure, a relation symbol P_i has, in a sense, two separate interpretations. For one, there are its clear instances $P_i^{\tilde{\mathfrak{A}},+}$. They can be determined, for example, by their similarity to paradigmatic instances of P_i, or, more likely when it comes to scientific terms, by the fulfillment of some sufficient condition. Then there are also the clear non-instances $P_i^{\tilde{\mathfrak{A}},-}$. These are determined, for example, by their similarity to paradigmatic non-instances of P_i, or by the failure to fulfill some necessary condition. Determining whether some tuple is in $P_i^{\tilde{\mathfrak{A}},+}$ is thus more or less unrelated to determining whether some tuple is in $P_i^{\tilde{\mathfrak{A}},-}$. (That a tuple is in $P_i^{\tilde{\mathfrak{A}},\circ}$ will typically only be determined by its being in neither $P_i^{\tilde{\mathfrak{A}},+}$ nor $P_i^{\tilde{\mathfrak{A}},-}$.) Given the difference in determining the members of $P_i^{\tilde{\mathfrak{A}},+}$ and of $P_i^{\tilde{\mathfrak{A}},-}$, it is natural to assign separate symbols of a language to these two concepts, say, P_i^+ and P_i^-.

In a partial structure, the interpretation $F_j^{\tilde{\mathfrak{A}}}$ of an n_j-place function symbol F_j can be seen as the clear instances of an n_j+1-ary relation. In analogy to the relation symbols in partial structures, it is natural to assign an n_j+1-place relation symbol F_j^+ to the concept that determines these clear instances. $F_j^{\tilde{\mathfrak{A}}}$ does not have a value if its argument is not in $C_{\tilde{\mathfrak{A}},j}$, and thus for every n_j+1-tuple not in the relation named by F_j^+, it is unknown whether it falls under the function or not. Thus there is no need for a relation symbol that names the clear non-instances of F_j.

Since constant symbols are interpreted in the usual way, this leads to a new language $\mathscr{L}' = \{P_i^+, P_i^-, F_j^+, c_k\}_{i\in I, j\in J, k\in K}$, chosen so that $\{P_i^+, P_i^-, F_j^+\}_{i\in I, j\in J} \cap \mathscr{L} = \varnothing$. And any partial structure for $\mathscr{L}$ determines a structure for $\mathscr{L}'$:

Definition 4.16. $\mathscr{L}'$-structure $\mathfrak{A}$ *corresponds* to partial $\mathscr{L}$-structure $\tilde{\mathfrak{A}}$ if and only if $|\mathfrak{A}| = |\tilde{\mathfrak{A}}|$, $P_i^{+\mathfrak{A}} = P_i^{\tilde{\mathfrak{A}},+}$ and $P_i^{-\mathfrak{A}} = P_i^{\tilde{\mathfrak{A}},-}$ for each $i \in I$, $F_j^{+\mathfrak{A}} = \{\bar{a}b \mid \bar{a} \in C_{\tilde{\mathfrak{A}},j}$ and $F_j^{\tilde{\mathfrak{A}}}(\bar{a}) = b\}$ for each $j \in J$, and $c_k^{\mathfrak{A}} = c_k^{\tilde{\mathfrak{A}}}$ for each $k \in K$.

In this definition, $\bar{a}b$ stands for the tuple $(a_1, \ldots, a_{n_j}, b) \in A^{n_j+1}$. Note that for every partial structure $\tilde{\mathfrak{A}}$ there is exactly one structure $\mathfrak{A}$ that corresponds to $\tilde{\mathfrak{A}}$. By introducing for each function symbol and each constant symbol of the partial structure a relation symbol that names all the clear non-instances of the function or constant symbol, respectively, one can again generalize partial structures similarly to §4.3.2. The notion of a structure corresponding to a partial structure thus provides a general way of defining a precise $\mathscr{B}$-language starting from vague terms.

Despite having two separate interpretations, the relation symbols P_i^+ and P_i^- are of course connected, since they are known to refer to instances and, respectively, non-instances of the same relation symbol P_i from $\mathscr{L}$. This connection, and the fact that over a restricted domain, F_j^+ is equivalent to a function F_j are thus background assumptions. They can therefore be described by primary statements in language $\mathscr{L}^* = \mathscr{L} \cup \mathscr{L}'$:

$$\begin{aligned} \Pi = \tilde{\Pi} &\cup \bigcup_{i \in I} \{\forall \bar{x}(P_i^+ \bar{x} \to P_i \bar{x}), \forall \bar{x}(P_i^- \bar{x} \to \neg P_i \bar{x})\} \\ &\cup \bigcup_{j \in J} \{\forall \bar{x} \forall y (F_j^+ \bar{x} y \to F_j \bar{x} = y)\} . \end{aligned} \tag{4.8}$$

In every structure $\mathfrak{A}$ that corresponds to a partial structure, relation $F_j^{+\mathfrak{A}}$ can provide a sufficient condition for function values because by definition 4.16, tuples in $F_j^{+\mathfrak{A}}$ differ in their last elements only if they also differ in one of their previous elements.

Since the structure $\mathfrak{A}$ corresponding to a partial $\mathscr{L}$-structure $\tilde{\mathfrak{A}}$ is itself an $\mathscr{L}'$-structure, Π cannot be true in $\mathfrak{A}$. However, Π may be true in an *expansion* of $\mathfrak{A}$ to $\mathscr{L}^*$, which differs from $\mathfrak{A}$ only in that it interprets the symbols in $\mathscr{L}^* - \mathscr{L}'$. With the help of corresponding structures, it is now possible to describe quasi-truth relative to $\tilde{\Pi}$:

Claim 4.14. *$\mathscr{L}$-sentence φ is quasi-true in partial $\mathscr{L}$-structure $\tilde{\mathfrak{A}}$ with respect to $\tilde{\Pi}$ if and only if the corresponding $\mathscr{L}'$-structure has an expansion $\mathfrak{C}$ such that*

$$\mathfrak{C} \vDash \{\varphi\} \cup \Pi . \tag{4.9}$$

Proof. '$\Leftarrow$': Let $\mathfrak{A}$ correspond to $\tilde{\mathfrak{A}}$ and $\mathfrak{C}$ be an expansion of $\mathfrak{A}$ such that $\mathfrak{C} \vDash \{\varphi\} \cup \Pi$. Then $\mathfrak{C}|_{\mathscr{L}} \vDash \{\varphi\} \cup \tilde{\Pi}$, $|\mathfrak{C}|_{\mathscr{L}}| = |\mathfrak{C}| = A$, and $P_i^{\tilde{\mathfrak{A}},+} = P_i^{+\mathfrak{A}} = P_i^{+\mathfrak{C}} \subseteq P_i^{\mathfrak{C}|_{\mathscr{L}}} \subseteq A^{m_i} - P_i^{-\mathfrak{C}} = A^{m_i} - P_i^{-\mathfrak{A}} = A^{m_i} - P_i^{\tilde{\mathfrak{A}},-}$ for each $i \in I$. Furthermore, for each $\bar{a} \in C_{\tilde{\mathfrak{A}},j}$, $F_j^{\mathfrak{C}|_{\mathscr{L}}}(\bar{a}) = b$ if $\bar{a}b \in F_j^{+\mathfrak{C}}$, and, since $F_j^{\mathfrak{C}|_{\mathscr{L}}}$ is a function, also *only* if $\bar{a}b \in F_j^{+\mathfrak{C}}$. Since further $F_j^{+\mathfrak{C}} = F_j^{+\mathfrak{A}}$, and $\bar{a}b \in F_j^{+\mathfrak{A}}$ if and only if $\bar{a} \in C_{\tilde{\mathfrak{A}},j}$ and $F_j^{\tilde{\mathfrak{A}}}(\bar{a}) = b$, it holds that $F_j^{\mathfrak{C}|_{\mathscr{L}}}|_{C_{\tilde{\mathfrak{A}},j}} = F_j^{\tilde{\mathfrak{A}}}$ for each $j \in J$. Finally, $c_k^{\mathfrak{C}|_{\mathscr{L}}} = c_k^{\mathfrak{A}} = c_k^{\tilde{\mathfrak{A}}}$. Thus $\mathfrak{C}|_{\mathscr{L}}$ is $\tilde{\mathfrak{A}}$-normal and hence φ is quasi-true in $\tilde{\mathfrak{A}}$.

'$\Rightarrow$': Let $\mathfrak{A}$ be the $\mathscr{L}'$-structure that corresponds to $\tilde{\mathfrak{A}}$ and let φ be quasi-true in $\tilde{\mathfrak{A}}$ with respect to $\tilde{\Pi}$. Then there is an $\mathscr{L}$-structure $\mathfrak{B}$ such that $\mathfrak{B} \vDash \tilde{\Pi} \cup \{\varphi\}$ and $P_i^{+\mathfrak{A}} = P_i^{\tilde{\mathfrak{A}},+} \subseteq P_i^{\mathfrak{B}} \subseteq A^{m_i} - P_i^{\tilde{\mathfrak{A}},-} = A^{m_i} - P_i^{-\mathfrak{A}}$ for each $i \in I$. Furthermore, $F_j^{\tilde{\mathfrak{A}}} = F_j^{\mathfrak{B}}|_{C_{\tilde{\mathfrak{A}},j}}$ and thus for each $j \in J$, $\bar{a} \in A^{n_j}$, and $b \in A$, $\bar{a}b \in F_j^{+\mathfrak{A}}$ only if

$F_j^{\mathfrak{B}}(\bar{a}) = b$. Finally, $c_k^{\mathfrak{A}} = c_k^{\tilde{\mathfrak{A}}} = c_k^{\mathfrak{B}}$ for each $k \in K$. Define the $\mathscr{L}^*$-structure $\mathfrak{C}$ so that $\mathfrak{C}|_{\mathscr{L}'} = \mathfrak{A}$ and $\mathfrak{C}|_{\mathscr{L}} = \mathfrak{B}$. Then $\mathfrak{C} \vDash \{\varphi\} \cup \Pi$. □

Somewhat shorter, φ is quasi-true in $\tilde{\mathfrak{A}}$ with respect to $\tilde{\Pi}$ if and only if its corresponding structure has an expansion in which $\{\varphi\} \cup \Pi$ is true.

In the new formalization of quasi-truth, the language $\mathscr{L}'$ is, in keeping with the basic motivation for partial structures, considered to be directly interpreted, while the interpretation of $\mathscr{L}^* - \mathscr{L}' = \{P_i, F_j\}_{i \in I, j \in J}$ is only given through the interpretation of $\mathscr{L}'$ and the primary statements Π. This reliance on a basic and an auxiliary vocabulary and their connection by a set of sentences distinguishes the Received View from other syntactic approaches.[31] In principle, all results from the Received View can therefore be used for partial structures. I want to present only one.

Claim 4.15. *If $\tilde{\Pi}$ and $\mathscr{L}$ are finite, then $\mathscr{L}$-sentence φ is quasi-true in partial $\mathscr{L}$-structure $\tilde{\mathfrak{A}}$ with respect to $\tilde{\Pi}$ if and only if for the corresponding $\mathscr{L}'$-structure $\mathfrak{A}$ it holds that*

$$\mathfrak{A} \vDash \mathsf{R}_{\mathscr{L}'}(\varphi \wedge \bigwedge \Pi) \ . \tag{4.10}$$

Proof. Since $\tilde{\Pi}$ and $\mathscr{L}$ are finite, so are Π and $\mathscr{L}^*$. Therefore, by claim 4.14, φ is quasi-true in $\tilde{\mathfrak{A}}$ if and only if $\mathfrak{A}$ has an expansion $\mathfrak{C}$ such that $\mathfrak{C} \vDash \varphi \wedge \bigwedge \Pi$. By lemma 2.4 (page 59), there is such an expansion if and only if $\mathfrak{A} \vDash \mathsf{R}_{\mathscr{L}'}(\varphi \wedge \bigwedge \Pi)$. □

Somewhat shorter, φ is quasi-true in $\tilde{\mathfrak{A}}$ with respect to $\tilde{\Pi}$ if and only if $\mathsf{R}_{\mathscr{L}'}(\varphi \wedge \bigwedge \Pi)$ is true in the structure corresponding to $\tilde{\mathfrak{A}}$.

The features of quasi-truth that follow from definition 4.12 can now also be recovered from claims 4.14 and 4.15. For example, that two incompatible sentences can both be quasi-true in the same partial structure follows from the fact that, given the primary statements Π, two incompatible $\mathscr{L}$-sentences can have compatible Ramsey sentences.

The differences between the definitions for partial homomorphisms and isomorphisms are analogous to the differences between the standard definitions of homomorphism and isomorphism between structures (Hodges 1993, 5), so that they can be easily discussed together:

[31] Incidentally, the sentences $\forall \bar{x}(R_i^+ \bar{x} \rightarrow \neg P_i^- \bar{x}), i \in I$ and $\forall \bar{x} \forall y \forall \bar{v} \forall w(F_j^+ \bar{x} y \wedge F_j^+ \bar{v} w \wedge \bigwedge_{1 \le r \le n_j} x_r = v_r \rightarrow y = w), j \in J$, which follow from Π and contain only basic terms, express that in a partial structure $\tilde{\mathfrak{A}}$, $P_i^{\tilde{\mathfrak{A}},+} \cap P_i^{\tilde{\mathfrak{A}},-} = \varnothing$ for all $i \in I$ and $F_j^{\tilde{\mathfrak{A}}}$ is a partial function for all $j \in J$. Since they are therefore basic presumptions of the formalism, they are good candidates for analytic sentences in $\mathscr{L}'$ (cf. Carnap 1952).

Claim 4.16. *Let $\mathfrak{A}$ correspond to $\tilde{\mathfrak{A}}$, and $\mathfrak{B}$ to $\tilde{\mathfrak{B}}$. Then f is a partial homomorphism/ partial isomorphism from $\tilde{\mathfrak{A}}$ to $\tilde{\mathfrak{B}}$ if and only if f is a homomorphism/isomorphism from $\mathfrak{A}$ to $\mathfrak{B}$.*

Proof. The proof for relations and constants is immediate. For functions, the following holds:

'$\Rightarrow$': For all $j \in J$, $\bar{a} \in A^{n_j}$, and $b \in A$, $\bar{a}b \in F_j^{+\mathfrak{A}}$ if and only if $\bar{a} \in C_{\tilde{\mathfrak{A}},j}$ and $F_j^{\tilde{\mathfrak{A}}}(\bar{a}) = b$. This holds only if/if and only if $f(\bar{a}) \in C_{\tilde{\mathfrak{B}},j}$ and $F_j^{\tilde{\mathfrak{B}}}(f(\bar{a})) = f(b)$, that is, $f(\bar{a})f(b) \in F_j^{+\mathfrak{B}}$.

'$\Leftarrow$': For all $j \in J$, $\bar{a} \in C_{\tilde{\mathfrak{A}},j}$ and $F_j^{\tilde{\mathfrak{A}}}(\bar{a}) = b$ if and only if $\bar{a}b \in F_j^{+\mathfrak{A}}$. This holds only if/if and only if $f(\bar{a})f(b) \in F_j^{+\mathfrak{B}}$, that is, $f(\bar{a}) \in C_{\tilde{\mathfrak{B}},j}$ and $F_j^{\tilde{\mathfrak{B}}}(f(\bar{a})) = f(b)$. □

Somewhat shorter, a mapping between two partial structures is a partial homomorphism/partial isomorphism if and only if it is a homomorphism/ isomorphism between their corresponding structures.

Claims 4.14 and 4.16 reduce the concepts of the partial structures approach to the model theory of first order logic, claims 4.15 and 4.16 reduce them to the model theory of second order logic. For example, since the truth value of a sentence of second order logic is conserved under isomorphisms, it follows from claims 4.15 and 4.16 that the quasi-truth-value of a sentence is conserved under partial isomorphisms.

4.4 Combining semantic and syntactic approaches

If van Fraassen (1980, 56) was right in his in his claim that syntactic concepts are "off the mark", solutions "to purely self-generated problems, and philosophically irrelevant", the preceding results would establish a *reductio ad empirismum logicum* of his own position, the partial structures approach, and semantic approaches in general. But insofar as these approaches have proven their merits (or can prove their merits once their generalizations within the Received View are taken into account), the inference has to go in the opposite direction: The tools developed within logical empiricism are more useful than its detractors have acknowledged.

The close connection between the approaches allows a comparison of more specific versions, for example approaches in first order predicate logic and approaches in first order model theory. Here, standard model theory has already led to major results, which only had to be put to work, for example by Przełęcki (1969) and then later, for example, by van Benthem (1982, 2011).

Taking the close relation seriously also allows identifying a problem in either approach if the problem has already been identified in the other. In general, those problems of syntactic approaches that do not stem from axiomatization in first order logic have their analogues in semantic approaches, as the above examples of the connection to the world and language dependence show. The conclusion by Chakravartty (2001, 326, §1) that "[r]ealism on the semantic view is by no means impossible, but faced with precisely those familiar, perennial difficulties of reference and correspondence that some semanticists think their approach does without" is another case in point. Given the translatability of semantic into syntactic descriptions and vice versa (§4.1.1) and the similarity in the solutions to connect either description to the world (§4.1.3), it would indeed be surprising if all semantic approaches differed from all syntactic approaches in their ontological commitments. Another example comes from theories that cannot be fruitfully formalized. Typically, those theories are considered a problem for syntactic approaches (see, for example Suppe 1974a, 63 and Beatty 1980, appendix 1), but the above results show that if a theory is not fruitfully formalizable in a syntactic approach, neither is it so formalizable in a semantic one. Of historical interest are the problems that led Hempel to abandon both the Received View and reliance on syntactic axiomatizations in general. He takes as a starting point of his criticism a defense of axiomatizations by Suppes (1968), which, so Hempel (1974, 248), argues for "[t]he importance of axiomatically formalizing scientific theories, much in the manner [...] envisaged by the standard construal". But, of course, Suppes (1968, 653) defends formalization in the sense "of a standard set-theoretical formulation", not in "the stricter conception of a first-order theory that assumes only elementary logic." Much of Hempel's criticism takes issue with the notion that one can analyze a scientific theory as an uninterpreted formal system, be it uninterpreted in a semantic or a syntactic approach. Hempel (1974, 251) concludes that the

> extensive theoretical use of antecedent terms appears to me to throw into question the conception of the internal principles of a theory as an axiomatized system whose postulates provide "implicit" definitions for its extralogical terms. [...] Hence, the theoretical "calculus" of a theory of this kind cannot be regarded as a strictly formalized, uninterpreted system, and the concepts of model theory cannot be applied to it without qualifications.

I have already argued that Hempel overstates his case, and indeed fundamentally misrepresents the Received View in his discussion, but it is important to see that his criticism applies to non-worldly pure structures as much as it applies to uninterpreted sets of sentences and was indeed explicitly directed at a semantic approach.

A more positive use of the close relation is the transfer of solutions from one view to the other. One example is the definitional expansion of theories to allow the identification of theories with formally different structures in semantic views (§4.1.2). As already noted, this goes far beyond the pure avoidance of assigning different names to the same sets. Other examples are the concept of a substructure in model theory, which captures the syntactic notion of a subvocabulary, the use of the Ramsey sentence by Sneed (1979), the syntactic description of empirical embeddings, and the description of partial truth by the truth of a Ramsey sentence. With respect to the relation between the Received View and van Fraassen's conception of scientific theories, Turney (1990, 449) concludes:

> We see now that there is a syntactic method, which is equivalent to his semantic method. The moral is this: The relevant distinction here is not between syntax and semantics. [...] It is between two ways of linking theory and observation: Correspondence rules versus embedding/implanting.

Of course, Turney assumes in this quote that correspondence rules cannot capture his conception of implanting, and that his conception of implanting captures van Fraassen's notion of empirical embedding. Neither is true, but the importance of Turney's point is this: If the difference between syntactic and semantic approaches is seen as one of formulation, it is possible to search for commonalities between the views and to transfer solutions from one approach to the other. On a meta-level, I therefore do hold the position of the critics of syntactic approaches: The language in which an analysis is phrased, whether it uses pure or indexed structures, structures, or structures and an object language, matters very little.

In the next two parts, I will provide examples of the power of syntactic approaches including the Received View and (again) their close proximity to semantic approaches.

Part II

Relations

Chapter 5

Preliminary remarks on criteria of empirical significance

Criteria of empirical significance are demarcation criteria meant to distinguish between those statements or terms that have some connection to empirical statements and those that do not. An early criterion suggested by Ayer was quickly shown to be trivial; it was followed by a slew of amendments and new trivialization proofs succinctly summarized and extended by Pokriefka (1983), who cuts out the middleman and proves the triviality of his amendment himself (Pokriefka 1984). The latest contributions to this "puncture-and-patch industry" (Lewis 1988b, §XII) are two criteria by Wright (1986, 1989) and trivialization proofs by Lewis (1988b, §IV, n. 12), Wright (1989, §II), and Yi (2001).

This history has "done a lot to discredit the very idea of delineating a class of statements as empirical" (Lewis 1988b, §I). In what Gemes (1998b, §1.1) calls "the problem of past failures", the failure of Ayer's early criterion and subsequent amendments is used for an enumerative induction to conclude that the problem of demarcation cannot be solved at all. As Lewis (1988b, 127, footnote removed) further notes, the many amendments have

> led to ever-increasing complexity and ever-diminishing contact with any intuitive idea of what it means for a statement to be empirical. Even if some page-long descendant of Ayer's criterion [provably admitted] more than the observation-statements and less than all the statements, we would be none the wiser. We do not want just any class of statements that is intermediate between clearly too little and clearly too much. We want the right class.

This also holds for criteria that do not amend Ayer's criterion, as their multitude suggests that they are little more than arbitrary bipartitions of the class of statements. Call this 'the charge of arbitrariness'.

After proving, almost in passing, that Ayer's own amendment of his early criterion is trivial, Church (1949, 53) concludes that "any satisfactory solution of the difficulty will demand systematic use of the logistic method". In this spirit, I will provide formalizations of the major criteria of empirical significance and analyze their logical structure. The first result of these analyses will be that the charge of arbitrariness is unfounded, for the non-trivial criteria are equivalent or bear strong inferential relations to each other and to concepts from definition- and measurement theory. I will also suggest a way to avoid the assumption that trivialized Ayer's criterion and its successor, thereby solving the problem of past failures.

These are already good reasons to look more closely at criteria of empirical significance, but there are others. For one, many criticisms of the criteria have seen rebuttals, mostly because they rely on misunderstandings of the criteria's intended applications. There is also still a *need* for criteria of empirical significance. Sometimes a criterion is needed to state very clearly what is not generally in dispute, as in Sober's discussions of the empirical significance of claims about a designer of life whose intentions and abilities are unknown (Sober 1999, 2007, 2008). In other cases, a generally accepted endeavor is put under scrutiny, like string theory (Smolin 2006, Woit 2006), fish stock assessment theories (Corkett 2002), or natural selection (Wassermann 1978). The empirical significance of more philosophical positions like theism (Diamond and Litzenburg 1975) or realism and antirealism (Sober 1990) have also been investigated.

To be successful, an explication of empirical significance should thus contribute to the solution of scientific and philosophical problems. But beyond this, an explication should also suggest entirely new research questions—this will be the subject of part III. In the end, these results will provide evidence that the search for criteria of empirical equivalence has been successful.

5.1 Methodological presumptions

The development of a criterion of empirical significance out of the vague and intuitive concept variously described as 'having empirical content', 'being connected to observations', 'being testable', and 'being empirically meaningful' amounts to an explication. Since explication is not the same as precisification (see §3.8.1), it would therefore not be helpful to rely on examples of intuitively clear cases of empirical significance or lack of empirical significance. I will instead rely on conditions of adequacy (see §2.3.1). Concepts are typically explicated in a restricted domain; Tarski, for instance, restricted himself to predicate logic when explicating 'truth', as did Carnap when explicating 'analytic'. Such a restriction is acceptable and indeed almost always necessary to attain any results at all (Martin 1952). It is therefore not a fundamental problem that the explicata discussed in the following assume a language of first or higher order predicate logic. Rather, in the spirit of Carnap (see §3.3, §3.8.2), the explicata should be seen as first steps towards the development of more general criteria. In other words, the criteria define empirical significance on the condition that the language is one of predicate logic. As argued in §4.1, this is not a more restrictive assumption than assuming the language of pure structures. Furthermore, the equivalences discussed here will suggest immediate generalizations beyond predicate logic. Finally, in the following I will discuss syntactic, model theoretic, and set theoretic criteria of empirical significance and the conditions under which they are equivalent.

More problematic than the use of predicate logic is that some of the criteria discussed in the following (the semantic ones, by the way) assume a bipartition of the non-logical vocabulary $\mathscr{V}$ as in the Received View. I have argued in §3.6 and §4.2 that a bipartition does not have to lead to an overly complex vocabulary, and that the assumption of a bipartition is not stronger or less natural than those by Suppes, van Fraassen, or in measurement theory. Furthermore, in keeping with artificial language philosophy, the vocabulary does not have to be partitioned into observational and theoretical terms, but only into in some sense basic and auxiliary terms, where the basic terms must only be unproblematic for the purposes at hand. According to Nielsen (1966, 15), for example, Flew's charge that theological statements are not falsifiable (Flew 1950) assumes that all and only "non-religious, straightforwardly empirical, factual statements" have primitive meaning. Flew thus does not consider it a requirement that $\mathscr{B}$-terms be observation terms, but only that $\mathscr{B}$-sentences be empirical or factual sentences. And Flew (1975, 274) claims that it is enough to assume that all and only statements about "anything which happens or which conceivably might happen in the ordinary world" have primitive meaning, so that $\mathscr{B}$-sentence only have to be about the ordinary world.[1] Furthermore, I will show in §6.8.1 how the restrictions that the criteria discussed here place on the descriptions of observations can be weakened.

There is, however, an important subtlety to be considered: In §2.8.3, I have assumed that the set of intended structures $\mathbf{N}_{\mathscr{B}}$ does not have to be a singleton set because of vagueness. Therefore, in general not every $\mathscr{B}$-sentence β has a determinate truth value in $\mathbf{N}_{\mathscr{B}}$, even under complete empirical information. In those cases, the truth or falsity of β is a matter of convention, and hence not empirically significant. As will become clear, however, those criteria that rely on a bipartition of the language presume that all $\mathscr{B}$-sentences are themselves empirically significant and that all $\mathscr{B}$-terms are precise. There are two ways to respond to this problem. The most obvious is to keep the $\mathscr{B}$-terms as they are and define the set of empirically significant $\mathscr{B}$-sentences as a subset of all $\mathscr{B}$-

[1]These are only illustrations: I especially do not think that Flew's circumscription of the $\mathscr{B}$-sentences is precise enough (see §9.1).

sentences. I do this in §6.8.1. The other, and technically simpler response is to introduce a new set of $\mathscr{B}$-terms that are not vague and therefore have only a single intended structure as element of $\mathbf{N}_{\mathscr{B}}$. I have developed and applied one way to arrive at a language without vague terms in §2.10.2 and §4.3.3, where the vague terms are included in the set of auxiliary terms, and their positive and negative extensions make up the new set of basic terms. This response is the technically more satisfying, and is also the one that leads to the historically prominent criteria of empirical significance.

Below, I will make a distinction between syntactic and semantic criteria of empirical significance based on whether the observations are described by sets of $\mathscr{B}$-sentences or by $\mathscr{B}$-structures. With $\mathscr{B}$-structures, observations can be described up to isomorphism, and with $\mathscr{B}$-sentences up to what I will call *syntactical equivalence*. Two structures $\mathfrak{A}$ and $\mathfrak{B}$ are syntactically equivalent ($\mathfrak{A} \equiv \mathfrak{B}$) if and only if their respective theories are equivalent ($\mathrm{Th}(\mathfrak{A}) \mathrel{\models\!\!\!\dashv} \mathrm{Th}(\mathfrak{B})$), that is, for all sentences σ, $\mathfrak{A} \models \sigma$ if and only if $\mathfrak{B} \models \sigma$. Thus in first order logic, syntactic equivalence is elementary equivalence.

5.2 On the explicanda

The criteria discussed in the following are not meant to determine the meaning of sentences as Ruja (1961), for instance, assumes in his critique, or the meaning of terms. Rynin (1957, 51–53) and Gemes (1998b, §1.5) argue in some detail that this is not the point of the criteria, but it is also obvious from their formal structure: The criteria are classificatory (so that a sentence or term can be empirically significant or not), while a criterion of meaning has to define a relation between sentences or, respectively, terms and meanings. More substantially, no amount of $\mathscr{B}$-sentences may be enough to give the meaning of a term if a distinction is made between the evidential basis for ascribing a theoretical term to a state and the reference of the term in that state (Feigl 1950, 48).

Pace Rynin (1957, 51), 'empirical significance' does not explicate 'meaningfulness', either, because the meaning of sentences and terms is generally accepted to be determined by both the sentence's or term's empirical import and the rules that govern its use with other sentences or terms (Carnap 1939, §25). Thus even a sentence or term not connected in the slightest to observation can be meaningful (cf. Sober 2008, 149–150). Whether there is more to the meaning of sentences and terms beyond their empirical import and relation to other sentences and terms depends on the status of semantic empiricism, which asserts the opposite (Rozeboom 1962, §II; Rozeboom 1970; Przełęcki 1969, §§5–6; Przełęcki 1974b, 402–403). This understanding of the criteria as criteria for the *empirical* meaningfulness of sentences is in line with Popper's notion of falsifiability as a demarcation criterion between empirical and non-empirical sentences (Popper 1935, §4, §9; cf. Carnap 1963c, §6.A).

Gemes (1998b, §1.4) argues that a criterion of empirical significance does not have to be a criterion of inductive confirmability. In connection with claim 6.7, I will point out previous results in the philosophy of science that show the need for a distinction between contexts in which only deductive inferences are possible and contexts in which only probabilistic inferences are possible. Accordingly, I will discuss criteria for the former contexts—deductive criteria of empirical significance (§6 and §7)—separately from criteria for the latter contexts—probabilistic criteria of empirical significance (§8). A major topic in the following will be the compatibility of deductive and probabilistic criteria in contexts that allow both deductive and probabilistic inferences. A sentence may then be called *empirically significant* (simpliciter) if and only if it is deductively empirically significant or probabilistically empirically significant.

Finally, I will restrict the discussion to descriptive, extensional sentences, and thus assume that the language whose well-formed sentences are to be tested for significance does not contain modal operators. This does not mean that the following discussion is of no value for, say, obligation- or necessity-statements, however, since those sentences may contain descriptive subformulas, a closure of which can be tested for significance (Marhenke 1950, 3). Furthermore, the criteria discussed here may allow for a generalization to languages that do contain modal operators.

5.3 General arguments about criteria of empirical significance

The typical argument against any kind of criterion of empirical significance is a pessimistic induction based on the problem of past failures: Since so far no adequate criterion has been found, it is argued, it is improbable that one will ever be found (cf. Hempel 1965b, Suarez 2000). Of course, this argument is no more than a plausibility consideration. Since there has not been a unified, detailed overview of the topic since Hempel's two essays (Hempel 1950, 1951)[2] and many critiques of suggested conditions are controversial themselves, the pessimistic induction is far from strong. Accordingly, it has been rejected by some, for example Feigl (1956) and Justus (2010).

A more general argument against a class of deductive criteria for terms has been given by Berkowitz (1979) based on a condition of adequacy suggested by Achinstein. Achinstein (1964, 99; 1968, 78, my formulation) proposes

Condition 1. If a set of sentences Σ is such that the occurrence of a term V_i in Σ suffices/does not suffice to guarantee that V_i is significant, then the occurrence of V_i in any set of sentences that is logically equivalent to Σ also suffices/does not

[2]Hempel (1965c) gives a shortened and edited combination of the two. Misak (1995) gives a non-technical discussion and focuses on the criteria suggested before Hempel's overview articles.

suffice to guarantee that V_i is significant.

As an example, Achinstein (1968, 78) notes that according to condition 1, the terms in σ are significant in a theory $\Sigma = \{\sigma \rightarrow \tau, \sigma\}$ only if they are also significant in the theory $\Sigma' = \{\sigma, \tau\}$.

Berkowitz (1979, 463, my notation) then states that many criteria of significance are based on the idea that "the occurrence of V_i in Σ suffices to guarantee that V_i is essential to the deduction of observational consequences", where 'essential' means that some observational (basic) consequences are not entailed without the sentences that contain V_i. Simplifying Berkowitz's argument somewhat, one can, for every set Σ of sentences, define a minimal set Ω of basic sentences entailed by Σ and rewrite $\Sigma \boxminus \Sigma \cup \Omega$. Thus no V_i that occurs in Σ but not in Ω can be empirically significant. Note that condition 1 has a natural analogue for criteria for the empirical significance of sentences, and that Berkowitz's trivialization proof has such an analogue, too, with the result that only basic sentences can be empirically significant.

Berkowitz counters his own argument by calling into question condition 1. To argue that a V_i is inessential for the derivation of a sentence σ from Σ by equivalently reformulating Σ and thus *using a set of sentences containing* V_i is, "in a sense, [...] begging the question" (Berkowitz 1979, 464). Indeed, a criterion developed by Gemes (1998a) explicitly disallows some equivalent reformulations of sets of sentences, and it is therefore doubtful that condition 1 can be considered uncontroversial.

A general argument *for* the existence of a criterion of significance is given by Sober (1999, 48) in terms of the testability of a hypothesis (cf. Sober 2008, 149):

> If it makes sense to say that an experiment does or does not test a given hypothesis, why is it suddenly misguided to ask whether any experiment *could* test the hypothesis? [...] If a set of observations provides a test of a proposition because it bears relation R to that proposition, then a proposition is testable when it is possible for there to be a set of observations that bears relation R to the proposition. Testing is to testability as dissolving is to solubility. If we can understand what testing is, we also should be able to understand what testability is.

The argument is, in terms of significance, the following: For each sentence σ and for each $\mathscr{B}$-sentence β, we know whether β tests σ. σ is significant if and only if there is a $\mathscr{B}$-sentence that tests σ. Therefore, we know for each sentence whether it is significant. However, Sober's inference assumes that we have a means of going through all observation sentences in a finite amount of time, that is, have some effective means of applying the criterion by which we know whether β tests σ. Sober's premise is also contentious, since it is not clear that for each sentence and each $\mathscr{B}$-sentence, it is unequivocally decidable whether the latter tests the former.

Sober may just have relied on an unfortunate turn of phrase, with the term 'understanding' suggesting some mental process that relates 'testing' to 'testability'. The term can be avoided completely: Sober (2008, 35) himself assumes that his concept of testing (and hence also his concept of testability) is an explicatum. Thus his claim can be rephrased as 'If we have an adequate and fruitful explicatum for 'testing', we also immediately have an adequate and fruitful explicatum for 'testability''. Specifically, given the relation between 'testing' and 'testability' that Sober assumes, a definition of 'observation ω tests Σ' immediately leads to the definition 'Σ is testable if and only if it is possible for there to be an observation ω such that ω tests Σ'.

Chapter 6

Deductive criteria for sentences

As discussed in chapter §5, the problem of past failures suggests that it is simply impossible to find an adequate demarcation criterion. And every new suggestion for a criterion will have to face the charge of arbitrariness, according to which not every non-trivial criterion is adequate. Rather, a criterion has to have the right relation to its explicandum. In this chapter, I will show that several criteria of empirical significance for sentences can overcome the charge of arbitrariness, since they are closely related to each other and the original explicandum.[1]

For the criteria are equivalent (§6.2.1) or nearly equivalent (§6.2.2) to falsifiability, which is also the non-trivial core of Ayer's criteria. Falsifiability in turn is closely connected to verifiability (§6.3). Falsifiability and verifiability are more inclusive than the (universally panned) criterion demanding both (§6.4), which is itself more inclusive than the criterion of strong $\mathscr{B}$-determinacy, suggested independently by Patrick Suppes, Marian Przełęcki, and David Lewis (§6.5). More inclusive than both falsifiability and verifiability is the criterion that demands either one, and which has been suggested by David Rynin in a syntactic and by Przełęcki in a semantic formulation. A criterion given by Carnap, once it is modified to avoid triviality, is a variant of this (§6.6). Falsifiability, verifiability, their disjunction, and strong $\mathscr{B}$-creativity thus make up the four major criteria of empirical significance.

The entailment relations between the four major criteria (summarized in figures 6.1 and 6.2) suggest the introduction of comparative concepts of empirical

[1]Parts of this chapter have been presented under the title "Empirical significance, Ramsey sentences, and the theory of definition" at the *PhDs in Logic II* conference at Tilburg University, The Netherlands, February 19, 2010. I thank the audience for helpful discussions. This chapter has also profited enormously from a reading group and further discussions at Tilburg University with Reinhard Muskens and Stefan Wintein.

significance, which provide a rebuttal to one of Hempel's criticisms of the concept of empirical significance (§6.7.1). And since the different formulations of each major criterion have been arrived at by different considerations, the formulations' equivalence allows a cumulative defense of each of them (§6.7.2).

Because of the equivalences, one can also choose based on expedience which formulation to generalize. I suggest two generalizations. The first weakens the criteria's presumptions about basic sentences (§6.8.1). I arrive at the the other by modifying an otherwise trivial criterion by Elliott Sober. This leads to a generalization of falsifiability that takes general background assumptions into account and is a further elaboration of criteria suggested previously by Carnap and Popper. This generalization transfers directly to all other criteria, and supplants the assumption at the heart of the trivial amendments of Ayer's criterion, thereby solving the problem of past failures (§6.8.2).

6.1 Preliminaries

All of the criteria in the following refer to a consistent set of analytic sentences or meaning postulates Π. Przełęcki (1974a, 345) argues that the meaning postulates $\Pi_{\mathscr{A}}$ for the auxiliary terms should be $\mathscr{B}$-conservative with respect to $\Pi_{\mathscr{B}}$, the meaning postulates for the basic terms. That is, $\Pi_{\mathscr{A}}$ should place no restrictions on $\mathscr{B}$-sentences or their interpretations beyond those given through $\Pi_{\mathscr{B}}$. I will not make this assumption, but rather generalize concepts and results where necessary. I have defended the viability of the concept of analyticity in §2. The assumption of a set of analytic sentences is also not necessarily a restriction, since Π may be empty. On the other hand, demanding $\Pi = \varnothing$ severely restricts the inferences that are possible—analytic entailment then coincides with logical entailment. If, for example, Π cannot even contain stipulative definitions, it is impossible to introduce terms like 'linear' and 'continuous', so that the inference from 'function f is linear' to 'function f is continuous' is impossible. I will further discuss the role of Π in the criteria in §6.8.2.

The criteria under discussion are meant to explicate empirical significance for sentences, not terms. Whether this is a restriction at all is a matter of debate. While Carnap (1956b) considers criteria for terms possible and perhaps even preferable to criteria for sentences (see also Hempel 1965c, §3), Przełęcki (1974a, 345–346), for example, considers such criteria misguided. But even if, *pace* Przełęcki, criteria for terms do turn out to be desirable, the criteria for sentences do not thereby become superfluous. Rather, they define empirical significance under the condition that the object under scrutiny is a sentence, not a term. I will discuss the relation of the criteria for sentences to the criteria for terms in §7.

As discussed in §5.2, I will assume in the following that the criteria under discussion are meant to be applied in contexts that allow only deductive inferences, and thus not as criteria of inductive confirmability. Under this assumption, some restrictions on the criteria are overly restrictive. Hempel's restriction of observational information to finite sets of molecular sentences (Hempel 1965c, §2) is the best example of this. In its stead, I will mainly rely on what Carnap sometimes calls the 'extended observation language', which contains all sentences that contain only logical and $\mathscr{B}$-terms (cf. Psillos 2000, 158–159). The language thus also includes all quantified sentences, and thus "empirical laws" or "empirical generalizations" (Carnap 1966, 225–227). I will revisit restrictions on the basic sentences in §6.8.1.

6.2 Falsifiability

6.2.1 Syntactic criteria

As I will discuss in §6.8.2, Hempel's formulation of the falsifiability criterion deviates from Popper's original criterion in at least one crucial respect, but it can serve as a good starting point for my discussion. Hempel (1965c, 106) states his "requirement of complete falsifiability in principle" like this:

> A sentence has empirical meaning if and only if its negation is not analytic and follows logically from some finite logically consistent class of observation sentences.

Since I am here not interested in criteria of confirmability, I will drop Hempel's requirement that the set of basic sentences be finite.[2] For two reasons, I will also allow the *analytic* entailment of the sentence's negation. First, analytic entailment is a simple generalization of logical entailment that can be undone by demanding that Π be empty. Second, only tautological $\mathscr{A}$-sentences follow logically from a consistent set of $\mathscr{B}$-sentences, and therefore no $\mathscr{A}$-sentences have empirical meaning according to Hempel's definition.[3] Finally, I will generalize the criterion for sentences to a criterion for sets of sentences because this allows the discussion of theories that cannot be finitely axiomatized and thus not be described in a single sentence. The generalization is straightforward: If τ is a sentence and Σ a set of sentences, then $\Sigma \vDash \neg\tau$ if and only if $\Sigma \cup \{\tau\} \vDash \bot$, where '$\bot$' is some contradiction. And in the second formula, the restriction to a singleton set is superfluous. With these modifications and my terminology, the criterion says that a set of sentences is empirically significant if and only if it is syntactically falsifiable

[2] A generalization of the criterion that can capture this requirement is given in §6.8.1.

[3] This focus on the empirical significance of sentences with $\mathscr{A}$-terms is not as ahistorical as it may seem: Carnap (1928a, §§61–67) already focused on the relation of all the sentences of a language to those sentences containing only a subset of its terms (the "basic relations"). Consider also Schlick's and Ayer's example sentence 'The Absolute enters into, but is itself incapable of, evolution and progress', whose alleged lack of content hinges on the term 'Absolute' (Ayer 1936, 36). With his focus on the untenability of induction, Popper (1935, §V) was arguably an exception (in this respect, see n. 2 above).

and not analytically false.[4]

Definition 6.1. A *set Ω of sentences falsifies* a set Σ of sentences if and only if $\Omega \cup \Sigma \cup \Pi \vDash \bot$.

Definition 6.2. A set Σ of $\mathscr{V}$-sentences is *syntactically falsifiable* if and only if it is falsified by a possible set of $\mathscr{B}$-sentences.

Note that the relation between falsification and falsifiability is analogous to that of testing and testability described by Sober: Falsifying is to falsifiability as testing is to testability (see §5.3). Since a falsifiable sentence cannot be analytic, the criterion of empirical significance could also be formulated as the demand that a sentence be syntactically falsifiable and analytically contingent.

As an example, assume that 'q' stands for 'foo', 'Ax' for 'x is bar', 'c' for 'the cat', and 'Bx' for 'is on the mat'. Then in most contexts, $B \in \mathscr{B}$ and $c \in \mathscr{B}$, and for $\Pi = \varnothing$, $A \in \mathscr{A}$ and $q \in \mathscr{A}$. The sentences $Bc \wedge Aq$ and $\forall x(Bx \wedge Ax)$ are then syntactically falsifiable because $\neg Bc \vDash \neg(Bc \wedge Aq)$ and $\exists x Bx \vDash \neg\forall x(Bx \wedge Ax)$. $\forall x Bx \vee \forall x Ax$ is not falsifiable, however, because the interpretation of A can always be chosen to encompass the whole domain, so that there is no $\mathscr{B}$-sentence that entails $\neg\forall x Ax$, and hence no $\mathscr{B}$ sentence that entails $\neg\forall x Bx \wedge \neg\forall x Ax \mathrel{\vDash\!\!\!\dashv} \neg(\forall x Bx \vee \forall x Ax)$.

Even though I have defined '$\mathscr{B}$-sentence' to be any sentence containing only $\mathscr{B}$-terms, syntactic falsifiability, like all other syntactic criteria in the following, only presumes that the $\mathscr{B}$-sentences form some distinguished set of sentences. When '$\mathscr{B}$-sentences' is defined in this way, the syntactic criteria are thus immediately generalized so that they do not rely on a bipartition of the vocabulary.

The criterion of falsifiability is typically introduced with the observation that few universally quantified sentences are entailed by molecular basic sentences, but their negations may be so entailed. But even assuming that most scientific laws can be given as universally quantified sentences, this purely formal observation is no justification of the criterion. The most important justification rather relies implicitly on the notion of $\mathscr{B}$-conservativeness, which is a necessary condition for explicit definitions (cf. Belnap 1993; Gupta 2009, §2.1).

Definition 6.3. A set Σ of $\mathscr{V}$-sentences is *syntactically $\mathscr{B}$-conservative* with respect to a set Δ of $\mathscr{V}$-sentences if and only if for any set Ω of $\mathscr{B}$-sentences and for any $\mathscr{B}$-sentence β, $\Omega \cup \Sigma \cup \Delta \vDash \beta$ only if $\Omega \cup \Delta \vDash \beta$.

A set of $\mathscr{V}$-sentences is *syntactically $\mathscr{B}$-creative* with respect to Δ if and only if it is not syntactically $\mathscr{B}$-conservative with respect to Δ.[5] If a logic is compact, $\Omega \cup \Sigma \cup \Delta \vDash \beta$ if and only if there is a finite set Ω' such that $\Omega' \cup \Sigma \cup \Delta \vDash \beta$. This

[4] As noted in §5.1, the qualifier 'syntactic' here does not refer to the use of syntactic deduction ('$\vdash$'), but to the syntactic description of empirical states (by sentences).

[5] Note again that 'syntactic' refers to the syntactic description of the observations. This terminology is essentially that of Przełęcki (1969, 52).

is equivalent to $\Sigma \cup \Delta \vDash \bigwedge \Omega' \to \beta$. Hence for first order logic, and if the set of basic sentences is closed under truth-functional composition,[6] Σ is syntactically $\mathscr{B}$-conservative relative to Δ if and only if for any $\mathscr{B}$-sentence β, $\Sigma \cup \Delta \vDash \beta$ only if $\Delta \vDash \beta$.

That the definition of a new term not in $\mathscr{B}$ must be $\mathscr{B}$-conservative encapsulates the idea "that the definition not have any consequences (other than those consequences involving the defined word itself) that were not obtainable already without the definition", as Belnap (1993, 123) puts it. Thus, a set that is syntactically $\mathscr{B}$-conservative with respect to Π sanctions no inferences between $\mathscr{B}$-sentences that are not already sanctioned by Π. In the following, $\mathscr{B}$-conservativeness simpliciter is understood to be $\mathscr{B}$-conservativeness with respect to Π.

Popper's justification of falsifiability essentially starts from $\mathscr{B}$-creativity because he demands "that the theory allow us to deduce, roughly speaking, more empirical singular statements than we can deduce from the initial conditions alone" (Popper 1935, 85). By assuming that the negation of a basic sentence is itself a basic sentence, he thus justifies his definition of falsifiability with the help of

Claim 6.1. *A set Σ of $\mathscr{V}$-sentences is syntactically falsifiable if and only if Σ is syntactically $\mathscr{B}$-creative with respect to Π.*

Proof. '$\Rightarrow$': If $\Omega \cup \Sigma \cup \Pi \vDash \bot$, then $\Omega \cup \Sigma \cup \Pi \vDash \beta$ for any basic sentence β. Since $\Omega \cup \Pi \nvDash \bot$, there is some β such that $\Omega \cup \Pi \nvDash \beta$.

'$\Leftarrow$': For β and Ω with $\Omega \cup \Sigma \cup \Pi \vDash \beta$ and $\Omega \cup \Pi \nvDash \beta$, $\Omega \cup \{\neg\beta\} \cup \Pi \nvDash \bot$ and $\Omega \cup \{\neg\beta\} \cup \Sigma \cup \Pi \vDash \bot$. □

The relation between falsifiability and $\mathscr{B}$-creativity provides a justification for Reichenbach's (and Nielsen's and Flew's) claim that the $\mathscr{B}$-sentences only need to be unproblematic, not observational (see §5.1): The theory of definition and the concept of $\mathscr{B}$-creativity are independent of the meaning of the $\mathscr{B}$-terms.

Sticking with the interpretation of $\mathscr{B}$-sentences as observational, a falsifiable sentence could be said to have empirical import, where "a sentence S has empirical import if from S in conjunction with suitable subsidiary hypotheses it is possible to derive observation sentences which are not derivable from the subsidiary hypotheses alone", as Hempel (1965c, 106) puts it (suitable subsidiary hypotheses for falsifiability being analytic and observational). It is one of the cruel jokes of philosophical terminology that he is describing Ayer's two criteria of verifiability. Given the close connection between Ayer's and Popper's criteria, it is unsurprising that the justification that Ayer provides for his criteria complements Popper's justification. Ayer (1936, 97–99) argues that the function of an empirical hypothesis is to predict experiences, and thus arrives at his first criterion of empirical significance, namely that "the mark of a genuine factual proposition [is] that some

[6] This is always the case if the $\mathscr{B}$-sentences are defined as all those containing only $\mathscr{B}$-vocabulary.

experiential propositions can be deduced from it in conjunction with certain other premises without being deducible from those other premises alone", where an experiential proposition "records an actual or possible observation" (Ayer 1946, 38–39).

Because no restriction is put on the "certain other premises", Ayer's first criterion is trivial in that it includes every non-analytic sentence (cf. Lewis 1988a). One way to avoid this triviality is to demand that the other premises be $\mathscr{B}$-sentences, which makes the criterion equivalent to $\mathscr{B}$-creativity. Instead, Ayer (1946, 13) proposes two definitions. The first stipulates that

> a statement is directly verifiable if it is either itself an observation-statement, or is such that in conjunction with one or more observation-statements it entails at least one observation-statement which is not deducible from these other premises alone [. . .].

If 'entailment' is understood as 'analytic entailment'[7] and the criterion is meant as a necessary and sufficient condition, this can be paraphrased as

Definition 6.4. A $\mathscr{V}$-sentence σ is *directly verifiable* if and only if σ is a $\mathscr{B}$-sentence or there is some set Ω of $\mathscr{B}$-sentences and a $\mathscr{B}$-sentence β such that $\Omega \cup \{\sigma\} \cup \Pi \vDash \beta$ and $\Omega \cup \Pi \nvDash \beta$.

Without any assumptions about the set of basic sentences, the next claim follows immediately:

Claim 6.2. *A $\mathscr{V}$-sentence σ is directly verifiable if and only if σ is a $\mathscr{B}$-sentence or is syntactically $\mathscr{B}$-creative with respect to Π.*

The condition that σ may be a $\mathscr{B}$-sentence is not redundant because σ may be analytic and therefore not $\mathscr{B}$-creative with respect to Π.

In his second definition, Ayer (1946, 13) proposes

> to say that a statement is indirectly verifiable if it satisfies the following conditions: first, that in conjunction with certain other premises [Γ] it entails one or more directly verifiable statements [β] which are not deducible from these other premises alone; and secondly, that these other premises do not include any statement that is not either analytic, or directly verifiable, or capable of being independently established as indirectly verifiable.

Since analytic entailment already allows the inclusion of Π in the premises of an inference, Π can be dropped from the auxiliary assumptions Γ. In the special case that Γ is a set of $\mathscr{B}$-sentences and β a $\mathscr{B}$-sentence as well, indirect verifiability

[7] This is what Ayer seems to do, since he calls translations from one language into another 'logically equivalent' (Ayer 1946, 6–7). Lewis (1988b, §II, fn. 5) gives an independent argument for reading Ayer in this way, but also notes that this entails some redundancies in Ayer's definitions.

reduces to direct verifiability (cf. Pokriefka 1983),[8] so that Γ can contain $\mathscr{B}$-sentences instead of directly verifiable sentences. Ayer's criterion can then be stated as

Definition 6.5. A $\mathscr{V}$-sentence σ is *indirectly verifiable* if and only if there is a set Γ of indirectly verifiable or $\mathscr{B}$-sentences and a sentence γ that is directly verifiable such that $\{\sigma\} \cup \Gamma \cup \Pi \vDash \gamma$ and $\Gamma \cup \Pi \nvDash \gamma$.

To avoid circularity, one would have to demand that the set of indirectly verifiable sentences is the *least* subset of the set of $\mathscr{V}$-sentences that fulfill the definiens of definition 6.5.[9] This new definition is recursive (cf. Moschovakis 1974, 1): The set of indirectly verifiable sentence is given by $\bigcup_{k=1}^{\infty} I_k$ with I_k as follows: I_1 is the set of sentences σ for which $\{\sigma\} \cup \Gamma \cup \Pi \vDash \gamma$ and $\Gamma \cup \Pi \nvDash \gamma$ with Γ a set of $\mathscr{B}$-sentences, and I_{l+1} is the set of sentences σ for which $\{\sigma\} \cup \Gamma \cup \Pi \vDash \gamma$ and $\Gamma \cup \Pi \nvDash \gamma$ with $\Gamma = \bigcup_{k=1}^{l} I_k$.

Church (1949) shows that for any sentence, as long as there are three logically independent $\mathscr{B}$-sentences, the sentence or its negation is indirectly verifiable, a trivialization that is possible even if γ is required to be a $\mathscr{B}$-sentence. This trivialization can be avoided by restricting both Γ and γ to $\mathscr{B}$-sentences, but this more exclusive version of indirect verifiability then again just amounts to $\mathscr{B}$-creativity or, equivalently, falsifiability.

In connection with his first criterion, Ayer (1936, 38) argues that a "hypothesis cannot be conclusively confuted any more than it can be conclusively verified", but that a sentence is verifiable "if it is possible for experience to render it probable" (Ayer 1936, 37). Ayer (1936, 99) then argues that "if an observation to which a given proposition is relevant conforms to our expectations, the truth of that proposition is confirmed. [Then] one can say that its probability has been increased." 'Probability' is here not used in its mathematical sense, but as a measure of our "confidence" in a proposition (Ayer 1936, 100). Thus although Ayer *justifies* his criterion with the purpose of theories, which according to him is the assertion of observation sentences, he *develops* his criterion under the assumption that an empirically significant sentence is one that can be confirmed or disconfirmed. Furthermore, he assumes that a sentence is confirmed if one of its consequences turns out to be true. This *prediction criterion of confirmation* is discussed and rejected by, for example, Hempel (1965f, §7).

Gemes (1998b, §1.4) discusses the historical importance of the assumptions that empirical significance is the same as confirmability and that confirmation can be explicated by purely formal means. He argues that the failure of the search for

[8] This holds even without the assumption needed for Church's trivialization proof given below, simply by restricting Γ and β; hence I consider it an innocent observation that does not just transform one trivial criterion into another one.

[9] Arguably, this is what Ayer means by his demand that the other premises be "capable of being independently established" as indirectly verifiable.

a criterion of empirical significance is inherited from the failure of purely formal criteria of confirmation: Purely formal criteria of confirmation allow the use of any terms whatsoever, and thus specifically terms that are not projectible given the other terms of the language (see §4.1.2). This is one reason why purely formal criteria of confirmation are not viable. Many criteria of empirical significance were based, like Ayer's, on some purely formal criterion of confirmation, and therefore have failed as well.

6.2.2 Semantic criteria

Syntactic $\mathscr{B}$-conservativeness has a semantic counterpart:

Definition 6.6. A set Σ of $\mathscr{V}$-sentences is *semantically $\mathscr{B}$-conservative* with respect to a set Δ of $\mathscr{V}$-sentences if and only if for each $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ for which there is a $\mathscr{V}$-structure $\mathfrak{B} \vDash \Delta$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$, there is also a $\mathscr{V}$-structure $\mathfrak{C} \vDash \Delta \cup \Sigma$ with $\mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$.

A set of $\mathscr{V}$-sentences is *semantically $\mathscr{B}$-creative* with respect to Δ if and only if it is not semantically $\mathscr{B}$-conservative with respect to Δ. Definition 6.6 is slightly more general than that given, for example, by Przełęcki (1974a, 345), so that it allows for any $\mathscr{V}$-sentence in Δ. A description of the generalization is given in appendix 6.11.1. Note that, like the other semantic definitions up to §6.7.1, this definition relies essentially on a bipartition of the vocabulary.

Assuming again that $B \in \mathscr{B}$ and $c \in \mathscr{B}$, $\Pi = \varnothing$, $A \in \mathscr{A}$, and $q \in \mathscr{A}$, the sentences $Bc \wedge Aq$ and $\forall x(Bx \wedge Ax)$ are semantically $\mathscr{B}$-creative because every $\mathscr{B}$-structure can be expanded to Π, but $\mathfrak{A}_{\mathscr{B}}$ with $B^{\mathfrak{A}_{\mathscr{B}}} = \{1, 2\}$ and $c^{\mathfrak{A}_{\mathscr{B}}} = 3$ cannot be expanded to a model of either sentence. $\forall x Bx \vee \forall x Ax$ is not $\mathscr{B}$-creative, however, because every $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ can be expanded to a model $\mathfrak{A}$ of $\forall x Bx \vee \forall x Ax$ by choosing $A^{\mathfrak{A}} = |\mathfrak{A}_{\mathscr{B}}|$.

As announced in §5.1, the difference between semantic and syntactic conservativeness lies in the precision of the empirical information, specifically in the difference between isomorphism ('$\simeq$') and syntactical equivalence ('$\equiv$'):[10]

Claim 6.3. *A set Σ of $\mathscr{V}$-sentences is syntactically $\mathscr{B}$-conservative with respect to Δ iff for each $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ for which there is a $\mathscr{V}$-structure $\mathfrak{B} \vDash \Delta$ with $\mathfrak{B}|_{\mathscr{B}} \equiv \mathfrak{A}_{\mathscr{B}}$, there is a $\mathscr{V}$-structure $\mathfrak{C} \vDash \Delta \cup \Sigma$ with $\mathfrak{C}|_{\mathscr{B}} \equiv \mathfrak{A}_{\mathscr{B}}$.*

Proof. '$\Rightarrow$': Assume $\mathfrak{A}_{\mathscr{B}}$ is syntactically equivalent to a structure that can be expanded to a model $\mathfrak{B}$ of Δ. Then choose $\Omega \cup \{\neg\beta\}$ equivalent to $\mathrm{Th}(\mathfrak{A}_{\mathscr{B}})$. It follows that $\mathfrak{B} \vDash \Omega \cup \{\neg\beta\} \cup \Delta$ and thus $\Omega \cup \Delta \nvDash \beta$. By syntactic $\mathscr{B}$-conservativeness, $\Omega \cup \Sigma \cup \Delta \nvDash \beta$, so there is a $\mathfrak{C} \vDash \Omega \cup \{\neg\beta\} \cup \Sigma \cup \Delta \mathrel{\text{⫤}\!\vDash} \mathrm{Th}(\mathfrak{A}_{\mathscr{B}}) \cup \Sigma \cup \Delta$. Thus there is a $\mathfrak{C} \vDash \Sigma \cup \Delta$ such that $\mathfrak{C}|_{\mathscr{B}} \equiv \mathfrak{A}_{\mathscr{B}}$.

[10]See §5.1. Because of claim 2.1, definition 6.6 could be formulated with '$\simeq$' instead of '$=$'.

'$\Leftarrow$': Let $\Omega \cup \Delta \not\models \beta$. Choose $\mathfrak{A} \models \Omega \cup \Delta \cup \{\neg\beta\}$; by assumption, there is a $\mathfrak{C} \models \Sigma \cup \Delta$ with $\mathfrak{C}|_{\mathscr{B}} \equiv \mathfrak{A}|_{\mathscr{B}}$ and thus $\mathfrak{C} \models \Omega \cup \Sigma \cup \Delta \cup \{\neg\beta\}$, so that $\Omega \cup \Sigma \cup \Delta \not\models \beta$. □

This suggests:

Claim 6.4. *A set Σ of $\mathscr{V}$-sentences is semantically $\mathscr{B}$-conservative with respect to Δ only if Σ is syntactically $\mathscr{B}$-conservative with respect to Δ. The converse does not hold in first order logic.*

Proof. '$\Rightarrow$': From claim 6.3 because $\mathfrak{A}|_{\mathscr{B}} = \mathfrak{B}|_{\mathscr{B}}$ only if $\mathfrak{A}|_{\mathscr{B}} \equiv \mathfrak{B}|_{\mathscr{B}}$.
'$\not\Leftarrow$': See appendix 6.11.2. □

Of course, the two criteria are equivalent in all languages in which syntactic equivalence amounts to isomorphism. A short overview of mainly philosophical treatments of the relation is given in appendix 6.11.2. Because of the difference between syntactic and semantic $\mathscr{B}$-conservativeness, it may not always be possible to bipartition the set of analytic sentences Π such that $\Pi_{\mathscr{A}}$ is semantically $\mathscr{B}$-conservative with respect to $\Pi_{\mathscr{B}}$: If Π is only syntactically conservative with respect to $\Pi_{\mathscr{B}}$, there are some $\mathscr{B}$-models of $\Pi_{\mathscr{B}}$ that cannot be expanded to models of Π, and there is no $\mathscr{B}$-sentence that excludes all and only those structures when added to $\Pi_{\mathscr{B}}$.

The analogy between syntactic and semantic $\mathscr{B}$-conservativeness suggests a semantic criterion of falsifiability analogous to syntactic falsifiability.

Definition 6.7. A *$\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ falsifies* a set Σ of $\mathscr{V}$-sentences if and only if for all $\mathfrak{C} \models \Pi$ with $\mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$, $\mathfrak{C} \not\models \Sigma$.

In other words, a structure $\mathfrak{A}_{\mathscr{B}}$ falsifies Σ if and only if Σ is false in every possible structure that is an expansion of $\mathfrak{A}_{\mathscr{B}}$.

Definition 6.8. A set Σ of $\mathscr{V}$-sentences is *semantically falsifiable* if and only if it is falsified by a possible $\mathscr{B}$-structure.

Now the following holds:

Claim 6.5. *A set Σ of $\mathscr{V}$-sentences is semantically falsifiable if and only if Σ is semantically $\mathscr{B}$-creative with respect to Π.*

Proof. Σ is semantically $\mathscr{B}$-creative with respect to Π iff there is an $\mathfrak{A}_{\mathscr{B}}$ that has an expansion $\mathfrak{B} \models \Pi$ (which is always the case since Π is consistent) and every expansion $\mathfrak{C} \models \Pi$ of $\mathfrak{A}_{\mathscr{B}}$ is such that $\mathfrak{C} \not\models \Sigma$. Because of claim 2.1, this holds iff Σ is semantically falsifiable. □

The relation between syntactic and semantic falsifiability is then given by claims 6.5, 6.4 and 6.1. Claim 2.1 will be used silently in the following, that is, I will argue as if a structure is possible if and only if it can be expanded to a model of Π.

David Lewis argues that one of his explications of 'partial aboutness' is closely connected to syntactic falsifiability. To see that it is even more closely connected to semantic falsifiability, consider first Lewis's explication of 'aboutness' as supervenience. According to Lewis (1988b, 136), a "statement is entirely about some subject matter iff its truth value supervenes on that subject matter. Two possible worlds which are exactly alike so far as that subject matter is concerned must both make the statement true, or else both make it false". Assuming that possible worlds are all and only those worlds in which all analytic sentences are true, and assuming that all statements can be expressed by sets of $\mathscr{V}$-sentences, there is a one-to-one mapping from possible worlds to $\mathscr{V}$-structures (cf. Kemeny 1963, §IV). Lewis does not explicate what it means for possible worlds to be "exactly alike" with respect to a subject matter (except that 'being exactly alike' is an equivalence relation), so I suggest identifying subject matters by the vocabulary used to describe them: Two possible worlds are exactly alike with respect to a subject matter $\mathscr{B}$ if and only if the reducts of their corresponding structures to $\mathscr{B}$ are identical. The identification of subject matters by their vocabulary is arguably what Dorr (2010) has in mind when suggesting that "Xs are more fundamental than Ys" only if it is possible to "introduce a language in which I can talk about Xs without even seeming to talk about Ys" (see §2.7). It is also arguably a central idea in Nagel's notion of reduction (Nagel 1951, 330), which requires that

> every term which occurs in the statements of [the reduced discipline] S_2 [...] must be either explicitly definable with the help of the vocabulary specific to the primary discipline [S_1] [...] or well-established empirical laws must be available with the help of which it is possible to state the sufficient conditions for the application of all expressions in S_2, exclusively in terms of expressions occurring in the explanatory principles of S_1.

Nagel's discussion thus suggests that different disciplines typically rely on different vocabularies (see §11.1). Fodor (1974, 98) explicitly assumes "that a science is individuated largely by reference to its typical predicates" (see §11.3). This assumption leads to

Definition 6.9. A set Σ of $\mathscr{V}$-sentences is *about subject matter* $\mathscr{B}$ if and only if for any $\mathscr{V}$-structures $\mathfrak{A} \models \Pi, \mathfrak{B} \models \Pi$ with $\mathfrak{A}|_{\mathscr{B}} = \mathfrak{B}|_{\mathscr{B}}$ it holds that $\mathfrak{A} \models \Sigma$ iff $\mathfrak{B} \models \Sigma$.

To distinguish aboutness more clearly from partial aboutness, I will also sometimes speak of sentences being *entirely about* a subject matter $\mathscr{B}$ when they are about a subject matter $\mathscr{B}$.

Lewis (1988b, §VII, footnote removed) suggests to weaken definition 6.9 based on an ordinary language analysis of the modifier 'partly':

> The recipe for modifying X by 'partly' is something like this. Think of the situation to which X, unmodified, applies. Look for an aspect of that situation that has parts, and therefore can be made partial. Make it partial—and there you have a situation to which 'partly X' could apply. If you find several aspects that could be made partial, you have ambiguity.

In this case, X stands for 'Statement S is about subject matter $\mathscr{B}$'. Lewis identifies four different aspects of the situation that have parts. The most obvious aspect is S itself, but considering parts of it leads Lewis (1988b, §XI) to a criterion that distinguishes between logically equivalent sentences. Another aspect is the subject matter $\mathscr{B}$. In order to arrive at a non-trivial criterion, Lewis (1988b, §IX) must assume that it is clear what it means for a subject matter to be "close-knit" and either "sufficiently large" or "sufficiently important". Clarifying these terms may, however, lead to an infinite regress, for instance if it turns out that a subject matter is close-knit if and only if the sufficiently large or important parts are partially about each other. Making the supervenience partial leads Lewis (1988b, §X) to a probabilistic conception of empirical significance, although I will argue in §6.6 that this is not the only option. Only his treatment of the content of a statement stays within the boundaries of predicate logic, if the above translation from modal semantics into model theory is assumed. Lewis (1988b, §VIII) defines the content of a statement as the set $\mathbf{C}$ of possible worlds that it excludes. In the model theoretic paraphrase, the content of a set Σ of sentences is thus given by $\mathbf{C}_\Sigma := \{\mathfrak{A} \mid \mathfrak{A} \vDash \Pi \text{ and } \mathfrak{A} \nvDash \Sigma\}$. The content of Σ is about subject matter $\mathscr{B}$ if and only if Σ itself is about subject matter $\mathscr{B}$, which is the case if and only if for any two $\mathfrak{B}, \mathfrak{C} \vDash \Pi$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{C}|_{\mathscr{B}}$, $\mathfrak{B} \in \mathbf{C}_\Sigma$ iff $\mathfrak{C} \in \mathbf{C}_\Sigma$. The parts of the content of Σ are then defined as the subsets of $\mathbf{C}_\Sigma$, which leads to

Definition 6.10. *Part of the content* of a set Σ of $\mathscr{V}$-sentences is *about subject matter* $\mathscr{B}$ if and only if there is a non-empty set of structures $\mathbf{F} \subseteq \mathbf{C}_\Sigma := \{\mathfrak{A} \mid \mathfrak{A} \vDash \Pi \text{ and } \mathfrak{A} \nvDash \Sigma\}$ such that for any two $\mathfrak{B} \vDash \Pi, \mathfrak{C} \vDash \Pi$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{C}|_{\mathscr{B}}$, $\mathfrak{B} \in \mathbf{F}$ iff $\mathfrak{C} \in \mathbf{F}$.

Lewis does not demand $\mathbf{F}$ to be non-empty, but without this restriction, part of the content of every sentence is about subject matter $\mathscr{B}$. If there is a way to capture any content (any set of possible worlds) by a sentence, Lewis (1988b, §VIII) notes, part of the content of a sentence is about subject matter $\mathscr{B}$ iff the sentence is syntactically falsifiable.[11] But Lewis's definition 6.10 is better compared to *semantic* falsifiability:

[11] To be more precise, since Lewis does not demand $\mathbf{F}$ to be non-empty, he can show that part of a statement's content is about subject matter $\mathscr{B}$ if and only if the statement is incompatible with some statement entirely about subject matter $\mathscr{B}$. But according to definition 6.9 and Lewis (1988b, 141) himself, contradictions are entirely about subject matter $\mathscr{B}$, and since contradictions are incompatible with every statement, this shows that his definition is trivial. Demanding $\mathbf{F}$ to be non-empty excludes contradictions.

Claim 6.6. *Part of the content of a set Σ of $\mathscr{V}$-sentences is about subject matter $\mathscr{B}$ if and only if Σ is semantically falsifiable.*

Proof. '$\Rightarrow$': Assume part $\mathbf{F} \subseteq \mathbf{C}_\Sigma$ of Σ's content is about subject matter $\mathscr{B}$. Define $\mathfrak{A}_\mathscr{B} := \mathfrak{A}|_\mathscr{B}$ for some $\mathfrak{A} \in \mathbf{F}$. Since $\mathfrak{A} \in \mathbf{F}$ and according to definition 6.10 either all $\mathfrak{B}$ with $\mathfrak{B}|_\mathscr{B} = \mathfrak{A}_\mathscr{B}$ are in $\mathbf{F}$ or none is, all such $\mathfrak{B}$ are in $\mathbf{F}$. Since all such $\mathfrak{B}$ are also in $\mathbf{C}_\Sigma$, $\mathfrak{B} \not\models \Sigma$, and the possible structure $\mathfrak{A}_\mathscr{B}$ falsifies Σ.

'$\Leftarrow$': Assume Σ is semantically falsified by $\mathfrak{A}_\mathscr{B}$. Define $\mathbf{F} := \{\mathfrak{B} \mid \mathfrak{B} \models \Pi \text{ and } \mathfrak{B}|_\mathscr{B} = \mathfrak{A}_\mathscr{B}\}$. Since $\varnothing \neq \mathbf{F} \subseteq \mathbf{C}_\Sigma$, part of Σ's content is about subject matter $\mathscr{B}$. $\square$

Because of claims 6.1, 6.5, and 6.6, the relation between syntactic falsifiability and Lewis's definition 6.10 is the same as that between syntactic and semantic $\mathscr{B}$-creativity, which is given in claim 6.3.

A sentence whose content is partly about subject matter $\mathscr{B}$ could also be said to have some basic or $\mathscr{B}$-content, and indeed this is essentially how Carnap (1928b, 327–328) described a criterion of meaningfulness at the time of the Vienna circle (see page 247). Decades later, he argued that, absent sentences already established as analytic, the $\mathscr{B}$-content of a sentence σ is given by its Ramsey sentence $\mathsf{R}_\mathscr{B}(\sigma)$ (Psillos 2000). As discussed in §2.8.2, $\mathsf{R}_\mathscr{B}(\sigma)$ plausibly describes σ's basic content. Now, a criterion of the meaning of a set of sentences cannot be a criterion of empirical significance (see §5.2). Analogously, a description of the $\mathscr{B}$-content of a set of sentences cannot be a criterion of empirical significance either.[12] Something weaker is needed, namely a criterion to determine when the basic content is non-empty. Since anything that is already entailed by the analytic sentences is not an empirical claim, this suggests

Definition 6.11. If Π can be finitely axiomatized, let $\tilde{\Pi}$ be this axiomatization. Then a $\mathscr{V}$-sentence σ has *$\mathscr{B}$-content* if and only if $\tilde{\Pi} \not\models \mathsf{R}_\mathscr{B}(\sigma \wedge \bigwedge \tilde{\Pi})$.

Under this definition, Carnap's later notion of $\mathscr{B}$-content squares well with the notion of falsifiability:

Claim 6.7. *If Π can be finitely axiomatized, then a $\mathscr{V}$-sentence σ has $\mathscr{B}$-content if and only if σ is semantically $\mathscr{B}$-creative with respect to Π.*

Proof. A sentence σ is Ramseyfied by substituting every $\mathscr{A}$-term $A_i, 1 \leq i \leq n$ in σ by a variable X_i and existentially quantifying over each X_i, leading to $\exists X_1 \ldots X_n \sigma[A_1/X_1, \ldots, A_n/X_n]$. Define $g : \{A_i\}_{1 \leq i \leq n} \to \{X_i\}_{1 \leq i \leq n}, A_i \mapsto X_i$.

'$\Leftarrow$': Assume that σ has no $\mathscr{B}$-content. Since $\mathsf{R}_\mathscr{B}(\sigma \wedge \bigwedge \tilde{\Pi})$ is a $\mathscr{B}$-sentence, $\tilde{\Pi} \models \mathsf{R}_\mathscr{B}(\sigma \wedge \bigwedge \tilde{\Pi})$ if and only if $\mathsf{R}_\mathscr{B}(\tilde{\Pi}) \models \mathsf{R}_\mathscr{B}(\sigma \wedge \bigwedge \tilde{\Pi})$. Then for any $\mathfrak{A}_\mathscr{B}$, $\mathfrak{A}_\mathscr{B} \models \mathsf{R}_\mathscr{B}(\bigwedge \tilde{\Pi})$ only if $\mathfrak{A}_\mathscr{B} \models \mathsf{R}_\mathscr{B}(\sigma \wedge \bigwedge \tilde{\Pi})$. Thus for any $\mathfrak{A}_\mathscr{B}$, if there is

[12] A reminder: If Σ is a finite set of sentences, I will sometimes write $\mathsf{R}_\mathscr{B}(\Sigma)$ instead of $\mathsf{R}_\mathscr{B}(\bigwedge \Sigma)$.

a satisfaction function ν mapping each variable $X_i, 1 \leq i \leq n$ to an extension of the same type over $|\mathfrak{A}_{\mathscr{B}}|$ such that $\mathfrak{A}_{\mathscr{B}}, \nu \vDash \bigwedge \tilde{\Pi}[A_1/X_1, \ldots, A_n/X_n]$, there is a satisfaction function ν' such that $\mathfrak{A}_{\mathscr{B}}, \nu' \vDash (\sigma \wedge \bigwedge \tilde{\Pi})[A_1/X_1, \ldots, A_n/X_n]$. Now assume that $\mathfrak{A}_{\mathscr{B}}$ can be expanded to a model $\mathfrak{B} \vDash \tilde{\Pi}$. $\mathfrak{B} = \langle |\mathfrak{B}|, f \rangle$ with domain $|\mathfrak{B}|$ and a function f that maps every $\mathscr{V}$-term to an extension of the same type in $|\mathfrak{B}|$. Since $\mathfrak{B}$ expands $\mathfrak{A}_{\mathscr{B}}$, $|\mathfrak{B}| = |\mathfrak{A}_{\mathscr{B}}|$. Thus any extension ν of $f|_{\{A_1,\ldots,A_n\}} \circ g^{-1}$ to all variables of the language is a satisfaction function such that $\mathfrak{A}_{\mathscr{B}}, \nu \vDash \bigwedge \tilde{\Pi}[A_1/X_1, \ldots, A_n/X_n]$. By assumption, there is then a satisfaction function ν' such that $\mathfrak{A}_{\mathscr{B}}, \nu' \vDash (\sigma \wedge \bigwedge \tilde{\Pi})[A_1/X_1, \ldots, A_n/X_n]$. Then any extension f of $\nu'|_{\{X_1,\ldots,X_n\}} \circ g$ to all $\mathscr{A}$-terms can be used to expand $\mathfrak{A}_{\mathscr{B}}$ to a model of $\bigwedge \tilde{\Pi} \wedge \sigma$, and therefore σ is semantically $\mathscr{B}$-conservative with respect to $\tilde{\Pi}$.

'$\Rightarrow$': Similar. □

Claim 6.7 generalizes lemma 2.4. The close connection between a theory's Ramsey sentence and the theory's falsifiability provides another reason to distinguish between deductive and probabilistic criteria of empirical significance. For Scheffler (1968, 273–274), Niiniluoto (1972), Tuomela (1973), and Raatikainen (2010) have argued in detail that in contexts that allow inductive inferences, a theory can be disconfirmed without its Ramsey sentence being false, so that falsification of a theory and its disconfirmation come apart. Insofar confirmation and disconfirmation of a theory are determined by probabilistic inferences, this means that one has to distinguish between criteria of empirical significance for contexts that allow only deductive inferences and criteria for contexts that allow only probabilistic inferences.

6.3 Verifiability

Another criterion of empirical significance that has been proposed very early on is that of syntactic verifiability (Hempel 1965c, 104). Modifying Hempel's formulation in a way analogous to his formulation of falsifiability leads to

Definition 6.12. A *set Ω of $\mathscr{V}$-sentences verifies* a set Σ of $\mathscr{V}$-sentences if and only if $\Omega \cup \Pi \vDash \Sigma$.

Definition 6.13. A set Σ of $\mathscr{V}$-sentences is *syntactically verifiable* if and only if there is a possible set Ω of $\mathscr{B}$-sentences that verifies Σ.

A set of sentences is then empirically significant if and only if it is analytically contingent and syntactically verifiable. Assuming again that $B \in \mathscr{B}$ and $c \in \mathscr{B}$, $\Pi = \varnothing$, $A \in \mathscr{A}$, and $q \in \mathscr{A}$, the unfalsifiable sentences $Bc \vee Aq$ and $\forall x Bx \vee \forall x Ax$ are verifiable because $Bc \vDash Bc \vee Aq$ and $\forall x Bx \vDash \forall x Bx \vee \forall x Ax$. On the other hand, the falsifiable sentence $\forall x(Bx \wedge Ax)$ cannot be verified, since no $\mathscr{B}$-sentence entails $\forall x Ax$, and thus no $\mathscr{B}$-sentences entails $\forall x Bx \wedge \forall x Ax \mathrel{\models\!\!\!\dashv} \forall x(Bx \wedge Ax)$.[13]

Hempel (1965c, 106) points out the following straightforward

Claim 6.8. *A $\mathscr{V}$-sentence σ is syntactically verifiable if and only if $\neg\sigma$ is syntactically falsifiable.*

The restriction to single sentences is essential, since there is no straightforward generalization of negation to arbitrary sets of sentences. It seems appropriate to also give a semantic version of verifiability.

Definition 6.14. A *$\mathscr{B}$-structure* $\mathfrak{A}_{\mathscr{B}}$ *verifies* a set Σ of $\mathscr{V}$-sentences if and only if for all $\mathfrak{C} \models \Pi$ with $\mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$, $\mathfrak{C} \models \Sigma$.

Definition 6.15. A set Σ of $\mathscr{V}$-sentences is *semantically verifiable* if and only if there is a possible $\mathscr{B}$-structure that verifies Σ.

And again, the following can easily be shown to hold:

Claim 6.9. *A $\mathscr{V}$-sentence σ is semantically verifiable if and only if $\neg\sigma$ is semantically falsifiable.*

The relations between syntactic and semantic falsifiability described in claims 6.3 and 6.4 therefore transfer to the verifiability of sentences. For sets of sentences, a relation analogous to claim 6.3 holds as well:

Claim 6.10. *A set Σ of $\mathscr{V}$-sentences is syntactically verifiable if and only if there is a possible $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ such that Σ is verified by each possible $\mathscr{B}$-structure syntactically equivalent to $\mathfrak{A}_{\mathscr{B}}$.*

Proof. '$\Rightarrow$': Assume that the possible set of $\mathscr{B}$-sentences Ω verifies Σ. Then for every $\mathfrak{B} \models \Omega \cup \Pi$, $\mathfrak{B} \models \Sigma$. Since Ω is possible, there is some such $\mathfrak{B}$. Choose $\mathfrak{A}_{\mathscr{B}} = \mathfrak{B}|_{\mathscr{B}}$. Then every $\mathfrak{C}$ with $\mathfrak{C}|_{\mathscr{B}} \equiv \mathfrak{A}_{\mathscr{B}}$ is such that $\mathfrak{C} \models \Omega$. Since for every possible $\mathscr{B}$-structure syntactically equivalent to $\mathfrak{A}_{\mathscr{B}}$, there is such a $\mathfrak{C}$, every possible $\mathscr{B}$-structure syntactically equivalent to $\mathfrak{A}_{\mathscr{B}}$ verifies Σ.

'$\Leftarrow$': Assume that every possible $\mathscr{B}$-structure syntactically equivalent to $\mathfrak{A}_{\mathscr{B}}$ verifies Σ. Choose $\Omega \mathrel{\models\!\!\!\dashv} \mathrm{Th}(\mathfrak{A}_{\mathscr{B}})$. Since $\mathfrak{A}_{\mathscr{B}}$ is possible, $\Omega \cup \Pi$ has a model, and thus Ω is possible. By assumption, $\mathfrak{B} \models \Omega \cup \Pi$ only if $\mathfrak{B} \models \Sigma$, and thus Ω verifies Σ. □

As in the case of falsifiability, semantic verifiability is like syntactic verifiability, except that the basic information is given by structures, not sets of sentences. Substituting in claim 6.10 'verifiable' by 'falsifiable' and 'verified' by 'falsified' results in a simple paraphrase of claim 6.3 that makes this analogy obvious. Furthermore, claims 6.9 and 6.7 entail the following:

[13] Thus we can never verify that everything is on the mat and is bar, because we can never verify that everything (or indeed anything) is bar.

Claim 6.11. *If Π can be finitely axiomatized, let $\tilde{\Pi}$ be this axiomatization. Then a $\mathscr{V}$-sentence σ is semantically verifiable if and only if $\tilde{\Pi} \nvDash \mathsf{R}_{\mathscr{B}}(\neg\sigma \wedge \bigwedge \tilde{\Pi})$.*

Even if the basic content of σ is considered to be the set of possible $\mathscr{B}$-sentences or $\mathscr{B}$-structures that *verify* σ, however, $\mathsf{R}_{\mathscr{B}}(\neg\sigma)$ (which I will call the reverse Ramsey sentence) is not the basic content of σ for $\Pi = \varnothing$. Rather, it can be shown similarly to the proof of claim 6.7 that the possible $\mathscr{B}$-structures that verify σ are the models of the negation of the reverse Ramsey sentence $\neg\mathsf{R}_{\mathscr{B}}(\neg\sigma)$. And this sentence is also analytically entailed by the same $\mathscr{B}$-sentences as σ, since $\beta \vDash \sigma$ if and only if $\neg\sigma \vDash \neg\beta$, which holds if and only if $\mathsf{R}_{\mathscr{B}}(\neg\sigma) \vDash \neg\beta$, and thus if and only if $\beta \vDash \neg\mathsf{R}_{\mathscr{B}}(\neg\sigma)$. The connection to claim 6.11 is rather that σ is verifiable if and only if its basic content is not empty; and σ's basic content is empty if $\neg\mathsf{R}_{\mathscr{B}}(\neg\sigma)$ has no models, that is, $\vDash \mathsf{R}_{\mathscr{B}}(\neg\sigma)$.

6.4 Falsifiability and verifiability

Calling a sentence empirically significant if and only if it is both falsifiable and verifiable ensures that the negation of any empirically significant sentence is also empirically significant. For this reason, Hempel (1965d, 122) considers a version of this criterion that allows only finite sets of molecular basic sentences, which he rejects as too strong. Rynin (1957, 51) also rejects such a finite version of this criterion. But Hempel's demand that the negation of a meaningless sentence be itself meaningless relies on nothing but intuition, an intuition that Rynin (1957, 55–56), for example, does not share. Hempel's consideration in favor of defining empirical significance as the conjunction of falsifiability and verifiability thus fails as a justification in artificial language philosophy because it relies crucially on an intuition; it fails in traditional, naturalized, and ordinary language philosophy because that intuition is not robust.

A real, though small, advantage of the criterion is that a sentence that is both verifiable and falsifiable is automatically analytically contingent, and therefore the criterion can be formulated without demanding analytic contingency explicitly. Probably the main reason for using this criterion is that it is a sufficient condition for empirical significance for both proponents of falsifiability and proponents of verifiability (see Kitts (1977) for an example of this kind of argument). This dialectical advantage and the slight convenience in formulation cannot, however, outweigh the criterion's lack of other justifications.[14]

[14]Note that I am not claiming that the criterion is inadequate. I am only claiming that so far, there have been no arguments in its favor.

6.5 Strong $\mathscr{B}$-determinacy

Given that the conjunction of falsifiability and verifiability had already been considered too strong a criterion of empirical significance by Hempel and Rynin, it may seem surprising that even stronger criteria have been suggested since. However, first, Hempel and Rynin reject criteria that allow only finite sets of molecular $\mathscr{B}$-sentences. Second, the stronger criteria have advantages not found in the conjunction of falsifiability and verifiability.

Przełęcki (1974a, §I) suggests a criterion of empirical significance for sentences that can easily be generalized to sets thereof:[15]

Definition 6.16. A *$\mathscr{B}$-structure* $\mathfrak{A}_{\mathscr{B}}$ *determines* a set Σ of $\mathscr{V}$-sentences if and only if for all $\mathscr{V}$-structures $\mathfrak{B}, \mathfrak{C} \vDash \Pi$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$ it holds that $\mathfrak{B} \vDash \Sigma$ iff $\mathfrak{C} \vDash \Sigma$.

Definition 6.17. A set Σ of $\mathscr{V}$-sentences is *strongly semantically $\mathscr{B}$-determined* if and only if it is determined by every possible $\mathscr{B}$-structure.

Since this definition includes analytically determined sentences, a set of sentences should be called empirically significant if and only if it is strongly semantically $\mathscr{B}$-determined and analytically contingent. The truth value of a strongly semantically $\mathscr{B}$-determined set Σ of sentences is fixed by any interpretation of the basic terms in any domain, because Σ is either true in all possible models that expand such a $\mathscr{B}$-structure, or it is false in all such models. Hence

Claim 6.12. *A set Σ of $\mathscr{V}$-sentences is strongly semantically $\mathscr{B}$-determined if and only if every possible $\mathscr{B}$-structure either falsifies or verifies Σ.*

As Przełęcki (1974a, 346–347) already notes, this definition is very exclusive. If, for example, the auxiliary term A_1 is conditionally defined by $\{\forall x[B_1 x \rightarrow (B_2 x \leftrightarrow A_1 x)]\} =: \Pi$, and '$B_1$', '$B_2$', and '$b$' are basic terms, then the possible structure $\mathfrak{A}_{\mathscr{B}} = \langle\{1,2\}, \{\langle B_1, \{1\}\rangle, \langle B_2, \{1\}\rangle, \langle b, 2\rangle\}\rangle$ does not determine $A_1(b)$. Therefore, $A_1(b)$ is not strongly semantically $\mathscr{B}$-determined. This is unsurprising, because, in Lewis's terminology, the definition includes only sentences that are (entirely) about subject matter $\mathscr{B}$:

Claim 6.13. *A set Σ of $\mathscr{V}$-sentences is strongly semantically $\mathscr{B}$-determined if and only if Σ is about subject matter $\mathscr{B}$.*

Proof. '$\Rightarrow$': Assume $\mathfrak{B}, \mathfrak{C} \vDash \Pi$, $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{C}|_{\mathscr{B}}$. Then $\mathfrak{B}|_{\mathscr{B}}$ is a possible $\mathscr{B}$-structure, and by assumption, $\mathfrak{B} \vDash \Sigma$ iff $\mathfrak{C} \vDash \Sigma$.

'$\Leftarrow$': Assume $\mathfrak{A}_{\mathscr{B}}$ is a possible $\mathscr{B}$-structure. For any two possible $\mathscr{V}$-structures $\mathfrak{B}$, $\mathfrak{C}$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$ and $\mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$, it holds that $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{C}|_{\mathscr{B}}$ and thus, by assumption, $\mathfrak{B} \vDash \Sigma$ iff $\mathfrak{C} \vDash \Sigma$. □

[15] Przełęcki (1969, 93) calls sentences that fulfill a special case of this criterion "strongly determined" (cf. Przełęcki 1974a, n. 2). Whence my choice of terminology.

That the criterion is relevant despite being exclusive is shown by a justification very attuned to the needs of the measuring scientist and questions of symmetry. Suppes (1959, 131) begins his justification with the idea that

> [a]n empirical hypothesis, or any statement in fact, which uses numerical quantities is empirically meaningful only if its truth value is invariant under the appropriate transformations of the numerical quantities involved.

The numerical quantities are functions, and transformations that lead only from one adequate function to another are appropriate (Suppes 1959, 132). To be adequate, a function has to fulfill the conditions of adequacy for the measurement it represents. Suppes (1959, 135) states the conditions for functions m representing mass measurement as

$$\begin{aligned} \Pi_{\text{mass}} := \{ & \forall x \forall y \, (x \precsim y \leftrightarrow mx \leq my), \\ & \forall x \forall y \, (m(x * y) = mx + my) \} \,, \end{aligned} \tag{6.1}$$

where '$\precsim$' stands for 'is at most as heavy as', '$*$' stands for physical combination, and x and y are silently understood to range over physical objects. Suppes (1959, 135) notes that "the functional composition of any similarity transformation φ with the function m yields a function $\varphi \circ m$ which also satisfies" Π_{mass}, where a similarity transformation in Suppes's sense is also called a positive linear transformation. Therefore, Suppes (1959, 138) suggests that

> a formula S [...] is empirically meaningful [...] if and only if S is satisfied in a model $\mathfrak{M}$ [...] when and only when it is satisfied in every model [...] related to $\mathfrak{M}$ by a similarity transformation.

To connect Suppes's criterion to Lewis's and thereby to Przełęcki's, note first that $\precsim$ and $*$ play the role of basic terms with some set of axioms $\Pi_{\mathscr{B}}$ (Suppes 1959, 135, n. 7), and m is the sole auxiliary term. Now let $\mathfrak{B}[m/\varphi \circ m^{\mathfrak{B}}]$ be the structure that $\mathfrak{B}$ becomes when m is interpreted by $\varphi \circ m^{\mathfrak{B}}$ instead of $m^{\mathfrak{B}}$. Suppes's criterion of adequacy can then be paraphrased like this:[16]

Definition 6.18 (Empirically meaningful statements about mass)**.** Assume the standard interpretation for arithmetical terms. Then a $\mathscr{V}$-sentence σ is *empirically meaningful* if and only if for any $\mathfrak{B} \vDash \Pi_{\mathscr{B}} \cup \Pi_{\text{mass}}$ and any $\mathfrak{C}$, if $\mathfrak{C} = \mathfrak{B}[m/\varphi \circ m^{\mathfrak{B}}]$ and φ is a positive linear transformation, then $\mathfrak{B} \vDash \sigma$ if and only if $\mathfrak{C} \vDash \sigma$.

Suppes justifies the demand that truth values have to be invariant under positive linear transformations on the grounds that all and only such transformations lead

[16]Przełęcki's paraphrase is slightly different, for one because he aims to prove its equivalence with definition 6.17, not definition 6.9, but also because his definition of $\mathscr{B}$-conservativeness is slightly less general (see appendix 6.11.1).

from one function m that fulfills Π_{mass} to another. This is basically what motivates strong $\mathscr{B}$-determinacy as well: Strongly $\mathscr{B}$-determined sentences are exactly those whose truth value is invariant under any transformation of models of Π that leaves their reduct to $\mathscr{B}$ invariant and the truth of Π invariant.

Claim 6.14. *Assume $\mathscr{B} = \{\precsim, *\}$, $\mathscr{A} = \{m\}$, $\Pi_{\mathscr{A}} = \Pi_{\text{mass}}$, and the standard interpretation for arithmetical terms (i. e., arithmetical terms are treated as logical constants). Then a $\mathscr{V}$-sentence σ is empirically meaningful according to definition 6.18 if and only if σ is about subject matter $\mathscr{B}$.*

Proof. Przełęcki (1974a, 349). □

In claim 6.14, the interpretations of '$+$' and '$\leq$' are assumed to be fixed by the standard interpretation of arithmetical terms. Przełęcki (1974a, 347–348) assumes that this is ensured by a semantic restriction on the possible structures. One could also ensure the standard interpretation with the usual axioms in second order logic, with '$+$' and '$\leq$' as auxiliary terms.

Suppes's conditions of adequacy determine admissible transformations for mass measurements, which in turn determine meaningful sentences about mass. Przełęcki's result shows that for these sentences, empirical meaningfulness can be defined equivalently without using admissible transformations. I now want to show that this is also possible for general sentences about measurements.

Essentially following Suppes and Zinnes (1963), Roberts and Franke (1976) define the general notion of meaningfulness just illustrated using the concepts of relational systems, measures, and scales. A *relational system* is a structure with p k_i-ary relations ($1 \leq i \leq p$) and q binary functions. A *measure* μ is defined as a homomorphism from one relational system $\mathfrak{E} = \langle |\mathfrak{E}|, \{\langle P_1, P_1^{\mathfrak{E}}\rangle, \ldots, \langle P_p, P_p^{\mathfrak{E}}\rangle\}, \{\langle \circ_1, \circ_1^{\mathfrak{E}}\rangle, \ldots, \langle \circ_q, \circ_q^{\mathfrak{E}}\rangle\}\rangle$, sometimes called '*empirical*', to another relational system $\mathfrak{F} = \langle |\mathfrak{F}|, \{\langle Q_1, Q_1^{\mathfrak{F}}\rangle, \ldots, \langle Q_p, Q_p^{\mathfrak{F}}\rangle\}, \{\langle *_1, *_1^{\mathfrak{F}}\rangle, \ldots, \langle *_q, *_q^{\mathfrak{F}}\rangle\}\rangle$, sometimes called '*formal*'. A *homomorphism* (an element of $\hom(\mathfrak{E}, \mathfrak{F})$) is a function $\mu : |\mathfrak{E}| \to |\mathfrak{F}|$ such that for all $a_1^{\mathfrak{E}}, \ldots, a_{k_i}^{\mathfrak{E}}, a^{\mathfrak{E}}, b^{\mathfrak{E}} \in |\mathfrak{E}|$ with $i = 1, \ldots, p$ and for all $j = 1, \ldots, q$ it holds that

$$P_i^{\mathfrak{E}}\left(a_1^{\mathfrak{E}}, \ldots, a_{k_i}^{\mathfrak{E}}\right) \text{ iff } Q_i^{\mathfrak{F}}\left(\mu(a_1^{\mathfrak{E}}), \ldots, \mu(a_{k_i}^{\mathfrak{E}})\right), \tag{6.2a}$$

$$\mu\left(a^{\mathfrak{E}} \circ_j^{\mathfrak{E}} b^{\mathfrak{E}}\right) = \mu(a^{\mathfrak{E}}) *_j^{\mathfrak{F}} \mu(b^{\mathfrak{E}}). \tag{6.2b}$$

The triple of an empirical relational system, a formal system, and a measure is then called a *scale*. Roberts and Franke (1976) argue that for questions of meaningfulness, the notion of an admissible transformation is (in my notation) best captured as follows:

> If $\langle \mathfrak{E}, \mathfrak{F}, \mu\rangle$ is a scale, then an *admissible transformation* ψ relative to $\mathfrak{E}$, $\mathfrak{F}$, and μ is any mapping of μ into a function $\psi(\mu) : |\mathfrak{E}| \to |\mathfrak{F}|$ such that $\psi(\mu)$ is also in $\hom(\mathfrak{E}, \mathfrak{F})$.

Their argument for this definition rests on the explication of 'meaningfulness' by Suppes and Zinnes (1963, 66), who suggest that a

> numerical statement is meaningful if and only if its truth (or falsity) is constant under admissible scale transformations of any of its numerical assignments[,]

where numerical assignments are measures.

The concept of a scale is defined by the relation between two structures. To capture it, like Suppes (1959) does, in a single structure $\mathfrak{A}$, one can define $\mathfrak{A}$ as having the structures $\mathfrak{E}$ and $\mathfrak{F}$ as relativized reducts. In this case, let $\mathfrak{A}$ have some domain $|\mathfrak{A}| \supseteq |\mathfrak{E}| \cup |\mathfrak{F}|$ and a vocabulary $\mathscr{A}$ containing the vocabularies $\mathscr{E}$ of $\mathfrak{E}$ and $\mathscr{F}$ of $\mathfrak{F}$, a function symbol f interpreted by the measurement μ, and two unary predicates E and F interpreted by $|\mathfrak{E}|$ and $|\mathfrak{F}|$, respectively. The relativized reduct $\mathfrak{A}|E_{\mathscr{E}} := \mathfrak{A}|E^{\mathfrak{A}}{}_{\mathscr{E}}$ is the substructure of $\mathfrak{A}|_{\mathscr{E}}$ whose domain is $E^{\mathfrak{A}} = |\mathfrak{E}|$. The relativization theorem then says that for every formula φ of $\mathscr{E}$ (or $\mathscr{F}$) and its relativization $\varphi^{(E)}$ (or $\varphi^{(F)}$, respectively) of $\mathscr{A}$, it holds that that $\mathfrak{E} \vDash \varphi$ ($\mathfrak{F} \vDash \varphi$) if and only if $\mathfrak{A} \vDash \varphi^{(E)}$ ($\mathfrak{A} \vDash \varphi^{(F)}$) (Hodges 1993, Theorem 5.1.1).

Now, let Π_{scale} determine the possible measurement scales, that is, the relativized reduct to F and $\mathscr{F}$ of every model of Π_{scale} is isomorphic to the formal structure $\mathfrak{F}$, and the class of relativized reducts to E and $\mathscr{E}$ of all models of Π_{scale} is the class of possible empirical structures. Since I am not assuming partial functions, but will need a substructure of $\mathfrak{A}$ with the domain $|\mathfrak{E}| \cup |\mathfrak{F}|$, define the extensions of the functions in $\mathfrak{E}$ to $|\mathfrak{A}|$ so that their restrictions to $|\mathfrak{F}|$ are full functions, and analogously for the functions in $\mathfrak{F}$. This is nothing but a technically convenient convention—since the values of a function $g^{\mathfrak{E}}$ can be freely chosen over $|\mathfrak{F}|$ (and *vice versa*), one can always choose the value of $g^{\mathfrak{E}}$ to be in $|\mathfrak{F}|$ whenever its arguments are in $|\mathfrak{F}|$ (and *vice versa*).

Restricting the domain of $\mu = f^{\mathfrak{A}}$ to $|\mathfrak{E}| = E^{\mathfrak{A}}$ results in a measure from $\mathfrak{E}$ to $\mathfrak{F}$ if and only if, first, the range of μ is $|\mathfrak{F}| = F^{\mathfrak{A}}$, and second, μ fulfills the conditions of adequacy (6.2). This is the case if and only if $\mathfrak{A} \vDash \Pi_{\text{adeq}}$ with

$$\begin{aligned}\Pi_{\text{adeq}} := \{\forall a(Ea \rightarrow Ffa)\} \cup \\ &\bigcup_{i=1}^{p} \left\{\forall a_1 \ldots \forall a_{k_i} \left[Ea_1 \wedge \cdots \wedge Ea_{k_i} \rightarrow \left(P_i a_1 \ldots a_{k_i} \leftrightarrow Q_i f a_1 \ldots f a_{k_i}\right)\right]\right\} \\ &\cup \bigcup_{j=1}^{q} \left\{\forall a \forall b \left(Ea \wedge Eb \rightarrow f(a \circ_j b) = fa *_j fb\right)\right\} .\end{aligned} \tag{6.3}$$

Π_{adeq} is a generalization of the conditions of adequacy Π_{mass} for mass measurements, with the relativization of the quantifiers to physical objects made explicit.

Again only to avoid partial functions, assume in the following that $f^{\mathfrak{A}}$ maps any element of $|\mathfrak{F}|$ to an element of $|\mathfrak{E}| \cup |\mathfrak{F}|$. All in all, $\mathfrak{A}$ is determined by a set Π_{scale} that entails Π_{adeq}, by the restrictions on $\mathfrak{F}$ and possible empirical structures, and by the additional restriction on the extensions of the functions in $\mathfrak{E}$ and $\mathfrak{F}$ discussed above. Note that $|\mathfrak{A}|$ can be a proper superset of $|\mathfrak{E}| \cup |\mathfrak{F}|$, and $\mathscr{A}$ can be a proper superset of $\mathscr{E} \cup \mathscr{F}$. This can ease the formalization of the relations and functions in $\mathfrak{E}$ and $\mathfrak{F}$ by allowing, for example, the language and objects of set theory.

By construction of Π_{scale}, any $\mathfrak{A} \models \Pi_{\text{scale}}$ fulfills the admissibility conditions for the relativized reduct $\mathfrak{A}|EF_{\mathscr{E}\mathscr{F}f}$ to $\mathscr{E}\mathscr{F}f := \mathscr{E} \cup \mathscr{F} \cup \{f\}$ and $EF := \lambda x(Ex \vee Fx)$ (Hodges 1993, 203), so that $\mathfrak{A}|EF_{\mathscr{E}\mathscr{F}f}$ exists. Because of the relativization theorem, $\mathfrak{A}|EF_{\mathscr{E}\mathscr{F}f} \models \Sigma$ if and only if $\mathfrak{A} \models \Sigma^{(EF)}$ for any $\mathscr{E}\mathscr{F}f$-sentence Σ.[17] Defining $\Sigma^{\mathfrak{A}}$ to be the set theoretic conditions on the extensions of the terms in Σ that have to hold for Σ to be true in $\mathfrak{A}$, one arrives at an equivalence between the truth of sentences in $\mathfrak{A}$ and the truth of set theoretic conditions for the scale $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$, where by construction $\mathfrak{A}|E_{\mathscr{E}} = \mathfrak{E}$ and $\mathfrak{A}|F_{\mathscr{F}} = \mathfrak{F}$: For any set Σ of $\mathscr{E}\mathscr{F}f$-sentences, $\mathfrak{A} \models \Sigma^{(EF)}$ if and only if $\Sigma^{\mathfrak{A}|EF_{\mathscr{E}\mathscr{F}f}}$ is true for the scale $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$.

The definition of admissible transformation argued for by Roberts and Franke (1976) can now be paraphrased as follows:

Definition 6.19. If $\mathfrak{A} \models \Pi_{\text{scale}}$, then an *admissible transformation φ relative to* $\mathfrak{A}$ is any mapping of $f^{\mathfrak{A}}$ into a function $\varphi(f^{\mathfrak{A}})$ such that $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \models \Pi_{\text{adeq}}$.

The explication of 'meaningfulness' by Suppes and Zinnes (1963) assumes the two concepts of a scale and an admissible transformation, and like the definition of meaningfulness for mass measurements by Suppes (1959), demands that a statement about $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$ be invariant under the admissible transformations of any adequate measure. This can be generalized to all $\mathscr{E}\mathscr{F}f$-sentences (rather than only their relativizations to EF):

Definition 6.20. A set Σ of $\mathscr{E}\mathscr{F}f$-sentences is *strongly invariant* if and only if for any $\mathfrak{A} \models \Pi_{\text{scale}}$ and any admissible transformation φ relative to $\mathfrak{A}$, it holds that $\mathfrak{A} \models \Sigma$ iff $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \models \Sigma$.

The restriction of definition 6.20 to relativizations of $\mathscr{E}\mathscr{F}f$-sentences to EF is indeed equivalent to the original definition by Suppes and Zinnes (1963):

Claim 6.15. *A set $\Sigma^{(EF)}$ of $\mathscr{E}\mathscr{F}f$-sentences is strongly invariant if and only if for any scale $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$, $\Sigma^{\mathfrak{A}|EF_{\mathscr{E}\mathscr{F}f}}$ with $\mathfrak{A} \models \Pi_{\text{scale}}$, $\mathfrak{A}|E_{\mathscr{E}} = \mathfrak{E}$, $\mathfrak{A}|F_{\mathscr{F}} = \mathfrak{F}$, and $f^{\mathfrak{A}}|_{|\mathfrak{E}|} = \mu$ as constructed above is meaningful according to Suppes and Zinnes (1963).*

Proof. '$\Rightarrow$': Assume that $\Sigma^{(EF)}$ is strongly invariant, that $\Sigma^{\mathfrak{A}|EF_{\mathscr{E}\mathscr{F}f}}$ is true for $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$, and that ψ is an admissible transformation for $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$. Then, by

[17] A reminder: $\Sigma^{(EF)} := \{\sigma^{(EF)} \mid \sigma \in \Sigma\}$, the set of the relativizations of the elements of Σ to EF.

construction of $\mathfrak{A}$, $\mathfrak{A} \models \Sigma^{(EF)}$. Now, ψ is an admissible transformation for $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$ only if $\psi\left(f^{\mathfrak{A}}|_{|\mathfrak{E}|}\right)$ fulfills equations 6.3. And then some extension φ of ψ with $\varphi(f^{\mathfrak{A}})|_{|\mathfrak{E}|} = \psi\left(f^{\mathfrak{A}}|_{|\mathfrak{E}|}\right)$ is admissible relative to $\mathfrak{A}$, so that by assumption $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \models \Sigma^{(EF)}$. Now $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \models \Pi_{\text{scale}}$ and $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]|E_{\mathscr{E}} = \mathfrak{A}|E_{\mathscr{E}} = \mathfrak{E}$, $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]|F_{\mathscr{F}} = \mathfrak{A}|F_{\mathscr{F}} = \mathfrak{F}$, and $f^{\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]}|_{|\mathfrak{E}|} = \varphi(f^{\mathfrak{A}})|_{|\mathfrak{E}|} = \psi(f^{\mathfrak{A}}|_{|\mathfrak{E}|}) = \psi(\mu)$, so that $\Sigma^{\mathfrak{A}|EF_{\mathscr{E}\mathscr{F}f}}$ is true for $\langle \mathfrak{E}, \mathfrak{F}, \psi(\mu) \rangle$. By an analogous reasoning, $\Sigma^{\mathfrak{A}|EF_{\mathscr{E}\mathscr{F}f}}$ is false for $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$ only if it is false for $\langle \mathfrak{E}, \mathfrak{F}, \psi(\mu) \rangle$, where ψ is any admissible transformation for $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$.

'$\Leftarrow$': Assume that $\Sigma^{\mathfrak{A}|EF_{\mathscr{E}\mathscr{F}f}}$ is meaningful for any scale, that $\mathfrak{A} \models \Sigma^{(EF)}$, and that φ is an admissible transformation relative to $\mathfrak{A}$. By construction, if $\mathfrak{A} \models \Sigma^{(EF)}$, then $\Sigma^{\mathfrak{A}|EF_{\mathscr{E}\mathscr{F}f}}$ is true for scale $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$ and, by assumption, for any scale $\langle \mathfrak{E}, \mathfrak{F}, \psi(\mu) \rangle$ with admissible ψ. Now, φ is admissible relative to $\mathfrak{A}$ only if $\varphi(f^{\mathfrak{A}})|_{|\mathfrak{E}|}$ fulfills equations 6.3 and thus ψ with $\psi(f^{\mathfrak{A}}|_{|\mathfrak{E}|}) := \varphi(f^{\mathfrak{A}})|_{|\mathfrak{E}|}$ is admissible for $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$. Thus $\Sigma^{\mathfrak{A}|EF_{\mathscr{E}\mathscr{F}f}}$ is true for $\langle \mathfrak{E}, \mathfrak{F}, \psi(\mu) \rangle$. Now $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \models \Pi_{\text{scale}}$ and $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]|E_{\mathscr{E}} = \mathfrak{A}|E_{\mathscr{E}} = \mathfrak{E}$, $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]|F_{\mathscr{F}} = \mathfrak{A}|F_{\mathscr{F}} = \mathfrak{F}$, and $f^{\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]}|_{|\mathfrak{E}|} = \varphi(f^{\mathfrak{A}})|_{|\mathfrak{E}|} = \psi(f^{\mathfrak{A}}|_{|\mathfrak{E}|}) = \psi(\mu)$, so that $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \models \Sigma^{(EF)}$. By analogous reasoning, $\mathfrak{A} \not\models \Sigma^{(EF)}$ only if $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \not\models \Sigma^{(EF)}$, where φ is any admissible transformation for $\mathfrak{A}$. □

Relative to the analytic sentences Π, every $\mathscr{B}$-structure determines a set of admissible transformations. Strong invariance universally quantifies on all possible $\mathscr{B}$-structures, and is therefore determined by the analytic sentences alone. Strong invariance is thus a symmetry relative to the analytic sentences. Now Przełęcki's result can be generalized:

Claim 6.16. *Assume $\mathscr{B} = \mathscr{E}$, $\mathscr{F} \cup \{f\} \subseteq \mathscr{A}$, and $\Pi = \Pi_{\text{scale}}$. Then a set Σ of $\mathscr{E}\mathscr{F}f$-sentences is strongly invariant if and only if Σ is about subject matter $\mathscr{B}$.*

Proof. '$\Leftarrow$': First, note that for any $\mathfrak{A} \models \Pi_{\text{scale}}$ and any admissible transformation φ relative to $\mathfrak{A}$, there is some $\mathfrak{B} \models \Pi_{\text{scale}}$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$ such that $\mathfrak{B}|_{\mathscr{B}\cup\{f\}} = \mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]|_{\mathscr{B}\cup\{f\}}$ (*), which can be shown as follows: By definition 6.19, $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \models \Pi_{\text{scale}}$, and since $f \notin \mathscr{B}$, $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$. Then choose $\mathfrak{B} := \mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]$.

Now assume that for any $\mathfrak{A}, \mathfrak{B} \models \Pi_{\text{scale}}$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$, $\mathfrak{A} \models \Sigma$ iff $\mathfrak{B} \models \Sigma$. Let $\mathfrak{C} \models \Pi_{\text{scale}}$ and φ be admissible relative to $\mathfrak{C}$. Then, because of (*), there is some $\mathfrak{D} \models \Pi_{\text{scale}}$ with $\mathfrak{D}|_{\mathscr{B}} = \mathfrak{C}|_{\mathscr{B}}$ and $\mathfrak{D} = \mathfrak{C}[f/\varphi(f^{\mathfrak{C}})]$. Therefore, by assumption, $\mathfrak{C}[f/\varphi(f^{\mathfrak{C}})] \models \Sigma$ iff $\mathfrak{C} \models \Sigma$.

'$\Rightarrow$': First, note that for any $\mathfrak{A} \models \Pi_{\text{scale}}, \mathfrak{B} \models \Pi_{\text{scale}}$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$, there is some transformation φ admissible relative to $\mathfrak{A}$ such that $\mathfrak{B}|_{\mathscr{E}\mathscr{F}f}$ is isomorphic to $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]|_{\mathscr{E}\mathscr{F}f}$ (**), which can be shown as follows: Since, by construction of Π_{scale}, $\mathfrak{A}|_{\mathscr{F}}$ is isomorphic to $\mathfrak{B}|_{\mathscr{F}}$, assume without loss of generality that $\mathfrak{A}|_{\mathscr{F}} =$

$\mathfrak{B}|_{\mathscr{F}}$. Now choose φ so that $\varphi(\mu) = f^{\mathfrak{B}}$ for every function μ. Then φ is admissible relative to $\mathfrak{A}$ because $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] = \mathfrak{B} \models \Pi_{\text{adeq}}$, and since $\mathscr{B} = \mathscr{E}$, $\mathfrak{B}|_{\mathscr{E}\mathscr{F}f} \simeq \mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]|_{\mathscr{E}\mathscr{F}f}$.

Now assume that $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$ and $\mathfrak{B} \models \Pi_{\text{scale}}$. By ($**$), there is some admissible φ such that $\mathfrak{B}|_{\mathscr{E}\mathscr{F}f} \simeq \mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]|_{\mathscr{E}\mathscr{F}f}$. Therefore, if Σ is strongly invariant, $\mathfrak{B} \models \Sigma$ iff $\mathfrak{A} \models \Sigma$. □

Like strong invariance, strong $\mathscr{B}$-determinacy is thus a symmetry relative to the analytic sentences Π.

To arrive at a syntactic version of strong $\mathscr{B}$-determinacy, it is helpful to look at the line of reasoning that led to definition 6.9. There, a set of sentences is taken to be about subject matter $\mathscr{B}$ if and only if its truth value is identical in any two worlds that are exactly alike so far as subject matter $\mathscr{B}$ is concerned. In connection with definitions 2.2 and 2.2, I described the difference between semantic and syntactic criteria as that between isomorphism and syntactic equivalence of $\mathscr{B}$-structures, which is borne out by claim 6.3 for falsifiability and claim 6.10 for verifiability. To arrive at an analogous relation for strong $\mathscr{B}$-determinacy, I thus suggest

Definition 6.21. A *set Γ of $\mathscr{V}$-sentences determines* a set Σ of $\mathscr{V}$-sentences if and only if $\Gamma \cup \Pi \models \Sigma$ or $\Gamma \cup \Sigma \cup \Pi \models \bot$.

Definition 6.22. A set Ω of $\mathscr{B}$-sentences is *maximal* if and only if for every $\mathscr{B}$-sentence β, $\Omega \cup \Pi \models \beta$ or $\Omega \cup \Pi \models \neg\beta$.

Then one can formulate

Definition 6.23. A set Σ of $\mathscr{V}$-sentences is *strongly syntactically $\mathscr{B}$-determined* if and only if it is determined by every possible and maximal set of $\mathscr{B}$-sentences.

As in the case of falsifiability and verifiability, the difference between syntactic and semantic strong $\mathscr{B}$-determinacy is that between isomorphism and syntactical equivalence:

Claim 6.17. *A set Σ of $\mathscr{V}$-sentences is strongly syntactically $\mathscr{B}$-determined if and only if for any $\mathscr{V}$-structures $\mathfrak{A}, \mathfrak{B} \models \Pi$ with $\mathfrak{A}|_{\mathscr{B}} \equiv \mathfrak{B}|_{\mathscr{B}}$, it holds that $\mathfrak{A} \models \Sigma$ iff $\mathfrak{B} \models \Sigma$.*

Proof. '$\Rightarrow$': Let $\mathfrak{A}, \mathfrak{B} \models \Pi$ and $\mathfrak{A}|_{\mathscr{B}} \equiv \mathfrak{B}|_{\mathscr{B}}$. Then $\mathfrak{A}, \mathfrak{B} \models \text{Th}(\mathfrak{B}|_{\mathscr{B}}) =: \Omega$. It is straightforward to show that Ω is maximal and possible. Thus, by assumption, $\Omega \cup \Pi \models \Sigma$ or $\Omega \cup \Sigma \cup \Pi \models \bot$. Thus $\mathfrak{A} \models \Sigma$ if and only if $\mathfrak{B} \models \Sigma$.

'$\Leftarrow$': Assume Ω is possible and maximal. Then for any $\mathfrak{A}, \mathfrak{B} \models \Omega \cup \Pi$, $\mathfrak{A}|_{\mathscr{B}} \equiv \mathfrak{B}|_{\mathscr{B}}$. Therefore, by assumption, $\mathfrak{A} \models \Sigma$ iff $\mathfrak{B} \models \Sigma$ and thus either all $\mathfrak{A} \models \Omega \cup \Pi$ are models of Σ or none is. Thus $\Omega \cup \Pi \models \Sigma$ or $\Omega \cup \Sigma \cup \Pi \models \bot$. □

This entails

Claim 6.18. *If a set Σ of $\mathscr{V}$-sentences is strongly syntactically $\mathscr{B}$-determined, then Σ is strongly semantically $\mathscr{B}$-determined.*

Proof. From claims 6.13 and 6.17 because $\mathfrak{A}|_{\mathscr{B}} = \mathfrak{B}|_{\mathscr{B}}$ only if $\mathfrak{A}|_{\mathscr{B}} \equiv \mathfrak{B}|_{\mathscr{B}}$. □

The relation of strong $\mathscr{B}$-determinacy to falsifiability and verifiability is given by

Claim 6.19. *Let Σ be a set of strongly syntactically (semantically) $\mathscr{B}$-determined $\mathscr{V}$-sentences. Then Σ is syntactically (semantically) falsifiable/verifiable if and only if Σ is not analytically true/false.*

Proof. '$\Rightarrow$': Immediate.

'$\Leftarrow$': Assume $\Pi \not\models \Sigma$. Then for some $\mathfrak{A}$, $\mathfrak{A} \models \Pi$ and $\mathfrak{A} \not\models \Sigma$. If Σ is syntactically $\mathscr{B}$-determined, then $\mathrm{Th}(\mathfrak{A}|_{\mathscr{B}}) \cup \Sigma \cup \Pi \models \bot$ because $\mathrm{Th}(\mathfrak{A}|_{\mathscr{B}}) \cup \Pi \not\models \Sigma$. Thus $\mathrm{Th}(\mathfrak{A}|_{\mathscr{B}})$ falsifies Σ. If Σ is semantically $\mathscr{B}$-determined, then for all $\mathfrak{B} \models \Pi$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$, $\mathfrak{B} \not\models \Sigma$. Thus $\mathfrak{A}|_{\mathscr{B}}$ falsifies Σ.

The proofs for verifiability are analogous. □

6.6 Weak $\mathscr{B}$-determinacy

Since Przełęcki considers strong semantic $\mathscr{B}$-determinacy too exclusive, he suggests a straightforward weakening of definition 6.17:

Definition 6.24. A set Σ of $\mathscr{V}$-sentences is *weakly semantically $\mathscr{B}$-determined* if and only if it is determined by a possible $\mathscr{B}$-structure.

The motivation for the criterion is clear: The truth value of a strongly semantically $\mathscr{B}$-determined sentence is fixed for any $\mathscr{B}$-structure, but there are many sentences whose truth values are fixed only for some structures. Przełęcki considers this enough to be empirically significant.

A connection to ordinary language can be found again starting from Lewis's notion of sentences about subject matter $\mathscr{B}$. The idea to take a sentence to be partially about subject matter $\mathscr{B}$ if it partially supervenes on subject matter $\mathscr{B}$ leads Lewis (1988b, §X) to a probabilistic notion of empirical significance, but I want to argue that his justification more plausibly leads to weak semantic $\mathscr{B}$-determinacy. Lewis (1988b, 149) argues that

> a statement is partly about a subject matter iff its truth value partially supervenes, in a suitably non-trivial way, on that subject matter. Let us say that the truth value of a statement *supervenes* on subject matter M *within* class X of worlds iff, whenever two worlds in X are M-equivalent, they give the statement the same truth value. [...] Supervenience within a [subclass X of all] worlds is partial supervenience.

Lewis needs the restriction to "suitable partial supervenience" to avoid trivialization, because if, say, it is possible for X to contain only one world, then any sentence σ partially supervenes on any M. To exclude such classes, Lewis demands that X contain a majority of the worlds in which σ is true and a majority of the worlds in which σ is false. To explicate the notion of 'majority' for worlds, he assumes that there is a suitable probability distribution over possible worlds and states that the condition is satisfied if and only if $\Pr(X \mid \sigma) > \frac{1}{2}$ and $\Pr(X \mid \neg\sigma) > \frac{1}{2}$. Under some additional assumptions, the notion of partial supervenience that results is equivalent to the standard probabilistic criterion that σ is empirically significant iff $\Pr(\sigma \mid \beta) \neq \Pr(\sigma)$ for some basic sentence β (see definition 8.2).

Lewis's notion of partial supervenience need not lead to a probabilistic criterion of empirical significance. He introduces the majority condition to avoid trivialization, but there is nothing in the concept of 'partial supervenience' itself that suggests the supervenience has to hold for the majority of σ worlds and $\neg\sigma$ worlds. It is much more in keeping with the goal of explicating *empirical* significance to place only empirical restrictions on X. The minimal requirement is thus that X be closed under empirical equivalence, such that for any world that is in X, every $\mathscr{B}$-equivalent world is also in X. This condition already avoids trivialization, does not require a probability distribution over possible world, and does not lead to complications when statements are taken to be expressed by *sets* of sentences, which are not generally easily negated. To partially supervene on subject matter $\mathscr{B}$, Σ thus has to be assigned the same truth value by all members of a set X closed under empirical equivalence.

Definition 6.25. A set Σ of $\mathscr{V}$-sentences *partly supervenes on subject matter* $\mathscr{B}$ if and only if there is some non-empty set X of possible $\mathscr{V}$-structures such that for any $\mathfrak{A} \in X$, all $\mathfrak{B} \vDash \Pi$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$ are in X, and for any $\mathfrak{A}, \mathfrak{B} \in X$ with $\mathfrak{A}|_{\mathscr{B}} = \mathfrak{B}|_{\mathscr{B}}$, $\mathfrak{B} \vDash \Sigma$ iff $\mathfrak{A} \vDash \Sigma$.

As announced, this is the same as weak semantic $\mathscr{B}$-determinacy:

Claim 6.20. *A set Σ of $\mathscr{V}$-sentences is weakly semantically $\mathscr{B}$-determined if and only if Σ partly supervenes on subject matter $\mathscr{B}$.*

Proof. If $\mathfrak{A}_{\mathscr{B}}$ is possible and determines Σ, choose X as the set of possible expansions of $\mathfrak{A}_{\mathscr{B}}$. If Σ partly supervenes on subject matter $\mathscr{B}$, then any $\mathfrak{A}|_{\mathscr{B}}$ with $\mathfrak{A} \in X$ is possible and determines Σ. □

Przełęcki (1974a, 347) points out that under his assumption that $\Pi_{\mathscr{A}}$ is $\mathscr{B}$-conservative with respect to $\Pi_{\mathscr{B}}$, definition 6.24 has a very conspicuous formulation: A $\mathscr{V}$-sentence σ is weakly semantically $\mathscr{B}$-determined if and only if $\{\sigma\} \cup \Pi_{\mathscr{A}}$ or $\{\neg\sigma\} \cup \Pi_{\mathscr{A}}$ is semantically $\mathscr{B}$-creative with respect to $\Pi_{\mathscr{B}}$. However, because not all sets of sentences are easily negated, this formulation is neither as general nor as conspicuous as

Claim 6.21. *A set Σ of $\mathscr{V}$-sentences is weakly semantically $\mathscr{B}$-determined if and only if Σ is semantically verifiable or semantically falsifiable.*

Proof. Assume Σ is semantically verifiable or semantically falsifiable. This holds iff there is a possible $\mathfrak{A}_{\mathscr{B}}$ such that Σ is true in all structures $\mathfrak{B} \models \Pi$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$ or false in all of them, that is, Σ has the same truth value in all of these structures. This is equivalent to Σ being weakly semantically $\mathscr{B}$-determined. □

With claims 6.5 and 6.8, this means that a sentence σ is weakly semantically $\mathscr{B}$-determined if and only if σ or $\neg\sigma$ is semantically $\mathscr{B}$-creative with respect to Π. This is Przełęcki's claim, reformulated using the generalized definition of $\mathscr{B}$-conservativeness (see appendix 6.11.1).

Considerations analogous to those leading to the definition of strong syntactic $\mathscr{B}$-determinacy lead to

Definition 6.26. A set Σ of $\mathscr{V}$-sentences is *weakly syntactically $\mathscr{B}$-determined* if and only if it is determined by some possible and maximal set of $\mathscr{B}$-sentences.

This definition relates to that of weak semantic $\mathscr{B}$-determinacy in the usual way:

Claim 6.22. *A set Σ of $\mathscr{V}$-sentences is weakly syntactically $\mathscr{B}$-determined if and only if there is some $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ such that for all structures $\mathfrak{B}, \mathfrak{C} \models \Pi$ with $\mathfrak{B}|_{\mathscr{B}} \equiv \mathfrak{C}|_{\mathscr{B}} \equiv \mathfrak{A}_{\mathscr{B}}$, it holds that $\mathfrak{B} \models \Sigma$ iff $\mathfrak{C} \models \Sigma$.*

Proof. '⇒': Choose $\Omega := \mathrm{Th}(\mathfrak{A}_{\mathscr{B}})$ and proceed as in the proof of claim 6.17.
'⇐': Choose some $\mathfrak{A}_{\mathscr{B}} \models \Omega$ and proceed as in the proof of claim 6.17 □

And analogously to the semantic case, the following holds:

Claim 6.23. *A set Σ of $\mathscr{V}$-sentences is weakly syntactically $\mathscr{B}$-determined if and only if Σ is syntactically verifiable or syntactically falsifiable.*

Proof. '⇒': Immediate.
'⇐': If Ω verifies or falsifies Σ, Ω is possible. Thus $\Omega \cup \Pi$ can be extended to a possible and maximal set of $\mathscr{B}$-sentences. □

As the disjunction of falsifiability and verifiability, weak syntactic $\mathscr{B}$-determinacy has occurred often in the history of philosophy, albeit repeatedly sailing under false colors. The illicit reflagging often occurred with the help of the prediction criterion of confirmation discussed in connection with Ayer's trivial definition 6.5 of indirect verifiability. For example, Carnap (1936, 435) calls the confirmation of a sentence S "directly reducible to a class C of sentences" if "S is a consequence of a finite subclass of C" (complete reducibility of confirmation) or "if the confirmation of S is not completely reducible to that of C but if there is an infinite subclass C' of C such that the sentences of C' are mutually independent and are consequences of S" (direct incomplete reducibility of confirmation). This definition is the first in a long chain that eventually leads to the requirement of confirmability, which "suffices as a formulation of the principle of empiricism" (Carnap 1937, 35). Carnap's terminology makes it clear that, like Ayer, he assumes the prediction criterion of confirmation (see also Gemes 1998b, §1.4).

Following the chain of definitions is tedious,[18] but significantly simplified when taking into account that it becomes trivial with the next link: Carnap (1936, 435) calls the confirmation of S

> *reducible* to that of [a class of sentences] C, if there is a finite series of classes $C_1, C_2, \dots, C_n$ such that the relation of directly reducible confirmation subsists 1) between S and C_1, 2) between every sentence of C_i and C_{i+1} ($i = 1$ to $n-1$), and 3) between every sentence of C_n and C.

It is then simple to prove

Claim 6.24. *If the class C of sentences allows the direct incomplete reducibility of at least one sentences γ, then the confirmation of every sentence σ is reducible to C.*

Proof. For any sentence σ, if γ is directly incompletely reducible to C, so is $\gamma \wedge \sigma$, which can therefore be in C_1. Then σ can be completely reduced to $C_1 := \{\gamma \wedge \sigma\}$ because $\{\gamma \wedge \sigma\} \vDash \sigma$ and $\{\gamma \wedge \sigma\}$ is a finite subset of itself. Thus the confirmation of σ is directly reducible to C_1, whose confirmation is directly reducible to C, and therefore the confirmation of σ is reducible to C. □

If a language contains infinitely many constants $\{c_i \mid i \in I\}$ for points in space-time, the sentence 'It will always be everywhere cold' is an incompletely directly reducible sentence γ, since the temperature at each point in space-time is logically independent from the temperature at any other and thus γ entails the infinite set of logically independent sentences $\Omega^* := \{\ulcorner\text{It is cold at } c_i\urcorner \mid i \in I\}$.

Since the reducibility of confirmation to a class of sentences is trivial, all other definitions that build on it collapse, too: The confirmation of a sentence S is reducible to a class of $\mathscr{B}$-predicates if the confirmation of S "is reducible [...] to a not contravalid sub-class of the class which contains the full sentences of the predicates of [$\mathscr{B}$] and the negations of these sentences" (Carnap 1936, 435–436); call such a sub-class a *confirmation class*. Full sentences are atomic sentences, and a contravalid sentence is incompatible with the laws of nature (Carnap 1936, 432–434).[19] Because of claim 6.24, if some confirmation class Ω allows the direct incomplete reducibility of at least one sentence γ, the confirmation of any sentence σ is reducible to Ω. (In the above example, Ω^* is a confirmation class for γ if

[18] That this holds for most definitions in the article may explain why, as far as I know, no concept introduced in "Testability and Meaning" besides that of reduction sentences has been used since.

[19] I will discuss the relevance of contravalidity in §6.8.2.

$\{c_i \mid i \in I\} \cup \{\lambda x(\text{It is cold at } x)\} \subseteq \mathscr{B}$.) Thus the confirmation of any sentence σ is reducible to $\mathscr{B}$. In that case σ is also confirmable, because a "sentence S is called *confirmable* [...] if the confirmation of S is reducible [...] to that of a class of observable predicates" (Carnap 1936, 456). Since nothing was assumed about σ, the principle of empiricism is then met by any sentence whatsoever.

The triviality of Carnap's general notion of reducibility leaves the direct reducibility of S to full sentences of $\mathscr{B}$ as the concept of confirmability, and this is just the disjunction of falsifiability and verifiability restricted to the class of atomic $\mathscr{B}$-sentences and their negations.

As shown above, Ayer's only non-trivial criterion of empirical significance is essentially equivalent to falsifiability. But in his first informal description of empirical significance, falsifiability and verifiability are on a par. Ayer (1936, 35) writes:

> We say that a sentence is factually significant to any given person, if, and only if, he knows how to verify the proposition which it purports to express—that is, if he knows what observations would lead him, under certain conditions, to accept the proposition as being true, or reject it as being false.

Within the vagaries of natural language, and as far as deductive inference is concerned, this is weak $\mathscr{B}$-determinacy. Since Ayer (1936, 37–38) rejects the idea that a sentence can be conclusively verified or falsified, he suggests his first definition of verifiability as a "weaker sense of verification". If, plausibly, this "weaker sense" is non-deductive, Ayer thus implicitly assumes the prediction criterion of confirmation.

In an early work, Carnap (1928b, 327–328) avoids the prediction criterion by leaving the concept of confirmation undefined. He writes:

> If a statement p expresses the content of an experience E, and if the statement q is either the same as p or can be derived from p and prior experiences, either through deductive or inductive arguments, then we say that q is "supported by" the experience E. [...] A statement p is said to have "factual content", if experiences which would support p or the contradictory of p are at least conceivable, and if their characteristics can be indicated.

Carnap's examples indicate that quantified $\mathscr{B}$-sentences describe conceivable experiences, so that in my terminology, Carnap considers a sentence to have factual content if and only if it is verifiable, falsifiable, confirmable or disconfirmable. In contexts that allow only deductive inferences, Carnap thus suggests to consider a sentence empirically significant if and only if it is weakly $\mathscr{B}$-determined.[20]

[20]In §8.7, I will argue that if confirmation is construed in terms of probabilities, the quote from Carnap amounts to an endorsement of a Bayesian criterion of empirical significance.

In a defense of criteria of empirical significance against the critique by Hempel (1950), Rynin (1957, 53) also suggests that a sentence be taken as significant if and only if it is either verifiable or falsifiable. For Rynin (1957, 51), this

> might constitute a kind of axiom of semantics, or at any rate some sort of adequacy requirement for a definition of 'meaningful statement'; I at any rate should consider it as self-evident that for a statement to be cognitively meaningful it must be possible for it to be true or false, that it have conditions of truth or falsity, hence necessary or sufficient truth conditions.

Of course, much in the quote hinges on these "conditions of truth or falsity". In his criterion, Rynin speaks of "ascertainable" truth conditions, and when discussing Hempel's critique of criteria of empirical significance, he notes that

> instead of talking of truth conditions [Hempel] prefers to formulate the verifiability principle in terms of relationships holding between the statements whose meaning is in question and what he calls "observation sentences", which I think it fair to treat as true statements affirming the occurrence of ascertainable states of affairs.—This difference in manner of formulation seems to me to be non-essential.

Apart from its restriction to molecular basic sentences (Hempel's "observation sentences"), Rynin's criterion is therefore equivalent to weak syntactic $\mathscr{B}$-determinacy.

Let me conclude this section with a puzzling observation that suggests that Hempel was not overly diligent in his dismissal of the search for a criterion of empirical significance. As mentioned above, Hempel (1965d, 122) considers the conjunction of falsifiability and verifiability as a criterion of empirical significance because it is symmetric under negation, but dismisses it as being too exclusive. Surprisingly, he discusses Rynin's article without mentioning Rynin's criterion, which is symmetric under negation and more inclusive than both falsifiability and verifiability. That is, he ignores a criterion that is symmetric under negation and more inclusive than the conjunction of verifiability and falsifiability (and even more inclusive than verifiability and falsifiability individually).

6.7 Import of the relations

Using the definitions above, one arrives at the notable number of equivalences and entailment relations shown in figures 6.1 and 6.2, with strong and weak $\mathscr{B}$-determinacy, falsifiability, and verifiability as the four major criteria of empirical significance. With this overview, it is now easy to consider the implications of the entailment relations and equivalences.

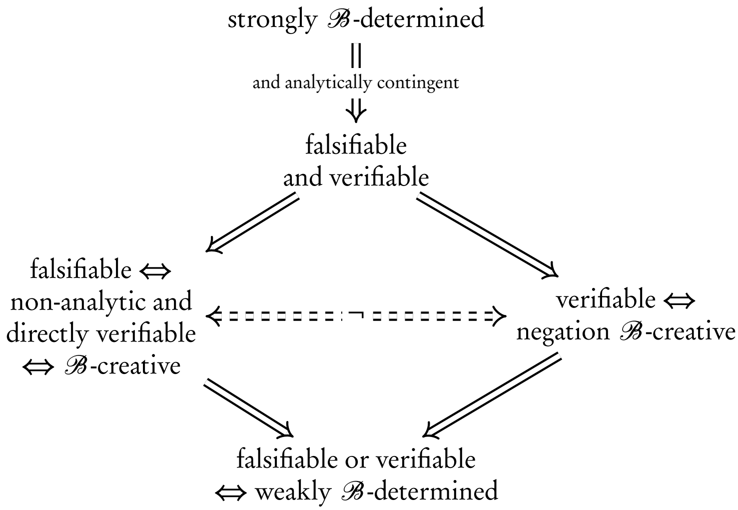

Figure 6.1: Relations between the syntactic definitions. The equivalence holds for direct verifiability and the negation of sets of sentences whenever the concepts are defined. A strongly $\mathscr{B}$-determined set of sentences is also weakly $\mathscr{B}$-determined even if not analytically contingent. Criteria of empirical significance typically also require that a set of sentences be analytically contingent.

6.7.1 Six comparative concepts of empirical significance

The entailment relations between the criteria show that there can be stronger and weaker criteria of empirical significance, and suggest that there may be criteria of comparative empirical significance. Hempel (1965c, 117) similarly states that "cognitive significance in a system is a matter of degree". He sees this as a reason to dispose of the concept altogether, and "instead of dichotomizing this array [of systems] into significant and non-significant systems" to compare systems of sentences by their precision, systematicity, simplicity, and level of confirmation. But this conclusion is unwarranted. For one, it is not clear what Hempel means when he states that cognitive significance "*is* a matter of degree". If cognitive significance is an explicatum, then it is whatever one decides it to be. If it is an explicandum, then deviating from it is not problematic. Perhaps Hempel intends to say that the best explicatum is one in which cognitive significance is a matter of degree, presumably because any dichotomy must be arbitrary. But this means that there is an explicatum, only it is not a classificatory one. This is nothing to be ashamed of, for Hempel (1952, §10) himself has argued that the move from a classificatory to a comparative concept is often a sign of an investigation's maturity (see also Hempel and Oppenheim 1936), as the explication of 'warm' by 'higher

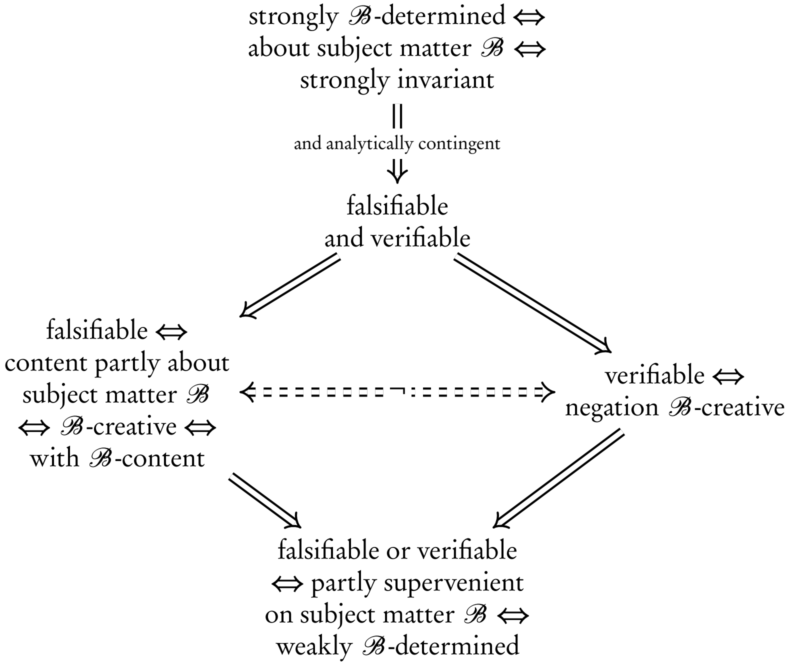

Figure 6.2: Relations between the semantic definitions. The equivalence holds for strong invariance, empirical content, and the negation of sets of sentences whenever the concepts are defined. A strongly $\mathscr{B}$-determined set of sentences is also weakly $\mathscr{B}$-determined even if not analytically contingent. Each of the nodes is entailed by its syntactic counterpart from figure 6.1. Criteria of empirical significance typically also demand that set of sentences be analytically contingent.

temperature than' illustrates (Carnap 1950b, §4, Hempel 1952, §10).[21]

As the split of strong and weak $\mathscr{B}$-determinacy into falsifiability and verifiability shows, a comparative explicatum for empirical significance will probably have to be partial, in that not all criteria can be compared with respect to their inclusiveness without further assumptions. Therefore I suggest

Definition 6.27. A set Σ of $\mathscr{V}$-sentences is *at least as syntactically (semantically) falsifiable/verifiable/$\mathscr{B}$-determined* as a set Γ of $\mathscr{V}$-sentences if and only if every possible set of $\mathscr{B}$-sentences (possible $\mathscr{B}$-structure) that falsifies/verifies/$\mathscr{B}$-determines Γ also falsifies/verifies/$\mathscr{B}$-determines Σ.

The partial order of the subset relation transfers to 'being at least as falsifiable/

[21] Indeed, Hempel (1950, 211) seems to take just this stance towards comparative criteria of empirical significance in an earlier work.

verifiable/$\mathscr{B}$-determined', in both its syntactic and its semantic guise, for each set Π of analytic sentences. 'At least as syntactically falsifiable' is called 'falsifiability of at least as high a degree' by Popper (1935, §33), who also notes that this order is partial (Popper 1935, §34).

There is a second reason why Hempel should not have dismissed the search for criteria of empirical significance so easily. For each set Π, each relation in definition 6.27 has natural greatest and, more importantly, least elements.

Claim 6.25. *A set Σ of $\mathscr{V}$-sentences is analytically false/analytically true/analytically false or analytically true if and only if Σ is at least as syntactically (semantically) falsifiable/syntactically (semantically) verifiable/syntactically $\mathscr{B}$-determined as any other set of $\mathscr{V}$-sentences.*

Proof. '$\Rightarrow$': Immediate.

'$\Leftarrow$': If Σ is not analytically false, it is not syntactically (semantically) at least as falsifiable as $\bot$. Analogously for verifiability and $\top$.

If Σ is neither analytically false nor analytically true, there are a structure $\mathfrak{A} \models \Pi \cup \Sigma$ and a structure $\mathfrak{B} \models \Pi$ with $\mathfrak{B} \not\models \Sigma$. Choose $\Gamma := \Omega := \mathrm{Th}(\mathfrak{A}|_{\mathscr{B}}) \cap \mathrm{Th}(\mathfrak{B}|_{\mathscr{B}})$. Then Ω determines Γ but not Σ. □

This shows that the comparative notions connect fruitfully to analyticity. Strong semantic $\mathscr{B}$-determinacy connects very straightforwardly to 'semantically more determinate than', because all and only sets of sentences semantically determined by every $\mathscr{B}$-structure are at least as semantically $\mathscr{B}$-determined as any other:

Claim 6.26. *A set Σ of $\mathscr{V}$-sentences is strongly semantically $\mathscr{B}$-determined if and only if Σ is at least as semantically $\mathscr{B}$-determined as any other set of $\mathscr{V}$-sentences.*

Falsifiability, verifiability, and weak $\mathscr{B}$-determinacy are immediately connected to their comparative counterparts:

Claim 6.27. *A set Σ of $\mathscr{V}$-sentences is not syntactically (semantically) falsifiable/verifiable/weakly $\mathscr{B}$-determined if and only if Σ is at most as syntactically (semantically) falsifiable/verifiable/$\mathscr{B}$-determined as any other $\mathscr{V}$-sentence.*

Proof. The claim holds for all criteria because only the empty set is a subset of every set. □

So the sentences that are not empirically significant according to the classical, classificatory criteria are the least elements of the criteria's comparative analogues. Therefore, even if Hempel is correct that empirical significance is a matter of degree, his conclusion that there cannot be an explicatum at all fails in two respects. First, empirical significance can be explicated by comparative concepts. Second, these comparative concepts have non-arbitrary least elements, so there is a natural way to dichotomize the array of sets of sentences into empirically significant and not empirically significant.

6.7.2 The justifications and implications of the criteria

Many of the circumscriptions of the explicandum of the criteria and the defense of the methodological presumptions in §5 and §6.1 provide arguments for the feasibility of a criterion of empirical significance. The relations between the criteria suggest that the criteria are already to a certain extent adequate.

For one, the equivalences of many of the criteria to falsifiability, verifiability, or $\mathscr{B}$-determinacy help to counter the charge of arbitrariness that Lewis (1988b, 127) and, presumably, Hempel (1965c, §4) have put forth. The equivalences suggest that the explicated notions are robust under a change of formalism from predicate logic to model theory to set theory, and a change of formulation within each formalism. This provides an argument analogous to (though less spectacular than) that for the successful explication of 'computability', in which the equivalence of different definitions is cited as evidence for their adequacy and sometimes for the truth of the Church-Turing thesis (Barker-Plummer 2011, Copeland 2008).

The equivalences also show that especially Lewis's charge, that many criteria of empirical significance have strayed too far from the explicandum, does not apply to the criteria discussed here: Strong semantic $\mathscr{B}$-determinacy is equivalent to aboutness, semantic falsifiability is equivalent to partial aboutness of content, and weak semantic $\mathscr{B}$-determinacy is arguably equivalent to partial supervenience, all of which are meant to capture the ordinary language notion of some claim being partially about some subject matter. And although verifiability is not equivalent to any of Lewis's criteria, it occurs with falsifiability in the disjunction that makes up weak $\mathscr{B}$-determinacy (claim 6.23). It is thus at least the link connecting two different notions of partial aboutness.

The close connection of the criteria to ordinary language may prompt another criticism: that the criteria are of little use in the sciences, the way the ordinary language notion of 'fish', which includes the likes of whales and dolphins, is of little use in biology (cf. Carnap 1950b, §3). If the sciences are taken to include mathematics, then the equivalence of falsifiability to $\mathscr{B}$-conservativeness already provides a rebuttal, for the notion of definition is essential in mathematics, and $\mathscr{B}$-conservativeness is essential for the notion of definition (cf. Belnap 1993). At least for sentences, claim 6.9 shows the relevance of verifiability by its connection to falsifiability through the negation of sentences, and claim 6.23 shows the relevance of weak $\mathscr{B}$-determinacy by its connection to both criteria through their disjunction. Lest one argue that it is only falsifiability that is really needed, I appeal to authority: Church (1949) only proves that every sentence *or its negation* is empirically significant according to Ayer's criterion of indirect verifiability. Assuming that Church's proof was the main reason for abandoning the criterion, this means that any criterion of empirical significance is too inclusive if it includes every sentence or its negation among the empirically significant sentences. Thus, because of claims 6.9 and 6.23, falsifiability would already be too inclusive if every sentence was falsifiable or verifiable, that is, weakly $\mathscr{B}$-determined.[22] Hence weak $\mathscr{B}$-determinacy is considered so closely related to falsifiability that the triviality of the former suffices as a reason to abandon the latter.

Without relying on the importance of mathematics, one can argue that definitions are similarly important in the natural sciences. Additionally, claim 6.14 and claims 6.15 and 6.16 show that at least strong $\mathscr{B}$-determinacy is important for the concepts of measurement because it generalizes strong invariance. Its close relation to weak $\mathscr{B}$-determinacy suggests that the latter criterion is important within measurement theory as well, in effect stating that a numerical statement is weakly $\mathscr{B}$-determinate if and only if its truth value is for some $\mathscr{B}$-structures invariant under the admissible transformations (cf. Przełęcki 1974a, 350). Similarly, one may demand that the statement be false or be true for all admissible transformations, thus arriving at a special case of falsifiability or verifiability in terms of admissible transformations and thus measurements.

Conversely, the equivalence to a special case of strong $\mathscr{B}$-determinacy protects strong invariance against the charge that it is *ad hoc*. There is always the possibility of being mislead by the special conditions of a context, in this case the features of measurements, but the equivalence shows that strong invariance is a special case of a much more generally motivated criterion.

In conclusion, the criteria are neither too far from the ordinary language notion, nor too far from the sciences. But there is also the more general charge of irrelevance. For even if a criterion is close to some intuitive notion and a generalization of some concept defined for scientific application, the intuitive notion and the scientific concept may be applicable in their domain but irrelevant. Such a charge becomes more difficult to sustain the more concepts rely on or connect to the criterion. The relations shown in figures 6.1 and 6.2 suggest what would be irrelevant as well if the criteria discussed here were irrelevant, and it is doubtful that the notion of meaningfulness in measurement, the notion of empirical content as explicated by the Ramsey sentence, and the notions of aboutness and partial aboutness are all of no use.

Finally, there is the problem of past failures. I have already noted that not all criteria have been shown to be trivial, and the equivalences between the criteria discussed here make it easy to show that none of the criteria listed in figures 6.1 and 6.2 is trivial: Assume $\mathscr{B} = \{B, b\}$, $\mathscr{A} = \{A_1, A_2\}$, and $\Pi = \{\forall x[Bx \leftrightarrow \neg A_1 x]\}$. Then $A_1 b$ is strongly syntactically $\mathscr{B}$-determined and analytically contingent, and thus empirically significant according to all the criteria; and $A_2 b$ is not weakly semantically $\mathscr{B}$-determined, and thus not empirically significant according to any of the criteria. It is also notable that none of the criteria are amendments of Ayer's criteria (direct verifiability, definition 6.4, being only the first half of Ayer's second criterion, definition 6.5 being the second half). Hence it is not utterly surprising that the induction on past failures itself fails when applied to the criteria discussed

[22] As I will show below, such a proof is impossible.

here.

The equivalences also provide a positive justification rather than a defense, because now the arguments in favor of each individual formulation turn out to be arguments for the same criterion. Thus Przełęcki's general argument and Suppes's measure theory-specific argument already lead to strong $\mathscr{B}$-determinacy, and Lewis's argument for aboutness adds additional support. Lewis's analysis of the term 'partly' also provides a justification for the differences between the major criteria: The underlying explicandum is ambiguous because 'partly about subject matter $\mathscr{B}$' is ambiguous. Falsifiability is accordingly supported by one specific disambiguation of 'partly about', but also by Ayer's and Popper's arguments, and finally by the arguments in favor of the Ramsey sentence as explication of 'empirical content'. My modification of Lewis's analysis of partial supervenience and Przełęcki's argument for weak $\mathscr{B}$-determinacy also support the same criterion. Verifiability, while historically not often defended by itself, again receives some justification through its role in weak $\mathscr{B}$-determinacy.

Furthermore, the relations between the criteria suggest that the criteria fulfill Carnap's desiderata for explications (§5.1). They are certainly more precise than the phrase 'empirically meaningful', and some of the formulations are fairly simple—at least, they are not "page-long", as Lewis (1988a, 127) feared. In fact, the equivalences allow the application of different formulations according to expedience. The equivalences to Lewis's ordinary language notions[23] suggest that the criteria are similar to their respective explicanda "in such a way that, in most cases in which the explicandum has been used, the explicatum can be used".[24] This leaves the central demand for fruitfulness, that is, for one, Hempel's demand for a comprehensive and sound theoretical system. The relations between the comparative and classificatory notions of empirical significance (claims 6.25 and 6.27) and the relations listed above to counter the charge of irrelevance are steps in that direction. The following section provides more evidence that the criteria allow the development of a comprehensive theoretical system, for it shows how they can be generalized with comparative ease. To be fruitful, an explicatum should furthermore suggest new directions of research. I will discuss two of those in §11 and §12.1.

6.8 Generalizations

The presumptions of the semantic criteria of empirical significance are comparably strong: They assume predicate logic, a bipartition of the vocabulary and a set Π containing only analytic sentences. As already noted, the syntactic criteria of em-

[23]As argued in §2.4, Lewis's purported ordinary language analyses are very likely themselves already explications, although Lewis's explicata are arguably fairly similar to their explicanda.

[24]See page 19. In §6.9, I discuss why the use of ordinary language intuitions may be tentatively justified in this case.

pirical significance can be defined with any distinguished set of basic sentences, and the equivalence proofs at most rely on the set being closed under truthfunctional composition. At least falsifiability has furthermore already been generalized to all systems of logic for which conservativeness has been defined, and there seems little reason to doubt that such a definition is always possible. In the following, I will suggest two further generalizations by weakening the restrictions on the basic sentences and the restrictions on Π.

6.8.1 General $\mathscr{B}$-sentences and structures

As discussed in §5.1, if the basic terms are vague, some *individual* $\mathscr{B}$-sentences are themselves not empirically significant, and thus should not be allowed, for example, for the empirically verification or falsification of other sentences. If the basic terms are vague, the set of basic sentences in the criteria of empirical significance therefore has to be restricted. Furthermore, many classical syntactic criteria allow only finite sets of molecular or atomic $\mathscr{B}$-sentences to falsify or verify a $\mathscr{V}$-sentence (Popper 1935, Hempel 1965c, Rynin 1957). Since there may be infinitely many non-equivalent molecular or atomic basic sentences (so that their union is infinite), this restriction cannot be captured by restricting the set of basic sentences. To accommodate these restrictions and others, I suggest to consider a set of sentences one of basic sentences if and only if it is in $\mathbf{\Omega}$, where $\mathbf{\Omega}$ can be determined as needed. Any element of $\mathbf{\Omega}$ will be called 'set of $\mathbf{\Omega}$-sentences'. This allows the following

Definition 6.28. A set Σ of $\mathscr{V}$-sentences is *syntactically falsifiable/verifiable/ weakly determined in* $\mathbf{\Omega}$ if and only if there is a possible set of $\mathbf{\Omega}$-sentences that falsifies/verifies/determines Σ.

If $\mathbf{\Omega}$ contains all sets of $\mathscr{B}$-sentences, definition 6.28 is equivalent to the conjunction of definitions 6.2, 6.13, and 6.23. To achieve a similarly versatile definition for semantic criteria of empirical significance, I suggest to use a set $\mathbf{\Omega}$ of sets of structures. Elements of $\mathbf{\Omega}$ will be called 'sets of $\mathbf{\Omega}$-structures'. Whether $\mathbf{\Omega}$ contains sets of structures or sets of sentences will be clear from context. For sets of $\mathbf{\Omega}$-structures, the subset relation is the analogue of the entailment relation for sets of sentences. In analogy to the compatibility of a set of sentences with Π, a set of $\mathscr{V}$-structures is possible if and only if its intersection with the models of Π is not empty. In other words, a set of $\mathscr{V}$-structures is possible if and only if one of its elements is possible according to definition 2.2. To define this notion also for structures of proper subsets of $\mathscr{V}$, I suggest

Definition 6.29. The possible subset of a set $\mathbf{S}$ of structures is the set of possible structures in $\mathbf{S}$. $\mathbf{S}$ is possible if and only if its possible subset is not empty.

This leads, in analogy to definitions 6.7, 6.14, and 6.16, to

Definition 6.30. A *set of structures determines* (*falsifies/verifies*) a set Σ of $\mathscr{V}$-sentences in $\mathbf{\Omega}$ if and only if all elements $\mathfrak{A}$ and $\mathfrak{B}$ ($\mathfrak{A}$) of its possible subset are such that $\mathfrak{A} \models \Sigma$ iff $\mathfrak{B} \models \Sigma$ ($\mathfrak{A} \models \Sigma$ / $\mathfrak{A} \not\models \Sigma$).

Finally, this suggests in analogy to definitions 6.8, 6.15, and 6.24

Definition 6.31. A set Σ of $\mathscr{V}$-sentences is *semantically falsifiable/verifiable/weakly determined in* $\mathbf{\Omega}$ if and only if there is a possible set of $\mathbf{\Omega}$-structures that falsifies/verifies/determines Σ.

If $\mathbf{\Omega}$ contains all sets of $\mathscr{B}$-structures, definition 6.31 is equivalent to the conjunction of definitions 6.8, 6.15, and 6.24. It is also possible to define a generalization of strong $\mathscr{B}$-determinacy:

Definition 6.32. A set Ω of $\mathbf{\Omega}$-sentences is maximal if and only if there is no possible set Γ of $\mathbf{\Omega}$-sentences such that $\Gamma \cup \Pi \models \Omega$ and $\Omega \cup \Pi \not\models \Gamma$.

If $\mathbf{\Omega}$ contains all sets of $\mathscr{B}$-sentences, definition 6.32 is equivalent to definition 6.22.

Definition 6.33. A set Ω of $\mathbf{\Omega}$-structures is maximal if and only if there is no possible set Γ of $\mathbf{\Omega}$-structures such that the possible subset of Γ is a proper subset of the possible subset of Ω.

As a generalization of definitions 6.17 and 6.23, I thus suggest

Definition 6.34. A set Σ of $\mathscr{V}$-sentences is *strongly semantically (syntactically) determined in* $\mathbf{\Omega}$ if and only if it is determined by every possible maximal set of $\mathbf{\Omega}$-sentences ($\mathbf{\Omega}$-structures).

If $\mathbf{\Omega}$ contains all sets of $\mathscr{B}$-sentences, this definition is equivalent to definition 6.23. If $\mathbf{\Omega}$ contains all singleton sets of $\mathscr{B}$-structures, this definition is equivalent to definition 6.17. Based on these concepts, one can construct generalized notions of comparative determinacy, falsifiability, and verifiability:

Definition 6.35. A set Σ of $\mathscr{V}$-sentences is *syntactically (semantically) at least as falsifiable/verifiable/determined in* $\mathbf{\Omega}$ as a set Γ of $\mathscr{V}$-sentences if and only if every possible set of $\mathbf{\Omega}$-sentences ($\mathbf{\Omega}$-structures) that falsifies/verifies/determines Γ in $\mathbf{\Omega}$ also falsifies/verifies/determines Σ in $\mathbf{\Omega}$.

Since $\mathbf{\Omega}$ is not further determined, it is only possible to prove a simple weak analogy to claim 6.25:

Claim 6.28. *If a set Σ of $\mathscr{V}$-sentences is analytically false/analytically true/analytically false or analytically true, then Σ is at least as syntactically (semantically) falsifiable/verifiable/determined in $\mathbf{\Omega}$ as any other set of $\mathscr{V}$-sentences.*

In complete analogy to claim 6.27, the following holds:

Claim 6.29. *A set Σ of $\mathscr{V}$-sentences is not syntactically (semantically) falsifiable/verifiable/weakly determined in $\mathbf{\Omega}$ if and only if Σ is at most as syntactically (semantically) falsifiable/verifiable/determined in $\mathbf{\Omega}$ as any other set of $\mathscr{V}$-sentences.*

Proof. The claim holds for all criteria because only the empty set is a subset of every set. □

Claims 6.28 and 6.29 show that the generalized comparative concepts, in partial analogy to the more specific ones, relate fruitfully to their classificatory counterparts. An exception is the notion of strong semantic determinacy in $\mathbf{\Omega}$. Unlike strong semantic $\mathscr{B}$-determinacy, but like strong syntactic $\mathscr{B}$-determinacy, it does not identify the greatest elements of its comparative counterpart. Thus the generalizations of strong syntactic and semantic $\mathscr{B}$-determinacy bring the concepts closer together. That the relation between strong $\mathscr{B}$-determinacy and comparative $\mathscr{B}$-determinacy disappears for their generalizations suggests that the relation is something of a fluke, and that only claim 6.28, as weak as it is, can be the basis of philosophical analyses. Fortunately, claim 6.27 generalizes neatly to claim 6.29 and thus suggests that *pace* Hempel, the least elements of the comparative notions provide fruitful criteria of empirical significance.

6.8.2 General background assumptions

With reference to Duhem (1914), Sober (2007, 5) notes that scientific theories, "on their own, do not make testable predictions. One needs to add 'auxiliary propositions' to the theories one wishes to test". Typically, these *supplementary sentences*[25] are not taken to be just analytic sentences, and therefore the above definitions would still exclude almost all scientific theories even though the criteria are non-trivial. Gemes (1998b, §1.2) calls this "the challenge from holism" and points out that the logical empiricists were acutely aware of it.[26] Indeed, Carnap's considerations (discussed in §3.6.5) show how many other theories may have to be considered to derive $\mathscr{B}$-sentences, depending on what terms are considered basic. Looking at figure 3.1 on page 90, for example, it is clear that a claim about electric fields leads to claims about colors one requires not only the theory of electromagnetism, but also at least a theory of color perception like the Young-Helmholtz trichromatic theory. Depending on the required detail of the $\mathscr{B}$-claims, one may even need elaborate theories about the functioning of retinal ganglion cells or the influence of language on color perception (see, for example, Winawer et al. 2007).

[25]In my terminology, 'auxiliary sentences' are $\mathscr{A}$-sentences, so I will speak of 'supplementary sentences' to avoid ambiguities.

[26]Gemes (1998b, §2) attempts to solve this problem for Ayer-type criteria with his concept of natural axiomatization. As I will argue below, this attempt is unnecessary because Ayer-type criteria in general are unnecessary.

One simple way to meet the challenge from holism is to consider the empirical significance of the union of the theory and the supplementary sentences. But this approach does not always allow inferring the empirical (non-)significance of a theory from the (non-)significance of the theory and its supplementary sentences. For a subset of a falsifiable or $\mathscr{B}$-determined set may itself not be falsifiable or $\mathscr{B}$-determined, respectively, and a subset of a non-verifiable or non-$\mathscr{B}$-determined set may itself be verifiable or $\mathscr{B}$-determined, respectively. This is a problem because a criterion of empirical significance is meant to determine whether a specific theory is significant, while this approach determines only the empirical significance of larger sets of theories, often spanning very disparate fields.

Ayer's definition of indirect verifiability is an attempt at meeting this challenge from holism in a different way, by defining empirical significance relative to a set Π containing not only analytic sentences, but also other empirically significant sentences. This kind of recursive definition[27] is suggestive given its success in the theory of definitions, for if a term A is definable in $\mathscr{B}$-terms, then any term definable in $\mathscr{B}$-terms and A is also definable in $\mathscr{B}$-terms alone, and if a set Σ of sentences is translatable into a set of $\mathscr{B}$-sentences, then any sentence translatable into a set of sentences containing only $\mathscr{B}$-terms and Σ is also translatable into a set of $\mathscr{B}$-sentences. The recursive definitions are thus equivalent to non-recursive definitions.[28] For strong syntactic $\mathscr{B}$-determinacy, this equivalence holds as well, because the truth value of a $\mathscr{B}$-determined set Σ of sentences is a function of the truth values of $\mathscr{B}$-sentences, and thus, if the truth value of any set of sentences is a function of the truth values of Σ and $\mathscr{B}$-sentences, it is also a function of the truth values of $\mathscr{B}$-sentences alone. But for any of the weaker criteria of empirical significance, this equivalence breaks down and thus the recursion brings with it the danger of triviality.

The problem of past failures clearly does not allow the conclusion that all criteria will fail. But the puncture-and-patch industry mentioned in §5 revolves exclusively around recursive criteria, most recursive criteria suggested so far are trivial, and no recursive criterion has been shown to be non-trivial. The correct inference to draw from the problem of past failures may thus be that there is no adequate *recursive* criterion of empirical significance.

Another reason to question the search for a recursive criterion is that recursive criteria do not seem to address the challenge from holism. In a recursive criterion, the supplementary sentences can contain any empirically significant sentence, even those that are known to be false. But the challenge from holism consists in the need for other *true* (or at least justified) sentences to evaluate a theory. Surprisingly enough given the amount of work put into recursive definitions of empirical significance, there have been few justifications of the assumption that a criterion

[27] As noted in connection with definition 6.5, Ayer's definition needs to be amended slightly to make it a proper recursive definition.

[28] Recursive definitions typically are (Moschovakis 1974, 1; Essler 1982, §16; Leivant 1994, §2.7), but in the cases considered here, there is not even the need to go to a higher order of the language.

of empirical significance should be recursive. In defense of his second criterion of empirical significance, Ayer (1946, 12) states that only by taking supplementary sentences into account can "hypotheticals" be rendered empirically significant. But Ayer then simply chooses empirically significant supplementary sentences without even considering justified supplementary sentences. As far as I know, this is the only justification that has been provided for recursive criteria.

In contradistinction, Sober (2008, 151, §2.14) suggests a criterion that uses justified supplementary sentences. He points to the triviality of Ayer's first criterion (see page 226) and, with "some fear [of] stumbling into the same old quagmire", suggests instead:

> Proposition P now has observational implications if and only if there exist true auxiliary assumptions A, and an observation statement O, such that (i) $P \wedge A$ entails O, but A by itself does not entail O, (ii) we now are justified in believing A, and (iii) the justification we now have for believing A does not depend on believing that P is true (or that it is false), and also does not depend on believing that O is true (or that it is false).

Sober argues for the requirement that A be justified independently of O as follows. Let P be any sentence, and O a justified true $\mathscr{B}$-sentence. Then $\neg P \vee O$ is true and justified, and if the choice $A := \{\neg P \vee O\}$ were allowed by the definition, P would have observational implications. In short, the criterion is trivial without the restriction.[29]

Sober's trivialization argument is incomplete because he assumes, but does not prove, that $\neg P \vee O \nvDash O$ for some O. Lewis (1988a) notes that if $\neg P \vee O \vDash O$ for all O, then P is a logical truth. Hence the restriction on A is justified under the assumption that not only logical truths should lack observational implications. The argument is also no strict proof as long as 'justification' and 'dependence of a justification' are not defined. Whether it is valid depends on whether, for example, O justifies $\neg P \vee O$ independently of $\neg O$, P, and $\neg P$ for any P.

More problematic is that Sober's criterion itself is arguably trivial: For, assume that P is any sentence and Q a true non-$\mathscr{B}$-sentence that entails O and is justified independently of O, $\neg O$, P, and $\neg P$. Then, in analogy to Sober's trivialization argument, $A := \{\neg P \vee Q\}$ is allowed by his definition, and thus, unless it is logically true, P has observational implications. Unlike Sober's trivialization argument, this argument assumes that the justification for a sentence does not always have to depend on *all* of the $\mathscr{B}$-sentences it entails, but this is a rather weak assumption about justification.

[29] This is an argument for the second conjunct of (iii). For the first conjunct, Sober (2008, 144–145) provides a trivialization argument only for his explication of 'P is testable against Q', so its relevance for the definition at hand is not immediately clear. Furthermore, I will argue in §8.3.1 that the argument fails.

Sober's and my trivialization arguments rely on the same trick: Some sentence (O or Q) is used to infer another ($\neg P \vee O$ or $\neg P \vee Q$) by irrelevant disjunction, but only the inferred sentence is included in A. One way of avoiding this specific trivialization is to disallow irrelevant disjunctions, for example by restricting justificatory inferences to relevant deductions as developed by Schurz (1991). I will suggest another way that is more suitable to the current case, starting from an observation that Schurz (1991, §2.2, parenthetical remark in the original) makes in his justification of relevant deduction. He notes that if one knows that A and tells someone else that $A \vee B$, one misleads the hearer because

> in practical speech situations the hearer assumes that if the speaker tells him a disjunction, say $A \vee B$, then the speaker's knowledge K_s about A and B is indeed *incomplete*, i. e. both $\neg A$ and $\neg B$ and thus also A and B are possible in K_s [because, given $K_s \vdash_L A \vee B$, $K_s \nvdash_L A/B$ implies $K_s \nvdash_L \neg B/\neg A$, respectively]. [...] The irrelevant conclusion together with this implicit assumption causes in the hearer an expectation which is not only irrelevant but *wrong*—

namely that it is possible that $\neg A$.[30] '$\vdash_L$' here refers to deductive inference in predicate logic.

My suggestion is to avoid this wrong expectation by disallowing the speaker to tell the hearer that $A \vee B$ but not A, if he knows that A. If the speaker asserts supplementary sentences Π, this leads to

Definition 6.36. Π is an *honest set* if and only if every $\varphi \in \Pi$ is a justified sentence, and Π also contains every sentence on which the justification of φ depends.

Somewhat less precisely, I will also speak of 'honest supplementary sentences' rather than 'honest sets of supplementary sentences'. One might paraphrase this requirement on Π more intuitively as the demand that Π contain *all* supplementary sentences that are accepted, either actually or in some counterfactual situation that is of interest. That is, an honest set is a set of sentences that, for all we know, could have been our set of background assumptions: It describes which set of justified beliefs we hold now (if Π is simply the set of all our currently justified beliefs), or which set we could have held before we got to hold all our current justified beliefs. The subjunctive here excludes false starts, that is, beliefs that at one point were justified but later became unjustified. This is plausible, as it is of little interest whether a theory is empirically significant under the assumption that the world is different from the way it in fact is (as far as we know). In this sense, an honest set is one step in the accumulation of our currently justified beliefs. Note that an honest set of supplementary sentences can be finite, because

[30] Specifically, the speaker is "liable to mislead" by "quietly and ostentatiously" (Grice 1975, 49) violating Grice's first maxim of quantity: "Make your contribution as informative as is required (for the current purposes of the exchange)" (Grice 1975, 45).

analytic sentences do not need justifications, and some sentences may be justified by entities that are not sentences (e. g. experiences). Coherentists may also argue that the sentences in a finite set can justify each other without primitively justified sentences.

To avoid trivialization of Sober's criterion, I therefore suggest the following modification:

Definition 6.37. A set Σ of $\mathscr{V}$-sentences *has basic implications* if and only if Σ is $\mathscr{B}$-creative with respect to an honest set containing all analytic sentences.

It is clear that in the intuitive paraphrase of definition 6.36, the counterfactual case is important because otherwise, an accepted theory could have no basic implications at all since anything it entails would (I assume) be justified and therefore in Π.

Definition 6.37 is syntactic falsifiability, with Π taken to be an honest set of justified or analytic sentences rather than just the set of analytic sentences. All other definitions discussed above, including the generalizations in §6.8.1, can be reinterpreted analogously. In the following, I will thus sometimes speak of falsifiability, $\mathscr{B}$-determinacy etc. *relative to* or *given* some set of sentences, in which case Π is assumed to be that set. This generalization of Π also leads to a generalization of all other concepts defined relative to Π, so that, for example, possibility, analytic truth and falsity, and contingency becomes possibility, truth, falsity, and contingency *relative to* or *given* Π (in which case I will obviously not speak of *analytic* truth, contingency, etc.).

Sober's definition and definition 6.37 differ beyond their restrictions on the supplementary sentences. Unlike Sober's criterion, $\mathscr{B}$-creativity is defined for set of sentences and includes sets that only allow new inferences from infinite sets of basic sentences (in the case of higher order logic). I do not think that Sober's criterion was designed with this distinction in mind, however, so this modification is only of a technical nature. Sober also does not assume that basic sentences are determined by their vocabulary. But as noted, this restriction can be accommodated in artificial languages, is not essential for the definition, and can be explicitly avoided by using the generalizations in §6.8.1. Finally, the definition does not include an explicit reference to the current point in time, because time dependence follows implicitly from the definition's dependence on justified supplementary sentences. Since the set of justified supplementary sentences changes over time, so does the set of sentences that are $\mathscr{B}$-creative with respect to them.

Ignoring these rather formal differences, definition 6.37 is at least as exclusive as Sober's definition: Condition (ii) is entailed by the demand that the supplementary sentences form an honest set. Condition (iii) for one excludes sentences P that are $\mathscr{B}$-creative relative to supplementary sentences whose justification depends on P. In definition 6.37, those sentences lack basic implications because they are already contained in the supplementary sentences, which therefore already entail everything that their conjunction with P entails. Condition (iii) furthermore excludes sentences P that only entail basic sentences O on which the justification of the supplementary sentences depends. In definition 6.37, these basic sentences are also contained in the supplementary sentences, so that they are already entailed by the supplementary sentences alone. There is also hope that definition 6.37 is strictly more exclusive than Sober's definition, and thus not trivial, for the use of honest supplementary sentences blocks the trivialization arguments given above. This is because in the arguments, $\neg P \vee O$ and $\neg P \vee Q$ are justified by O and Q, respectively. If O or, respectively, Q are included in Π, Π alone entails O.

Historically, the interpretations of Π have been varied. Przełęcki's definitions explicitly assume Π to contain all and only analytic sentences. Rynin seems to define his criterion with analytic sentences in mind, and, as argued above, Ayer's definition of direct verifiability relies on analytic inferences, too. Lewis's definitions rely on the concept of possible worlds, and whether these are the models of the analytic sentences is up for discussion. Suppes is silent on the matter. Carnap (1936, 443), on the other hand, considers basic sentences possible only if they are compatible with the laws of physics, that is, only if they are not contravalid. And in another passage, Carnap (1935b, 11) writes:

> A proposition P which is not directly verifiable can only be verified by direct verification of propositions deduced from P together with other already verified propositions.

This is almost Ayer's definition of indirect verifiability, except for one crucial difference: The "other propositions" are not only required to be verifiable, but actually verified. Unlike Ayer, Carnap does not define 'verifiability' recursively, but rather relative to a set of justified propositions.[31] Popper (1935, §3, emphasis changed) is very explicit about the role of justified sentences in his conception of falsification:[32]

> [T]here is the testing of the theory by way of empirical applications of the conclusions which can be derived from it.
>
> [...] With the help of other statements, *previously accepted*, certain singular statements—which we may call 'predictions'—are deduced from the theory [...]. Next we seek a decision as regards these (and other) derived statements by comparing them with the results of practical applications and experiments. [I]f the decision is negative, or in other words, if the conclusions have been falsified, then their falsification also falsifies the theory from which they were logically deduced.

[31]If one assumes the prediction criterion of confirmation, however, Carnap's definition is still trivial: For any P and any two $\mathscr{B}$-sentences $\beta \nvdash \gamma$, where β is true, $\{(P \to \gamma) \wedge \beta\} \vdash \beta$ and is thus verified by β. Since $\{P, (P \to \gamma) \wedge \beta\} \vdash \gamma$ while $(P \to \gamma) \wedge \beta \nvdash \gamma$, P is indirectly verifiable.

[32]Lakatos (1974, 106–107) also stresses that for Popper, falsifiability is relative to background assumptions, and gives further references to Popper's remarks on the topic.

With the inclusion of previously accepted sentences in Π, the viability of Popper's criterion depends on the conception of acceptance, as do the viability of definition 6.37 and Sober's criterion. Indeed, apart from the specific restriction (iii) on supplementary sentences, Sober's criterion is essentially Popper's falsifiability criterion. But requirement (iii), if taken to be the only restriction on the set of accepted sentences, trivializes Sober's criterion. Therefore it seems fair to say that Sober's criterion, insofar as it is successful, is anticipated by Popper.

In a review, Nott (1959) concludes about Popper's dissolution of the problem of induction:

> One cannot help feeling that if [*The Logic of Scientific Discovery*] had been translated as soon as it was originally published[,] philosophy in this country might have been saved some detours. Professor Popper's thesis has that quality of greatness that, once seen, it appears simple and almost obvious.

This is the right conclusion, but the wrong thesis. It is doubtful that Popper's dissolution of the problem of induction is indeed simple and obvious (cf. Salmon 1967, §II.3). But the preceding analysis suggests that if Popper's falsifiability criterion, rather than Ayer's, had been the basis of further research into the problem of demarcation, philosophy might have been saved the "sorry history of unintuitive and ineffective patches" (Lewis 1988b, §I) that discredited the very idea of empirical significance. Of course, Carnap's verifiability criterion or his criterion of direct reducibility would have been a good starting point as well. In the end, it may have just been the technical allure of recursive definitions what sidetracked philosophy for so long.

6.9 Conditions of adequacy

In §6.7.2, I relied on Lewis's definition of partial aboutness to argue that the criteria of empirical significance are similar to their explicandum. However, when discussing Carnap's condition of similarity (§2.3.1), I suggested that conditions of adequacy should be used instead of Carnap's condition that most uses of the explicanda have to be captured by the explicata. Reliance on Lewis's ordinary language intuition may still be preliminarily justified, because there are very few if any well-supported conditions of adequacy for criteria of empirical significance. Lewis's intuitions may therefore be tentatively used as a proxy, with the hope that an explicatum that fits with the ordinary language intuitions is likely to meet the conditions of adequacy that one would place on an explicatum.

An additional problem with the conditions of adequacy is that the intended use for a criterion of empirical significance differs from author to author and from time to time. The early logical empiricists, for example, intended the criterion to identify meaningless sentences, while Popper intended his criterion to distinguish between empirical and non-empirical sentences, both of which can be meaningful (Popper 1935, §4; 1963, §II).

The problems of considering criteria of empirical significance as criteria of meaningfulness are illustrated by the well-known discussion by Hempel (1965b, p. 102). He demands that "[i]f under a given criterion of cognitive significance a sentence σ is non-significant, then so must be all truth-functional compound sentences in which σ occurs nonvacuously as a component."

Condition 2. If $\mathscr{V}$-sentence σ is not empirically significant, then so is every sentence that is logically equivalent only to sentences in which σ occurs.

The *non-vacuous occurrence* of σ in a sentence τ is here taken to mean that τ is not logically equivalent to a sentence in which σ does not occur. Hempel states that condition 2 has to be met because "if $[\sigma]$ cannot be significantly assigned a truth value, then it is impossible to assign truth values to the compound sentences containing $[\sigma]$." Rynin (1957, 55) argues that this is incorrect, since in $\sigma \vee \upsilon$, for example, υ may be true, and hence, since a disjunction is true if either of its disjuncts is true, so is $\sigma \vee \upsilon$. Since a true sentence is significant, $\sigma \vee \upsilon$ is therefore significant.

Hempel could try avoiding Rynin's conclusion by pointing out that Rynin uses a very specific definition of the truth of a disjunction. If one takes the term 'truth-functional' literally, that is, assumes that the truth value of a disjunction is determined by a function $f : \{\text{true}, \text{false}\} \times \{\text{true}, \text{false}\} \to \{\text{true}, \text{false}\}$, then the truth value of $\sigma \vee \upsilon$ is indeed undefined if the truth value of σ is undefined. It is then unclear, however, what 'logically equivalent' means in condition 2. One can take the position that two sentences τ and ϱ are logically equivalent if and only if they are *syntactically* deducible from each other. However, in that case Rynin's counterargument is applicable as well. For if syntactic deduction is defined for all sentences, it holds that $\upsilon \vdash \sigma \vee \upsilon$, so that from the truth of υ, the truth of $\sigma \vee \upsilon$ follows.[33] More generally, if a sentence can be meaningful without being empirically significant, most of the criticisms by Hempel (1965c) are unjustified (Hempel 1965d; Sober 2008, 149–150), and thus I will ignore them henceforth.

An uncontroversial condition of adequacy, on the other hand, is

Condition 3. The concept of empirical significance is not trivial.

A trivial definition, one that includes all or no objects of the domain, cannot be a good explicatum for a concept that is meant to include some, but not all objects of the domain. At the very least, a trivial explicatum is uninformative. The most effective criticisms of criteria of empirical significance have at their center trivialization proofs. Most criticisms of Ayer-type criteria are of this kind. I have already noted that Church's criticism, the most famous one, does not show that

[33] This assumes that deductive inference is truth-preserving.

every sentence is indirectly verifiable. Thus there can be plausible strengthened conditions of adequacy that are similar to the exclusion of trivial criteria.

For one, there is already a partition of sentences into analytically determinate and indeterminate ones. According to artificial language philosophy, the analytically determined sentences are *chosen* to be so determined, and thus they do not contain any empirical information. Accordingly analytically determinate sentences were explicitly excluded from the set of empirically significant sentences. If now a criterion where to render all analytically contingent sentences empirically significant, it would thus add nothing to this demand. Similarly, if some set Π of sentences is accepted as an honest set, a criterion that renders empirically significant all sentences that are neither entailed nor incompatible with Π adds nothing to our assumption that Π is true. This suggests

Condition 4. Not all contingent sentences are empirically significant.

'Contingent' is here understood like 'analytically contingent' (definition 2.3), except that Π may be an honest set of supplementary sentences. Lewis (1988a) relies on the weaker demand that not all analytically contingent or analytically false sentences are empirically significant.

The counterpart to analytically determined sentences are, in a sense, basic sentences, because they are defined to be sentences whose truth value is empirically determined. A criterion that includes only basic sentences in the set of empirically significant sentences would thus again be uninformative, leading to

Condition 5. Not only $\mathscr{B}$-sentences are empirically significant.

Pokriefka (1984), for example, relies on condition 5 to criticize his own criterion. As pointed out in §6.2.1, Popper (1935, 85; cf. Hansson 2008b, §4.2) and Ayer (1936, 97) argue for their criteria of empirical significance based on

Condition 6. All and only sets of sentences that assert basic sentences are empirically significant.

The assumption here is that—at least as far as *deductive* assertion is concerned—a set of sentences *asserts basic sentences* if and only if it is $\mathscr{B}$-creative. I will discuss a similar consideration by Sober (2008) in §8.4. Quine (1951, 41) also thinks of "the conceptual scheme of science as a tool, ultimately, for predicting future experience in the light of past experience", which entails that every scientific theory, if it is to be useful, must fulfill condition 6.

For deductive inference, syntactic and semantic $\mathscr{B}$-creativity are a clear and virtually undisputed criteria for the making of $\mathscr{B}$-assertions. Things are much more complicated in the case of probabilistic inference. Given the need for compatible criteria of empirical significance for deductive and probabilistic contexts (see §5.2), it will be helpful to keep the following in mind:

Definition 6.38. Two sets of $\mathscr{V}$-sentences Σ and Λ are *deductively empirically equivalent relative to* supplementary sentences Π if and only if for every set Ω of basic sentences and every basic sentence β, $\Omega \cup \Sigma \cup \Pi \vDash \beta$ iff $\Omega \cup \Lambda \cup \Pi \vDash \beta$.

This definition assumes that empirical states are described in $\mathscr{B}$-sentences, not $\mathscr{B}$-structures. An analogous semantic definition is straightforward, but will not be needed in the following. It is now easy to establish

Claim 6.30. *Set Σ of $\mathscr{V}$-sentences is $\mathscr{B}$-conservative relative to Π if and only if Σ and $\top$ are deductively empirically equivalent relative to Π.*

Proof. Σ is deductively empirically equivalent to $\top$ if and only if for every set Ω and sentence β either $\Omega \cup \Sigma \cup \Pi \vDash \beta$ and $\Omega \cup \Pi \vDash \beta$, or $\Omega \cup \Sigma \cup \Pi \nvDash \beta$ and $\Omega \cup \Pi \nvDash \beta$, that is, either $\Omega \cup \Pi \vDash \beta$ or $\Omega \cup \Sigma \cup \Pi \nvDash \beta$. This is the case if and only if Σ is $\mathscr{B}$-conservative relative to Π. □

As far as deductive assertions are concerned, one can thus reformulate condition 6 as follows:

Condition 7. All and only sets of sentences that are not empirically equivalent to a tautology are empirically significant.

The hope is then that with a definition of probabilistic empirical equivalence, condition 6 is equivalent to condition 7 as well. Flew (1950, 258) goes as far as calling every sentence a tautology that does not deductively assert basic sentences, but this is clearly to strong: ‘Borogroves are mimsy’ is not a tautology, but on account of containing two undefined terms, does not make any basic assertions.

* * *

Trivially, according to condition 6, and thus condition 7, $\mathscr{B}$-creativity is up to logical equivalence the only adequate criterion of empirical significance as far as deductive assertions are concerned. If the negation of a basic sentence is itself a basic sentence, claim 6.1 thus establishes that falsifiability is adequate as well. $\mathscr{B}$-creativity further fulfills the non-triviality condition 3, as established in §6.7.2 using the analytic sentence $\Pi = \{\forall x(Bx \leftrightarrow \neg A_1 x)\}$, the $\mathscr{B}$-creative sentence $A_1 b$ and the $\mathscr{B}$-conservative sentence $A_2 b$. The example also shows that $\mathscr{B}$-creativity fulfills the stronger condition 4 because $A_2 b$ is analytically contingent, and condition 5 because $A_1 b$ is not a basic sentence. It is fortunate that $\mathscr{B}$-creativity fulfills all other conditions of adequacy as well (excluding the disputed one by Hempel). For otherwise, there would be no concept with the intended properties at all.

Definition 6.37 defines ‘having basic implications’ as $\mathscr{B}$-creativity relative to honest sets, and thus as an absolute (one-place) concept. Since in the example, Π contains only (and for the purposes of the example all) analytic sentences, it is an honest set. Thus the example shows that definition 6.37 fulfills condition 5. It is simple to show that no analytic sentence has basic implications, and thus that condition 3 is met. Because of the reliance on the undefined term 'justified' in definition 6.36, however, it is impossible to disprove that there is for each analytically indeterminate sentence σ some honest set Γ such that σ is $\mathscr{B}$-creative relative to Γ. Thus it cannot be shown that condition 4 is met, and the triviality of Sober's criterion might suggest that the new criteria will suffer the same fate as Ayer's. However, there are important differences between Ayer's criterion and the new ones. First, the criteria's core ideas—falsifiability, verifiability, and $\mathscr{B}$-determinacy relative to a set Π—are not trivial. Therefore, in the case of a trivialization proof, one can always fall back on falsifiability, verifiability, or $\mathscr{B}$-determinacy, and find a new restriction on the supplementary sentences Π. (Without a criterion of justification, empirical significance can be defined in a precise way only *relative* to a set Π, and it has to be decided on a case-by-case basis whether Π is acceptable.) Second, and in keeping with my criticism of recursive definitions of empirical significance, Π is not assumed to be determined by the criterion itself. This blocks trivialization proofs that rely on recursive definitions. From these two differences follows a third: Problems can only occur with the definition of 'honest set', and amendments of the criteria can accordingly be restricted to this definition. In this sense, falsifiability, verifiability, and $\mathscr{B}$-determinacy are already good criteria of empirical significance. What is missing is a good criterion of justification for sets of sentences.

* * *

Whether a set of sentences indeed has to assert basic sentences to be empirically adequate is still a matter of discussion. For one, in some situations it may be sufficient to establish that a sentences is true, even if it does not assert a basic sentence. If, to take a somewhat frivolous example, the question is whether a bet is won or lost, it can be enough if a sentence is verifiable.[34] Thus verifiability and weak $\mathscr{B}$-determinacy may still have a role to play. In the end, the conclusion may well be that *empirical significance* has to be split up into different concepts, depending on their intended use. If, for example, the goal is prediction, $\mathscr{B}$-creativity and falsifiability are adequate criteria.

6.10 Conclusion

The belief that the search for a criterion of empirical significance has been a failure is usually based on the problem of past failures. I have given an alternative view on this search, mostly based on criteria that have not been shown to be trivial. I have

[34]There is, however, the question who would bet that the sentence is false if the best outcome would be that the sentence is not shown to be true.

already argued that the criteria successfully explicate their explicandum (§6.7.2), specifically with respect to their fruitful connection to each other, comparative criteria of empirical significance, Ramsey sentences, and concepts from measurement and definition theory. This result counters the charge of arbitrariness.

More generally, the equivalences show that various different frameworks (measurement theory, aboutness, Ramsey sentences, some semantic approaches, some syntactic approaches) allow the definition of plausible criteria of empirical significance. Because of the inferential relations between the criteria, it is easy to see when an analysis or justification of one criterion transfers to another. The relations also allow for a more informed search for generalizations of the criteria. The generalization of the concepts of $\mathscr{B}$-sentences and $\mathscr{B}$-structures suggested in this article reduces the conceptual presumptions of the criteria, and thus allows them to be applied in more circumstances. The generalization from analytic to general supplementary sentences similarly allows the criteria's application in more contexts, thereby solving the problem of holism, and additionally shows a way to avoid the problem of past failures.

6.11 Appendix

6.11.1 A generalization of Przełęcki's definition of $\mathscr{B}$-conservativeness

Definition 6.6 of semantic $\mathscr{B}$-conservativeness is slightly more general than Przełęcki's, who suggests

Definition 6.39 (Przełęcki)**.** A set Σ of $\mathscr{V}$-sentences is *semantically $\mathscr{B}$-conservative with respect to* Δ if and only if for each $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}} \models \Delta$ there is a $\mathscr{V}$-structure $\mathfrak{C} \models \Delta \cup \Sigma$ with $\mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$.

Semi-formally, this can be paraphrased as

$$\forall \mathfrak{A}_{\mathscr{B}} [\mathfrak{A}_{\mathscr{B}} \models \Delta \Rightarrow \exists \mathfrak{C}(\mathfrak{C} \models \Delta \cup \Sigma \wedge \mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}})] , \tag{6.4}$$

which is obviously restricted to $\mathscr{B}$-sentences in Δ due to '$\mathfrak{A}_{\mathscr{B}} \models \Delta$' in the antecedent of the implication. An equivalent reformulation is

$$\forall \mathfrak{A}_{\mathscr{B}} [\exists \mathfrak{B}(\mathfrak{B}|_{\mathscr{B}} \models \Delta \wedge \mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}) \Rightarrow \exists \mathfrak{C}(\mathfrak{C} \models \Delta \cup \Sigma \wedge \mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}})] . \tag{6.5}$$

If the restriction on $\mathfrak{B}$ in '$\mathfrak{B}|_{\mathscr{B}} \models \Delta$' is dropped, '$\mathfrak{B} \models \Delta$' is defined for any set Δ of $\mathscr{V}$-sentences, which leads to definition 6.6.

6.11.2 The relation of syntactic and semantic $\mathscr{B}$-conservativeness

That syntactic $\mathscr{B}$-conservativeness does not entail semantic $\mathscr{B}$-conservativeness in first order logic is known in model theory as the difference between elementary and pseudoelementary classes (Hodges 1993, §5.2). Elementary classes are simply those whose members are models of a set Γ of first order sentences ($\{\mathfrak{A} \mid \mathfrak{A} \vDash \Gamma\}$), while pseudoelementary classes are those whose members are reducts, to a fixed subvocabulary, of the members of an elementary class ($\{\mathfrak{A}|_{\mathscr{B}} \mid \mathfrak{A} \vDash \Gamma\}$). Recall that $\Lambda|_{\mathscr{B}}$ is the set of $\mathscr{B}$-consequences of Λ, that is, the set of $\mathscr{B}$-sentences entailed by Λ.

Claim 6.31. *A set Σ of $\mathscr{V}$-sentences is syntactically $\mathscr{B}$-conservative relative to Δ if and only if $\{\mathfrak{A}_{\mathscr{B}} \mid \mathfrak{A}_{\mathscr{B}} \vDash \Delta|_{\mathscr{B}}\} = \{\mathfrak{A}_{\mathscr{B}} \mid \mathfrak{A}_{\mathscr{B}} \vDash (\Sigma \cup \Delta)|_{\mathscr{B}}\}$. Σ is semantically $\mathscr{B}$-conservative relative to Δ if and only if $\{\mathfrak{A}|_{\mathscr{B}} \mid \mathfrak{A} \vDash \Delta\} = \{\mathfrak{A}|_{\mathscr{B}} \mid \mathfrak{A} \vDash \Sigma \cup \Delta\}$.*

Proof. The proof for syntactic $\mathscr{B}$-conservativeness follows from claim 6.3, the proof for semantic $\mathscr{B}$-conservativeness is immediate. □

Mal'tsev introduced pseudoelementary classes in 1941 (see Hodges 1993, 207, 260). Przełęcki (1969, 52–53) was possibly the first to mention the distinction between syntactic and semantic $\mathscr{B}$-creativity in connection with philosophy of science, specifically with respect to questions of concept formation and the theory of definition. He does mention a proof, personally communicated by C. C. Chang, that semantic $\mathscr{B}$-creativity *relative to the empty set* does not entail syntactic $\mathscr{B}$-creativity relative to the empty set. Przełęcki and Wójcicki (1971, §1) reproduce a proof by Łoś (1955):

Proof of claim 6.4, second half. Let

$$\mathfrak{A}_{\mathscr{B}} := \langle \mathbb{N}, +, \cdot, 0, 1 \rangle \tag{6.6}$$

be the standard model of arithmetic and define $\Delta := \mathrm{Th}(\mathfrak{A}_{\mathscr{B}})$. Δ is therefore complete. Define $\mathscr{A} := \{P\}$ for the predicate symbol P and

$$\sigma := P0 \wedge \forall x[Px \rightarrow Px+1] \wedge \exists x \neg Px \,. \tag{6.7}$$

$\{\sigma\} \cup \Delta$ is consistent, but there are models of Δ, for example $\mathfrak{A}_{\mathscr{B}}$, that cannot be expanded to a model of $\{\sigma\} \cup \Delta$. □

Following Sneed (1971), subsequent discussions of the distinction were often phrased in terms of the Ramsey-eliminability of theoretical terms (cf. Rynasiewicz 1983, §1; van Benthem 1978). With a reference to Przełęcki (1969), van Benthem (1982, §2.1.3) gives a proof with a fairly intuitive natural language interpretation:

Proof of claim 6.4, second half. Let Δ be a complete axiomatization of the theory of 0 and successor:

$$\begin{aligned} \Delta :=& \{\neg\exists x\, sx = 0, \forall x \forall y(sx = sy \rightarrow x = y), \forall x(x \neq 0 \rightarrow \exists y\, x = sy)\} \cup \\ & \bigcup_{n \in \mathbb{N}} \{\neg\exists x_0 \exists x_1 \exists x_2 \ldots \exists x_{n-1} \exists x_n \\ & \qquad (sx_0 = x_1 \wedge sx_1 = x_2 \wedge \cdots \wedge sx_{n-1} = x_n \wedge sx_n = x_0)\} \,. \end{aligned} \tag{6.8}$$

Δ can be understood as the theory that time proceeds without loops by a one-to-one successor function s starting from 0. Define $\mathscr{A} := \{\prec, E\}$ with a 'before' relation, an 'early' predicate, and a set of sentences stating the following: 0 is early, the successor of an early time is also early, each time is earlier than some late (not early) time, and any time later than a late time is itself late:

$$\sigma := E0 \land \forall x(Ex \to Esx) \land \forall x \exists y(x \prec y \land \neg Ey) \land \forall x \forall y(x \prec y \land \neg Ex \to \neg Ey)\,. \tag{6.9}$$

Any finite subset of $\{\sigma\} \cup \Delta$ has a model, and by compactness, $\{\sigma\} \cup \Delta$ itself has a model. However, $\langle \mathbb{N}, S, 0\rangle$ is a model of Δ that cannot be expanded to a model of $\{\sigma\} \cup \Delta$. □

The difference between semantic and syntactic falsifiability was recently again discussed in connection with Ramsey sentences by Ketland (2004, 297–299) and Demopoulos (2011, §4).

Chapter 7

Deductive criteria for terms

Because of the problem of past failures for deductive criteria for the empirical significance of sentences, a number of philosophers have tried to define a deductive criterion for terms,[1] in the hope that a vocabulary of empirically significant terms allows only the construction of empirically significant sentences. Carnap tried to arrive at such a criterion in the *Aufbau*, "Testability", and "Theoretical concepts", where this last attempt spawned its own short industry of punctures and patches (Achinstein 1964, Berkowitz 1979, Van Cleve 1971, Kaplan 1975, Creath 1976). Less well known are the criteria by Wójcicki (1966), Rozeboom (1962), and Luce (1978). I will again show a variety of relations between he different criteria for terms (summarized in figures 7.1 and 7.2), thereby countering the charge of arbitrariness. I will also discuss the relation of the criteria for terms to the criteria for sentences.

7.1 Preliminaries

7.1.1 Methodological presumptions

Many criteria in the following define empirical significance only for relations. I have chosen not to generalize these criteria because, first, any set of sentences containing functions and constants is definitionally equivalent to a set of sentences containing only relations and sentences expressing appropriate uniqueness conditions. Second, the exclusion of theoretical functions and constants avoids technically important but philosophically less relevant complications.

While many criteria for the empirical significance of sentences have been

[1] In this chapter, it will be especially important to keep in mind that, in line with the tradition in philosophy of science, but not logic, I use 'terms' synonymous with 'non-logical constants' (see §1.2). Accordingly, relations are specific terms.

defined relative to a set of analytic sentences, all criteria for terms have been defined relative to a theory Θ, and thus heed the credo by Stegmüller (1973, 24) "that is notorious insofar as it was *not* pronounced by *any* scholastic *or* early modern philosophers, although it should have been pronounced a long time ago: 'termini sine theoria nihil valent'"[2]. Θ therefore takes over the role of the supplementary assumptions Π, so that specifically, only set of $\mathscr{B}$-sentences that are compatible with Θ are possible and only $\mathscr{V}$-structures $\mathfrak{A} \vDash \Theta$ with $\mathfrak{A}|_{\mathscr{B}} \in \mathbf{M}_{\mathscr{B}}$ are possible, that is, are in $\mathbf{M}$. $\mathbf{M}_{\mathscr{B}}$ is here again the set of possible $\mathscr{B}$-structures as defined in §2.8.1. Analogous to definition 2.2 and claim 2.1, it then holds that a structure is a model of Θ if and only if it is isomorphic to a structure that can be expanded to a $\mathscr{V}$-structure that is possible given Θ. In the following, I will silently assume that Θ is consistent, such that at least one structure is possible, and also not distinguish between isomorphic structures. Because of claim 2.1, this means that I will call all structures with expansions to models of Θ 'possible'.

7.1.2 The problem of the explicandum

Analogous to sentences, which can be meaningful because of their analytic content, a term can be meaningful not only because of its connections to observational terms, but also because of its connections (within Θ) to non-observational terms. Some of the criteria below will be distinguished exactly on these grounds.

A variety of conditions of adequacy demand that empirically significant terms stand in a specific relation to empirically significant sentences, for example that any sentence that contains only empirically meaningful terms be itself empirically meaningful. These conditions are problematic if, like Hempel (1965c, §3), one sees a criterion for terms as a means to define a criterion for sentences. For then the condition of adequacy is empty until one develops a criterion for the empirical significance of sentences independently from the criteria for terms, at which point the latter are not needed anymore to develop such a criterion. In fact, one of the results of my discussion will be that the criteria for the empirical significance of terms are sorely lacking a clear purpose. Carnap's criterion specifically will turn out to *rely* on a criterion for the empirical significance of sentences, which renders it particularly unsuited as a basis for a criterion for sentences. In a final piece of evidence that the criteria are solutions in search of a problem, the analysis of the suggested criteria will lead to some suggestions on how to apply them.

7.2 Definability and its variants

In the heyday of criteria of empirical significance, Carnap (1928a, §38) suggested that a term $A_i \in \mathscr{A}$ is empirically significant (and, even stronger, meaningful) if

[2]Presumably: 'Words mean nothing without a theory', although 'terminus' is usually translated as 'boundary' or 'end'.

and only if it can be explicitly defined in $\mathscr{B}$-terms.[3] Carnap (1928a, §67, §122) does not intend these definitions to follow from the meanings of the terms outside of any empirical theory, but rather from the regularities that are described by empirical theories (cf. Carnap 1967a, ix; 1963c, 945). In other words, he claims that scientific theories *entail* these explicit definitions.[4]

Definition 7.1. A relation P_i is *$\mathscr{B}$-definable* in Θ if and only if there is a $\mathscr{B}$-formula β such that

$$\Theta \models \forall \bar{x}[P_i \bar{x} \leftrightarrow \beta(\bar{x})] \tag{7.1}$$

If all and only $\mathscr{B}$-definable relations are considered empirically significant, empirically significant theoretical relations are inessential for the formulation of scientific theories in a very straightforward sense:

Claim 7.1. *If σ is a $\mathscr{V}$-sentence of $\mathscr{B}$-terms*[5] *and $\mathscr{B}$-definable relations, then σ* can be translated *into a $\mathscr{B}$-sentence by Θ, that is, there is a $\mathscr{B}$-sentence β such that $\Theta \models \sigma \leftrightarrow \beta$.*

Proof. If P_i is $\mathscr{K}$-definable in Θ, then for every $\mathscr{K} \cup \{P_i\}$-sentence σ there is a $\mathscr{K}$-sentence κ such that $\Theta \models \sigma \leftrightarrow \kappa$ (Essler 1982, 103). Therefore, if the $\mathscr{A}$-relations in σ are $\{P_{i_1}, \ldots, P_{i_{k+1}}\}$, σ can be translated into a $\mathscr{B} \cup \{P_{i_1}, \ldots, P_{i_k}\}$-sentence σ_k, and for $1 \leq l \leq k$, σ_l can be translated into a $\mathscr{B} \cup \{P_{i_1}, \ldots, P_{i_{l-1}}\}$-sentence σ_{l-1}, with σ_0 being a $\mathscr{B}$-sentence. □

Therefore, under the assumption of Θ, any sentence that contains only $\mathscr{B}$-definable theoretical relations or $\mathscr{B}$-terms is but a paraphrase of a $\mathscr{B}$-sentence. Assuming with artificial language philosophy that only $\mathscr{B}$-terms are directly interpreted, $\mathscr{B}$-definability is the only criterion that ensures the precise interpretation of any empirically significant relation:

Claim 7.2. *A relation P_i is $\mathscr{B}$-definable in Θ only if for all $\mathfrak{A}, \mathfrak{B} \models \Theta$ with $\mathfrak{A}|_{\mathscr{B}} = \mathfrak{B}|_{\mathscr{B}}$, $P_i^{\mathfrak{A}} = P_i^{\mathfrak{B}}$. The converse holds in first order logic and for finitely axiomatizable Θ—but not all infinite Θ—also in higher order logic.*

Proof. '⇒': If there is a $\mathscr{B}$-formula β with $\Theta \models \{\forall \bar{x}[P_i \bar{x} \leftrightarrow \beta(\bar{x})]\}$, then the extensions of P_i and β are the same in all models of Θ due to the definition of satisfaction in a model.

[3] Carnap (1928a, §38) also discusses the need for "definitions in use". As far as terms (i. e., non-logical constants) are concerned, these are equivalent to explicit definitions because of the eliminability theorems (cf. Gupta 2009, §2.3).

[4] The definitions of the definability of constant or function symbols additionally contain uniqueness conditions for the constants and function values, respectively. The conditions are philosophically interesting because they introduce restrictions on the sets of sentences in which constant and function symbols can be defined (Essler 1982, §14, §15; Hodges 1993, 59), but also introduce technical subtleties that would lead the current discussion to far afield.

[5] A last cautionary reminder: 'Terms' indeed means 'non-logical constants'.

'$\Leftarrow$': This holds in first order logic because of Beth's theorem (Hodges 1993, theorem 6.6.4). For higher order logic, substitute all $\mathscr{A}$-relations $\bar{P}$ in $\Theta(\bar{B},\bar{P})$ by variables. By lemma 2.4, $\forall \bar{x}[P_i \bar{x} \leftrightarrow \exists \bar{X}(\Theta(\bar{B},\bar{X}) \wedge X_i \bar{x})]$ is an explicit definition of P_i in $\mathscr{B}$ (cf. Shapiro 2000, §6.6.3; Tarski 1935, theorem 3). This procedure is, assuming a language that does not allow infinite conjunctions, obviously not possible for infinite sets Θ. Shapiro (2000, §6.6.3) shows that for some infinite Θ, there are undefinable relations with unique expansions (cf. Leivant 1994, 258). $\square$

According to claim 7.2, then, demanding the $\mathscr{B}$-definability of P_i amounts to the requirement that there is a $\mathscr{B}$-formula that is extensionally equivalent to P_i in all models of Θ. Therefore, assuming that the models of Θ are the (possibly counterfactual) contexts in which a theoretical relation P_i of Θ can be used, and assuming that the extensional equivalence of two formulas in all (counterfactual) contexts is the same as the intensional equivalence of the terms, $\mathscr{B}$-definability amounts to the intensional equivalence of P_i with a $\mathscr{B}$-formula (cf. Belnap 1993, 135).

Against Carnap himself, Goodman (1963, 554–555) argues that Carnap's explicit definitions (Carnap 1928a) should be read extensionally, that is, as factually, but not subjunctively true. As a criterion of empirical significance, this would amount to the demand that a relation be equivalent to a $\mathscr{B}$-formula in one specific (the actual) model of Θ. Goodman points out that this reading would counter any claims that Carnap's approach is an extreme form of reductionism that could be based on the intensional equivalence claims. Of course, there are other ways to define empirical significance that avoid intensional equivalence. Luce (1978, §4), for example, suggests to consider a relation P_i over the domain of a relational first order $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ to be meaningful if and only if it is invariant under all of $\mathfrak{A}_{\mathscr{B}}$'s automorphisms. And he notes that in some cases, so invariant relations are still not $\mathscr{B}$-definable (Luce 1978, 6). As I am interested in theories, I will consider P_i meaningful if and only if it is so invariant in all models of Θ (Luce's definition can be recovered if it is assumed that Θ determines $\mathfrak{A}_{\mathscr{B}}$ up to isomorphism). Θ is here assumed to contain only relations.

Definition 7.2. A relation P_i is *$\mathscr{B}$-symmetric* in a relational first order theory Θ if and only if for each $\mathfrak{A} \vDash \Theta$, $P_i^{\mathfrak{A}}$ is invariant under all automorphisms of $\mathfrak{A}|_{\mathscr{B}}$.

Luce connects $\mathscr{B}$-symmetry with dimensional invariance (roughly, he shows under some additional assumptions that a law is dimensionally invariant if and only if all its relations are $\mathscr{B}$-symmetric). Its connection to $\mathscr{B}$-definability can be elucidated starting from a simple observation. For a $\mathscr{B}$-definable relation P_i, there is one $\mathscr{B}$-formula that is equivalent to P_i in all models of Θ. Exchanging the existential and the universal quantifier weakens the concept in a transparent way:

Definition 7.3. A relation P_i is *$\mathscr{B}$-definable in every model of Θ* if and only if for each $\mathfrak{A} \vDash \Theta$ there is a $\mathscr{B}$-formula β such that $\mathfrak{A} \vDash \forall \bar{x}\left[P_i \bar{x} \leftrightarrow \beta(\bar{x})\right]$.

A relation $\mathscr{B}$-definable in every model of Θ is thus extensionally equivalent to some (not necessarily the same) $\mathscr{B}$-formula in every model of Θ, and this is already a generalization of $\mathscr{B}$-symmetry (since Θ is here not assumed to be relational):

Claim 7.3. *Let Θ be a relational first order theory in vocabulary $\mathscr{B} \cup \{P_i\}$. Then P_i is $\mathscr{B}$-definable in every model of Θ if and only if P_i is $\mathscr{B}$-symmetric in Θ.*

Proof. Hodges (1993, corollary 10.5.2). □

A corollary of claim 7.3 further elucidates the relation of $\mathscr{B}$-definability in each model (and thus of $\mathscr{B}$-symmetry) to $\mathscr{B}$-definability.

Definition 7.4. A relation P_i is *piecewise $\mathscr{B}$-definable* in Θ if and only if there is a set $\{\beta_j\}_{1 \leq j \leq n}$ of $\mathscr{B}$-formulas such that

$$\Theta \vDash \bigvee_{j=1}^{n} \forall \bar{x} \left[P_i \bar{x} \leftrightarrow \beta_j(\bar{x}) \right] . \tag{7.2}$$

Claim 7.4. *In first order logic, a relation P_i is $\mathscr{B}$-definable in every model of Θ if and only if P_i is piecewise $\mathscr{B}$-definable in Θ.*

Proof. Hodges (1993, corollary 10.5.3). □

Through claim 7.1, $\mathscr{B}$-definability ensures translatability into $\mathscr{B}$-terms, which can be taken to be a very strong criterion of empirical significance for sentences.[6] $\mathscr{B}$-definability in all models amounts to a slightly weaker criterion of empirical significance for sentences.

Claim 7.5. *If every $\mathscr{A}$-relation in the $\mathscr{V}$-sentence σ is $\mathscr{B}$-definable in every model of Θ, then σ is strongly semantically $\mathscr{B}$-determined by Θ.*

Proof. A $\mathscr{B}$-formula φ is true either in both or neither of any two models of Θ with the same $\mathscr{B}$-reduct under the same variable assignment. For an induction on the complexity of formulas, assume first that $\varphi\bar{x} = P_i\bar{x}$. Then there is, for every model of Θ, some $\mathscr{B}$-formula β that is coextensive with $\varphi\bar{x}$. Under the same variable assignment, either both or neither of any two models of Θ with the same $\mathscr{B}$-reduct are therefore models of $\varphi\bar{x}$. Thus, for the inductive step, there is a variable assignment under which $\varphi\bar{x}$ is true in a model of Θ if and only if there is a variable assignment (the same) under which $\varphi\bar{x}$ is true in any other model of Θ with the same $\mathscr{B}$-reduct. The inductive step for universal quantification is analogous, and for the logical connectives immediate. Since σ is a formula without free variables, variable assignments are irrelevant. □

[6] Even if, for example following Carnap (1958, 273), $\mathscr{B}$ is assumed to be closed under definability, claim 7.6 below establishes that translatability does not run afoul of condition of adequacy 5 in §6.9, however.

Because of claim 7.1, every sentence that contains only $\mathscr{B}$-definable or $\mathscr{B}$-terms can be translated into a $\mathscr{B}$-sentence. Thus $\mathscr{B}$-definability provides a sufficient condition for a strong connection of a sentence to $\mathscr{B}$-sentences and thus also $\mathscr{B}$-structures. However, neither $\mathscr{B}$-definability nor $\mathscr{B}$-definability in a model provide a *necessary* condition for translatability. Taking the *non-vacuous occurrence* of P_i in a sentence σ to mean that σ is not logically equivalent to a sentence in which P_i does not occur, it is easy to prove

Claim 7.6. *There are sentences σ that can be translated into $\mathscr{B}$-sentences by Θ and in which relations occur non-vacuously that are neither $\mathscr{B}$-definable in Θ nor $\mathscr{B}$-definable in every model of Θ.*

Proof. Choose $\Theta \mathrel{\vDash\!\!\dashv} \{\forall x(Bx \to P_i x)\}$ and $\sigma \mathrel{\vDash\!\!\dashv} Bb \wedge P_i b$. Then P_i is not $\mathscr{B}$-definable in every model and thus not $\mathscr{B}$-definable, and σ is not logically equivalent to a sentence in which P_i does not occur. But σ is equivalent to Bb given Θ, and can thus be translated into a $\mathscr{B}$-sentence. □

7.3 Reducibility and its variants

According to Carnap (1936, 440), explicit definability is too strict a criterion for empirical significance because he claims (but does not prove) that disposition concepts cannot be explicitly defined. To ensure the empirical significance of disposition concepts, which he takes to be indispensable in science, Carnap (1935a, §8) suggests to define empirical significance as reducibility.

Definition 7.5. Relation P_i is *reducible to $\mathscr{B}$* in Θ if and only if there is some $\mathscr{B}$-formula β such that

$$\Theta \nvDash \neg\exists\bar{x}\beta(\bar{x})\,, \tag{7.3}$$

as well as

$$\Theta \vDash \forall\bar{x}[\beta(\bar{x}) \to P_i\bar{x}] \tag{7.4}$$

or

$$\Theta \vDash \forall\bar{x}[\beta(\bar{x}) \to \neg P_i\bar{x}]\,. \tag{7.5}$$

In either case, I will call β a *reduction formula* for P_i, and the two conditions (7.4) and (7.5) *reduction sentences*. It is far from clear that reducibility to $\mathscr{B}$ suffices to analyze the meaning of disposition concepts (cf. Belnap 1993, 136–137; Malzkorn 2001, §2.1). But meaning differs from empirical meaningfulness (see §5.2), and since only the latter is explicated by empirical significance, reducibility to $\mathscr{B}$ may still be a criterion of empirical significance. For one, it is obvious that every $\mathscr{B}$-definable relation is also reducible to $\mathscr{B}$ (with the two reduction formulas being contradictories and thus at least one of them fulfilling the existence condition (7.3)). Reducibility is thus a straightforward weakening of a criterion of empirical significance that is usually considered too strong.

Wójcicki (1966, 84) proceeds from Carnap's considerations towards another criterion of empirical significance. He first suggests

Definition 7.6. The *positive extension of P_i relative to Θ* in a $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ is

$$P_i^{\mathfrak{A}_{\mathscr{B}},\Theta,+} = \bigcap\{P_i^{\mathfrak{B}} \mid \mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}} \text{ and } \mathfrak{B} \vDash \Theta\} . \tag{7.6}$$

The *negative extension of P_i relative to Θ* in in a $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ is

$$P_i^{\mathfrak{A}_{\mathscr{B}},\Theta,-} = \bigcap\{\complement P_i^{\mathfrak{B}} \mid \mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}} \text{ and } \mathfrak{B} \vDash \Theta\} . \tag{7.7}$$

Since the positive and the negative extension of P_i relative to Θ are intersections of normal extensions of P_i, $P_i^{\mathfrak{A}_{\mathscr{B}},\Theta,+}$ and $P_i^{\mathfrak{A}_{\mathscr{B}},\Theta,-}$ are disjoint extensions of the same type as P_i, and thus positive and, respectively, negative extensions in the sense of Przełęcki (see §2.8.3). Accordingly, Wójcicki (1966, 87) draws attention to the relation between empirical significance and first order vagueness. In extreme cases of vagueness, a relation has empty positive and negative extensions.

Definition 7.7. A relation P_i is *completely $\mathscr{B}$-vague in Θ* if and only if for all $\mathfrak{A} \vDash \Theta$, $P_i^{\mathfrak{A}|_{\mathscr{B}},\Theta,+} = \varnothing$ and $P_i^{\mathfrak{A}|_{\mathscr{B}},\Theta,-} = \varnothing$.

Claim 7.2 shows that no definable relation is completely $\mathscr{B}$-vague. A first relation between complete $\mathscr{B}$-vagueness and reducibility is straightforward.

Claim 7.7. *P_i is reducible to $\mathscr{B}$ in Θ only if P_i is not completely vague in Θ.*

Proof. If P_i is reducible to $\mathscr{B}$ in Θ, then there is an $\mathscr{B}$-formula β such that $\Theta \nvDash \neg\exists\bar{x}\beta(\bar{x})$ so that for some $\mathfrak{A} \vDash \Theta$, $\beta^{\mathfrak{A}}$ is nonempty, and either $\Theta \vDash \forall\bar{x}[\beta(\bar{x}) \to \neg P_i\bar{x}]$ or $\Theta \vDash \forall\bar{x}[\beta(\bar{x}) \to P_i\bar{x}]$. In the latter case, for every $\mathfrak{B} \vDash \Theta$, $\beta^{\mathfrak{B}} \subseteq P_i^{\mathfrak{B}}$. Therefore, for any $\mathfrak{B} \vDash \Theta$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$, $P_i^{\mathfrak{B}} \neq \varnothing$. The proof for $\Theta \vDash \forall\bar{x}[\beta(\bar{x}) \to \neg P_i\bar{x}]$ and $\complement P_i^{\mathfrak{B}} \neq \varnothing$ is analogous. □

Claim 7.2 provides a semantic criterion for $\mathscr{B}$-definability, and analogously, claim 7.7 provides a semantic criterion for the reducibility to $\mathscr{B}$. If P_i is $\mathscr{B}$-definable and Θ is finite, $\forall\bar{x}\left(P_i\bar{x} \leftrightarrow \exists\bar{X}[\Theta(\bar{B},\bar{X}) \wedge X_i\bar{x}]\right)$ is a $\mathscr{B}$-definition of P_i, and is entailed by Θ (see the proof of claim 7.2). To arrive at an analogous result for reducibility to $\mathscr{B}$, one can use a suggestion by Martin (1966), who defines P_i's Ramsey constant $R^{\Theta}_{\mathscr{B},P_i}$ for finite Θ as $\forall\bar{x}\left(R^{\Theta}_{\mathscr{B},P_i}\bar{x} \leftrightarrow \exists\bar{X}[\Theta(\bar{B},\bar{X}) \wedge X_i\bar{x}]\right)$.[7] Even though Martin's use of Ramsey constants is flawed (Scheffler 1968, 272–273; Bohnert 1971b), the concept itself is not useless; for instance, for any $\mathscr{B}$-definable relation P_i, $R^{\Theta}_{\mathscr{B},P_i}$ is P_i's definiens in $\mathscr{B}$. To arrive at reduction sentences for

[7] Technically, a Ramsey constant $R^{\Theta}_{\mathscr{B},P_i}$ can thus be seen as the generalized predicate $\lambda x\exists\bar{X}[\Theta(\bar{B},\bar{X}) \wedge X_i\bar{x}]$.

P_i, define the reverse Ramsey constant $Я^{\Theta}_{\mathscr{B},P_i}$ by $\forall\bar{x}\,(Я^{\Theta}_{\mathscr{B},P_i}\bar{x} \leftrightarrow \exists\bar{X}\,[\Theta(\bar{B},\bar{X}) \wedge \neg X_i\bar{x}])$. That the concepts of the (reverse) Ramsey constant will figure in the reduction sentences for P_i is suggested by the close relation between the positive and negative extensions of P_i and its Ramsey constants:

Claim 7.8. *For finite Θ and every $\mathfrak{A}$, $\left(\neg Я^{\Theta}_{\mathscr{B},P_i}\right)^{\mathfrak{A}} = P_i^{\mathfrak{A}|_{\mathscr{B}},\Theta,+}$ and $\left(\neg R^{\Theta}_{\mathscr{B},P_i}\right)^{\mathfrak{A}} = P_i^{\mathfrak{A}|_{\mathscr{B}},\Theta,-}$.*

Proof. $\neg Я^{\Theta}_{\mathscr{B},P_i}\bar{x} \mathrel{\models\!\!\!\dashv} \lambda\bar{x}\,(\forall\bar{X}[\Theta(\bar{B},\bar{X}) \rightarrow X_i\bar{x}])$. Thus, if $\mathfrak{A} \not\models \Theta$, $\left(\neg Я^{\Theta}_{\mathscr{B},P_i}\right)^{\mathfrak{A}} = |\mathfrak{A}| = P_i^{\mathfrak{A}|_{\mathscr{B}},\Theta,+}$. If $\mathfrak{A} \models \Theta$, then $\left(\neg Я^{\Theta}_{\mathscr{B},P_i}\right)^{\mathfrak{A}} = \lambda\bar{x}\,(\forall\bar{X}[\Theta(\bar{B},\bar{X}) \rightarrow X_i\bar{x}])^{\mathfrak{A}} = \bigcap\left\{P_i^{\mathfrak{B}} \mid \mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}} \text{ and } \mathfrak{B} \models \Theta\right\} = P_i^{\mathfrak{A}|_{\mathscr{B}},\Theta,+}$ by lemma 2.4. The proof for $\neg R^{\Theta}_{\mathscr{B},P_i}$ and $P_i^{\mathfrak{A}|_{\mathscr{B}},\Theta,-}$ is analogous. □

From claims 7.2 and 7.8 it follows that if P_i is $\mathscr{B}$-definable, then $\forall\bar{x}\,(Я^{\Theta}_{\mathscr{B},P_i}\bar{x} \leftrightarrow \neg R^{\Theta}_{\mathscr{B},P_i}\bar{x})$. Claim 7.8 has a simple corollary:

Corollary 7.9. *For finite Θ, a relation P_i is completely $\mathscr{B}$-vague in Θ if and only if $\Theta \models \forall\bar{x}Я^{\Theta}_{\mathscr{B},P_i}\bar{x}$ and $\Theta \models \forall\bar{x}R^{\Theta}_{\mathscr{B},P_i}\bar{x}$.*

Proof. From definition 7.6 and claim 7.8. □

The Ramsey sentence $\mathsf{R}_{\mathscr{B}}(\Theta)$ of a finite theory Θ entails the same $\mathscr{B}$-sentences that Θ entails, and the negation of the reverse Ramsey sentence, $\neg\mathsf{R}_{\mathscr{B}}(\neg\bigwedge\Theta)$, is entailed by the same $\mathscr{B}$-sentences that entail Θ (see §6.3). In a nice analogy, the Ramsey constant of P_i entails the same $\mathscr{B}$-formulas as P_i when assuming Θ,[8] and the negation of the reverse Ramsey constant is entailed by the same $\mathscr{B}$-formulas as P_i when assuming Θ:

Lemma 7.10. *For finite Θ and any $\mathscr{B}$-formula β*

$$\Theta \models \forall\bar{x}[\beta(\bar{x}) \rightarrow P_i\bar{x}] \textit{ if and only if } \models \forall\bar{x}[\beta(\bar{x}) \rightarrow \neg Я^{\Theta}_{\mathscr{B},P_i}\bar{x}]\,, \tag{7.8}$$

and

$$\Theta \models \forall\bar{x}[\beta(\bar{x}) \rightarrow \neg P_i\bar{x}] \textit{ if and only if } \models \forall\bar{x}[\beta(\bar{x}) \rightarrow \neg R^{\Theta}_{\mathscr{B},P_i}\bar{x}]\,. \tag{7.9}$$

Proof. I will prove the first conjunct of the lemma. The proof of the second conjunct is analogous.

'⇒': Assume that the structure $\mathfrak{A}$ and the variable assignment ν are such that $\mathfrak{A},\nu \not\models \neg Я^{\Theta}_{\mathscr{B},P_i}\bar{x}$. Then $\mathfrak{A},\nu \not\models \forall\bar{X}\left[\Theta(\bar{B},\bar{X}) \rightarrow X_i\bar{x}\right]$, that is, for some $\mathfrak{B} \models \Theta$ with

[8] Formula (7.9) gives the contrapositive of this claim.

$\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$ it holds that $\mathfrak{B}, \nu \nvDash P_i\bar{x}$. Hence $\mathfrak{B}, \nu \nvDash \beta(\bar{x})$. Therefore, since β is a $\mathscr{B}$-formula and $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$, it holds that $\mathfrak{A}, \nu \nvDash \beta(\bar{x})$. By contraposition, every $\mathfrak{A} \vDash \Theta$ is such that $\mathfrak{A} \vDash \forall\bar{x}[\beta(\bar{x}) \rightarrow \neg Я^{\Theta}_{\mathscr{B},P_i}\bar{x}]$.

'$\Leftarrow$': Assume that $\mathfrak{A}$ and the variable assignment ν are such that $\mathfrak{A}, \nu \vDash \beta(\bar{x})$. Then $\mathfrak{A}, \nu \vDash \forall\bar{X}\left[\Theta(\bar{B}, \bar{X}) \rightarrow X_i\bar{x}\right]$, that is, for every $\mathfrak{B} \vDash \Theta$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$ it holds that $\mathfrak{B}, \nu \vDash P_i\bar{x}$. Hence, specifically, if $\mathfrak{A} \vDash \Theta$, then $\mathfrak{A}, \nu \vDash P_i\bar{x}$. □

I can now give a necessary and sufficient condition for the reducibility to $\mathscr{B}$:

Claim 7.11. *For finite Θ, P_i is reducible to $\mathscr{B}$ in Θ only if $\Theta \nvDash \forall\bar{x}Я^{\Theta}_{\mathscr{B},P_i}\bar{x}$ or $\Theta \nvDash \forall\bar{x}R^{\Theta}_{\mathscr{B},P_i}\bar{x}$. If higher order sentences are also observational, the converse holds as well.*

Proof. '$\Rightarrow$': Immediately from claims 7.7 and 7.9.

'$\Leftarrow$': In higher order logic, $\neg Я^{\Theta}_{\mathscr{B},P_i}\bar{x}$ is a $\mathscr{B}$-formula. Since trivially $\forall\bar{x}\left(\neg Я^{\Theta}_{\mathscr{B},P_i}\bar{x} \rightarrow \neg Я^{\Theta}_{\mathscr{B},P_i}\bar{x}\right)$, by lemma 7.10, $\Theta \vDash \forall\bar{x}\left(\neg Я^{\Theta}_{\mathscr{B},P_i}\bar{x} \rightarrow P_i\bar{x}\right)$. And if $\Theta \nvDash \forall\bar{x}Я^{\Theta}_{\mathscr{B},P_i}\bar{x}$, then $\Theta \nvDash \neg\exists\bar{x}\neg Я^{\Theta}_{\mathscr{B},P_i}\bar{x}$. Analogously for $\neg R^{\Theta}_{\mathscr{B},P_i}$. □

In their discussion of partial structures, Bueno (1997, 592–593) and French and Ladyman (1999, 105) leave out the primary statements (the analogue of the penumbral connections for vague terms, see definitions 2.4 and 4.11), the sentences in the object language that have to be true in every $\mathfrak{A}$-normal structure, in part to illustrate the irrelevance of the vocabulary for the representation of scientific knowledge. As Bueno (1997, 593, n. 5) puts it, "the partial relations presented in the [partial structure] seem to be enough to represent (relevant) aspects of our knowledge of its domain".[9] Treating partial relations as $\mathscr{A}$-relations, this suggests

Definition 7.8. A relation P_i is *relevant in Θ* if and only if for some $\mathfrak{A} \vDash \Theta$, $\langle P_i^{\mathfrak{A}|_{\mathscr{B}},\Theta,+}, P_i^{\mathfrak{A}|_{\mathscr{B}},\Theta,-}\rangle \neq \langle\varnothing,\varnothing\rangle$.

The definition captures Bueno's claim because without penumbral connections, a relation with empty positive and negative extensions can be interpreted in any way whatsoever. It therefore does not represent any knowledge about the domain. It is now trivial to prove

Claim 7.12. *A relation P_i is relevant in Θ if and only if P_i is not completely $\mathscr{B}$-vague.*

Starting from Bueno's assumption that all relevant scientific knowledge can be represented in partial structures, the claims 7.11, 7.9, and 7.12 can be seen as

[9]Bueno further claims that if the primary statements turn out to be relevant, they could be incorporated into the background assumptions. However, he does not specify how this would, formally, differ from not leaving them out in the first place.

a justification of Carnap's criterion of empirical significance as given by definition 7.5. For they amount to the claim that an auxiliary relation is reducible only if (in higher order logic also if) it can be represented in a partial structure. And with Bueno's claim that the partial relations suffice to represent relevant aspects of our knowledge of a partial structure's domain, they therefore entail that reducible relations suffice to represent our relevant knowledge.

There is a possible objection to Carnap's criterion, however: The relation $\lambda xy(x = y)$ or, for that matter, $\lambda x(x = x)$ are not completely $\mathscr{B}$-vague, but are not in any interesting way related to observations. Thus Wójcicki (1966, 87) suggests

Definition 7.9. In first order logic and for a relational theory Θ, relation P_i is in Θ *$\mathscr{B}$-vague up to logical relations* if and only if in all $\mathfrak{A} \vDash \Theta$, the positive and negative extensions of P_i are invariant under all permutations of $|\mathfrak{A}|$.

This definition does not only include relations that, like identity and self-identity, are permutation invariant in all models of Θ and thus logical constants according to one line of thought (cf. MacFarlane 2009, §5). It also includes relations that change under permutations in some models of Θ, but whose definite positive and negative extension is permutation invariant.

Claim 7.13. *For finite Θ, relation P_i is vague up to logical relations if and only if in every model of Θ, $\text{Я}^{\Theta}_{\mathscr{B},P_i}$ and $R^{\Theta}_{\mathscr{B},P_i}$ can be defined by logical constants.*

Proof. From claim 7.3, since a relation is in all models $\mathfrak{A}$ of Θ invariant under all permutations on $|\mathfrak{A}|$ if and only if it is $\varnothing$-symmetric in Θ. □

It is clear from claims 7.13 and 7.3 that some $\mathscr{B}$-definable and $\mathscr{B}$-symmetric relations are vague up to logical relations. Since the converse holds as well, there is no straightforward inferential relation between vagueness up to logical relations and the other two criteria.

The relation between the criteria for terms of this section and the criteria for sentences is complicated. For instance, a sentence σ whose terms meet the criteria of this section can fail to meet the weakest deductive criterion of empirical significance for sentences:

Claim 7.14. *For some finite sets Θ of sentences and sentences σ whose relations are reducible to $\mathscr{B}$ in Θ and not vague up to logical relations, σ is not weakly semantically $\mathscr{B}$-determined relative to Θ.*

Proof. Choose $\mathscr{B} = \{B, b\}$, $\mathscr{A} = \{P_1, P_2, P_3, P_4\}$, $\Theta \mathrel{\vDash\!\dashv} \{\forall x(Bx \rightarrow P_1 x \land P_2 x), \forall x(\neg Bx \rightarrow P_3 x \land P_4 x)\}$ and $\sigma \mathrel{\vDash\!\dashv} (P_1 b \land P_3 b) \lor (\neg P_2 b \land \neg P_4 b)$. Then all terms in σ are reducible to $\mathscr{B}$ and not vague up to logical relations. But σ is not semantically falsifiable relative to Θ according to claim 6.7:

$$\vDash \forall x[Bx \to \lambda y(y=y)x \land Bx] \land \forall x[\neg Bx \to \lambda y(y=y)x \land Bx] \\ \land \left([\lambda y(y=y)b \land \lambda y(y=y)b] \lor [\neg Bb \land \neg Bb]\right) \tag{7.10a}$$

$$\vDash \exists \bar{X}\left(\forall x[Bx \to X_1x \land X_2x] \land \forall x[\neg Bx \to X_3x \land X_4x] \\ \land [(X_1b \land X_3b) \lor (\neg X_2b \land \neg X_4b)]\right) \tag{7.10b}$$

$$\vDash \mathsf{R}_{\mathscr{B}}(\Theta \cup \{\sigma\})\,. \tag{7.10c}$$

σ is also not semantically verifiable relative to Θ according to claim 6.11:

$$\vDash \forall x[Bx \to Bx \land \lambda y(y=y)x] \land \forall x[\neg Bx \to \neg Bx \land \lambda y(y=y)x] \\ \land [\neg Bb \lor Bb] \land [\lambda y(y=y)b \lor \lambda y(y=y)b] \tag{7.11a}$$

$$\vDash \exists \bar{X}\left(\forall x[Bx \to X_1x \land X_2x] \land \forall x[\neg Bx \to X_3x \land X_4x] \\ \land [\neg X_1b \lor \neg X_3b] \land [X_2b \lor X_4b]\right) \tag{7.11b}$$

$$\vDash \mathsf{R}_{\mathscr{B}}(\Theta \cup \{\neg\sigma\})\,. \tag{7.11c}$$

By claim 6.21, σ is not weakly semantically $\mathscr{B}$-determined relative to Θ. □

Conversely, some sentences that contain exclusively relations that fail to meet any of the criteria in this section are *translatable* into $\mathscr{B}$-sentences, and thus also strongly $\mathscr{B}$-determined.

Claim 7.15. *For some finite sets Θ of sentences and sentences σ whose relations are not reducible to $\mathscr{B}$ in Θ and vague up to logical relations, σ can be translated into a $\mathscr{B}$-sentence by Θ.*

Proof. Choose $\Theta \coloneqq \{\forall x(P_1x \leftrightarrow Bx) \lor \forall x(P_2x \leftrightarrow Bx)\}$ and $\sigma \coloneqq \exists x\forall y(P_1y \to y=x) \lor \exists x\forall y(P_2y \to y=x)$. Then P_1 and P_2 are not reducible and are vague up to logical relations, but $\Theta \vDash \sigma \leftrightarrow \exists x\forall y(By \to y=x)$. □

The criteria in this section therefore provide neither a necessary nor a sufficient criterion for any of the criteria for sentences, at least if the condition is to hold for all sentences and if it is to be based solely on the status of the sentences' terms.[10]

7.4 $\mathscr{B}$-dependence and its variants

For a $\mathscr{B}$-reducible term, there is at least one object in the domain that is in its positive extension or its negative extension. But, for example, there is no such object for an auxiliary term A connected to a basic term by $\Theta = \{\forall x(Ax \leftrightarrow Bx) \lor \forall x(Ax \leftrightarrow \neg Bx)\}$, even though the extension of A cannot be freely chosen. That

[10] Cf. definition 7.19 below.

there are such restrictions on the interpretation of terms necessitates penumbral connections for vague terms. $\mathscr{B}$-dependence and its variants are criteria that determine whether there are any such penumbral restrictions on the extension of a term. I will first review a criterion by Wójcicki (1966) and prove its equivalence to a slight modification of a criterion by Carnap (1956b). Then I will compare the two criteria to a slightly but crucially more inclusive criterion by Rozeboom (1962).

7.4.1 Wójcicki

Wójcicki (1966, 94) gives a definition for the empirical significance of auxiliary relations that can also be phrased for all auxiliary terms.

Definition 7.10. The set of *$\mathscr{B}$-possible extensions of $V_i \in \mathscr{V}$* given $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ and Θ is

$$\mathbf{V_i}^{\mathfrak{A}_{\mathscr{B}},\Theta} = \{V_i^{\mathfrak{C}} \mid \mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}} \text{ and } \mathfrak{C} \models \Theta\} . \tag{7.12}$$

Definition 7.11. The term $A_i \in \mathscr{A}$ is *$\mathscr{B}$-dependent* in Θ if and only if there are $\mathscr{B}$-structures $\mathfrak{A}_{\mathscr{B}}$ and $\mathfrak{B}_{\mathscr{B}}$ with $|\mathfrak{A}_{\mathscr{B}}| = |\mathfrak{B}_{\mathscr{B}}|$ such that

$$\varnothing \neq \mathbf{A_i}^{\mathfrak{A}_{\mathscr{B}},\Theta} \neq \mathbf{A_i}^{\mathfrak{B}_{\mathscr{B}},\Theta} \neq \varnothing . \tag{7.13}$$

Definition 7.11 can be simplified as follows:

Claim 7.16. *A term $A_i \in \mathscr{A}$ is $\mathscr{B}$-dependent in Θ if and only if there are models $\mathfrak{A}, \mathfrak{B} \models \Theta$ with $|\mathfrak{A}| = |\mathfrak{B}|$ such that*

$$A_i^{\mathfrak{B}} \notin \mathbf{A_i}^{\mathfrak{A}|_{\mathscr{B}},\Theta} \tag{7.14}$$

Proof. There are models $\mathfrak{A}, \mathfrak{B} \models \Theta$ with $|\mathfrak{A}| = |\mathfrak{B}|$ and $A_i^{\mathfrak{A}} \notin \mathbf{A_i}^{\mathfrak{B}|_{\mathscr{B}},\Theta}$ iff there is some $\mathfrak{D} \models \Theta$ with $\mathfrak{D}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$ such that $A_i^{\mathfrak{A}} \notin \mathbf{A_i}^{\mathfrak{B}|_{\mathscr{B}},\Theta}$, and thus $\varnothing \neq \mathbf{A_i}^{\mathfrak{D}|_{\mathscr{B}},\Theta} \neq \mathbf{A_i}^{\mathfrak{B}|_{\mathscr{B}},\Theta} \neq \varnothing$ for some $\mathfrak{D}, \mathfrak{B} \models \Theta$ with $|\mathfrak{D}| = |\mathfrak{A}| = |\mathfrak{B}|$. □

The relation between $\mathscr{B}$-dependence and vagueness up to logical constants is given by

Claim 7.17. *Let Θ be a relational first order theory. If P_i is not vague up to logical relations, then P_i is $\mathscr{B}$-dependent in Θ.*

Proof. Let $\mathfrak{A} \models \Theta$. If P_i is not vague up to logical relations, then $P_i^{\mathfrak{A}|_{\mathscr{B}},\Theta,+}$ or $P_i^{\mathfrak{A}|_{\mathscr{B}},\Theta,-}$ is different from both $|\mathfrak{A}|$ and $\varnothing$. Without loss of generality, assume that $P_i^{\mathfrak{A}|_{\mathscr{B}},\Theta,+} \neq |\mathfrak{A}|$ and $P_i^{\mathfrak{A}|_{\mathscr{B}},\Theta,+} \neq \varnothing$. Choose $c \in P_i^{\mathfrak{A}|_{\mathscr{B}},\Theta,+}$ and $d \notin P_i^{\mathfrak{A}|_{\mathscr{B}},\Theta,+}$, and choose $\mathfrak{B}$ such that all its extensions are those of $\mathfrak{A}$, except that c and d are systematically switched. Then $\mathfrak{B} \simeq \mathfrak{A}$ and thus $|\mathfrak{A}| = |\mathfrak{B}|$ and $\mathfrak{B} \models \Theta$. Since $c \in P_i^{\mathfrak{A}}$, $c \notin P_i^{\mathfrak{B}}$. Therefore, since $c \in P_i^{\mathfrak{A}|_{\mathscr{B}},\Theta,+} = \bigcap \mathbf{P_i}^{\mathfrak{A}|_{\mathscr{B}},\Theta}$, it holds that $P_i^{\mathfrak{B}} \notin \mathbf{P_i}^{\mathfrak{A}|_{\mathscr{B}},\Theta}$. By claim 7.16, P_i is $\mathscr{B}$-dependent. □

However, there are relations that are not $\mathscr{B}$-dependent even though they are not completely $\mathscr{B}$-vague.

Claim 7.18. *For some Θ and P_i, P_i is not completely $\mathscr{B}$-vague but also not $\mathscr{B}$-dependent.*

Proof. Choose $\Theta = \{\forall x P_1 x\}$, $\{B\} = \mathscr{B}$. Then $P_1^{\mathfrak{A}|_{\mathscr{B}},\Theta,+} = |\mathfrak{A}|$ for all $\mathfrak{A} \models \Theta$, so that P_1 is not completely vague. But $\mathbf{P_i}^{\mathfrak{A}|_{\mathscr{B}},\Theta} = \{|\mathfrak{A}|\}$ for every $\mathfrak{A} \models \Theta$, so that for all $\mathfrak{A}, \mathfrak{B} \models \Theta$ with $|\mathfrak{A}| = |\mathfrak{B}|$, $P_i^{\mathfrak{B}} = |\mathfrak{B}| = |\mathfrak{A}| \in \{|\mathfrak{A}|\} = \mathbf{P_i}^{\mathfrak{A}|_{\mathscr{B}},\Theta}$. □

For finite sets of sentences, Wójcicki's definition has a simple formulation in terms of Ramsey sentences that is based on

Lemma 7.19. *A_i is $\mathscr{B}$-dependent with respect to Θ if and only if for some $\mathfrak{A}$ and some $\mathfrak{B}, \mathfrak{C} \models \Theta$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$ and $\mathfrak{C}|_{\{A_i\}} = \mathfrak{A}|_{\{A_i\}}$, there is no $\mathfrak{D} \models \Theta$ such that $\mathfrak{A}|_{\mathscr{B}\cup\{A_i\}} = \mathfrak{D}|_{\mathscr{B}\cup\{A_i\}}$.*

Proof. A_i is $\mathscr{B}$-dependent with respect to Θ iff for some $\mathfrak{E}, \mathfrak{C} \models \Theta$ with $|\mathfrak{E}| = |\mathfrak{C}|$, $\mathbf{A_i}^{\mathfrak{E}|_{\mathscr{B}},\Theta} \neq \mathbf{A_i}^{\mathfrak{C}|_{\mathscr{B}},\Theta}$, that is, iff there is some $\mathfrak{B} \models \Theta$, $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{E}|_{\mathscr{B}}$ such that for all $\mathfrak{D} \models \Theta$, $\mathfrak{D}|_{\mathscr{B}} = \mathfrak{C}|_{\mathscr{B}}$ only if $\mathfrak{D}|_{\{A_i\}} \neq \mathfrak{B}|_{\{A_i\}}$. This holds iff there are some $\mathfrak{B}, \mathfrak{C} \models \Theta$ with $|\mathfrak{B}| = |\mathfrak{C}|$ such that for no $\mathfrak{D} \models \Theta$, $\mathfrak{D}|_{\mathscr{B}} = \mathfrak{C}|_{\mathscr{B}}$ and $\mathfrak{D}|_{\{A_i\}} = \mathfrak{B}|_{\{A_i\}}$.

'$\Rightarrow$': For $\mathfrak{B}$ and $\mathfrak{C}$ as above, choose $\mathfrak{A}|_{\mathscr{B}} = \mathfrak{C}|_{\mathscr{B}}$ and $\mathfrak{A}|_{\{A_i\}} = \mathfrak{B}|_{\{A_i\}}$. Then there is no $\mathfrak{D} \models \Theta$ with $\mathfrak{D}|_{\mathscr{B}} = \mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$ and $\mathfrak{D}|_{\{A_i\}} = \mathfrak{B}|_{\{A_i\}} = \mathfrak{A}|_{\{A_i\}}$, that is, no $\mathfrak{D} \models \Theta$ with $\mathfrak{D}|_{\mathscr{B}\cup\{A_i\}} = \mathfrak{A}|_{\mathscr{B}\cup\{A_i\}}$.

'$\Leftarrow$': Let $\mathfrak{A}$ be such that there are $\mathfrak{B}, \mathfrak{C} \models \Theta$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$ and $\mathfrak{C}|_{\{A_i\}} = \mathfrak{A}|_{\{A_i\}}$, and for all $\mathfrak{D} \models \Theta$, $\mathfrak{D}|_{\mathscr{B}\cup\{A_i\}} \neq \mathfrak{A}|_{\mathscr{B}\cup\{A_i\}}$. Then $|\mathfrak{B}| = |\mathfrak{C}|$, and for all $\mathfrak{D} \models \Theta$, $\mathfrak{D}|_{\mathscr{B}} \neq \mathfrak{A}|_{\mathscr{B}}$ or $\mathfrak{D}|_{\{A_i\}} \neq \mathfrak{A}|_{\{A_i\}}$, that is, there is no $\mathfrak{D} \models \Theta$ with $\mathfrak{D}|_{\mathscr{B}} = \mathfrak{B}|_{\mathscr{B}}$ and $\mathfrak{D}|_{\{A_i\}} = \mathfrak{C}|_{\{A_i\}}$. □

Now it is straightforward to describe definition 7.11 in terms of Ramsey sentences:

Claim 7.20. *For finite Θ, $A_i \in \mathscr{A}$ is $\mathscr{B}$-dependent in Θ if and only if*

$$\mathsf{R}_{\mathscr{B}}(\Theta) \wedge \mathsf{R}_{\{A_i\}}(\Theta) \not\models \mathsf{R}_{\mathscr{B}\cup\{A_i\}}(\Theta)\,. \tag{7.15}$$

Proof. From lemmas 7.19 and 2.4. □

7.4.2 Carnap

Carnap (1956b) suggests a criterion that, roughly, demands that the $\mathscr{A}$-term A_i feature essentially in the derivation of an observation sentence. It seems to have escaped notice that for his proof that the criterion is not too exclusive, Carnap assumes that A_i can be primitively interpreted. Using $\mathscr{B}$-structures instead of observation sentences, I will show that under this assumption, Carnap's criterion is equivalent to $\mathscr{B}$-dependence.

The following is a special case of Carnap's definition (Carnap 1956b, 51):

Definition 7.12. A term A_i is *directly syntactically Carnap-significant* with respect to $\mathscr{B}$ and Θ if and only if the following holds: $A_i \in \mathscr{A}$ and there are a $\mathscr{B}$-sentence β and an $\{A_i\}$-sentence α_{A_i}, such that

1. $\alpha_{A_i} \wedge \Theta \nvDash \bot$,
2. $\alpha_{A_i} \wedge \Theta \vDash \beta$, and
3. $\Theta \nvDash \beta$.

Considering claim 7.22, Carnap-significance is very close to Rozeboom's definition 7.14: A relation is not given effective meaning if and only if, when it occurs in a $\mathscr{B} \cup \{A_i\}$-sentence that is semantically $\mathscr{B}$-conservative relative to the empty set, the sentence is also semantically $\mathscr{B}$-conservative relative to Θ. A relation is not Carnap-significant if and only if, when it occurs in an $\{A_i\}$-sentence that is compatible with Θ, the sentence is also syntactically $\mathscr{B}$-conservative relative to Θ. Thus, instead of semantic $\mathscr{B}$-conservativeness, specifically of a $\mathscr{B} \cup \{A_i\}$-sentence, Carnap-significance relies on syntactic conservativeness and consistency with Θ.

One major problem with definition 7.12 as it is stated is that, in contradiction to Carnap's intent (Carnap 1956b, 53, see §3.1), it is not logically weaker than definition 7.5. For assume

$$\Theta = \{\exists x_1 x_2[x_1 \neq x_2 \wedge \forall y(y = x_1 \vee y = x_2)], \tag{7.16a}$$
$$\exists x[Bx \wedge \forall y(By \rightarrow x = y)], \tag{7.16b}$$
$$\forall x(Px \longleftrightarrow Bx)\} . \tag{7.16c}$$

P is explicitly defined through a $\mathscr{B}$-term, but the only sentences that contain only P are either incompatible with Θ or are syntactically $\mathscr{B}$-conservative relative to Θ.

The solution to this apparent inconsistency in Carnap's claims is that Carnap treats mathematical constants as logical constants and allows for mathematical constants to have physical meaning and appear as arguments of auxiliary relations. When Carnap (1956b, 59, my notation) argues that definition 7.12 is not too narrow, he considers a specific example in which one might *think* that it is too narrow, and argues that in this case,

> there must be a possible distribution of values of M for the space-time points of the region a', which is compatible with Θ [...]. Let

> 'F' be a logical constant, designating a mathematical function which represents such a value distribution. Then we take the following sentence as α_{A_i}: 'For every space-time point in a', the value of A_i is equal to that of F.' This sentence α_{A_i} is compatible with Θ. [...] Then α_{A_i} contains 'A_i' as the only descriptive term [...]. [...] β is logically implied [...], according to our assumption, [...] by $\alpha_{A_i} \wedge \Theta$.

Carnap thus assumes that all mathematical terms are logical terms and can be identified with theoretical terms. This assumption seems to lead to a host of problems. For one, if two theoretical terms have the same values, they are identified with the same function and are thus identical, which may lead to trouble if they are related to different observation terms. It may thus be difficult to individuate theoretical terms, and may require a reformulation of many scientific theories (assuming that such a reformulation is even possible).

Without looking at this specific assumption about $\mathscr{A}$-terms, Carnap's proof relies on the possibility of giving an interpretation $A_i^{\mathfrak{A}}$ to A_i so that every model $\mathfrak{C}$ with that interpretation is a model of β, that is, $\mathfrak{C} \not\models \neg\beta$. Instead of observation sentence $\neg\beta$, take a set of possible observation sentences, and assume that the set can determine possible $\mathscr{B}$-structures up to isomorphism. Assuming further that $\mathfrak{A}$ also determines the domain of $\mathfrak{C}$ allows moving from Carnap's syntactic definition to

Definition 7.13. A term $A_i \in \mathscr{A}$ is *directly semantically Carnap-significant* with respect to $\mathscr{B}$ and Θ if and only if there are structures $\mathfrak{A}, \mathfrak{B} \models \Theta$ with $|\mathfrak{A}| = |\mathfrak{B}|$ such that for all $\mathfrak{C} \models \Theta$ with $\mathfrak{C}|_{\{A_i\}} = \mathfrak{A}|_{\{A_i\}}$, $\mathfrak{C}|_{\mathscr{B}} \neq \mathfrak{B}|_{\mathscr{B}}$.

Direct semantic Carnap-significance is equivalent to $\mathscr{B}$-dependence:

Corollary 7.21. *A term $A_i \in \mathscr{A}$ is directly semantically Carnap-significant with respect to $\mathscr{B}$ and Θ if and only if A_i is $\mathscr{B}$-dependent with respect to Θ.*

Proof. By claim 7.16, a term is $\mathscr{B}$-dependent iff there are models $\mathfrak{A}, \mathfrak{B} \models \Theta$ with $|\mathfrak{A}| = |\mathfrak{B}|$ such that $A_i^{\mathfrak{B}} \notin \mathbf{A}_\mathrm{i}^{\mathfrak{A}|_{\mathscr{B}},\Theta}$. This holds iff for all $\mathfrak{C} \models \Theta$ with $\mathfrak{C}|_{\{A_i\}} = \mathfrak{B}|_{\{A_i\}}$, $\mathfrak{B}|_{\mathscr{B}} \neq \mathfrak{A}|_{\mathscr{B}}$, which holds iff for all $\mathfrak{C} \models \Theta$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$, $\mathfrak{C}|_{\{A_i\}} \neq \mathfrak{B}|_{\{A_i\}}$. □

7.4.3 Rozeboom

Rozeboom (1962, 354, postulate 7) gives

Definition 7.14. A term $A_i \in \mathscr{A}$ is *given effective meaning* with respect to $\mathscr{B}$ by a finite set Θ of sentences if and only if

$$\exists \bar{X} \Theta(\bar{B}, \bar{X}) \not\models \forall X_i \exists \bar{X}_{-i} \Theta(\bar{B}, \bar{X}) \,. \tag{7.17}$$

The following holds:

Claim 7.22. *A term $A_i \in \mathscr{A}$ is given effective meaning by a finite set Θ if and only if*

$$\mathsf{R}_{\mathscr{B}}(\Theta) \not\models \mathsf{R}_{\mathscr{B}\cup\{A_i\}}(\Theta)\,. \tag{7.18}$$

Proof. $\exists\bar{X}\Theta(\bar{B},\bar{X}) \models \forall X_i \exists \bar{X}_{-i}\Theta(\bar{B},\bar{X})$ iff for all $\mathfrak{A}$ with a $\mathfrak{B} \models \Theta$ such that $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$, it holds that for every $\mathfrak{C}$ with $\mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$, $\mathfrak{C}|_{\mathscr{B}\cup\{A_i\}} = \mathfrak{A}|_{\mathscr{B}\cup\{A_i\}}$. By lemma 2.4, this is the case iff $\mathsf{R}_{\mathscr{B}}(\Theta) \models \mathsf{R}_{\mathscr{B}\cup\{A_i\}}(\Theta)$. □

Since $\mathsf{R}_{\mathscr{B}\cup\{A_i\}}(\Theta) \models \mathsf{R}_{\mathscr{B}}(\Theta)$, claim 7.22 establishes that A_i is not given effective meaning if and only if it occurs vacuously in $\mathsf{R}_{\mathscr{B}\cup\{A_i\}}(\Theta)$, that is, there is a logically equivalent formulation of $\mathsf{R}_{\mathscr{B}\cup\{A_i\}}(\Theta)$ that does not contain A_i. A term that is not given effective meaning is of no use for the assertion of basic sentences, since adding Θ as set of supplementary sentences does not lead to any new basic assertions:

Claim 7.23. *If $A_i \in \mathscr{A}$ is not given effective meaning by the finite set Θ, then any $\mathscr{B}\cup\{A_i\}$-sentence σ that is semantically $\mathscr{B}$-conservative relative to $\varnothing$ is semantically $\mathscr{B}$-conservative relative to Θ.*

Proof. Since Θ is by assumption finite, σ is $\mathscr{B}$-conservative relative to Θ iff $\mathsf{R}_{\mathscr{B}}(\Theta) \models \mathsf{R}_{\mathscr{B}}(\Theta \wedge \sigma)$. Since A_i is not given effective meaning and σ is a $\mathscr{B}\cup\{A_i\}$-sentence, $\mathsf{R}_{\mathscr{B}}(\Theta) \wedge \sigma \mathrel{\models\!\!\!\dashv} \mathsf{R}_{\mathscr{B}\cup\{A_i\}}(\Theta \wedge \sigma)$. Since σ is semantically $\mathscr{B}$-conservative relative to $\varnothing$, $\models \mathsf{R}_{\mathscr{B}}(\sigma)$. Thus by claim 2.3, $\mathsf{R}_{\mathscr{B}}(\Theta) \models \mathsf{R}_{\mathscr{B}}\big(\mathsf{R}_{\mathscr{B}}(\Theta)\big) \models \mathsf{R}_{\mathscr{B}}\big(\mathsf{R}_{\mathscr{B}}(\Theta) \wedge \mathsf{R}_{\mathscr{B}}(\sigma)\big) \models \mathsf{R}_{\mathscr{B}}\big(\mathsf{R}_{\mathscr{B}}(\Theta) \wedge \sigma\big) \models \mathsf{R}_{\mathscr{B}}\big(\mathsf{R}_{\mathscr{B}\cup\{A_i\}}(\Theta \wedge \sigma)\big) \models \mathsf{R}_{\mathscr{B}}(\Theta \wedge \sigma)$. □

With claim 6.5, claim 7.23 states that for an A_i that is not given effective meaning by Θ, a $\mathscr{B}\cup\{A_i\}$-sentence σ is semantically falsifiable relative to Θ only if σ is semantically falsifiable relative to no supplementary assumptions.

Definition 7.14 can be generalized to infinite theories Θ with the help of $\mathscr{B}$-possible extensions (definition 7.10). Let $\top$ be a tautology.

Definition 7.15. The term $A_i \in \mathscr{A}$ is *$\mathscr{B}$-restricted* by Θ if and only if there is a model $\mathfrak{A} \models \Theta$ such that

$$\mathbf{A_i}^{\mathfrak{A}|_{\mathscr{B}},\Theta} \neq \mathbf{A_i}^{\mathfrak{A}|_{\mathscr{B}},\top}\,. \tag{7.19}$$

A relation that is not $\mathscr{B}$-restricted will be called *$\mathscr{B}$-unrestricted*. Because of claim 2.1, definition 7.15 could be equivalently rephrased by restricting the structures $\mathfrak{A}$ to elements of $\mathbf{M}$. $\mathscr{B}$-restrictedness indeed generalizes *giving effective meaning*:

Claim 7.24. *For finite Θ, the term $A_i \in \mathscr{A}$ is given effective meaning by Θ if and only if A_i is $\mathscr{B}$-restricted by Θ.*

Proof. From claim 7.22 and lemma 2.4. □

Lemma 7.19 provides the means to describe the relation between $\mathscr{B}$-dependence, $\mathscr{B}$-restrictedness, and *giving effective meaning* via

Claim 7.25. *A_i is $\mathscr{B}$-dependent in Θ only if it is $\mathscr{B}$-restricted by Θ. For finite Θ, A_i is $\mathscr{B}$-dependent only if it is given effective meaning by Θ.*

Proof. In lemma 7.19, there is no $\mathfrak{D} \vDash \Theta$ with $\mathfrak{D}|_{\{A_i\}} = \mathfrak{A}|_{\{A_i\}} = \mathfrak{C}|_{\{A_i\}}$ and $\mathfrak{D}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}} = \mathfrak{B}|_{\mathscr{B}}$. Thus $A_i^{\mathfrak{C}} \notin \mathbf{A_i}^{\mathfrak{B}|_{\mathscr{B}},\Theta}$ and hence $\mathbf{A_i}^{\mathfrak{A}|_{\mathscr{B}},\Theta} \neq \mathbf{A_i}^{\mathfrak{A}|_{\mathscr{B}},\top}$. The claim for finite Θ follows from 7.24. □

Giving effective meaning is a strictly more inclusive concept than *not being completely vague*:

Claim 7.26. *For finite Θ, if relation $P_i \in \mathscr{A}$ is not completely $\mathscr{B}$-vague in Θ, then P_i is given effective meaning with respect to $\mathscr{B}$ by Θ. The converse does not hold.*

Proof. Assume that P_i is not given effective meaning with respect to $\mathscr{B}$ by Θ. Thus every model $\mathfrak{A}$ of $\mathsf{R}_{\mathscr{B}}(\Theta)$ is also a model of $\mathsf{R}_{\mathscr{B}\cup\{P_i\}}(\Theta)$. Specifically, if $\mathfrak{A} \vDash \Theta$, then any $\mathfrak{B}$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$ is a model of $\mathsf{R}_{\mathscr{B}\cup\{P_i\}}(\Theta)$. Thus choose $\mathfrak{B}$ such that $P_i^{\mathfrak{B}} = \complement P_i^{\mathfrak{A}}$. Then there is a $\mathfrak{C} \vDash \Theta$ with $\mathfrak{C}|_{\mathscr{B}\cup\{P_i\}} = \mathfrak{B}|_{\mathscr{B}\cup\{P_i\}}$ and thus $P_i^{\mathfrak{C}} \cap P_i^{\mathfrak{A}} = \complement P_i^{\mathfrak{C}} \cap \complement P_i^{\mathfrak{A}} = \varnothing$. Thus $P_i^{\mathfrak{A},\Theta,+} = P_i^{\mathfrak{A},\Theta,-} = \varnothing$ for any $\mathfrak{A} \vDash \Theta$, and thus P_i is completely vague.

$\Theta = \{\forall \bar{x}(P_i\bar{x} \leftrightarrow B\bar{x}) \vee \forall \bar{x}(P_i\bar{x} \leftrightarrow \neg B\bar{x})\}$ with basic term B and auxiliary term P_i gives effective meaning to P_i even though P_i is completely vague. □

Thus *giving effective meaning* and $\mathscr{B}$-restrictedness are the most inclusive criteria for the empirical significance of terms (see figure 7.1).

7.5 Sets of empirically significant terms

The idea that criteria of empirical significance have to be recursive has stayed with philosophy of science at least since Ayer. Accordingly, Carnap attempted to extend reducibility and Carnap-significance to include a recursion, even though on the face of it, both criteria can be applied indiscriminately to any term that occurs in Θ. In this section, I will first argue that Carnap's recursive criteria are too weak, and suggest instead a simple generalization of $\mathscr{B}$-dependence (and thus direct semantic Carnap-significance) to sets of terms. Not $\mathscr{B}$-dependent sets of terms will turn out to have still some very slight relation to the interpretations of $\mathscr{B}$-terms, a relation that is absent in sets of $\mathscr{B}$-isolated terms, which will be introduced thereafter. A $\mathscr{B}$-vacuous set of terms finally will be one that is not at all restricted in its interpretations.

As a recursive definition of reducibility, Carnap (1935a, 446) gives

$$
\begin{tikzcd}[row sep=large, column sep=large]
\mathcal{B}\text{-definable} \arrow[d, Rightarrow] & \\
\begin{array}{c} \mathcal{B}\text{-definable in every model} \\ \Leftrightarrow \mathcal{B}\text{-symmetric} \Leftrightarrow \\ \text{piecewise } \mathcal{B}\text{-definable} \end{array} \arrow[d, Rightarrow] & \\
\begin{array}{c} \text{not vague up to logical relations} \Leftrightarrow \\ \text{Я}^{\Theta}_{\mathcal{B},P_i} \text{ or } R^{\Theta}_{\mathcal{B},P_i} \text{ not } \varnothing\text{-definable} \\ \text{in every model} \end{array} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \begin{array}{c} \mathcal{B}\text{-dependent} \Leftrightarrow \\ \text{directly semantically} \\ \text{Carnap-significant} \end{array} \arrow[d, Rightarrow] \\
\begin{array}{c} \text{reducible to } \mathcal{B} \Leftrightarrow \\ \text{not completely } \mathcal{B}\text{-vague} \Leftrightarrow \\ \Theta \not\models \forall \bar{x}\, \text{Я}^{\Theta}_{\mathcal{B},P_i} \bar{x} \text{ or } \Theta \not\models \forall \bar{x} R^{\Theta}_{\mathcal{B},P_i} \bar{x} \\ \Leftrightarrow \text{relevant} \end{array} \arrow[r, Rightarrow] & \begin{array}{c} \text{given effective meaning} \\ \Leftrightarrow \mathcal{B}\text{-restricted} \end{array}
\end{tikzcd}
$$

Figure 7.1: Entailment relations between the deductive criteria for the empirical significance of terms. Some of the relations hold only for special cases.

Definition 7.16. Relation P_i is *chain-reducible* to $\mathcal{B}$ in Θ if and only if there are relations $P_{i_1}, \ldots, P_{i_k}$ with $P_{i_k} = P_i$ such that for $1 < l \leq k$, P_{i_l} is reducible to $\mathcal{B} \cup \{P_{i_1}, \ldots, P_{i_{l-1}}\}$, and P_{i_1} is reducible to $\mathcal{B}$.

Since definition 7.16 is relative to a theory Θ, not all terms will be chain-reducible to $\mathcal{B}$, that is, the definition will not be trivial in the way that the recursive criteria for sentences are. But there are good reasons to think that definition 7.16 is much too weak:

Claim 7.27. *Some Θ contain relations that are chain-reducible to $\mathcal{B}$, but are not reducible to $\mathcal{B}$ and are not given effective meaning.*

Proof. Choose $\Theta = \{\forall x(Bx \to P_1 x), \forall x(\neg Bx \wedge P_1 x \to P_2 x)\}$. Then $\Theta \not\models \neg\exists x\, Bx$, so that P_1 is reducible to $\mathcal{B}$, and $\Theta \not\models \neg\exists x(\neg Bx \wedge P_1 x)$, so that P_2 is reducible to $\mathcal{B} \cup \{P_1\}$ and thus chain-reducible to $\mathcal{B}$. But P_2 is not given effective meaning with respect to $\mathcal{B}$ by Θ because

$$\models \forall x[Bx \to \lambda y(By \vee P_2 y)x] \wedge \forall x[\neg Bx \wedge \lambda y(By \vee P_2 y)x \to P_2 x] \tag{7.20a}$$

$$\models \exists X_1[\forall x(Bx \to X_1 x) \wedge \forall x(\neg Bx \wedge X_1 x \to P_2 x)] \tag{7.20b}$$

$$\models \mathsf{R}_{\mathcal{B}\cup\{P_2\}}(\Theta)\,. \tag{7.20c}$$

Therefore $\mathsf{R}_{\mathscr{B}}(\Theta) \vDash \mathsf{R}_{\mathscr{B}\cup\{P_2\}}(\Theta)$, and by claim 7.22, P_i is not given effective meaning. By claims 7.26 and 7.7, P_i is also not $\mathscr{B}$-reducible. □

Since some terms that are not given effective meaning are chain-reducible, one may have chain-reducible terms that are not restricted in their interpretation at all. And thanks to claim 7.23, any sentence that is not already $\mathscr{B}$-creative does not become $\mathscr{B}$-creative either when its auxiliary terms are chain-reduced by the introduction of a specially constructed Θ.

§7.4.2 contains only the first step of Carnap's recursive definition of empirical significance. The full definition is as follows (Carnap 1956b, 51):

Definition 7.17. A term A_i is *syntactically Carnap-significant relative to* the class $\mathscr{K}$ of terms with respect to $\mathscr{B}$ and Θ if and only if $\mathscr{K} \subseteq \mathscr{A}$, $A_i \in \mathscr{A}$, $A_i \notin \mathscr{K}$, and there are a $\mathscr{B}$-sentence β, an $\{A_i\}$-sentence α_{A_i}, and a $\mathscr{K}$-sentence $\alpha_{\mathscr{K}}$ such that

1. $\alpha_{A_i} \cup \alpha_{\mathscr{K}} \cup \Theta \nvDash \bot$,
2. $\alpha_{A_i} \cup \alpha_{\mathscr{K}} \cup \Theta \vDash \beta$, and
3. $\alpha_{\mathscr{K}} \cup \Theta \nvDash \beta$.

Definition 7.18. A term A_i is *syntactically Carnap-significant* with respect to $\mathscr{B}$ and Θ if and only if there is a sequence $A_{i_1}, \ldots, A_{i_{n-1}}, A_{i_n} = A_i \subseteq \mathscr{A}$ of relations such that A_{i_1} is directly syntactically Carnap-significant and for every $k, 1 < k \leq n$, A_{i_k} is syntactically Carnap-significant relative to $\{A_{i_l} \mid 1 \leq l \leq k\}$ with respect to $\mathscr{B}$ and Θ.

Given that Carnap-significance was meant to be weaker than reducibility, it is perhaps not surprising that relations that are not given effective meaning can be Carnap-significant (with Carnap's assumptions for his proof that Carnap-significance is not too narrow):

Claim 7.28. *Assuming that there is a term*[11] *c so that Ac is still an $\{A_i\}$-sentence for an $\mathscr{A}$-term A_i, some Θ contain relations that are Carnap-significant, but are not given effective meaning.*

Proof. Choose $\mathscr{B} = \{B_1, B_2, b\}$, $\mathscr{A} = \{P_1, P_2\}$, and $\Theta = \{\forall x(P_1 x \to B_1 b) \land \forall x(P_1 x \land P_2 x \to B_2 b)\}$. Then $P_1 c \land \Theta \vDash B_1 b$ and $\Theta \nvDash B_1 b$, so that P_1 is directly syntactically Carnap-significant. Furthermore, $P_1 c \land P_2 c \land \Theta \vDash B_2 b$ and $P_1 c \land \Theta \nvDash B_2 b$, so that P_1 is syntactically Carnap-significant. But

$$\vDash \forall x[\lambda y(y \neq y)x \to B_1 b] \land \forall x[\lambda y(y \neq y)x \land P_2 x \to B_2 b] \tag{7.21a}$$

$$\vDash \exists X_1[\forall x(X_1 x \to B_1 b) \land \forall x(X_1 x \land P_2 x \to B_2 b)] \tag{7.21b}$$

$$\vDash \mathsf{R}_{\mathscr{B}\cup\{P_2\}}(\Theta)\,, \tag{7.21c}$$

[11] In this one case, the term is also a term according to logical terminology.

so that P_2 is not given effective meaning. □

With a criterion of empirical significance for terms, Carnap defines a criterion for the significance of sentences:

Definition 7.19. $\mathscr{V}$-sentence σ is *syntactically Carnap-significant* relative to Θ if and only if every term in σ is syntactically Carnap-significant or a $\mathscr{B}$-term.

The following holds, however:

Corollary 7.29. *There are $\mathscr{B}$-conservative sentences relative to Θ that are syntactically Carnap-significant relative to Θ and $\mathscr{B}$-conservative sentences relative to Θ that contain only terms that are chain-reducible relative to Θ.*

Proof. From claim 7.23 and claims 7.27 and 7.28, respectively. □

Carnap (1935a, theorem 7) proves a relation between chain-reducible relations and the reducibility of sentences (see §6.6) containing them. Since his definition of the reducibility of sentences is trivial, corollary 7.29 is not particularly surprising as far as chain-reducibility is concerned.

In a response to Van Cleve (1971) and Kaplan (1975), Creath (1976) suggests criteria of empirical significance for terms and sentences that are weaker than syntactic Carnap-significance. Since it seems that Carnap-significance is already to weak, I will not discuss his criteria. Considering the simplification of $\mathscr{B}$-dependence by claim 7.16, it seems more promising to generalize $\mathscr{B}$-dependence by generalizing the concept of a *$\mathscr{B}$-possible extension* of terms (definition 7.10) to whole sets of terms:

Definition 7.20. The set of *$\mathscr{B}$-possible $\mathscr{I}$-structures* given $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ and Θ is

$$\boldsymbol{\mathscr{I}}^{\mathfrak{A}_{\mathscr{B}},\Theta} = \left\{ \mathfrak{C}|_{\mathscr{I}} \;\middle|\; \mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}} \text{ and } \mathfrak{C} \vDash \Theta \right\} . \tag{7.22}$$

Unlike $\mathscr{B}$-possible extensions, $\mathscr{B}$-possible $\mathscr{I}$-structures include the domain of $\mathfrak{A}_{\mathscr{B}}$. But since $\mathscr{B}$-possible extensions are defined relative to a $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ as well, this nothing but a technical convenience—definitions 7.15 and 7.11 as well as claim 7.16, which rely on $\mathscr{B}$-possible extensions, could also have been formulated with $\mathscr{B}$-possible $\{A_i\}$-structures. With definition 7.10 generalized, $\mathscr{B}$-dependence (as reformulated by claim 7.16) is also easily generalized:

Definition 7.21. The set $\mathscr{I} \subseteq \mathscr{A}$ of terms is *$\mathscr{B}$-dependent* with respect to $\mathscr{B}$ and Θ if and only if there are models $\mathfrak{A}, \mathfrak{B} \vDash \Theta$ with $|\mathfrak{A}| = |\mathfrak{B}|$ such that

$$\mathfrak{B}|_{\mathscr{I}} \notin \boldsymbol{\mathscr{I}}^{\mathfrak{A}|_{\mathscr{B}},\Theta} . \tag{7.23}$$

For finite sets Θ, definition 7.21 again has a neat formulation in terms of Ramsey sentences:

Claim 7.30. *For finite Θ, a set $\mathscr{I} \subseteq \mathscr{A}$ is $\mathscr{B}$-dependent with respect to Θ if and only if*

$$\mathsf{R}_{\mathscr{B}}(\Theta) \wedge \mathsf{R}_{\mathscr{I}}(\Theta) \not\models \mathsf{R}_{\mathscr{B}\cup\mathscr{I}}(\Theta)\,. \tag{7.24}$$

Proof. Analogous to the proof of claim 7.20. □

If a set $\mathscr{I}$ of $\mathscr{A}$-terms is not $\mathscr{B}$-dependent, its interpretation in a $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ is thus solely restricted by Θ and $|\mathfrak{A}_{\mathscr{B}}|$, and not at all by the interpretation of the $\mathscr{B}$-terms. However, $\mathscr{I}$ may subsequently become $\mathscr{B}$-dependent because of changes in Θ that are not ostensibly about any term in $\mathscr{I}$ at all:

Claim 7.31. *There is a set Θ of sentences and a set $\mathscr{I}$ of terms that is not $\mathscr{B}$-dependent in Θ, but $\mathscr{B}$-dependent in $\Theta \cup \Delta$, where Δ is a set of $\mathscr{V} - \mathscr{I}$-sentences.*

Proof. Choose $\mathscr{B} = \{B\}$, $\mathscr{A} = \{A_1, A_2\}$, $\Theta = \{\forall x(A_1 x \leftrightarrow A_2 x)\}$, $\Delta = \{\forall x(Bx \leftrightarrow A_1 x)\}$, and $\mathscr{I} = \{A_2\}$. □

Wójcicki (1966, definition 6) suggests a sufficient condition for empirical non-significance for sets of terms where such a surreptitious restriction of the interpretations is not possible:

Definition 7.22. A set $\mathscr{I}$ of $\mathscr{A}$-terms is *$\mathscr{B}$-isolated* in Θ if and only if for any two models $\mathfrak{A}, \mathfrak{B}$ of Θ, every $\mathfrak{C}$ with $\mathfrak{C}|_{\mathscr{V}-\mathscr{I}} = \mathfrak{A}|_{\mathscr{V}-\mathscr{I}}$ and $\mathfrak{C}|_{\mathscr{I}} = \mathfrak{B}|_{\mathscr{I}}$ is also a model of Θ.

Claim 7.32. *The set $\mathscr{I} \subseteq \mathscr{A}$ of terms is not $\mathscr{B}$-isolated in Θ if and only if there are models $\mathfrak{A}, \mathfrak{B} \models \Theta$ with $|\mathfrak{A}| = |\mathfrak{B}|$ such that*

$$\mathfrak{B}|_{\mathscr{I}} \notin \boldsymbol{\mathscr{I}}^{\mathfrak{A}|_{\mathscr{V}-\mathscr{I}},\Theta}\,. \tag{7.25}$$

Proof. Straightforward from definitions 7.22 and 7.20. □

It is thus clear that no $\mathscr{B}$-isolated set of terms is $\mathscr{B}$-dependent. And unlike in the case of $\mathscr{B}$-dependence, a $\mathscr{B}$-isolated set $\mathscr{I}$ of terms can only cease to be $\mathscr{B}$-isolated because of changes in Θ that are explicitly about terms in $\mathscr{I}$:

Claim 7.33. *If $\mathscr{I}$ is $\mathscr{B}$-isolated in Θ, then $\mathscr{I}$ is $\mathscr{B}$-isolated in $\Theta \cup \Delta$ for any set Δ of $\mathscr{V} - \mathscr{I}$-sentences.*

Proof. Assume $\mathscr{I}$ is $\mathscr{B}$-isolated in Θ and $\mathfrak{A}, \mathfrak{B} \models \Theta \cup \Delta$. Then $\mathfrak{A}, \mathfrak{B} \models \Theta$ and thus $\mathfrak{C}$ with $\mathfrak{C}|_{\mathscr{V}-\mathscr{I}} = \mathfrak{A}|_{\mathscr{V}-\mathscr{I}}$ and $\mathfrak{C}|_{\mathscr{I}} = \mathfrak{B}|_{\mathscr{I}}$ is a model of Θ. Since Δ contains only $\mathscr{V} - \mathscr{I}$-sentences and $\mathfrak{C}|_{\mathscr{V}-\mathscr{I}} = \mathfrak{A}|_{\mathscr{V}-\mathscr{I}} \models \Delta$, $\mathfrak{C} \models \Theta \cup \Delta$. □

As was to be expected, $\mathscr{B}$-isolation has for finite sets of sentences a paraphrase in terms of Ramsey sentences:

Claim 7.34. *For finite Θ, a set $\mathscr{I}$ of $\mathscr{A}$-terms is $\mathscr{B}$-isolated in Θ if and only if*

$$\mathsf{R}_{\mathscr{V}-\mathscr{I}}(\Theta) \wedge \mathsf{R}_{\mathscr{I}}(\Theta) \vDash \Theta \,. \tag{7.26}$$

Proof. Immediate from lemma 2.4. □

Wójcicki further gives

Definition 7.23. A *term $A_i \in \mathscr{A}$ is $\mathscr{B}$-isolated* in Θ if and only if it is a member of a set of terms that is $\mathscr{B}$-isolated in Θ.

A relation between $\mathscr{B}$-isolation and $\mathscr{B}$-dependence is given by

Claim 7.35. *If a relation is $\mathscr{B}$-isolated, then it is not $\mathscr{B}$-dependent.*

Proof. Wójcicki (1966, proof of theorem 5.1). □

$\mathscr{B}$-isolation is related to sentences by

Claim 7.36. *Assume $\mathscr{I}$ is $\mathscr{B}$-isolated in Θ. If a $\mathscr{B} \cup \mathscr{I}$-sentence σ is semantically $\mathscr{B}$-conservative relative to $\varnothing$ and compatible with Θ, then σ is semantically $\mathscr{B}$-conservative relative to Θ.*

Proof. Let $\mathfrak{A}_{\mathscr{B}}$ be such that there is a $\mathfrak{B} \vDash \Theta$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$. By assumption, there is a $\mathfrak{C} \vDash \Theta \cup \{\sigma\}$. Again by assumption, for $\mathfrak{D}$ with $\mathfrak{D}|_{\mathscr{B}} = \mathfrak{B}|_{\mathscr{B}}$ and $\mathfrak{D}|_{\mathscr{I}} = \mathfrak{C}|_{\mathscr{I}}$, $\mathfrak{D} \vDash \Theta \cup \{\sigma\}$. Thus $\mathfrak{A}_{\mathscr{B}}$ can be expanded to a model of $\Theta \cup \{\sigma\}$. □

Furthermore, the following holds:

Claim 7.37. *Some $\mathscr{B}$-isolated terms are $\mathscr{B}$-restricted.*

Proof. Choose $\Theta = \{\forall x Bx, \forall x Ax\}$ with $\mathscr{B} = \{B\}$ and $\mathscr{A} = \{A\}$. Since $\mathsf{R}_{\mathscr{B}}(\Theta) \mathrel{\vDash\!\!\!\dashv} \forall x Bx$ and $\mathsf{R}_{\mathscr{A}}(\Theta) \mathrel{\vDash\!\!\!\dashv} \forall x Ax$, A is $\mathscr{B}$-isolated by claim 7.34, but $\mathscr{B}$-restricted by claims 7.22 and 7.24. □

To arrive for $\mathscr{B}$-restrictedness at an analogue to $\mathscr{B}$-isolation, I suggest

Definition 7.24. A set $\mathscr{I}$ *occurs $\mathscr{B}$-vacuously* (*is $\mathscr{B}$-vacuous*) in Θ if and only if for each $A_i \in \mathscr{I}$ and $\mathfrak{A} \vDash \Theta$,

$$\mathbf{A_i}^{\mathfrak{A}|_{\mathscr{V}-\mathscr{I}},\Theta} = \mathbf{A_i}^{\mathfrak{A}|_{\mathscr{V}-\mathscr{I}},\top} \,. \tag{7.27}$$

That the relation between $\mathscr{B}$-dependence and $\mathscr{B}$-isolation is analogous to the relation between $\mathscr{B}$-restrictedness and definition 7.24 becomes clear when comparing claim 7.20 with claim 7.34, and claim 7.22 with

Claim 7.38. *For finite Θ, a set $\mathscr{I}$ of $\mathscr{A}$-terms is $\mathscr{B}$-*vacuous *in Θ if and only if*

$$\mathsf{R}_{\mathscr{V}-\mathscr{I}}(\Theta) \vDash \Theta \,. \tag{7.28}$$

$$\mathscr{B}\text{-dependent} \Longrightarrow \text{not } \mathscr{B}\text{-isolated} \Longrightarrow \text{not } \mathscr{B}\text{-vacuous}$$

Figure 7.2: Entailment relations between the deductive criteria for the empirical significance of sets of terms.

Proof. Similar to the proof of claim 7.22. □

If (and only if) a set $\mathscr{I}$ of terms is $\mathscr{B}$-vacuous in Θ, there is a set $\Theta' \models\!\mid \Theta$ of sentences that does not contain terms from $\mathscr{I}$; in other words, the terms of $\mathscr{I}$ occur only trivially: Any set of terms can be so introduced into any set of sentences simply by equivalent paraphrase. Clearly, every $\mathscr{B}$-vacuous term is $\mathscr{B}$-unrestricted and also $\mathscr{B}$-isolated. The relation to sentences is given by

Claim 7.39. *Let $\mathscr{I}$ occur $\mathscr{B}$-vacuously in Θ. If a $\mathscr{B} \cup \mathscr{I}$-sentence σ is semantically $\mathscr{B}$-conservative relative to $\varnothing$, then σ is $\mathscr{B}$-conservative relative to Θ.*

Proof. As the proof for claim 7.36, except without the assumption that σ is compatible with Θ. □

In summary (see figure 7.2), if a set $\mathscr{I}$ of $\mathscr{A}$-terms is not $\mathscr{B}$-dependent, its interpretation is not restricted by the interpretation of $\mathscr{B}$-terms, but only by the domain and by Θ. If $\mathscr{I}$ is $\mathscr{B}$-isolated, its interpretation is just as unrestricted, and will even stay so if $\mathscr{V} - \mathscr{I}$-sentences are added to Θ. Finally, if $\mathscr{I}$ is $\mathscr{B}$-vacuous, its interpretation is only restricted by the domain. Of course, since the interpretation of each term in a $\mathscr{B}$-vacuous set $\mathscr{I}$ can be freely chosen, $\mathscr{I}$ can also be interpreted in every domain. Having thus arguably reached the apex of irrelevance, it is time to discuss the role these different criteria can play.

7.6 The points of the criteria

Most deductive criteria for the empirical significance of sentences were defined relative to the set of analytic sentences. Ayer attempted to take non-analytic sentences into account essentially by, first, defining empirical significance relative to an arbitrary set Θ of sentences, and second, restricting the sentences that can be members of Θ. It is this second step that resulted in the triviality of Ayer's own criterion and many of its successors. All criteria for terms that I know of, on the other hand, define empirical significance relative to an arbitrary consistent set of sentences. It thus seems that all criteria for terms are attempts to solve a problem significantly simpler than the one the criteria for sentences were supposed to solve. This mismatch becomes especially obvious in Carnap's criterion (definitions 7.21, 7.17, and 7.18), in which Carnap (1956b, 49) *assumes* that the conditions 1–3 for α_{A_i} ensure that α_{A_i} "makes a difference for the prediction of an observable event". And the conditions are simply that α_{A_i} be compatible with and syntactically $\mathscr{B}$-creative relative to Θ, so that at the core of Carnap's criterion lies the assumption that $\mathscr{B}$-creativity is the correct deductive criterion of empirical significance for sentences. But on the basis of his criterion for terms, Carnap then defines a *new* criterion for sentences (definition 7.19), again *relative to* Θ. It is thus wholly unclear why one cannot just keep syntactic $\mathscr{B}$-creativity as a criterion for sentences.

Assuming with Hempel (1965c, §3) that all criteria for terms are meant to lead to criteria for sentences via an analogue of Carnap's definition 7.19, all the criteria for terms therefore beg the question that Ayer's recursive criterion was meant to answer. And if the question is just what a criterion for sentences *relative to some given set of sentences* looks like, then there are already good answers without the detour over criteria for terms (as argued in §6.8.2).

Of course, a criterion for terms may still be used as a means to determine whether some specific criterion for sentences is fulfilled. However, the criteria for terms provide only necessary or sufficient conditions. A sentence that contains only $\mathscr{B}$-definable terms is translatable into $\mathscr{B}$, and a sentence containing only terms definable in every model is strongly semantically $\mathscr{B}$-dependent. But there are sentences with terms neither definable nor definable in every model that can be translated into $\mathscr{B}$-sentences. Thus the strongest criteria for terms provide sufficient but no necessary conditions for the empirical significance of sentences according to the strongest criteria for sentences. The weakest criteria provide necessary but no sufficient criteria for the weakest criteria for sentences: A sentence containing only terms that are not given effective meaning is not $\mathscr{B}$-creative, but even sentences that contain only $\mathscr{B}$-dependent terms can fail to be weakly $\mathscr{B}$-determined. Reducibility and its variants provide neither necessary nor sufficient conditions for any criterion for sentences.

Given this state of affairs, deductive criteria of empirical significance for terms seem like solutions in search of a problem. A good start for finding out what the criteria were meant to achieve in the first place should be their justifications when first introduced, but these turn out to be mostly raw intuitions. When introducing his last criterion, for example, Carnap (1956b, VI) relies on his description of the explicandum and two criteria of adequacy that he does not defend. In his discussion of reducibility, Carnap (1935a, 1936) stops at stating his definitions.

This, of course, does not mean that there cannot be a justification for any of the criteria. For example, even though reducibility does not relate in an interesting way to any of the criteria for sentences, it may be considered partly justified by Peirce's pragmatist maxim (Peirce 1878, II):

> Consider what effects, which might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of those effects is the whole of our conception of the object.

And Peirce (1878, III) clarifies:

> Let us illustrate this rule by some examples; and, to begin with the simplest one possible, let us ask what we mean by calling a thing hard. Evidently that it will not be scratched by many other substances. The whole conception of this quality, as of every other, lies in its conceived effects.

Thus a concept consists of the effects that follow from the concept applying to an object (cf. Hookway 2010, §2). If what one conceives based on a theory is what is entailed by that theory, then an effect described by β belongs to a concept P introduced by a theory Θ if and only if for an object c, $\Theta \vDash Pc \rightarrow \beta$. Since the effect may depend on the object to which the concept applies, this means $\Theta \vDash \forall \bar{x}[P\bar{x} \rightarrow \beta(\bar{x})]$, so that, for example, it holds for every object x that x is not scratched by many objects if x is hard. The formula also allows for concepts relating different objects so that, for example, one may say for any x and y that x scratches y if x is harder than y. Since the effect should not be trivially connected to the concepts, one should demand that $\Theta \nvDash \forall \bar{x}\beta(\bar{x})$.

Therefore a concept is non-empty only if condition (7.3) of the definition 7.5 of reducibility is met. From claim 7.10 and specifically formula (7.9), it follows that the concept of a term P consists of its Ramsey constant $R^{\Theta}_{\mathscr{B},P}$ for finite Θ, and from claim 7.11 it follows that a concept is non-empty only if $\Theta \nvDash \forall \bar{x} R^{\Theta}_{\mathscr{B},P}\bar{x}$. Peirce's justifications of the pragmatist maxim (cf. Hookway 2010, §2) thus amount to a partial defense of Carnap's criterion of reducibility. The problem with the pragmatist maxim again seems to be that, as in the case of reducibility to $\mathscr{B}$, some sentences that contain only non-empty concepts have no $\mathscr{B}$-implications, as the example $\Theta \boxminus \{\forall x(Px \rightarrow Bx)\}$, $\sigma \boxminus \neg Pb$ shows.

Given the dearth of justifications and the lack of need for a criterion for terms, I expect it to be more fruitful to investigate the features of the criteria to find out what they might be good for. Claim 7.20, for example, points to a somewhat subtle difference. While a term that is $\mathscr{B}$-unrestricted can be assigned any interpretation whatsoever, a term that is not $\mathscr{B}$-dependent can be assigned its interpretation no matter what the interpretation of the $\mathscr{B}$-terms is like. In other words, to determine whether an interpretation fits a term that is not $\mathscr{B}$-dependent, one only needs to know the formalism Θ. To determine whether an interpretation fits a $\mathscr{B}$-unrestricted term, one does not need to know anything at all.

In this respect, the criteria for sets of terms are even more interesting. If a set $\mathscr{I}$ of terms occurs $\mathscr{B}$-vacuously in Θ, there is, in a sense, absolutely nothing gained by including $\mathscr{I}$ in the formulation of Θ. In fact, for any set $\mathscr{J}$ of terms not already occurring in Θ, there is a new theory $\Lambda \boxminus \Theta$ in which the terms of $\mathscr{J}$ occur vacuously. Specifically, in every model of Θ, every term in a vacuously occurring set $\mathscr{I}$ can be interpreted in any way whatsoever. A set $\mathscr{I}$ of $\mathscr{B}$-isolated terms, on the other hand, can be interpreted in any way whatsoever as long as the interpretations respect the relations between the terms of $\mathscr{I}$ as determined by Θ. In §4.1.3, corollary 4.7 was used to argue for the need for possible structures to connect sets of sentences to the world. Specifically, since a set of sentence alone can never distinguish between isomorphic structures, any set of the right cardinality can be made into a model of the sentences. In the case of $\mathscr{B}$-isolated terms, this means that given *any* model $\mathfrak{A} \vDash \Theta$ and a single intended interpretation $\mathfrak{N}_{\mathscr{V}-\mathscr{I}}$ of the $\mathscr{V}-\mathscr{I}$-terms, any bijection from $|\mathfrak{A}|$ to $\left|\mathfrak{N}_{\mathscr{V}-\mathscr{I}}\right|$ defines an interpretation of the $\mathscr{I}$-terms on $\mathfrak{N}_{\mathscr{V}-\mathscr{I}}$. Put somewhat more precisely, if a $\mathscr{V}-\mathscr{I}$-structure $\mathfrak{A}$ can be expanded to a model of Θ at all, the interpretation of the $\mathscr{I}$-terms is at best determined up to permutations of $\left|\mathfrak{N}_{\mathscr{V}-\mathscr{I}}\right|$. An analogous result also holds for sets of terms that are not $\mathscr{B}$-dependent: If a $\mathscr{B}$-structure $\mathfrak{N}_{\mathscr{B}}$ can be expanded to a model of Θ, the interpretations of the $\mathscr{I}$-terms are at best determined up to permutations of $|\mathfrak{N}_{\mathscr{B}}|$.

This discussion of $\mathscr{B}$-isolated and not $\mathscr{B}$-dependent sets of terms is very reminiscent of Demopoulos's argument against the Carnap sentence discussed in §2.10.3. And while Demopoulos's argument is invalid when applied to $\mathscr{A}$-terms in general, it can be applied verbatim to sets $\mathscr{I}$ of terms that are $\mathscr{B}$-isolated or not $\mathscr{B}$-dependent. For Demopoulos (2003, 387, my notation) argues that for a single intended $\mathscr{B}$-structure $\mathfrak{N}_{\mathscr{B}}$ and any model $\mathfrak{A} \vDash \Theta$, it is possible

> to extend the partial interpretation $[\mathfrak{N}_{\mathscr{B}}]$ to the theoretical vocabulary of Θ [here: $\mathscr{I}$] by letting each predicate of its theoretical vocabulary denote the image in N of its interpretation in $\mathfrak{A}$ under any one–one correspondence between A and N.

This construction is exactly the one given above. Demopoulos (2008, 381) further argues that the Carnap sentence

> is incapable of accurately representing the truth of theoretical claims because it takes their truth to collapse into satisfiability in a sufficiently large domain. This is hardly what we take the truth of theoretical claims to consist in, since we characteristically—and rightly—distinguish them from those of pure mathematics.

And unlike for $\mathscr{A}$-terms in general, Demopoulos's criticism applies directly to sets of terms that are not $\mathscr{B}$-dependent and *a fortiori* to $\mathscr{B}$-isolated sets of terms. Thus such sets of terms should be considered *mathematical*.

According to claims 7.31, it is possible to render a set $\mathscr{I}$ of terms $\mathscr{B}$-dependent by adding sentences to Θ that do not contain any $\mathscr{I}$-terms. And according to claim 7.33, this is impossible for $\mathscr{B}$-isolated sets of terms. Thus there can be two kinds of mathematical terms in a theory—one that is and will always stay mathematical absent direct reinterpretation, and one that may become non-mathematical in the course of the development of the theory. It may be just this difference in kind that has led to disagreements about the mathematical status of some concepts in empirical theories.

In conclusion, it seems that the suggested criteria for the empirical significance of terms are neither needed nor particularly helpful in determining whether some sentence is connected to basic sentences. On the other hand, it seems that almost by chance, some of the criteria provide a means to identify and distinguish between mathematical terms. For the further discussion of criteria of empirical significance, criteria for terms thus do not seem relevant. However, I will come back to them when discussing reduction and concept formation.

Chapter 8

Probabilistic criteria for sentences

Given the questionable status of deductive criteria for terms, I will not try to develop probabilistic criteria for terms. As a starting point for my discussion of probabilistic criteria for sentences, I will rely on the most recent previous discussion, that by Elliott Sober. Over the last two decades, Sober (1990, 1999, 2007, 2008) has developed and defended a contrastive probabilistic criterion of empirical significance called 'testability'. To evaluate Sober's criterion, I briefly discuss his objections to falsifiability (§8.1.1) and to a Bayesian definition of empirical significance (§8.1.2). To arrive at a precise and consistent definition of Sober's criterion, I discuss its implications, some of them unwanted (§8.2.1), and develop an interpretation of inequalities between possibly undefined probabilities that arguably follows from Sober's assumptions (§8.2.2). I will then argue that one of Sober's restrictions is wholly unjustified, (§8.3.1) and, more importantly, the other one is not strict enough to avoid a trivialization of the criterion (§8.3.2). In light of these problems, I suggest two modified versions of Sober's criterion, a relative one that is provably non-trivial, and—on the basis of honest supplementary sentences (definition 6.36)—an absolute one to which the trivialization proof for Sober's criterion does not apply (§8.3.3). However, even with these modifications, the resulting criterion does not fulfill all the conditions of adequacy that Sober himself endorses or relies on (§§8.4–8.5). Rather, the conditions uniquely determine a non-contrastive criterion that is consistent with falsifiability when both criteria can be applied (§8.6). Dropping a contentious condition of adequacy also allows the Bayesian criterion (§8.7).

8.1 Two criteria of empirical significance and their problems

Sober (2008, 154) conjectures that his criterion is “a step forward from the failed proposals of the logical positivists”, but this is misleading because the logical positivists wanted to distinguish between meaningful and meaningless sentences (cf. Carnap 1963c, §6.A). Sober (2010, 1), on the other hand, does not consider his definition of testability to provide a criterion of empirical significance precisely because “ ‘[e]mpirical significance’ suggests that a sequence of terms has meaning iff it is empirically testable”, a position to which he does not subscribe. Rather, Sober (2008, 149–150) argues, meaningfulness is a semantic concept, while testability is epistemic. And furthermore:

> It seems clear that meaningfulness and testability are different. I suppose that the sentence “undetectable angels exist” is untestable, but the sentence is not meaningless gibberish. We know what it says, what logical relations it bears to other statements, and we can discuss whether it is knowable; none of this would be possible if the string of words literally made no sense.

Therefore Sober is rather improving on the demarcation criterion by Popper, who wanted to distinguish empirical from non-empirical statements (see §5.2). I will follow Sober in the search for a demarcation criterion for empirical statements, but not in his choice of terminology. I think that ‘empirical significance’ differs from ‘meaningfulness’ enough to avoid confusion, and, as a technical term, is clearly meant as a placeholder for a concept that is yet to be explicated.

8.1.1 Falsifiability

In line with his search for a demarcation criterion, Sober (2007; 2008, §2.8) introduces his criterion as avoiding the problems of Popper’s falsifiability criterion. One problem with falsifiability, Sober (2008, 130, cf. 151) notes, is that no purely probabilistic statement is falsifiable: “Consider a simple example: the statement that a coin has a probability of .5 of landing heads each time it is tossed [. . .] is testable, but it does not satisfy Popper’s criterion”. To avoid this result, Popper (1935, ch. VIII) generalizes his criterion, considering a theory Θ falsified even if an event occurs that is possible but very improbable according to $\Theta \cup \Pi$, where Π here are the supplementary sentences. But this generalization runs into problems as well, as Sober (2002) explains: If a theory entails a basic sentence, then it allows a deductive inference via *modus ponens*, $\ulcorner\{X, X \rightarrow Y\} \vDash Y\urcorner$: The assumption of the theory X and the fact that it implies Y entail Y. Popper justifies his criterion of falsifiability by using the implication of a basic sentence in a *modus tollens*, $\ulcorner\{\neg Y, X \rightarrow Y\} \vDash \neg X\urcorner$. The justification of the probabilistic generalization of his criterion would thus have to rely on a probabilistic version of *modus tollens*, in which a theory X is false or at the very least improbable if $\Pr(Y \mid X)$ is high and Y is false. But as Sober (2002, 69–70) points out (cf. Sober 2008, §1.4):

> There is a "smooth [t]ransition" between probabilistic and deductive *modus ponens*; the minor premiss ("X") either ensures that Y is true, or makes Y very probable, depending on how the major premiss is formulated. In contrast, there is a radical discontinuity between probabilistic and deductive modus tollens. The *minor premiss* ("not-Y") guarantees that X is false in the one case, but has no implications whatever about the probability of X in the other.

Therefore, while Popper is right to infer from the fact that a theory entails a basic sentence that the theory is falsifiable, he cannot infer the same from the fact that it assigns a high probability to a basic sentence (and thus a low probability to the sentence's negation). Thus the generalization of falsifiability to probabilistic theories has not been justified.

8.1.2 Bayesian empirical significance

A more successful criterion of empirical significance for probabilistic theories has been suggested within Bayesianism, the position that non-deductive inferences should follow the rules of the probability calculus, and specifically that the confirmation of scientific theories should follow Bayes's theorem,[1]

$$\Pr(\Theta \mid \Omega \cup \Pi) = \frac{\Pr(\Omega \mid \Theta \cup \Pi)\Pr(\Theta \mid \Pi)}{\Pr(\Omega \mid \Theta \cup \Pi)\Pr(\Omega \mid \{\neg \bigwedge \Theta\} \cup \Pi)} \,. \tag{8.1}$$

$\Pr(\Theta \mid \Pi)$ is the probability of the theory given only the supplementary sentences, that is, before the set Ω of basic sentences is taken into account, and thus called the prior probability (of Θ). $\Pr(\Theta \mid \Omega \cup \Pi)$ is the probability of Θ after Ω is taken into account, and hence called the posterior probability.[2] $\Pr(\Omega \mid \Theta \cup \Pi)$ and $\Pr(\Omega \mid \{\neg \bigwedge \Theta\} \cup \Pi)$ are the likelihoods of Θ and $\{\neg \bigwedge \Theta\}$, respectively (for Ω).[3] Here and in the following I will assume that for all theories Θ and supplementary sentences Π it holds that $\Pr(\Theta \cup \Pi) > 0$, and thus specifically that Θ and Π are compatible ($\Theta \cup \Pi \not\models \bot$).

[1] I will assume that probabilities 'Pr' are assigned to sets of sentences, and in arguments for 'Pr' identify sentences and their singleton sets.

[2] Sober (2008, 8) calls $\Pr(\Theta)$ the prior and $\Pr(\Theta \mid \Omega)$ the posterior probability and discusses Bayesianism without supplementary sentences. But he also argues that $\Pr(\Omega \mid \Theta)$, which would be used to determine $\Pr(\Theta \mid \Omega)$, is, unlike $\Pr(\Omega \mid \Theta \cup \Pi)$, almost never defined. Prior and posterior probabilities therefore have to be defined relative to supplementary sentences, lest Bayesianism be empty.

[3] If Θ does not have a finite axiomatization, one also has to find an axiomatization that expresses that Θ is false (see n. 16 below.), and which can thus be substituted for $\neg \bigwedge \Theta$.

In Bayesianism, the confirmation of a theory Θ is defined as follows (Howson and Urbach 1993, §7a; Sober 2008, 15):

Definition 8.1. Set Ω of basic sentences *confirms Θ relative to Π in Bayesianism* if and only if

$$\Pr(\Theta \mid \Omega \cup \Pi) > \Pr(\Theta \mid \Pi)\,. \tag{8.2}$$

Ω disconfirms Θ relative to Π in Bayesianism if and only if

$$\Pr(\Theta \mid \Omega \cup \Pi) < \Pr(\Theta \mid \Pi)\,. \tag{8.3}$$

For Θ to be actually confirmed or disconfirmed, Ω has to be true. Ω then *tests Θ relative to Π in Bayesianism* if and only if Ω confirms or disconfirms Θ relative to Π in Bayesianism.

Sober (1999, 48) states for any relation R:

> If a set of observations provides a test of a proposition because it bears relation R to that proposition, then a proposition is testable when it is possible for there to be a set of observations that bears relation R to the proposition. Testing is to testability as dissolving is to solubility.

Since definition 8.1 determines what it is to test a theory, it therefore also determines a criterion of testability (cf. Sober 2008, 150)[4]:

Definition 8.2. *Basic sentences are relevant for* theory Θ *relative to* supplementary sentences Π if and only if there is a possible set of basic sentences Ω such that[5]

$$\Pr(\Theta \mid \Omega \cup \Pi) \neq \Pr(\Theta \mid \Pi)\,. \tag{8.4}$$

The sense of 'possible' in this definition and in the rest of this and the following three chapters (on intelligent design and confirmation) will be that a set of sentences is possible if and only if it is compatible with Π, where Π is the set of supplementary sentences in the specific context.

Sober (2008, 150, 24–30) rejects the Bayesian definitions of confirmation, disconfirmation, and testability. He argues that if Θ is, say, the theory of general relativity, it is well-nigh impossible to assign a probability to Θ, or assess the likelihood of $\{\neg \bigwedge \Theta\}$, so that $\Pr(\Theta \mid \Pi)$, $\Pr(\Omega \mid \{\neg \bigwedge \Theta\} \cup \Pi)$, and $\Pr(\Theta \mid \Omega \cup \Pi)$ are often undefined. Bayes theorem (8.1) then becomes unusable, and definition 8.2 very questionable.

---

[4] Sober (2008, 144–145) does not mention supplementary sentences in his discussion (and rejection) of the definition, but is very explicit about their relevance.

[5] I will always silently assume that for any occurring conditional probability $\Pr(\Lambda \mid \Gamma)$, $\Pr(\Gamma) \neq 0$.

8.2 Sober's criterion of empirical significance

8.2.1 A precise formulation

Sober (2008, 152) suggests a criterion of empirical significance that avoids all of the problems discussed so far. Unlike falsifiability, his criterion does not render all probabilistic theories empirically non-significant and does not rely on a faulty probabilistic generalization of *modus tollens* for its justification; unlike the Bayesian criterion, it does not rely on the probabilities of whole theories or on likelihoods of the negations of theories. The criterion is the following:[6]

> Hypothesis Θ can now be tested against hypothesis Λ if and only if there exist true auxiliary assumptions Π and an observation statement Ω such that (i) $\Pr(\Omega\,|\,\Theta \cup \Pi) \neq \Pr(\Omega\,|\,\Lambda \cup \Pi)$, (ii) we now are justified in believing Π, and (iii) the justification we now have for believing Π does not depend on believing that Θ is true or that Λ is true and also does not depend on believing that Ω is true (or that it is false).

For the reason given in n. 25 on page 257, I will speak of 'supplementary sentences' rather than 'auxiliary assumptions' in the following. Further, to identify Sober's notion of testability as a technical notion, I will refer to it as 'contrastive testability'.

Contrastive testability as defined in the above quote has some lacunae that I will point out and amend in this section. For example, Sober (2008, 151) remarks that in his definition of contrastive testability, the "word 'now' marks the fact that whether a proposition has observational implications depends on the rest of what we are justified in believing, and that can change". However, the use of the indexical 'now' does not define testability relative to time in general, but relative to the specific time of the utterance. Thus 'Θ can now be tested against Λ' is defined, but 'Θ will be testable against Λ within a decade' is not. 'Within a decade, the utterance "Θ can now be tested against Λ" will be true', on the other hand, is defined. To avoid such cumbersome formulations, one can define contrastive testability as the three-place predicate 'Θ can at time t be tested against Λ'. This achieves Sober's intention more explicitly.[7]

The other indexical term of the definition, 'we', is as crucial as 'now', as Sober's quote shows. Which beliefs are justified not only changes in time, but also changes from group to group. For example, Sober (1999, 49) discusses at length the special role that observations play in the justification of beliefs, and states: "The fact that dinosaurs, in a sense, are observable entities, while quarks, in a sense, are not, is irrelevant. The point is that we have *actually* observed neither". Since what has actually been observed differs from group to group and even from person to

[6] Here and in the rest of the chapter, I will always silently substitute my own notation in formulas.

[7] As in §6.8.2, I will suppress the reference to a specific time later on; but I want to note it explicitly at least once.

person, the term ‘we’ cannot be avoided in Sober’s definition. However, because ‘we’ occurs only in the definiens, Sober’s criterion for testability is technically not a definition but rather a claim—and a false one at that. This is because the criterion violates the demand that in an explicit definition, any free variable of the definiens must also occur free in the definiendum, and thus Sober’s criterion is creative (cf. Belnap 1993, 139): If for two theories Θ and Λ, one referent of ‘we’ fulfills the definiens at t, the definiendum applies to Θ and Λ at t. But then the definiendum applies to Θ and Λ at t no matter the referent of ‘we’, and thus any referent fulfills the definiens for Θ and Λ at t. Hence according to Sober’s criterion, it holds for any Ω, Θ, Λ, and Π that fulfill (i): If one group is justified in believing Π independently of Θ, Λ, Ω, and the falsity of Ω, then every group is.[8] To avoid this unintended implication, testability must be defined as both relative to time and relative to a group of people. It is thus a four-place predicate.

As it stands, restriction (iii) on the supplementary sentences sounds like the demand that the justification of Π must not depend on the fact that the truth of Θ, Λ or Ω or the falsity of Ω is content of our beliefs. But very few statements are justified by the having of a belief, so that condition (iii) would be almost empty if this was meant. The restriction is therefore probably better expressed as the demand that the justifications for Π must not depend on the fact that the belief in the truth of Θ, Λ, or Ω or the falsity of Ω *is justified*. For convenience, I will mostly drop the reference to beliefs in the following, and speak of justified sentences, rather than justified beliefs in the truth of propositions expressed by sentences.

Finally, the condition (i) on the likelihoods of Θ and Λ needs to be elucidated, given that Sober’s critique of the Bayesian criterion of empirical significance assumes that some likelihoods are undefined. *Prima facie*, one would expect that Θ and Λ cannot be tested against each other if and only if $\Pr(\Omega \mid \Theta \cup \Pi) = \Pr(\Omega \mid \Lambda \cup \Pi)$ for all Ω and Π that fulfill conditions (ii) and (iii). But this would mean that the lack of contrastive testability is transitive for any theories that are not used to justify their own or each other’s supplementary sentences. For assume that Θ cannot be contrastively tested against Λ, and Λ cannot be contrastively tested against Δ. Then for all Ω, $\Pr(\Omega \mid \Theta \cup \Pi) = \Pr(\Omega \mid \Lambda \cup \Pi)$ and $\Pr(\Omega \mid \Lambda \cup \Pi) = \Pr(\Omega \mid \Delta \cup \Pi)$, so that for all Ω, $\Pr(\Omega \mid \Theta \cup \Pi) = \Pr(\Omega \mid \Delta \cup \Pi)$. Thus Θ cannot be contrastively tested against Δ. This transitivity is incompatible with Sober’s remark that it is not clear that intelligent design (ID, see §9) “can be tested against the Epicurean hypothesis that a mindless chance process gave vertebrates their eyes (or, for that matter, against the evolutionary hypothesis that the process of evolution by natural selection did the work)” (Sober 2008, 148). Assuming that the chance hypothesis can be tested against evolutionary theory (ET), if ID cannot be tested against either one, the lack of testability is not transitive. The solution

[8] Keeping Sober’s indexical formulation, it holds for any Ω, Θ, Λ, and Π that fulfill (i): If for one group the claim ‘We are now justified in believing Π independently of Θ, Λ, Ω, or the falsity of Ω’ is true, then the claim is true for every group.

to this puzzle is that Sober (2010, 2–3) interprets the inequality as true if and only if both likelihoods are defined and different. This interpretation, however, plays havoc with classical logic, for $p \neq q \mathrel{\vDash\!\dashv} \neg\, p = q$. Therefore, if the likelihoods p and q are defined and different, while the likelihood a is undefined, it follows from Sober's interpretation of the inequality that $p = a$ and $a = q$, while $p \neq q$. To avoid such inconsistencies, it is probably best to treat undefined likelihoods separately in the definition.

Finally, Sober (2008, 148) assumes that the basic sentences that are relevant for testability are those that are possible given our background knowledge (cf. Sober 1999, 48–49). These considerations lead to

Definition 8.3. Theory Θ *is contrastively testable against* theory Λ if and only if there are supplementary sentences Π and a set of basic sentences Ω that are possible given our background knowledge such that

(I) $\Pr(\Omega \mid \Theta \cup \Pi)$ and $\Pr(\Omega \mid \Lambda \cup \Pi)$ are defined,

(II) $\Pr(\Omega \mid \Theta \cup \Pi) \neq \Pr(\Omega \mid \Lambda \cup \Pi)$,

(III) Π is justified, and

(IV) the justification of Π

a) does not depend on Θ or Λ being justified and

b) does not depend on the belief or the disbelief in Ω being justified.

One could reformulate definition 8.3 to include a reference to times and groups of people, that is, define 'Θ can be tested against Λ *at time* t *by group* g' by relativizing 'justification' (and possibly 'dependence') to t and g. In similar cases, especially when the supplementary sentences are simply the background assumptions, these relativizations are typically suppressed because it is clear that the background assumptions and generally the set of justified sentences can change over time and from group to group. Thus I will do likewise.

Sober calls his criterion simply 'testability', but the qualifier 'contrastive' distinguishes it clearly from the ordinary language term and emphasizes that, atypically, the empirical significance of one theory is defined relative to another. It may seem problematic to explicate a one-place predicate like 'makes observational assertions' by a two-place predicate like contrastive testability. Frege (1918, 291), for example, objects to the explication of 'truth' as a correspondence relation on the grounds that the first is a one-place, the second a two-place predicate. However, many successful explications involve a change of the logical structure, as the explication of 'warm' by 'warmer than' and finally 'temperature' illustrates (Carnap 1950b, §4). As already noted, Hempel (1952, §10) argues that the move from a classificatory to a comparative concept is often a sign of an investigation's maturity, and his criticism of empirical significance (Hempel 1965c, 117) is best seen as an argument for a comparative explicatum for empirical significance.

But unlike 'warmer than', contrastive testability is symmetric: The definiens is invariant up to logical equivalence under exchange of Θ and Λ. Thus contrastive testability does not provide a means to decide which of two theories is what could be called 'more empirically significant'. And this may be a problem. What is more, in some passages Sober himself uses 'testability' like a one-place predicate. For instance, he claims that 'Undetectable angels exist' is untestable and that 'This coin has probability of .5 of landing heads' is testable, which is, strictly speaking, meaningless for a two-place predicate like contrastive testability. And both claims are important for Sober's line of argument, since he relies on the first to argue that testability is different from meaningfulness, and on the second to argue that falsifiability is not an adequate criterion of empirical significance. Thus even Sober seems to rely tacitly on a concept of empirical significance that is not captured by contrastive testability. I will come back to this in §8.6.2.

8.2.2 Interpreting inequalities between probabilities

In the discussion of Sober's interpretation of the inequality in his criterion (§8.2.1), it has already become apparent that dealing with undefined probabilities is not an entirely straightforward matter. And even though definition 8.4 treats the case of undefined likelihoods explicitly, it still involves some lacunae. Specifically, if one of the likelihoods is not defined, it is not obvious how to treat the inequality (8.6), and not entirely obvious how to treat the whole definition. For the definition is logically a conjunction with the inequality as a conjunct, and it is unclear whether a conjunction with one undefined conjunct is undefined as well. It would thus be desirable if the inequality were never undefined, so that the usual rules of logic can apply. Luckily, Sober's assumptions arguably entail just such an interpretation of the inequality.

It is uncontentious that $\Pr(\Sigma \mid \Theta)$ is defined when Θ assigns a real-valued probability to Σ. But as Sober himself states when arguing for the need for supplementary sentences, theories alone often do not assert anything, and thus do not assign a real-valued probability to any basic sentence. And even with supplementary sentences Π, no theory will make assertions about everything.[9] Rather, $\Theta \cup \Pi$ restricts the set of reals from the unit interval that can be assigned to *some* set Σ of sentences to a subset of the unit interval, possibly to one specific value $x \in [0, 1]$, while the set of reals for some other sets of sentences may remain unrestricted. The conditional probability $\Pr(\Sigma \mid \Theta \cup \Pi)$ can then either always be read as the set of reals that Σ can be assigned under the assumption of $\Theta \cup \Pi$,

[9] More precisely: For any even remotely plausible theory Θ, $\Theta \cup \Pi$ can be complete in $\mathscr{B}$ only for very restricted vocabularies $\mathscr{B}$, where Θ is complete in $\mathscr{B}$ if and only if for all sets Ω of basic sentences, $\Pr(\Omega \mid \Theta \cup \Pi) = x, x \in [0, 1]$. It is thus always possible to expand $\mathscr{B}$ by well-interpreted basic terms so that Θ fails to be complete in $\mathscr{B}$.

Values of likelihoods		$\Pr(\Omega \mid \Theta \cup \Pi) \neq \Pr(\Omega \mid \Lambda \cup \Pi)$								
$\Pr(\Omega \mid \Theta \cup \Pi)$	$\Pr(\Omega \mid \Lambda \cup \Pi)$	1	2	3	4	5	6	7	8	9
x	y	$T\ldots$								
x	x	$F\ldots$								
x	U	T	T	T	F	F	F	U	U	U
U	U	T	F	U	T	F	U	T	F	U

Table 8.1: The nine possible interpretations of the inequality in definition 8.5 depending on the values of the probabilities, where x, y with $x \neq y$ are acceptable values for likelihoods, 'T' stands for 'true', 'F' stands for 'false', and 'U' stands for 'undefined'.

from $[0,1]$ to proper subsets thereof down to the singleton set $\{x\}, x \in [0,1]$. Or $\Pr(\Sigma \mid \Theta \cup \Pi)$ may be read as defined only when it is a set of some specific kind considered acceptable (e. g., a proper sub-interval of $[0,1]$ or a singleton set), and undefined in all other cases.

Depending on the treatment of formulas that contain undefined terms, the second reading of conditional probabilities leads to different interpretations of the inequality (8.6), given in table 8.1. Sober seems to assume the validity of classical logic, so that $\neg\sigma$ is false if and only if σ is true and tautologies are always true. This excludes some of the possible interpretations: Considering again the case where p and q are defined and different, while a is undefined, it is clear that interpretations 4–6 (5 being Sober's) are inconsistent because they lead to $p = a$, $a = q$, and $p \neq q$. When a is undefined, interpretations 1, 3, 7, and 9 do not render $a \neq a$ false, and thus they are also excluded. Interpretation 8 is excluded if one demands that classical logic be truth-preserving and at least one disjunct of a true disjunction be true. For then, if p and q are defined and identical, while a and b are undefined, $p = q$ and $a = b$ are true, and entail $a = p \vee b \neq q$, which according to interpretation 8 has two undefined disjuncts. The remaining interpretation 2 can be seen as following from the introduction of the special value 'undefined' for probability-terms.

Under these assumptions, there are thus two possible interpretations of the inequality:

1. When all sets of reals are acceptable, the inequality is true if and only if the two sets differ. Otherwise, it is false.

2. When some sets of reals are unacceptable, the inequality is true if and only if its two sides are defined and different, or one side is defined and the other one is not. Otherwise, the inequality is false. (Interpretation 2 in table 8.1)

It is clear that the inequality is true more often for interpretation 1 than for interpretation 2, since in interpretation 1 it is true whenever the set on one side differs from the set on the other side, but also when there is a difference between two sets that are unacceptable under the second reading of the likelihoods. It is also clear that Sober does not subscribe to interpretation 1, since in that case, there are no undefined likelihoods. In fact, he developed his concept of contrastive testability under the assumption that only singleton sets are acceptable (Sober 2010, 3). With these two readings of probability and the corresponding interpretations of the inequality, definition 8.3 is now indeed defined in all cases because the inequality is always either true or false.

8.3 The restrictions on the supplementary sentences

Definition 8.3 is formally correct, but is nonetheless problematic. Specifically, I will argue that there are serious problems both with restriction IVa (that the supplementary assumptions be independent of the tested theories) and IVb (that the supplementary assumptions be independent of the basic sentences). Sober (1999, 54) introduces supplementary assumptions into the definition because "hypotheses rarely make observational predictions on their own; they require supplementation by auxiliary assumptions if they are to be tested" (cf. Sober 2007, 5–6; Sober 2008, 144).[10] But this "raises the question of which auxiliary assumptions we should use to render a theory testable. What makes an auxiliary assumption 'suitable'?" (Sober 2008, 144). Restrictions IVa and IVb are answers to these questions.

8.3.1 Dependence on the theories

Sober (2008, 145) at one point simply states that the need for restriction IVa is obvious, and elaborates in another passage that without it, his criterion would beg the question (Sober 2007, 6). But this, at least, is not obvious. Arguments must not in general allow their conclusion among their premises (that is, beg the question). For otherwise every claim could be shown to be true by including it in the premises of the argument, and the concept of an argument would be trivial. But even without restriction IVa, it is not possible to simply assume that Θ can be tested against Λ when the criterion is applied. In fact, I want to show that IVa is often ineffective or redundant, and in general lacks a justification.

There are a number of cases in which restriction IVa is ineffective. Obviously, two theories that are not contrastively testable *without* IVa cannot be contrastively testable *with* IVa. For if there are no supplementary sentences Π that fulfill IVb, there are also no supplementary sentences that fulfill IVb and IVa. And the converse usually also holds: Two theories that are contrastively testable without

[10]Since Sober does not use 'prediction' to refer exclusively to claims about the future, I will treat it as synonymous with 'assertion'.

IVa usually do not fail to be so with IVa. It is typical, for example, that (i) theories Θ and Λ are incompatible, and (ii) the supplementary sentences used to determine the likelihoods of Θ at most depend on Θ, not on the competing theory Λ (and vice versa). Now, for restriction IVa to do any work, there has to be a set Ω of $\mathscr{B}$-sentences such that there are no supplementary sentences that fulfill conditions I–IV, but there are justified supplementary sentences Π that fulfill restriction IVb, $\Pr(\Omega \mid \Theta \cup \Pi) \neq \Pr(\Omega \mid \Theta \cup \Pi)$, and both likelihoods are defined. By assumption (ii), the supplementary sentences are of the form $\Pi \mathrel{\dashv\vDash} \Pi_1 \cup \Pi_2$, where Π_1 does not depend on Λ and Π_2 does not depend on Θ. Thus $\bigwedge \Theta \to \bigwedge \Pi_1$ and $\bigwedge \Lambda \to \bigwedge \Pi_2$ are justified without assuming Θ or Λ. Therefore, $\Pi' := \{(\bigwedge \Theta \to \bigwedge \Pi_1) \wedge (\bigwedge \Lambda \to \bigwedge \Pi_2)\}$ fulfills restriction IVa, and since by assumption (i), $\Theta \cup \Pi' \mathrel{\dashv\vDash} \Theta \cup \Pi$ and $\Lambda \cup \Pi' \mathrel{\dashv\vDash} \Lambda \cup \Pi$, $\Pr(\Omega \mid \Theta \cup \Pi') \neq \Pr(\Omega \mid \Lambda \cup \Pi')$ so that Θ can be tested against Λ.[11]

I have just argued that restriction IVa is often ineffective. Another problem is that it is also redundant in the important cases.[12] To see this, note that restriction III demands that Π be justified, and that a sentence whose justification depends on another sentence B be justified only if B is justified. (Giving up this relation between 'justified' and 'depend' would render Sober's restriction IV altogether empty.) Thus a supplementary sentence whose justification depends on Θ fulfills the definiens of definition 8.3 only if Θ itself is justified, and analogously for Λ. Typically, however, the question of empirical significance does not even come up for theories that are already justified. Indeed, Sober assumes that a theory is confirmed only if it has been tested, and this is possible only if it is testable. Assuming that only confirmed theories are justified, restriction IVa therefore goes beyond restriction III *only* when the question of empirical significance has already been answered. Of course, one may want to justify or confirm an already justified or confirmed theory *further*, but this is then not a question of contrastive testability any more.[13]

This general argument is not countered by the example that Sober (2008, 145) adduces to show the need for restriction IVa. In it, he envisions Jones being tried for a murder, with a size 12 shoe print, cigar ash, and .45 Colt shells found at the crime scene. When considering whether Smith may be the culprit instead, Sober notes, one must not simply conclude that the evidence favors Smith over Jones on

---

[11] Since this inference relies on the sentences '$\bigwedge \Theta \to \bigwedge \Pi_1$' and '$\bigwedge \Lambda \to \bigwedge \Pi_2$', it assumes that Θ, Λ, Π_1, and Π_2 are finite. The assumption can be somewhat alleviated in two ways. First, if Θ, Λ, Π_1, and Π_2 are not finite but can be axiomatized by the finite sets $\Theta^\dagger$, $\Lambda^\dagger$, $\Pi_1^\dagger$, and $\Pi_2^\dagger$, respectively, the sentences '$\bigwedge \Theta^\dagger \to \bigwedge \Pi_1^\dagger$' and '$\bigwedge \Lambda^\dagger \to \bigwedge \Pi_2^\dagger$' can be used instead. Second, it may be possible to equivalently reformulate Π as $\Pi^* \cup \Pi_1 \cup \Pi_2$, where Π^* does not depend on either Θ or Λ and can be infinite. Then Π' can be defined as $\Pi^* \cup \{\bigwedge \Theta \to \bigwedge \Pi_1\} \cup \{\bigwedge \Lambda \to \bigwedge \Pi_2\}$.

[12] In those cases, it is thus irrelevant that it is ineffective.

[13] There is also the question how any theory or supplementary sentence could ever be justified *simpliciter*, rather than confirmed contrastively against some other theory. I will come back to this problem in §10.2.

the basis of the *assumption* that Smith is a Colt-owning smoker with size 12 feet, while Jones is not.

First note that in this example the question is which theory can be inferred from the evidence, not which observations are asserted by the theory; that is, the example revolves around a question of confirmation, not empirical significance. More important in the following is that the belief about Smith's shoe size would be excluded from the supplementary assumptions even without restriction IVa. This is because, first, the belief that Smith is the murderer (Θ) is itself not justified, and thus cannot justify anything. Second, Smith's murdering someone does not allow any conclusion about her shoe size. This conclusion also requires the belief that there was a size 12 shoe print at the crime scene (Ω). In other words, the justification of the supplementary assumption cannot depend on Θ, for then it would be excluded from Π by restriction III, and it is excluded by restriction IVb anyway, because its justification depends on Ω.

Restriction IVa is included in definition 8.3 for more serious reasons than fictitious murder trials with careless jurors. It is meant to address an argument in defense of the contrastive testability of ID against ET that Sober (1999, 65, note removed) describes as follows (cf. Sober 2008, 143–146):

> [A]dvocates of the design argument should not be confident that they know what characteristics God would have wanted to give to organisms on earth if he had created them. Creationists may be tempted to respond to this challenge simply by inspecting the life we see around us and saying that God wanted to create *that*. After all, if life is the result of God's blueprint, can't we infer what the blueprint said by seeing what the resulting edifice looks like? [But you] can't just *assume* that God created organisms, and you also can't *assume* that if God created organisms he would have made them with such-and-such characteristics.

Analogously to the murder trial, the justification of the supplementary sentence about God's intentions in the creationists' argument depends both on the assumption that God exists and the assumption of the basic sentence that life is as we see it around us, like "that".[14] Therefore it is excluded from Π by restriction III because it is not justified until the belief in God is justified. And if a description of life as we see it around us is given as the basic sentence Ω for which the likelihoods of ID and ET differ, then the supplementary sentence is also excluded by restriction IVb for dependence on Ω.

8.3.2 Dependence on basic sentences

Sober (2008, 145, my notation) justifies restriction IVb as follows (cf. Sober 2007,

[14] In disanalogy to the murder trial, the question in this case is indeed which basic sentences the theory asserts, not which theory the basic sentences confirm.

6):

> If Ω is true, so is the disjunction "either Θ is false or Ω is true". If you use this disjunction as your auxiliary assumption Π_1, then it turns out that the conjunction $\bigwedge\Theta \wedge \bigwedge\Pi_1$ entails Ω. This allows Θ to make a prediction about Ω even when Θ has nothing at all to do with Ω. The same ploy can be used to obtain auxiliary assumptions Π_2 so that the conjunction $\bigwedge\Lambda \wedge \bigwedge\Pi_2$ also entails Ω. Using propositions Π_1 and Π_2 as auxiliary assumptions leads to the conclusion that the two hypotheses Θ and Λ both have likelihoods of unity.

As it stands, this argument proves nothing about the relevance of restriction IVb for the definition of contrastive testability, since it only shows that for one specific supplementary assumption, $\Pi \boxminus \Pi_1 \cup \Pi_2$, both theories' likelihoods are 1. But to show that Θ cannot be tested against Λ, their likelihoods have to be identical for *all* supplementary assumptions that fulfill restrictions III and IVa. (Furthermore, if the goal was to arrive at the *same* likelihood for both theories, $\Pi \boxminus \Omega$ would achieve the same result.)

But the ingenuity of the choice of Π_1 is exactly that, if Θ and Λ are completely unrelated to Ω, the likelihood of $\Theta \cup \Pi_1$ is 1, while the likelihood of $\Lambda \cup \Pi_1$ is not. Reconceptualized in this way, Sober's case for restriction IVb is a typical trivialization proof, since it shows that without it, any two theories can be tested against each other. The argument has four tacit assumptions, however. First, a set σ of sentences (here: $\neg\bigwedge\Theta \vee \bigwedge\Omega$) logically entailed by a set Σ of justified sentences (here: Ω) is also justified, since otherwise σ might be excluded by restriction III. Second, σ depends for its justification only on Σ, for otherwise, σ might be excluded by restriction IVa or, implausibly, for its dependence on $\neg\bigwedge\Omega$ by restriction IVb. Third, $\Pr\big(\Omega \,|\, \Lambda \cup \{\neg\bigwedge\Theta \vee \bigwedge\Omega\}\big) \neq 1$. Finally, $\Pr\big(\Omega \,|\, \Lambda \cup \{\neg\bigwedge\Theta \vee \bigwedge\Omega\}\big)$ must be defined, for otherwise condition I is not fulfilled.

The fourth tacit assumption is probably false, since neither Θ nor Λ are related to Ω. By considering additional supplementary sentences Π, however, one can arrive at a modification of the proof that has plausible premises. Let Π be such that it is unrelated to Θ and Λ, and $\Pr(\Omega \,|\, A)$ is defined. Then a plausible fourth tacit assumption is that for any σ unrelated to Θ and Λ, $\Pr(\Omega \,|\, \Lambda \cup \{\neg\bigwedge\Theta \vee \sigma\} \cup \Pi)$ is therefore defined as well: Conjoining Λ with Π does not render the conditional probability of Ω undefined, since Λ is not related to Ω. Conjoining $\neg\bigwedge\Theta \vee \sigma$ with $\Lambda \cup \Pi$ arguably does not render the conditional probability of Ω undefined, either, because the inferences one can draw from $\neg\bigwedge\Theta \vee \sigma$ are weaker than those that one can draw from Θ, and Θ is already unrelated to Ω. Choosing $\sigma = \bigwedge\Omega$ and incorporating Π in all premises now allows giving a corrected version of Sober's proof, which can also be recovered from the proof of claim 8.1 below.

Sober does not show why the reference to $\neg\bigwedge\Omega$ (the disbelief in Ω) in restriction IVb is necessary, but his trivialization proof can be repeated for $\neg\bigwedge\Omega$. Since $\Pr(\Omega \,|\, \Theta \cup \Pi) = 1 - \Pr(\neg\bigwedge\Omega \,|\, \Theta \cup \Pi)$, and analogously for Λ instead of Θ, Π can be justified with Ω, while the likelihoods of Θ and Λ would differ for $\neg \bigwedge \Omega$. This proof assumes that the negation of a basic sentence is itself a basic sentence, which is fairly uncontroversial. It is not only tacitly assumed by Popper (see claim 6.1) and Sober (e. g. 1999, n. 14), but also fulfilled by the most common restrictions on basic sentences: It holds if all and only sentences with a specific non-logical vocabulary are basic (cf. Psillos 2000, 158–159), if all and only molecular sentences with a specific vocabulary are basic (cf. Carnap 1937, §23), and if all and only sentences are basic whose quantifiers are relativized to observable objects (as discussed by Carnap 1956b, §II and assumed by van Fraassen—see §4.2). A sentence could also be considered basic if and only if it is about subject matter $\mathscr{B}$, and according to Lewis (1988b, 140–141), if a sentence is about observations, so is its negation. All these restrictions even entail that the set of observation sentences is closed under truth-functional composition.

While restriction IVb is necessary to avoid trivialization of definition 8.3, it is not sufficient. Specifically, any two theories can be tested against each other as long as one of them can be finitely axiomatized:

Claim 8.1. *Let Θ, Λ, Ω, and Π be sets of sentences and σ a sentence such that*

1. *Ω is a set of basic sentences,*
2. $\sigma \vDash \Omega$,
3. *σ and Π are justified independently of Ω, the falsity of Ω, Θ, and Λ,*
4. *σ and Π are unrelated to Θ and Λ,*
5. $\Pr(\Omega \mid \Pi)$ *is defined, and*
6. $\Pr(\Omega \mid \Lambda \cup \{\neg \bigwedge \Theta\} \cup \Pi) \neq 1$.

Given Sober's tacit assumptions in the defense of restriction IVb, Θ and Λ can then be tested against each other.

Proof. Choose Ω, σ, and Π such that conditions 1–6 hold. Since σ and Π are justified, so is $\Pi^* \boxminus \{\neg \bigwedge \Theta \vee \sigma\} \cup \Pi$ by Sober's first tacit assumption. It follows from Sober's second tacit assumption that, since the justifications of σ and Π do not depend on Ω, its falsity, Θ, or Λ, neither does the justification of Π^*. Therefore Π^* fulfills restrictions III and IV of definition 8.3.

Now, from $\Pr(\Omega \mid \Lambda \cup \{\neg \bigwedge \Theta\} \cup \Pi) \neq 1$ and $\sigma \vDash \Omega$ it follows that $\Pr\big(\Omega \mid \Lambda \cup$

$\{\neg\bigwedge\Theta\vee\sigma\}\cup\Pi)\neq 1$:

$$\Pr(\Omega\,|\,\Lambda\cup\{\neg\bigwedge\Theta\vee\sigma\})\cup\Pi) =$$
$$\frac{\Pr(\Omega\cup\Lambda\cup\{\neg\bigwedge\Theta\}\cup\Pi)+\Pr(\Omega\cup\Lambda\cup\{\sigma\}\cup\Pi)-\Pr(\Omega\cup\Lambda\cup\{\neg\bigwedge\Theta\wedge\sigma\})}{\Pr(\Lambda\cup\{\neg\bigwedge\Theta\}\cup\Pi)+\Pr(\Lambda\cup\{\sigma\}\cup\Pi)-\Pr(\Lambda\cup\{\neg\bigwedge\Theta\wedge\sigma\}\cup\Pi)}$$
$$=\frac{\Pr(\Omega\wedge\Lambda\cup\{\neg\bigwedge\Theta\}\cup\Pi)+c}{\Pr(\Lambda\cup\{\neg\bigwedge\Theta\}\cup\Pi)+c}\,. \tag{8.5}$$

The last term equals 1 if and only if $\Pr\big(\Omega\,|\,\Lambda\cup\{\neg\bigwedge\Theta\}\cup\Pi\big)=1$. Since $\Pr(\Omega\,|\,\Pi)$ is defined, $\Pr(\Omega\,|\,\Lambda\cup\{\neg\bigwedge\Theta\vee\sigma\}\cup\Pi)=\Pr(\Omega\,|\,\Lambda\cup\Pi^{*})$ is defined by the fourth tacit assumption. Since furthermore $\Pr(\Omega\,|\,\Theta\cup\Pi^{*})=1$, it holds that $\Pr(\Omega\,|\,\Theta\cup\Pi^{*})\neq\Pr(\Omega\,|\,\Lambda\cup\Pi^{*})$, where both probabilities are defined. Θ and Λ can therefore be contrastively tested against each other. □

Note that for $\Pr(\Theta\cup\Lambda\cup\Pi)=0$, condition 6 simplifies to '$\Pr(\Omega\,|\,\Lambda\cup\Pi)\neq 1$'. The corrected version of the trivialization proof that Sober uses to justify restriction IVb can be recovered by dropping σ in condition 3 and choosing $\sigma \models\!\!\!\dashv \bigwedge\Omega$. Then condition 2 is trivially true, 3 amounts to Sober's restriction IV, 4 and 5 are the antecedents of the fourth tacit assumption, and 6 is equivalent to the third tacit assumption.

Conditions 1–3 are impossible to fulfill if a justification can proceed only deductively from basic sentences, because then the justification of a sentence depends on every basic sentence it entails. However, since Sober's criterion is meant to be applicable to inductive theories, it is plausible that supplementary sentences can also be inductively justified. In that case, it is easy to find sentences Ω, σ, and Π that fulfill all the requirements. For instance, let Π express that 1 out of 10 vases of some kind breaks when dropped from a specific height. Let furthermore σ express that a specific vase of that kind does not break when dropped a hundred times from a that height, and Ω express that the vase does not break on the hundredth drop. Then σ is justified independently of Ω when the vase is dropped 99 times without breaking, so that Ω, σ, and Π fulfill conditions 1–6 for any two theories that are not related to vases. Since according to Sober hypotheses rarely make observational predictions on their own, that includes almost all theories. But even for two theories that make assertions about vases, it should not be difficult to find other basic sentences that neither they nor their negations assert with probability 1, but that can be asserted by enumerative induction.

Considering the somewhat problematic status of the fourth tacit assumption, the result of this section can be seen as a dilemma: Either the tacit assumptions (especially the fourth) hold or they do not hold. If they do not hold, Sober's restriction IVb has not been justified. If they do, the restriction is to weak to avoid trivialization.

8.3.3 New definitions

Sober's restriction IVa is unjustified where it is not redundant or ineffective, and restrictions III and IV together are to weak to avoid trivialization. Clearly, the search for general restrictions on the supplementary sentences poses a host of subtle problems. To bracket these problems, I suggest

Definition 8.4. Theory Θ can be *contrastively tested against* theory Λ *relative to* supplementary sentences Π if and only if there exists a set Ω of basic sentence such that $\Pr(\Omega\,|\,\Theta \cup \Pi)$ and $\Pr(\Omega\,|\,\Lambda \cup \Pi)$ are defined and

$$\Pr(\Omega\,|\,\Theta \cup \Pi) \neq \Pr(\Omega\,|\,\Lambda \cup \Pi)\,. \tag{8.6}$$

This definition is not trivial: Choose $\Pi = \varnothing$, two non-basic, non-equivalent sentences σ and σ', and, for some set Ω of basic sentences, $\Theta \mathrel{\models\!\!\!\!\dashv} \{\sigma, \Pr(\Omega) = p\}$ and $\Lambda \mathrel{\models\!\!\!\!\dashv} \{\sigma', \Pr(\Omega) = q\}$ for some probabilities p and q. Then Θ and Λ are never equivalent, and Θ can be contrastively tested against Λ if and only if $p \neq q$, so that many contingent theories can and many contingent theories cannot be tested against each other relative to Π. Definition 8.4 makes it necessary, however, to decide on a case-by-case basis which supplementary sentences are suitable. This may be a good preliminary strategy, because often the suitable supplementary sentences are reasonably clear. For instance, often the suitable supplementary sentences are those that could feature as background assumptions.

Eventually, of course, it would be helpful to have a general criterion for suitable supplementary sentences and define absolute contrastive testability as contrastive testability relative to suitable supplementary sentences. To this end, I suggest the following. The proof of claim 8.1 is a modification of the trivialization proof for Sober's criterion of 'having observable implications' given in §6.8.2 and leads to a similar diagnosis. Sober's proof and that of claim 8.1 rely on the possibility of including a sentence ($\neg \bigwedge \Theta \vee \bigwedge \Omega$ or $\neg \bigwedge \Theta \vee \sigma$) in Π that is justified by another one ($\bigwedge \Omega$ or σ) that is itself not included in Π. Both trivialization proofs can therefore be blocked by explicating 'suitable supplementary sentences' as 'honest set of supplementary sentences' (see definition 6.36). Note that this definition only uses concepts that already occur in Sober's definition 8.3 of contrastive testability.

Definition 6.36 allows a modification of Sober's criterion of testability as follows:

Definition 8.5. Theory Θ is *absolutely contrastively testable against* theory Λ if and only if Θ can be tested against Λ relative to an honest set of supplementary sentences containing all analytic sentences.

The proof that relative contrastive testability is not trivial also shows that some theories are absolutely contrastively testable against each other, for $\varnothing$ is an honest set according to definition 6.36, assuming that there are no relevant analytic sentences. I will not attempt to prove that there are two sets of sentences that fail to be absolutely contrastively testable, because this would amount to finding two sets that are not contrastively testable relative to any honest set. The proof is immediate for equivalent sets of sentences, but impossible for other pairs of sets without more precise notions of justification and dependence.[15]

Definition 8.5 is at least as exclusive as Sober's definition 8.3, however. The restriction of the supplementary sentences in definition 8.5 to honest sets entails restriction III of definition 8.3. And while the restriction to honest sets does not entail restriction IVb, it precludes all trivializations precluded by that restriction: Two theories Θ and Λ fail to be contrastively testable because of restriction IVb only if for any σ whose inclusion in Π would lead to differing likelihoods for some Ω, the justification of σ depends on Ω or the falsity of Ω. In that case, IVb ensures that Θ and Λ are not contrastively testable. The restriction of Π to honest sets leads to the same result, because if the justification of σ depends on Ω (or its falsity) and Π is honest, then $\Omega \subseteq \Pi$ ($\Delta \subseteq \Pi$ for some set Δ of sentences expressing the falsity of Ω).[16] Thus $\Pr(\Omega \mid \Theta \cup \Pi) = 1 = \Pr(\Omega \mid \Lambda \cup \Pi)$ (or $\Pr(\Omega \mid \Theta \cup \Pi) = 0 = \Pr(\Omega \mid \Lambda \cup \Pi)$). As an example, take the sentence $\neg \bigwedge \Theta \vee \bigwedge \Omega$ of Sober's trivialization proof. Restriction IVb excludes $\neg \bigwedge \Theta \vee \bigwedge \Omega$ from the supplementary sentences Π, so that the likelihoods of Θ and Λ for Ω cannot differ because of $\neg \bigwedge \Theta \vee \bigwedge \Omega$. The restriction to honest sets, on the other hand, leads to the inclusion of Ω in Π, so that the likelihoods do not differ, either. Unlike restriction IVb, the restriction to honest sets also leads to identical likelihoods if $\neg \bigwedge \Theta \vee \sigma$ is justified by a sentence $\sigma \vDash \Omega$, thereby precluding the proof of claim 8.1. Specifically, premise 6 will be false because $\Pi \vDash \Omega$.

Since it is not clear in which case restriction IVa is meant to preclude trivialization, or in general, which problem it is meant to solve, I cannot show that definition 6.36 can fulfill the role of restriction IVa. Given the restriction's questionable role and justification, this should not be considered a drawback of definition 8.5. If there is a justification for restriction IVa, however, one can modify definition 8.5 by defining contrastive testability as contrastive testability relative to an honest set that does not include Θ or Λ. This restriction entails restriction IVa.

That the notion of an honest set plausibly explicates the notion of possible background assumptions (see §6.8.2) provides a justification of definition 8.5 that is independent of the trivialization proofs: If relative contrastive testability is a good explication of empirical significance, then absolute contrastive testability explicates what it means for a theory to be empirically significant in our current epistemic situation (as determined by our background assumptions) or a possible situation on our way to our current epistemic situation. The independent justification of definition 8.5 is not only relevant because the proof of claim 8.1 rests, like Sober's justification of restriction IVb, on somewhat contentious assumptions. More

[15]Equivalent sets are excluded from the conclusion of claim 8.1 because of premise 6 and the assumption that for any conditional probability $\Pr(B \mid C)$, $\Pr(C) \neq 0$ (see n. 5).

[16]If Ω is finite, then $\Delta \mathrel{\Dashv\vDash} \{\neg \bigwedge \Omega\}$. In general, Δ has to be such that $\mathfrak{A} \vDash \Delta$ if and only if $\mathfrak{A} \nvDash \Omega$.

importantly, it justifies the hope that the definition is right, while the amendments in light of the trivialization proofs at best allow the claim that the modified definitions are not obviously wrong. Of course, definition 8.5 might still allow the proof that any two non-equivalent theories can be absolutely tested against each other. In response to such a proof, one can fall back on definition 8.4 until a better explication of 'suitable' is found than definition 6.36. In general, any results that determine which supplementary sentences are suitable, or which assumptions are possible background assumptions, can be used directly as a substitute for definition 6.36 (see §6.9).

8.4 Conditions of adequacy

Excluding the problematic conditions of adequacy postulated by Achinstein (page 217) and Hempel (page 264), I have discussed the following conditions of adequacy for deductive criteria (see §6.9):

Condition 3. The concept of empirical significance is not trivial.

Condition 4. Not all contingent sentences are empirically significant.

Condition 5. Not only $\mathscr{B}$-sentences are empirically significant.

Condition 6. All and only sets of sentences that assert basic sentences are empirically significant.

Condition 7. All and only sets of sentences that are not empirically equivalent to a tautology are empirically significant.

In this section, I will argue that Sober's assumptions and his intended application of the criterion lead to six conditions of adequacy. Many of these relate empirical significance to concepts that rely on inferences and therefore have a deductive and a probabilistic formulation. This is because deductive inference (entailment: $\Sigma \vDash \Phi$) clearly does not generalize probabilistic inference. For instance, from $\Pr(\Sigma) = 1$ and $\Pr(\Phi \mid \Sigma) = q$, one can probabilistically infer that $\Pr(\Phi) = q$, and this inference, related to Sober's (Update)-rule discussed below, cannot be captured by assigning truth values to Φ and Σ. But probabilistic inference also does not generalize deductive inference. For assume that the domain has infinite cardinality. Then it may be that $\Pr(\Phi \mid \Sigma) = 1$, but there are cases in which Σ is true and Φ is false. This happens, for example, when the domain is the interval $[0, 2]$ with a uniform probability distribution, Σ is '$x \leq 1$', and Φ is '$x < 1$' (cf. Feller 1971, 33–34). This difference between the deductive and the probabilistic concept of inference generally leads to differences between the deductive and probabilistic formulations of the conditions of adequacy. I have discussed conditions of adequacy for deductive criteria in §6.9. Here I will just discuss their probabilistic counterparts and their relations to Sober's positions.

Condition 3 holds for deductive criteria just as it holds for probabilistic ones. Since Sober intends to distinguish between theories that are worthy to be pursued and theories that are not, his criterion must not be trivial, and he implicitly relies on this condition of adequacy when arguing for restriction IVb. More specifically, he relies on the condition that not all non-analytic sentences should be empirically significant. Condition 4, that not all contingent sentences should be empirically significant, can be considered to be a strengthening of this. This condition is the basis for my claim in §8.3.3 that the proof of the non-triviality of absolute contrastive testability needs more precise definitions of 'justification' and 'dependence', and an assumption behind claim 8.1 (see n. 15). Condition 5, that not only $\mathscr{B}$-sentences should be empirically significant is uncontroversial also for probabilistic criteria, but will not be of importance in the following.

As noted, Ayer (1936, 97) argues for condition 6, the demand that all and only empirically significant sets of sentences make basic assertions. Sober (2008, 130) states that "a testable statement makes predictions, either by deductively entailing that an observation will occur or by conferring a probability on an observational outcome." Thus for Sober empirical significance is a sufficient condition for making basic assertions. Let this be condition (i). Sober also subscribes to the converse of condition (i), as can be seen from his claims that "[t]he problem with the hypothesis of intelligent design is [...] that it doesn't predict much of anything" (Sober 2008, §2.15) and that his "criticism of the design argument might be summarized by saying that the design hypothesis is untestable" (Sober 2008, 148).[17] Since Sober (2008, §2.12) infers the lack of empirical significance from the lack of basic assertions, his criticism of ID *relies* only on condition (i). However, Sober's criticism of Popper's falsifiability criterion does seem to rest on the converse of condition (i) for probabilistic assertions: 'This coin has probability of .5 of landing heads each time it is tossed' makes a probabilistic assertion, and its lack of falsifiability is a reason for Sober to reject Popper's criterion. This seems to assume that every theory that makes probabilistic assertions is empirically significant.

Sober (2008, 52, n. 29) further states two relations between deductive empirical significance and the making of deductive basic assertions:

> If a true observation sentence entails Θ [...] you can conclude without further ado that Θ is true; this is just *modus ponens*. And if Θ entails Ω and Ω turns out to be false, you can conclude that Θ is false [...]; this is just *modus tollens*.

These are two sufficient conditions for empirical significance, namely (ii) entailment *by* a basic sentence and (iii) entailment *of* a basic sentence. Condition (iii) is

[17] When Sober (1999, 54) states that "hypotheses rarely make observational predictions on their own; they require supplementation by auxiliary assumptions if they are to be tested", he similarly seems to be treating testability and the making of observational assertions simply as synonymous.

the converse of condition (i) for deductive assertions (cf. Sober 1999, 72, n. 14). Therefore, according to Sober all and only theories that make basic assertions are empirically significant.

Condition (ii), however, is incompatible with condition (i): For any sentence σ and basic sentence Ω, $\Omega \vDash \Omega \vee \sigma$, that is, $\Omega \vee \sigma$ is empirically significant according to condition (ii). But let σ be such that it does not make basic assertions, that is, for any basic sentence Ω', $\sigma \nvDash \Omega'$, and σ does not confer any probability on Ω'. Then, as a matter of logic, $\Omega \vee \sigma \nvDash \Omega'$, so $\Omega \vee \sigma$ does not make deductive assertions. $\Omega \vee \sigma$ also does not confer a probability on any basic sentence, since the inferences one can draw from $\Omega \vee \sigma$ are weaker then those that one can draw from σ, and σ already does not allow assigning a probability to any basic sentence. Thus $\Omega \vee \sigma$ does not make any basic assertions and is therefore not empirically significant according to condition (i), which is incompatible with condition (ii). On pain of inconsistency, Sober therefore has to choose whether all theories entailed by basic sentences are empirically significant or whether all theories that are empirically significant make basic assertions. Given that his core argument against ID is that ID fails to make assertions, I take it that he would choose the latter.[18]

Claim 6.30 establishes that for deductive criteria, condition 6 entails condition 7, the demand that the criterion should exclude all theories that are empirically equivalent to tautologies. This suggests that condition 7 is therefore also a plausible demand for probabilistic criteria, especially since adding a tautology to any set of sentences does not change the probabilities that can be assigned to the other sentences in the set. Out of caution, one may treat the empirical non-equivalence to a tautology as a necessary, but not as a sufficient condition for empirical significance in the probabilistic case.

For Sober, probabilistic criteria of empirical significance also have some new conditions of adequacy, for example

Condition 8. The criterion should not rely on the probabilities of whole theories or likelihoods of the negations of whole theories.

Sober (2008, 24–30) argues that for many theories Θ the probabilities $\Pr(\Theta \mid \Pi)$, $\Pr(\Theta \mid \Omega \cup \Pi)$, and $\Pr(\Omega \mid \{\neg \bigwedge \Theta\} \cup \Pi)$ are undefined (cf. Sober 1990, §III). A criterion that relies on these probabilities would therefore be unusable in many cases.

Condition 9. The criterion should be equivalent to an adequate Bayesian criterion of empirical significance whenever all occurring probabilities are defined.

Since Bayesianism relies on probabilities of whole theories and likelihoods of negations of whole theories, Sober rejects it as a general method of scientific

[18] Note that the claim "There is an intelligent designer" is equivalent to "There is a human designer or there is a non-human designer" and thus analytically entailed by a basic sentence like "There are humans who design". Arguably, however, "There is a non-human designer" does not make an basic assertion, so that "There is an intelligent designer" does not either.

inference. With respect to likelihoodism, Sober (2008, 37) notes (cf. Sober 2008, 32):

> The likelihoodist is happy to assign probabilities to hypotheses when the assignment of values to priors and likelihoods can be justified by appeal to empirical information. Likelihoodism emerges as a statistical philosophy distinct from Bayesianism only when this is not possible.

Since there are criteria of empirical significance that have been developed within Bayesianism, this suggests that a probabilistic criterion of empirical significance should be equivalent to one of these Bayesian criteria whenever all probabilities are defined. This Bayesian criterion should, of course, fulfill all criteria of adequacy other than 8.

Condition 10. The probabilistic criterion should contain as a special case an adequate criterion of deductive empirical significance that relies only on *modus ponens*.

Sober (2002, 69–70) sees a smooth transition between probabilistic and deductive *modus ponens*. More specifically, Sober (2008, 50, my notation) points out the following:

$$\text{(Update)} \quad \begin{array}{l} \mathrm{Pr}_{\text{then}}(\Theta \mid \Omega) \text{ is very high} \\ \Omega \\ \Omega \text{ is all the evidence we have gathered between then and now.} \\ \hline \mathrm{Pr}_{\text{now}}(\Theta) \text{ is very high} \end{array}$$

> This is nothing other than the rule of updating by strict conditionalization. (Update) is a sensible rule, and it also has the property of being a generalization of deductive modus ponens.

As argued above (page 316), (Update) is not, strictly speaking, a generalization of *modus ponens*. But at least when all and only sentences with probability 1 are certain, deductive and probabilistic inference coincide. This can be put more precisely as follows. Each structure $\mathfrak{A}$ of a language $\mathcal{V}$ of predicate logic assigns a truth value to each sentence in $\mathcal{V}$. If $\mathrm{Pr}_{\mathfrak{M}}$ is defined as the function that assigns 1 to all sentences true in $\mathfrak{M}$ and 0 to all sentence false in $\mathfrak{M}$, then $\mathrm{Pr}_{\mathfrak{M}}$ is a probability assignment (see §8.9, claim 8.9). Call such probability assignments *truth value-like*. For truth value-like probability assignments, probabilistic inferences and deductive inferences coincide: The possible values of $\mathrm{Pr}(\Sigma \mid \Phi)$ are restricted to 0 and 1, and $\mathrm{Pr}(\Sigma \mid \Phi) = 1$ for all probability distributions if and only if $\Phi \vDash \Sigma$ (as always assuming that $\mathrm{Pr}(\Phi) \neq 0$; see §8.9, claim 8.11). Truth value-like probabilities may be assigned by fiat, but they also occur more or less naturally when there are no regularities whatsoever, so that no probabilities can be assigned to sentences that are not known to be true and thus have probability 1 or known to be false and thus have probability 0.[19]

In this sense, then, there can be a smooth transition between probabilistic and deductive inference. Given that all and only theories that make deductive or probabilistic assertions must be empirically significant by condition of adequacy 6, there must then also be a smooth transition between any criterion of probabilistic empirical significance and a criterion of deductive empirical significance that uses the implications of the theory only in a *modus ponens*. As I will say, the probabilistic criterion must contain as a special case a deductive criterion that relies only on *modus ponens*. Of course, the deductive criterion should fulfill all those conditions of adequacy that also have purely deductive formulations, that is, conditions 3, 6, and 7. To fulfill condition 6, it is enough for the deductive criterion to include all and only theories that make deductive assertions, because it is impossible that it could include theories that make only probabilistic assertions. Analogously, it is enough if the criterion excludes all theories that are deductively empirically equivalent to a tautology to meet condition 7.

Independently of any smooth transition in the case of *modus ponens*, it is clear that the criterion of empirical significance simpliciter should be a generalization of an adequate deductive criterion. Thus, when deductive and probabilistic inference coincide, the probabilistic criterion must not include theories that the deductive criterion excludes. For if it did, these theories would be included by the criterion of empirical significance simpliciter, and thus this criterion would not generalize the deductive criterion, but rather contradict it.

I have argued in §6.9 (page 266) that all and only $\mathscr{B}$-creative sentences make deductive observations. Thus $\mathscr{B}$-creativity is, up to equivalence, the only deductive syntactic criterion that meets all of Sober's conditions of adequacy that apply to deductive criteria. Since it furthermore relies only on *modus ponens*, condition 10 can be formulated as

Condition 11. When the probability assignment is truth value-like, the probabilistic criterion should be equivalent to $\mathscr{B}$-creativity.

It is thus especially fortuitous that syntactic $\mathscr{B}$-creativity is equivalent to Sober's criterion of *having observational implications* when the question of suitable supplementary assumptions is bracketed. Since I will assume in the following that empirical claims are described by sets of sentences, '$\mathscr{B}$-creativity' will always be short for 'syntactic $\mathscr{B}$-creativity'.

* * *

While I have tried to mention supporting arguments for these conditions of adequacy when possible, some of them remain controversial; especially condition 8

[19] This is arguably the case in Popper's approach to induction (cf. Salmon 1967, §II.3).

would be challenged by Bayesians. But these conditions all follow from Sober's basic assumptions or apply to Sober's criterion because of its intended application. Of course, it may be that these conditions of adequacy are incompatible, so that some have to be given up. This is the case for conditions (i) and (ii) discussed under condition of adequacy 6. But a criterion of empirical significance that is to be applied as Sober intends should fulfill as many of these conditions as possible.

8.5 Contrastive testability and the conditions of adequacy

By Sober's own standards, a probabilistic criterion of empirical significance must meet conditions of adequacy 4, 6, 7, 8, 9, and 11. Condition of adequacy 4 is that not all contingent sentences can be empirically significant, and Sober's definition 8.3 of contrastive testability does not meet this condition. As argued in §8.3.3, however, definition 8.4 of relative contrastive testability and, arguably, definition 8.5 or absolute contrastive testability (using honest sets) do.

Though non-trivial, contrastive testability fails to meet the two most important conditions of adequacy, conditions 6 and 7. That some theories that do not make probabilistic assertions and are probabilistically empirically equivalent to tautologies are contrastively testable can be inferred from an example that Sober (1999, n. 24) attributes, in a different context, to Greg Mougin:[20]

> Let $\Theta =$ God created the eye, $\beta =$ Jones is pregnant, $\Pi =$ Jones is sexually active, and $\Lambda =$ Jones used birth control. It is possible to test Θ against Λ; given independently attested background assumptions Π, β favors Θ over Λ.

In the example, the basic sentence β is assigned one probability by the background assumptions alone (since Θ is not about Jones at all), and another by the conjunction of the background assumptions and Λ. Now choose $\Theta \models\!\!\!\dashv \top$. Then Θ does not make any assertions and hence no basic ones, and it has trivially as much empirical content as a tautology. But Θ can still be contrastively tested against Λ, both relative to Π and absolutely, since the justification of Π does not depend on β, $\neg\beta$, Θ, or Λ.

By design, contrastive testability does not rely on prior probabilities or the likelihoods of the negation of theories and thus meets condition of adequacy 8. Contrastive testability fails to meet condition 9 simply because so far, no Bayesian criterion of empirical significance has been suggested that is equivalent to contrastive testability when all occurring probabilities are defined. Specifically, relative

[20] Unlike in the example by Salmon (1971, 29–88), it is this time not John Jones who is using birth control, but his wife.

contrastive testability is not equivalent to the typical Bayesian criterion of empirical significance, definition 8.2, as is clear from the logical structures of the two concepts.

In principle, a probabilistic two-place predicate may contain a deductive one-place predicate as a special case. For example, if the probability assignments are truth value-like, ‘$\Pr(\Omega \mid \Theta) = .5 \vee \Pr(\Omega \mid \Lambda) = 1$’ is equivalent to ‘$\Lambda \vDash \Omega$’ because the first argument, Θ, becomes irrelevant. In this sense, the two-place predicate of contrastive testability could therefore contain the one-place predicate of falsifiability as a special case. But since contrastive testability is symmetric, either both or neither of its two arguments are irrelevant for truth value-like assignments and thus it cannot meet condition 11.

Θ can be tested against Λ if and only if their defined likelihoods differ for at least one basic sentence. If only singleton sets of probabilities are acceptable (as Sober (2010, 3) assumed when developing his criterion), this means that at least with respect to one basic sentence, one of the two theories must be wrong. Arguably, then, contrastive testability explicates what it means for two theories to be probabilistically empirically incompatible for the special case that only singleton sets of likelihoods are acceptable. This is borne out by the comparison with

Definition 8.6. Theories Θ and Λ are *deductively empirically incompatible* relative to supplementary sentences Π if and only if there is a set Ω of basic sentences and a basic sentence β such that $\Omega \cup \Theta \cup \Pi \vDash \beta$ and $\Omega \cup \Lambda \cup \Pi \vDash \neg\beta$.

Claim 8.2. *Let Θ and Λ be deductive theories, let all probability assignments be truth value-like, and let the set $\{0, 1\}$ be unacceptable as a value of a likelihood. Then Θ can be tested against Λ relative to Π if and only if Θ and Λ are deductively empirically incompatible relative to Π.*

Proof. Θ can be tested against Λ if and only if there are Ω and β such that the likelihood of one theory for β given Ω is 0, while the other one is 1. Without loss of generality, assume $\Pr(\beta \mid \Omega \cup \Theta \cup \Pi) = 1$ and $\Pr(\Omega \mid \Lambda \cup \Pi) = 0$, that is, $\Pr(\neg\beta \mid \Omega \cup \Lambda \cup \Pi) = 1$. By claim 8.11 (see §8.9), this holds if and only if $\Omega \cup \Theta \cup \Pi \vDash \beta$ and $\Omega \cup \Lambda \cup \Pi \vDash \neg\beta$. □

Thus, if only singleton sets are acceptable as values of likelihoods, then, as demanded by condition of adequacy 11, contrastive testability contains as a special case a criterion for deductive theories that relies only on *modus ponens*. However, since it is not $\mathscr{B}$-creativity, it is the wrong one.

8.6 Explicating probabilistic empirical significance

Contrastive testability does not meet all criteria of adequacy, but that might just be because the criteria cannot all be met at once. I will argue that this is not so by suggesting a criterion of empirical significance that does meet all the conditions. First, however, I want to discuss briefly an intuitively attractive but inadequate criterion.

One may think of defining that a theory is not empirically significant if and only if it cannot be tested against *any* theory. But this definition is inordinately inclusive. For assume that Θ is not such that all assertions from suitable supplementary sentences become undefined, that is, $\Pr(\Omega \mid \Pi)$ and $\Pr(\Omega \mid \Theta \cup \Pi)$ are defined (though possibly identical) for some Ω and some suitable Π. Then Θ is empirically significant if there is any Λ such that $\Pr(\Omega \mid \Lambda \cup \Pi)$ is defined and different from $\Pr(\Omega \mid \Pi)$. For if $\Pr(\Omega \mid \Pi) = \Pr(\Omega \mid \Theta \cup \Pi)$, Θ can be tested against Λ, and if $\Pr(\Omega \mid \Pi) \neq \Pr(\Omega \mid \Theta \cup \Pi)$, Θ can be tested against any tautology. The premises of this argument are commonly true, for example, according to Sober, if Θ is 'God created the eye', Ω is 'Jones is pregnant', and Π is 'Jones is sexually active', for then Λ can be 'Jones used birth control'. Choosing $\Theta \vDash\!\dashv \{\top\}$, the argument shows that tautologies are empirically significant, which runs afoul of conditions of adequacy 6 and 7. It is also straightforward to show that conditions 9 and 11 are not met.

8.6.1 Probabilistic empirical equivalence

A more promising path to a criterion of probabilistic significance leads through the criterion of probabilistic empirically equivalence. To show that contrastive testability (definition 8.3) does not fulfill condition 7 it was sufficient to produce one contrastively testable theory that is probabilistically empirically equivalent to a tautology. To arrive at a criterion of empirical significance that provably fulfills 7, however, it is necessary to define probabilistic empirical equivalence.

Luckily, it is possible to explicate condition of adequacy 7 in line with Sober's position, for he states that "empirically equivalent theories have identical likelihoods" for any observation (Sober 1990, 399). Treating the case of undefined likelihoods explicitly, this leads directly to

Definition 8.7. Theories Θ and Λ are *probabilistically empirically equivalent relative to* supplementary sentences Π if and only if for all basic sentences Ω,

(I) $\Pr(\Omega \mid \Theta \cup \Pi)$ and $\Pr(\Omega \mid \Lambda \cup \Pi)$ are not defined or

(II) $\Pr(\Omega \mid \Theta \cup \Pi)$ and $\Pr(\Omega \mid \Lambda \cup \Pi)$ are defined and $\Pr(\Omega \mid \Theta \cup \Pi) = \Pr(\Omega \mid \Lambda \cup \Pi)$.

Note that condition I is redundant given the interpretations of sentences with probabilities laid out in §8.2.2. As defined, probabilistic empirical equivalence contains deductive empirical equivalence as a special case:

Claim 8.3. *Let Θ and Λ be deductive theories and let all probability assignments be truth value-like. Then Θ and Λ are probabilistically empirically equivalent relative to Π if and only if Θ and Λ are deductively empirically equivalent relative to Π.*

Proof. If all probability assignments are truth value-like, then interpretation 1 and interpretation 2 of inequality (8.6) are equivalent, independently of whether $\{0,1\}$ is an acceptable set of probabilities. For if $\{0,1\}$ is an acceptable set, the interpretations are trivially equivalent; if $\{0,1\}$ is not acceptable, the inequality is false if and only if both likelihoods have the value $\{0\}$, $\{1\}$, or $\{0,1\}$/undefined. Otherwise, the inequality is true. Therefore, it suffices to prove the claim for interpretation 1.

'$\Rightarrow$': $\Omega \cup \Sigma \cup \Pi \vDash \beta$ iff $\Omega \cup \{\neg\beta\} \cup \Sigma \cup \Pi \vDash \bot$. By claim 8.10, this holds iff $\Pr(\Omega \cup \{\neg\beta\} \cup \Sigma \cup \Pi) = 0$, that is $\Pr(\Omega \cup \{\neg\beta\} \mid \Sigma \cup \Pi) = 0$. By assumption, $0 = \Pr(\Omega \cup \{\neg\beta\} \mid \Lambda \cup \Pi) = \Pr(\Omega \cup \{\neg\beta\} \cup \Lambda \cup \Pi)$, that is, by claim 8.10, $\Omega \cup \Lambda \cup \Pi \vDash \beta$.

'$\Leftarrow$': $\Pr(\Omega \mid \Sigma \cup \Pi) = 0$ iff $\Pr(\Omega \cup \Sigma \cup \Pi) = 0$ and thus, by claim 8.10, iff $\Omega \cup \Sigma \cup \Pi \vDash \bot$. Since $\bot$ is a $\mathscr{V}$-sentence, this holds by assumption iff $\Omega \cup \Lambda \cup \Pi \vDash \bot$, and thus by claim 8.10 iff $\Pr(\Omega \mid \Lambda \cup \Pi) = 0$. By claim 8.11, $\Pr(\Omega \mid \Sigma \cup \Pi) = 1$ iff $\Sigma \cup \Pi \vDash \Omega$, which holds iff for every $\beta \in \Omega$, $\Sigma \cup \Pi \vDash \beta$. This is by assumption the case iff for every $\beta \in \Omega$, $\Lambda \cup \Pi \vDash \beta$, that is, $\Lambda \cup \Pi \vDash \Omega$ and thus by claim 8.11 iff $\Pr(\Omega \mid \Lambda \cup \Pi) = 1$. □

Another reason to consider definition 8.7 a good explication of probabilistic empirical equivalence is that, if all occurring probabilities are defined, it bears the same relation to the Bayesian criterion of empirical significance given in definition 8.2 as the criterion of deductive empirical equivalence bears to falsifiability: If two theories are deductively empirically equivalent, then either both or neither are deductively empirically significant (see claim 8.12). Analogously, the following holds:

Claim 8.4. *If all occurring probabilities are defined and Θ is probabilistically empirically equivalent to Λ relative to supplementary sentences Π, then, relative to Π, basic sentences are relevant for Θ if and only if basic sentences are relevant for Λ.*

Proof. If Ω is any set of basic sentences for which $\Pr(\Omega \mid \Theta \wedge A) = \Pr(\Omega \mid \Lambda \cup \Pi)$, then $\Pr(\Omega \mid \Theta \cup \Pi) = \Pr(\Omega \mid \Pi)$ if and only if $\Pr(\Omega \mid \Lambda \cup \Pi) = \Pr(\Omega \mid \Pi)$. Therefore $\Pr(\Theta \mid \Pi) = \Pr(\Lambda \mid \Omega \cup \Pi)$ if and only if $\Pr(\Lambda \mid A) = \Pr(\Lambda \mid \Omega \cup \Pi)$ (claim 8.13). Thus basic sentences are relevant for both Θ and Λ or for neither. □

8.6.2 Probabilistic basic assertions

According to claim 6.30, a theory is $\mathscr{B}$-creative, that is, makes deductive basic assertions, if and only if it is not deductively empirically equivalent to a tautology. This suggests that conditions of adequacy 6 and 7 are in fact equivalent, so that a theory makes probabilistic basic assertions relative to supplementary sentences Π if and only if it is not probabilistically empirically equivalent to a tautology relative to Π according to definition 8.4. This leads to

Definition 8.8. Theory Θ *makes probabilistic basic assertions relative to* Π if and only if there exists a set Ω of basic sentences such that

(I) exactly one of $\Pr(\Omega \mid \Theta \cup \Pi)$ and $\Pr(\Omega \mid \Pi)$ is defined or

(II) $\Pr(\Omega \mid \Theta \cup \Pi)$ and $\Pr(\Omega \mid \Pi)$ are defined and $\Pr(\Omega \mid H \cup \Pi) \neq \Pr(\Omega \mid \Pi)$.

Again, condition I is redundant given the interpretations of sentences with probabilities laid out in §8.2.2. In the following, I will not state explicitly conditions that immediately follow from these interpretations explicitly in the following.

Claim 8.5. *Θ makes probabilistic basic assertions relative to Π according to definition 8.8 if and only if Θ is not probabilistically empirically equivalent to a tautology relative to Π according to definition 8.7.*

Proof. Let $\ulcorner Bx \urcorner$ stand for $\ulcorner x$ is a set of basic sentences$\urcorner$, $\ulcorner Dxy \urcorner$ stand for $\ulcorner \Pr(x \mid y \cup \Pi)$ is defined$\urcorner$, and $\ulcorner Exyz \urcorner$ for $\ulcorner \Pr(x \mid y \cup \Pi) = \Pr(x \mid z \cup \Pi) \urcorner$. Then it is straightforward to prove that

$$\begin{aligned} &\exists x \big(Bx \wedge [(Dxy \leftrightarrow \neg Dxz) \vee (Dxy \wedge Dxz \wedge \neg Exyz)]\big) \\ &\qquad \models\mkern-12mu\dashv \neg \forall x \big(Bx \rightarrow [(\neg Dxy \wedge \neg Dxz) \vee (Dxy \wedge Dxz \wedge Exyz)]\big) \, . \qquad (8.7) \end{aligned}$$

Since $\Pr(\Omega \mid \top \cup \Pi) = \Pr(\Omega \mid \Pi)$, the claim follows. □

With the help of definition 6.36 of an honest set, one can give

Definition 8.9. Theory Θ *makes probabilistic basic assertions* if and only if Θ makes probabilistic basic assertions relative to an honest set Π of supplementary sentences.

It may be considered problematic that a theory Θ makes basic assertions if assuming Θ makes it impossible to assign a probability to a basic sentence that otherwise would be assigned a probability by the supplementary sentences. To render such theories empirically non-significant, the biconditional in condition I of definition 8.8 could be made into a conjunction. However, it is very plausible that for a theory Θ that makes no predictions, $\Pr(\Omega \mid \Theta \cup \Pi)$ is defined whenever $\Pr(\Omega \mid \Pi)$ is defined. For one, if $\Pr(\Omega \mid \Pi)$ is defined and asserts a specific frequency of events, then, given Π, one should expect that frequency. But if $\Pr(\Omega \mid \Theta \cup \Pi)$ is undefined, one should expect that there is no such frequency, and this expectation is plausibly a basic assertion. An example would be the prediction that under specific circumstances, some law fails that was assumed to hold universally. But even if $\Pr(\Omega \mid \Pi)$ is not asserting a frequency but a single case probability, there is a significant pragmatic difference between being able to assign a probability to an event and not being able to do so. For the former case demands decisions under risk, while the latter demands decisions under uncertainty.

Definitions 8.8 and 8.9 rely only on concepts that Sober uses himself, and should therefore be conceptually unproblematic for him. He does not discuss the domains of applicability of the concepts, but with one exception, the domains can just be assumed to be the same for definitions 8.8 and 8.9 as they are for contrastive testability. The exception is the term $\Pr(\Omega\,|\,\Pi)$. Sober could argue that the concept of a likelihood cannot be applied to tautologies because the supplementary sentences themselves assign probabilities to no or too few basic sentences. Sober (2008, 29–30) in fact briefly discusses $\Pr(\Omega)$, but not in connection with supplementary sentences. The discussion therefore does clearly not apply to definition 8.8, and it does not apply to definition 8.9 because Θ is testable if there are *some*, not necessarily tautological, suitable supplementary sentences Π such that $\Pr(\Omega\,|\,\Theta \cup \Pi)$ differs from $\Pr(\Omega\,|\,\Pi)$. Sober's original definition 8.3 and definition 6.36 of an honest set put no restrictions on individual elements of Π except that they be justified (in Sober's definition, independently of a specific basic sentence and two specific theories). Therefore whole theories can be included in the supplementary sentences. Since Sober introduces supplementary sentences to allow for actual scientific practice, and assertions made by scientific theories in fact often rely on other scientific theories as supplementary sentences, such an inclusion obeys letter and spirit of Sober's criterion. Since scientific theories Θ are supposed to make basic assertions, $\Pr(\Omega\,|\,\Theta \cup \Pi)$ will often be defined. And the inclusion of Θ into the supplementary sentences is then just the notational change to $\Pr(\Omega\,|\,\Pi^*)$ with $\Pi^* \mathrel{\vert\!\!=\!\!\vert} \Theta \cup \Pi$.

It now follows from condition of adequacy 6 that all and only theories that fulfill definition 8.9 (8.8) are probabilistically empirically significant (relative to supplementary sentences Π). Without any basic conceptual problems, a theory can then be defined to be empirically significant (relative to Π) if and only if it it makes basic assertions (relative to Π), that is, if and only if it makes probabilistic basic assertions (relative to Π) or it makes deductive basic assertions (relative to Π).

As argued in §6.9, $\mathscr{B}$-creativeness fulfills all appropriate conditions of adequacy demanded by Sober. I now want to show that definition 8.8 of probabilistic empirical significance does, too. That it is non-trivial and thus meets condition of adequacy 3 is easily shown since sentences without any terms that occur in basic sentences are not testable relative to $\varnothing$. And in the example with Jones's pregnancy, the theory that Jones uses birth control (Λ) has a different likelihood in conjunction with the supplementary sentence Π that Jones is sexually active than Π alone, and therefore Λ makes probabilistic assertions relative to Π. This example also shows that there are positive instances of sentences that make absolute probabilistic assertions. As in the case of absolute falsifiability, it is impossible to prove that there is a non-tautological theory that makes no probabilistic assertions relative to any honest set without more precise notions of justification and dependence; thus the non-triviality of definition 8.9 cannot be shown at present.

As discussed in §6.9, this is primarily a problem of the notion of justification, and for probabilistic criteria of empirical significance, this means that even if definition 8.9 turns out to be inadequate, any adequate definition of 'suitable set of supplementary sentences' can be used to turn definition 8.8 into an absolute notion.

That analogues of restrictions III and IV of definition 8.3 cannot be substituted for the restriction to honest sets can be shown as follows:[21] Let the basic sentence Ω and the supplementary sentences Π be such that $\Pr(\Omega \mid \Pi)$ is defined. For instance, Π might again express that 1 out of 10 vases break when dropped and Ω that a specific vase does not break. Let again σ be justified without Ω, but entail Ω. Then Ω, σ, and Π again fulfill conditions III and IV for any theory that is not related to vases. Since σ and Π are justified independently of Ω, the falsity of Ω, and Θ, so is $\Pi^* \mathrel{\vDash\!\dashv} \{\neg \bigwedge \Theta \vee \sigma\} \cup \Pi$. Since $\Pr(\Omega \mid \Pi^*) \neq \Pr(\Omega \mid \Theta \cup \Pi^*)$, Θ would make predictions if restrictions III and IV of definition 8.3 were substituted for the restriction to honest sets.

Luckily, definition 8.9 is stricter than it would be with analogues of restrictions III and IV. The restriction of the supplementary sentences in definition 8.5 to honest sets entails restriction III. And while the restriction to honest sets does not entail restriction IVb, it precludes all trivializations precluded by that restriction: A theory Θ fails to make probabilistic assertions because of restriction IVb only if for any σ whose inclusion in Π would lead to $\Pr(\Omega \mid \Theta \cup \Pi) \neq \Pr(\Omega \mid \Pi)$ for some Ω, the justification of σ depends on Ω or the falsity of Ω. In that case, IVb ensures that Θ makes no probabilistic assertion. The restriction of Π to honest sets leads to the same result, because if the justification of σ depends on Ω (or its falsity) and Π is honest, then $\Omega \subseteq \Pi$ (or some set Δ expressing the falsity of Ω).[22] Thus $\Pr(\Omega \mid \Theta \cup \Pi) = 1 = \Pr(\Omega \mid \Pi)$ (or $\Pr(\Omega \mid \Theta \cup \Pi) = 0 = \Pr(\Omega \mid \Pi)$). Unlike restriction IVb, the restriction to honest sets also leads to identical likelihoods if $\{\neg \bigwedge \Theta \vee \sigma\} \in \Pi$ is justified by a sentence $\sigma \vDash \Omega$, thereby precluding the trivialization proof in the previous paragraph. The restriction to honest sets also precludes any trivialization that restriction IVa could preclude, for if an element of Π depends on Θ, $\Theta \subseteq \Pi$, so that $\Pr(\Omega \mid \Theta \cup \Pi) = \Pr(\Omega \mid \Pi)$ for all Ω.

Definitions 8.8 and 8.9 trivially fulfill condition of adequacy 6. Because of claim 8.5, they meet condition 7 as well. Since condition 7 only states that empirical equivalence to a tautology is a sufficient condition for empirically non-significance, 7 is also met if the biconditional of condition I in definition 8.8 is substituted by a conjunction, so that more theories are empirically non-significant.

By design, definitions 8.8 and 8.9 fulfill condition of adequacy 8. Condition of adequacy 9 is met because of

Claim 8.6. *If all occurring probabilities are defined, then Θ makes probabilistic basic assertions if and only if basic sentences are relevant for Θ.*

[21] The analogues of the restrictions feature only one theory Θ rather than two theories Θ and Λ.
[22] See note 16.

Proof. For all sets Ω of basic sentences, $\Pr(\Omega \mid \Theta \cup \Pi) \neq \Pr(\Omega \mid \Pi)$ if and only if $\Pr(\Theta \mid \Omega \cup \Pi) \neq \Pr(\Theta \mid \Pi)$ (claim 8.13). Therefore there is an Ω such that $\Pr(\Omega \mid \Theta \cup \Pi) \neq \Pr(\Omega \mid \Pi)$ if and only if there is an Ω such that $\Pr(\Theta \mid \Omega \cup \Pi) \neq \Pr(\Theta \mid \Pi)$. □

Definition 8.8 fulfills condition of adequacy 11, because it generalizes $\mathscr{B}$-creativity (see definition 6.3):

Claim 8.7. *Let Θ be a deductive theory and let all probability assignments be truth value-like. Then Θ makes probabilistic basic assertions relative to Π if and only if Θ is contingent and $\mathscr{B}$-creative relative to Π.*

Proof. Since interpretation 1 and interpretation 2 of the inequality in condition (II) are equivalent, it suffices to prove the claim for interpretation 1. Θ does not make probabilistic assertions relative to Π if and only if for every set Ω of basic sentences and every basic sentence β, $\Omega \cup \Theta \cup \Pi$ restricts the probability for β to the same set of values as Π. This is the case if and only $\Omega \cup \Theta \cup \Pi$ restricts the probability to 1 iff $\Omega \cup \Pi$ does. By claim 8.11, this holds if and only if $\Omega \cup \Theta \cup \Pi$ and $\Omega \cup \Pi$ entail the same basic sentences for every Ω, that is, if and only if Θ is $\mathscr{B}$-conservative relative to Π. If Θ makes probabilistic basic assertions, it is by assumption possible and thus, since it is also $\mathscr{B}$-creative, contingent. □

Note that the proof also holds if the biconditional of condition I in definition 8.8 is substituted by a conjunction, because then Θ makes no basic assertions if and only if for all Ω and β, $\Omega \cup \Theta \cup \Pi$ restricts the probabilities to the same set of values as $\Omega \cup \Pi$, or $\Omega \cup \Pi$ restricts the probabilities more than $\Omega \cup \Theta \cup \Pi$. But the latter is impossible since $\Omega \cup \Pi$ restricts the probabilities of a basic sentence to $\{0\}$ or $\{1\}$ only if $\Omega \cup \Theta \cup \Pi$ does. If the suitable supplementary sentences for falsifiability are given by honest sets, condition 11 is also fulfilled by definition 8.9.

Therefore, definition 8.8 and arguably definition 8.9 fulfill all conditions of adequacy that Sober wants a criterion of empirical significance to meet. Additionally, they also make it possible to evaluate one theory, ID for example, independent of another one like ET. Finally, the claims in which Sober uses testability as a one-place predicate (see §8.2.1) are not only meaningful, but also correct: ‘Undetectable angels exist’ arguably makes no basic assertions relative to any honest set of sentences, and ‘This coin has a probability of .5 of landing heads each time it is tossed’ makes basic assertions, for it assigns a probability of .5 to a basic sentence relative to $\varnothing$.

8.7 Bayesian relevance

The conclusion that $\mathscr{B}$-creativity and ‘making probabilistic assertions’ are the only two adequate criteria of empirical significance rests crucially on the assumption that only those criteria are adequate that identify all and only sets of sentences that make deductive or probabilistic basic assertions. I have noted in §6.9 that one may also be content with a criterion that identifies those sets Θ of sentences that can be confirmed or disconfirmed. For deductive criteria, this leads to weak syntactic $\mathscr{B}$-determinacy. To arrive at a criterion of empirical significance, I had suggested to include the demand that Θ be contingent relative to Π. The resulting deductive criterion relies only on *modus ponens* and is a special case of Bayesian relevance (definition 8.2):

Claim 8.8. *Let Θ be a deductive theory and let all probability assignments be truth value-like. Then basic sentences are relevant for theory Θ relative to supplementary sentences Π if and only if Θ is contingent and weakly syntactically $\mathscr{B}$-determined relative to Π.*

Proof. $\mathscr{B}$-sentences are relevant for Θ iff for some set Ω of basic sentences that is possible relative to Π (i) $\Pr(\Theta \mid \Pi) \neq 1$ and $\Pr(\Theta \mid \Omega \cup \Pi) = 1$ or (ii) $\Pr(\Theta \mid \Pi) \neq 0$ and $\Pr(\Theta \mid \Omega \cup \Pi) = 0$. By claim 8.11, (i) holds for some such Ω iff $\Pi \not\models \Theta$ and $\Omega \cup \Pi \models \Theta$ (and hence $\Pi \cup \Theta \not\models \bot$), which is the case iff Θ is contingent and verifiable relative to Π. By claim 8.10, (ii) holds for some such Ω iff $\Theta \cup \Pi \not\models \bot$ and $\Omega \cup \Theta \cup \Pi \models \bot$. This holds iff Θ is falsifiable relative to Π and furthermore contingent: Otherwise, since Ω is possible relative to Π, $\Omega \cup \Pi \mathrel{\models\!\!\!\dashv} \Omega \cup \Theta \cup \Pi \models \bot$.

By claim 6.23, Θ is contingent and either verifiable or falsifiable relative to Π iff Θ is contingent and weakly syntactically $\mathscr{B}$-determined relative to Π. □

Thus in the probabilistic case, the choice between theories that predict and theories whose truth or falsity is empirically accessible amounts to the choice between theories that make probabilistic basic assertions (definition 8.8) and theories for which basic sentences are relevant (definition 8.2). Which of the two criteria one chooses will, of course, depend on their intended application.

In §6.6, I have quoted an informal elucidation by Ayer (1936, 35) according to which a sentence is factually significant if and only if it can be confirmed or disconfirmed, and an informal definition by Carnap (1928b, 327–328) according to which a sentence has factual content if and only if it can be confirmed or disconfirmed. In the context of deductive inference, this meant that such sentences have to be weakly $\mathscr{B}$-determined; but both Ayer and Carnap meant their descriptions to apply to non-deductive inference as well, in which case their descriptions lead to Bayesian relevance. Later, Carnap (1963c, §5) presented as a thesis of empiricism the

> *Principle of confirmability*. If it is in principle impossible for any conceivable observational result to be either confirming or disconfirming evidence for a linguistic expression A, then expression A is devoid of cognitive meaning.

With reference to Carnap (1963c, §5), Skyrms (1984, 14–15) suggests the following:

> A confirmability theory of empirical meaningfulness might hold that a proposition has empirical meaning if and only if some evidence weighs for or against it; i. e., if its credibility on that evidence is greater than or less than its credibility unconditionally.

Note that unlike Carnap, Skyrms is very explicit about explicating *empirical* meaningfulness, not meaningfulness simpliciter. Since Skyrms assumes the standard probabilistic explication of confirmation and disconfirmation (definition 8.1), he describes Bayesian relevance as a criterion of empirical significance. Skyrms (1984, 14) then uses Bayes's theorem (8.1) to state that by "Bayes' theorem, a bit of evidence is relevant to the hypothesis if and only if the hypothesis is relevant to the evidence in the same sense (positive or negative)." This reformulation is only allowed if all occurring probabilities are defined, and by claim 8.6 leads to the position that all and only sentences that make probabilistic basic assertions are empirically significant.

Although sometimes mentioned in connection with Bayesian relevance,[23] Reichenbach's use of probabilities to define empirical significance differs radically from it. Reichenbach (1938, 54, emphasis removed) states the

> First principle of the probability theory of meaning: a proposition has meaning if it is possible to determine a weight, i. e., a degree of probability, for the proposition.

Thus for Reichenbach, Sober's claim that many theories of science cannot be assigned probabilities would entail that these theories have no meaning. Specifically, according to Reichenbach, the Bayesian relevance criterion of empirical significance cannot even be applied to a sentence that has no meaning.

8.8 Conclusion

Contrastive testability fails as a criterion of probabilistic empirical significance because it fails to meet four criteria of adequacy that follow from Sober's position and the intended application of contrastive testability. Given that contrastive testability is not an adequate criterion of empirical significance, I have suggested to consider a theory empirically significant if and only if it makes basic assertions, because this definition fulfills all six of Sober's conditions of adequacy, and in particular contains both $\mathscr{B}$-creativity and the Bayesian criterion of empirical significance as special cases. The criterion could be called a synthesis, as it is acceptable for falsificationists, Bayesianists, and likelihoodists alike. If one does not consider it necessary that a theory make basic assertions, however, one can use the Bayesian criterion even when not all probabilities are defined. This criterion contains weak $\mathscr{B}$-determinacy as a special case.

[23] For Sober (2008, 150, n. 26), it warrants a 'cf.', Skyrms (1984, 122, n. 21) calls it a "somewhat different" formulation.

8.9 Additional proofs

Claim 8.9. *For every language $\mathscr{L}$ and every $\mathfrak{M}$, $\mathrm{Pr}_{\mathfrak{M}} : \mathfrak{P}\mathscr{L} \to \{0,1\}, \mathrm{Pr}_{\mathfrak{M}}(\Sigma) = 1 \Leftrightarrow \mathfrak{M} \models \Sigma$ is a probability assignment.*

Proof. Show that for all $\Sigma, \Xi \in \mathscr{L}$ and any $\mathfrak{M}$ it holds:

1. $\mathrm{Pr}_{\mathfrak{M}}(\Sigma) \geq 0$,
2. $\mathrm{Pr}_{\mathfrak{M}}(\top) = 1$, and
3. if Σ and Ξ are finite and $\Sigma \cup \Xi \models \bot$, $\mathrm{Pr}_{\mathfrak{M}}(\bigwedge \Sigma \vee \bigwedge \Xi) = \mathrm{Pr}_{\mathfrak{M}}(\Sigma) + \mathrm{Pr}_{\mathfrak{M}}(\Xi)$.

1 and 2 are immediate. 3 holds because for $\Sigma \cup \Xi \models \bot$, $\mathfrak{M} \not\models \Sigma$ or $\mathfrak{M} \not\models \Xi$, so that $\mathrm{Pr}_{\mathfrak{M}}(\bigwedge \Sigma \vee \bigwedge \Xi) = 1$ if and only if either $\mathfrak{M} \models \Sigma$ or $\mathfrak{M} \models \Xi$ but not both, which holds if and only if $\mathrm{Pr}_{\mathfrak{M}}(\Sigma) = 1$ or $\mathrm{Pr}_{\mathfrak{M}}(\Xi) = 1$ but not both, that is, $\mathrm{Pr}_{\mathfrak{M}}(\Sigma) + \mathrm{Pr}_{\mathfrak{M}}(\Xi) = 1$. □

Claim 8.10. *For any sets Ξ, Σ of $\mathscr{V}$-sentences, $\Xi \cup \Sigma \models \bot$ if and only if for all $\mathfrak{M}$, $\mathrm{Pr}_{\mathfrak{M}}(\Xi \cup \Sigma) = 0$.*

Proof.

$$\begin{aligned} \Xi \cup \Sigma \models \bot &\Leftrightarrow \forall \mathfrak{M}(\mathfrak{M} \not\models \Xi \cup \Sigma) && (8.8)\\ &\Leftrightarrow \forall \mathfrak{M}(\mathrm{Pr}_{\mathfrak{M}}(\Xi \cup \Sigma) = 0 && (8.9) \end{aligned}$$

□

Claim 8.11. *For any sets Ξ, Σ of $\mathscr{V}$-sentences, $\Xi \models \Sigma$ if and only if for all $\mathfrak{M}$ it holds: If $\mathrm{Pr}_{\mathfrak{M}}(\Xi) \neq 0$ then $\mathrm{Pr}_{\mathfrak{M}}(\Sigma \mid \Xi) = 1$.*

Proof.

$$\begin{aligned} \Xi \models \Sigma &\Leftrightarrow \forall \mathfrak{M}\left[\mathfrak{M} \models \Xi \Rightarrow \mathfrak{M} \models \Sigma\right] && (8.10)\\ &\Leftrightarrow \forall \mathfrak{M}\left[\mathrm{Pr}_{\mathfrak{M}}(\Xi) = 1 \Rightarrow \mathrm{Pr}_{\mathfrak{M}}(\Sigma) = 1\right] && (8.11)\\ &\Leftrightarrow \forall \mathfrak{M}\left[\mathrm{Pr}_{\mathfrak{M}}(\Xi) \neq 0 \Rightarrow \mathrm{Pr}_{\mathfrak{M}}(\Sigma \mid \Xi) = \frac{\mathrm{Pr}_{\mathfrak{M}}(\Sigma)}{\mathrm{Pr}_{\mathfrak{M}}(\Sigma \cup \Xi)} = 1\right] && (8.12) \end{aligned}$$

□

Claim 8.12. *If the negation of a basic sentence is again a basic sentence, and Θ is deductively empirically equivalent to Λ relative to Π, then, relative to Π, Θ is $\mathscr{B}$-creative if and only if Λ is $\mathscr{B}$-creative.*

Proof. Assume that for all basic sentences β and sets of basic sentence Ω, $\Theta \cup \Omega \cup A \vDash \beta$ if and only if $\Lambda \cup \Omega \cup A \vDash \beta$. Then, for all Ω and β, $\Theta \cup \Omega \cup A \vDash \beta$ and $\Omega \cup A \nvDash \beta$ if and only if $\Theta \cup \Omega \cup A \vDash \beta$ and $\Omega \cup A \nvDash \beta$. Thus there are Ω and β such that $\Theta \cup \Omega \cup A \vDash \beta$ and $\Omega \cup A \nvDash \beta$ if and only if there are Ω and β such that $\Theta \cup \Omega \cup A \vDash \beta$ and $\Omega \cup A \nvDash \beta$. This means that Θ is $\mathscr{B}$-creative relative to Π if and only if Λ is $\mathscr{B}$-creative relative to Π. □

Claim 8.13. *If* $\Pr(\Theta \mid \Pi)$ *is defined, then* $\Pr(\Theta \mid \Omega \wedge A) = \Pr(\Theta \mid \Pi)$ *if and only if* $\Pr(\Omega \mid \Theta \cup \Pi) = \Pr(\Omega \mid \Pi)$.

Proof. The claim follows immediately from

$$\frac{\Pr(\Theta \mid \Omega \cup \Pi)}{\Pr(\Theta \mid \Pi)} = \frac{\Pr(\Omega \mid \Theta \cup \Pi)}{\Pr(\Omega \mid \Pi)} . \tag{8.13}$$

□

Part III

Applications

Chapter 9

Intelligent Design

With the results of part II, I can now tackle problems that rely on the demarcation of empirically significant sets of sentences. One such problem is the question of the status of the theory of intelligent design (ID), specifically as a competitor of evolutionary theory (ET). The current discussion of this problem is logically related to the discussion about theism and its empirical significance. This is because according to some, the statement 'God exists' may be translated as '[T]here exists necessarily a person without a body (i. e., a spirit) who necessarily is eternal, perfectly free, omnipotent, omniscient, perfectly good, and the creator of all things' (Swinburne 2004, 7),[1] which entails but is not entailed by 'An intelligent designer exists'. Analogously, 'God created the biological organisms' entails but is not entailed by 'An intelligent designer created the biological organisms'.

The view that God is a specific designer is controversial (Diamond 1975a, 39–43; Diamond 1975b), but has a long tradition. Mackie (1982, 1), for example, uses Swinburne's definition verbatim, and Sobel (2004, §6) suggests that the central properties of God according to the "common conception of traditional theology" are "omnipotence, omniscience, perfect goodness, and being the Creator and Sustainer of the universe".[2] An exception is Tooley (1975, 483), who, although he discusses traditional theism, defines God as "the one person who, though he can act within the world and can communicate with man, is neither dependent upon the world nor simply a part of it, but rather transcends the realm of human existence, and who, in addition, is morally perfect, omnipotent, omniscient, eternal, and incorporeal". Since Tooley does not define God to *actually* act or have acted in this world, according to his definition the existence of God arguably does not entail the

[1]With a reference to his discussion of the trinity, Swinburne adds in a footnote: "In understanding God as *a* person, while being fair to the Judaic and Islamic view of God, I am oversimplifying the Christian view."

[2]Nielsen (1966, 13) lists further definitions according to which God is a designer.

existence of a designer. Of course, the statement that God created the biological organisms still entails that an intelligent designer created the biological organisms. Theological positions that do not construe theism as having any empirical impact or factual content of course do not easily relate to ID.

Many proponents of ID claim that the adaptations of biological organisms are evidence for the existence of an intelligent designer. This inference is weaker than the classic version of the teleological argument for the existence of God, also called the argument for (or to) design or, somewhat misleadingly, the argument from design (Mackie 1982, 133). This version, by Paley (1802), relies mostly on the adaptations of biological organisms as evidence, but concludes both that an intelligent designer exists (Paley 1802, §XXIII) and that the designer has the properties of God (Paley 1802, §§XXIV–XXVI). While the evidence for some of these properties are of an astronomical or chemical nature (Paley 1802, §§XXI–XXII), the conclusions that Paley draws from the biological evidence alone go far beyond the existence of some designer. Since Paley's conclusion is logically stronger than the conclusion of the ID argument, it is in greater danger of disconfirmation. The omnipotence and benevolence of God, for example, are in tension with the existence of evil (Mackie 1982, §9; Sobel 2004, §XII) and the maladaptations found in biological organisms (Mackie 1982, 138; Sobel 2004, §7.1). The existence of an intelligent designer is not disconfirmed by such evidence, since a not further specified intelligent designer may fall short in power or benevolence.

On the other hand, there are two challenges to theism that are at least as and generally more urgent for logically weaker theories, and thus for ID. According to the falsifiability challenge, theism makes no assertions whatsoever because it is compatible with any consistent empirical statement. According to the translatability challenge, there is a set of empirical statements with the same cognitive content as theism, and since empirical statements cannot contain references to God, any such reference is devoid of cognitive content. However, even assuming that theism is intended to be an empirical theory, it does not have to meet the challenges in their current form, for both rely on assumptions that are controversial, to put it mildly. In §9.1, I will therefore review the falsifiability challenge and argue that its controversial assumptions can be avoided if theistic utterances are directly charged with lack of empirical assertions. I will review the translatability challenge in §9.2 and argue that its problematic assumptions are avoided by a related challenge, according to which theism cannot replace a competing non-theistic theory if the competing theory has not been disconfirmed, and makes all the empirical assertions of theism, but not vice versa. ID then faces both of these challenges as well.

As already discussed, Elliot Sober (1999, 2007, 2008) argues that for probabilistic theories, the criterion of falsifiability should be replaced by contrastive testability, and he charges ID with a lack of contrastive testability. In §9.3.1, I show that according to Sober's own claims, ID is testable because it is falsifiable.

I will then argue that Sober's argument for ID's lack of contrastive testability is invalid, and that ID is in fact contrastively testable (§9.3.2). Sober's challenge to ID, I suggest, inconsistently combines the modified falsifiability challenge with the modified translatability challenge. In the next step, I aim to assess whether ID in fact meets either of the modified challenges by first elucidating a variety of different and incompatible suggested definitions of ID (§9.4). Under one very plausible definition, ID succumbs to the modified falsifiability challenge, and this entails that it is analytically false under other definitions (§9.5.1). Under yet another definition, ID succumbs to the modified translatability challenge (§9.5.2). The modified falsifiability challenge is a plausible necessary condition for scientific theories, and the modified translatability challenge is a plausible necessary condition for scientific research. In §9.6, I show that an argument by Laudan that there can be no criterion for 'being scientific' is not valid.

9.1 The falsifiability challenge

The falsifiability challenge (the name is due to Tooley 1975, 485) is motivated by the impression that theism has suffered the "death by a thousand qualifications", where a "fine brash hypothesis" is weakened until it fails to assert anything (Flew 1950, 258). Statements like 'God created the world' and 'God loves us as a father loves his children' (Flew 1950, 258), for example, seem to assert, among other things, that there are no maladaptations in organisms and nothing bad will ever happen to us (unless, maybe, we deserve it). But even though these assertions are false, many theists do not give up the original statements, and rather claim that God's creation and love differ from human creations and love. These modifications qualify the original statements to an extent that it becomes unclear whether the statements assert anything at all. For Flew (1950, 259), this suggests a challenge to his fellow disputants:

> I therefore put to the succeeding symposiasts the simple central question, 'What would have to occur or to have occurred to constitute for you a disproof of the love of, or of the existence of, God?'

Flew (1950, 258–259) justifies the challenge by arguing that to assert a statement is to deny its negation, so that someone who does not deny any statements also does not assert any.

Nielsen (1966, 15) points out that every statement meets the falsifiability challenge if there is no restriction on which statements one may deny, and suggests that Flew has to assume a restriction to "non-religious, straightforwardly empirical, factual statements", which thus play the role of the basic sentences. Flew (1975, 274) claims that Nielsen's restriction is too exclusive, and suggests a restriction to "anything which happens or which conceivably might happen in the ordinary world". This restriction is arguably too vague, since one may think, *pace* Tooley

(1975, 483), that God exists in the ordinary world. I will just call the statements that the restriction allows 'observation statements', which is traditional. Their exact nature will not be important until later, and Nielsen's remarks may serve as a guide until then.

Thus the falsifiability challenge rests on the assumption that a theory makes an assertion only if it is incompatible with an observation statement. But this is debatable because some assertions may just not be observational, and it is manifestly false for probabilistic theories. Thus there is no reason to assume that only falsifiable theories make observational assertions, and no reason why a theory would have to meet the falsifiability challenge.

Tooley (1975, §III) suggests to modify the falsifiability challenge so that proponents of a theory are not challenged to show its falsifiability, but rather its testability, that is, its confirmability or disconfirmability. Confirmation and disconfirmation here rely on some concept of induction to be specified. But Tooley's strategy confuses an indicator (falsifiability) with the property it indicates (making assertions). Generalizing falsifiability to testability makes sense only if this at the same time generalizes the criterion from an indicator of theories that make deductive assertions to an indicator of theories that make deductive or probabilistic assertions. Otherwise, the original motivation for the challenge is lost. Flew motivates the falsifiability challenge by noting that, although theistic statements *prima facie* make assertions, many theists use common words in an uncommon way, thereby weakening the theistic statements. Flew's challenge is accordingly a question about the personal beliefs of his fellow disputants, and the goal is to elucidate what assertions the theistic statements make *according to them*, and accordingly what the words 'create' and 'love' mean *according to them* when it comes to God. It is an interesting psychological point that it may be easier for humans to identify a theory's assertions when they are asked to think about what the theory denies rather then what it asserts. But as far as a statement is independent of the vagaries of the human psyche, such a reformulation is not necessary. If the question is whether a statement makes assertions, it is enough to find a criterion that determines just that. Thus Flew's falsifiability challenge is but the demand that theistic claims be $\mathcal{B}$-creative, reformulated with the help of claim 6.1.[3] When generalizing the falsifiability challenge, one thus has to generalize $\mathcal{B}$-creativity, not falsifiability.

That a statement makes an assertion only if it makes an observational assertion is a very controversial assumption, but as Nielsen's consideration shows, some restriction to the assertions is necessary to arrive at a non-trivial challenge. Since according to some, theism is not meant to make observational assertions (cf. Diamond 1975b), theism may not have to meet any challenge that demands them. ID, on the other hand, is meant to be an empirical theory, and thus has to make

[3] Flew hence motivates the falsifiability challenge to theism in the same way that Popper motivates his falsifiability criterion in general (see page §225).

observational assertions. With $\mathscr{B}$-sentences as observational sentences, ID thus has to be $\mathscr{B}$-creative or make probabilistic basic assertions. Call this the *modified falsifiability challenge*.

With definitions 6.3 and 8.9, the falsifiability challenge can be modified to allow for probabilistic theories without giving up on its basic idea: The challenge for ID is to make observational assertions (simpliciter), that is, either deductive or probabilistic ones. This challenge is justified because it is plausible that every empirical theory must make some observational assertions. Laudan (1983b, 37) claims to have given a counterexample of a respectable scientific theory that makes no observational assertions, but the relevant passage of the article he refers to (Laudan 1982, 17) contains only the point that theories often rely on supplementary sentences for their observational assertions. Since definitions 6.3 and 8.9 allow for supplementary sentences, this is not an objection to the modified falsifiability challenge.

While Ayer and Popper consider the making of observational assertions to be a necessary and sufficient condition for being a scientific theory (see §6.9), Laudan (1982, 18) and Rothbart (1982, 95) note that this is clearly too strong. 'Anne got her head stuck in a drainpipe' makes an observational assertion, but is no scientific theory in that the statement does not, for example, offer the systematic economy or heuristic fertility that Hempel (1965g) expects from a theory.[4] This feature of scientific theories also poses a problem for the translatability challenge.

9.2 The translatability challenge

Tooley (1975, 489–490) provides both the name 'translatability challenge' and its most explicit formulation—the formulation by Ayer (1936, 114–120), although the basis of Tooley's, is comparatively freewheeling. Assuming that only the set of analytic statements is suitable as the set Π of supplementary sentences, the argument is roughly as follows: Only a statement that makes deductive observational assertions has cognitive content, and two statements that in conjunction with Π entail the same observation statements have the same cognitive content. It follows that if a theological statement Θ has cognitive content, the set Ω of observation statements entailed by $\Theta \cup \Pi$ has some, namely the same cognitive content as Θ. Since the term 'God' in any theological statement Θ is intended to refer to something that is not observational, 'God' does not occur in Ω. Therefore a theological statement Θ either has no cognitive content at all, or the specific theological content of Θ, the reference to God, is not cognitive.

There are a number of problems with this argument, for example the apparent irrelevance of Θ having *any* cognitive content, and the problem of explicating the notion of 'reference'. I want to restrict my discussion to the simple point that it is

[4] I have not explicated 'theory', but Hempel's demand is plausible in that it suggests, for example, that a theory can provide understanding, psychologically understood (cf. de Regt 2009).

doubtful that two theories that entail the same observation statements have the same cognitive content. For one, the theories may assign different probabilities to some observation statement without entailing it.[5] Furthermore, it is doubtful that $\Theta \cup \Pi$ can be replaced by Ω without cognitive loss. Hempel (1965g, 222), for example, argues that non-observational statements are necessary for "inductive explanatory and predictive use and [. . .] systematic economy and heuristic fertility". Niiniluoto (1972) and Tuomela (1973) give overviews of the discussion about the need for non-observational statements, essentially supporting Hempel's conclusion. Thus there is no reason why theism would have to meet the translatability challenge, because it is implausible that a theory can be replaced without significant loss by the observation statements it entails.

On the other hand, a theory may be replaced by another *theory* without losing the theoretical virtues that can come with the use of non-observational statements. Such a replacement clearly comes with a cognitive loss, however, if the replaced theory is not disconfirmed and makes all the observational assertions of its replacement, while the replacing theory does not make all the observational assertions of the replaced theory. For then, the replacement would result in the systematization of fewer observations. Call this the *modified translatability challenge*. Rothbart (1982, 99) suggests that a necessary condition for being scientific is to be worthy of experimental testing, which entails that a "hypothesis must encapsulate its rival's successes". Taking the successes to be the observational assertions of a theory that has not been disconfirmed, this is the modified translatability challenge. Rothbart's notion of 'scientific' shifts the focus from individual theories to sets of theories: Theories are scientific relative to other theories, so that a theory may become unscientific if a new theory is developed. Considering Rothbart's motivation (that a theory is scientific when it is worth testing), one may want to see the modified translatability challenge not as a criterion for scientific *theories*, but rather as a criterion for scientific *research*. The pursuit of a theory that only makes empirical assertions that a more general theory also makes is arguably not scientific.

Many contemporary theistic theories do not construe God's existence as competing with any empirical theories (cf. Mackie 1982, §8; Sobel 2004, §VII), so that they are not intended as replacements. ID, on the other hand, is meant to replace evolutionary theory (ET), and since ET is not disconfirmed, this is impossible if ET makes all the observational assertions of ID, while ID fails to make all the observational assertions of ET.

This modification of the translatability challenge does not rely on any controversial assumptions about the cognitive content of theories, but does need a concept of 'making all the observational assertions of'. Deductively, this can be expressed by

[5] As with the falsifiability challenge, Tooley (1975, 505–506) uses an intentionally underspecified notion of testability to account for probabilistic inferences. And as in the case of the falsifiability challenge, this confuses the indicator of a property (being tested by the same observation statements) with the property itself (making the same observational assertions).

Definition 9.1. Theory Θ *makes all the deductive $\mathscr{B}$-assertions of* Λ relative to Π if and only if for all possible sets Ω of observation sentences and all observation sentences β

$$\Omega \cup \Lambda \cup \Pi \vDash \beta \text{ only if } \Omega \cup \Theta \cup \Pi \vDash \beta \,. \tag{9.1}$$

The following holds:

Claim 9.1. *Theory Θ makes all the deductive $\mathscr{B}$-assertions of Λ relative to Π if and only if Θ is at least as syntactically falsifiable as Λ.*

Proof. '$\Leftarrow$': Assume $\Omega \cup \Lambda \cup \Pi \vDash \beta$. Then $\Omega \cup \{\neg\beta\}\Lambda \cup \Pi \vDash \bot$. By assumption, $\Omega \cup \{\neg\beta\} \cup \Lambda \cup \Pi \vDash \bot$. Thus $\Omega \cup \Theta \cup \Pi \vDash \beta$.

'$\Rightarrow$': Assume $\Omega \cup \Lambda \vDash \bot$. Since $\bot$ is a $\mathscr{B}$-sentence, $\Omega \cup \Theta \vDash \bot$ by assumption. □

The proof only assumes that the negation of a $\mathscr{B}$-sentence is again a $\mathscr{B}$-sentence. The probabilistic version is given by

Definition 9.2. Theory Θ *makes all the probabilistic $\mathscr{B}$-assertions of* Λ relative to Π if and only if for all sets Ω of observation sentences

$$\Pr(\Omega \,|\, \Lambda \cup \Pi) = \Pr(\Omega \,|\, \Pi) \text{ or } \Pr(\Omega \,|\, \Lambda \cup \Pi) = \Pr(\Omega \,|\, \Theta \cup \Pi) \,. \tag{9.2}$$

Claim 9.2. *Let Θ and Λ be deductive theories and let all probability assignments be truth value-like. Then Θ makes all probabilistic $\mathscr{B}$-assertions of Λ relative to Π if and only if Θ makes all deductive $\mathscr{B}$-assertions of Λ relative to Π*

Proof. '$\Leftarrow$': $\Pr(\Omega|\Lambda \cup \Pi) = 1$, $\Pr(\Omega|\Lambda \cup \Pi) = 0$, or $\Pr(\Omega|\Lambda \cup \Pi) = \{0,1\}/$ $\Pr(\Omega|\Lambda \cup \Pi)$ is undefined. Without loss of generality, assume that in the last case, $\Pr(\Omega|\Lambda \cup \Pi)$ is undefined. Proof by cases:

If $\Pr(\Omega|\Lambda \cup \Pi) = 1$, then by claim 8.11 $\Theta \cup \Pi \vDash \Omega$ and by assumption $\Theta \cup \Pi \vDash \Omega$. Thus by claim 8.11, $\Pr(\Omega|\Theta \cup \Pi) = 1 = \Pr(\Omega|\Lambda \cup \Pi)$.

If $\Pr(\Omega|\Lambda \cup \Pi) = 0$, then $\Pr(\Omega \cup \Lambda \cup \Pi) = 0$ and by claim 8.10 $\Omega \cup \Lambda \cup \Pi \vDash \bot$. Since $\bot$ is a $\mathscr{B}$-sentence, by assumption $\Omega \cup \Theta \cup \Pi \vDash \bot$. by claim 8.10, $\Pr(\Omega|\Theta \cup \Pi) = 0 = \Pr(\Omega|\Lambda \cup \Pi)$.

If $\Pr(\Omega|\Lambda \cup \Pi)$ is undefined, $\Pr(\Omega|\Lambda \cup \Pi) \neq 1$ and $\Pr(\Omega|\Lambda \cup \Pi) \neq 0$. By claims 8.11 and 8.10, $\Lambda \cup \Pi \nvDash \Omega$ and $\Omega \cup \Lambda \cup \Pi \nvDash \bot$. Therefore $\Lambda \cup \Pi \nvDash \Omega$ and $\Lambda \cup \Pi \cup \Omega \nvDash \bot$. By claims 8.11 and 8.10, $\Pr(\Omega|\Theta \cup \Pi) \neq 1$ and $\Pr(\Omega|\Theta \cup \Pi) \neq 0$, so that $\Pr(\Omega|\Theta \cup \Pi)$ is undefined as well.

'$\Rightarrow$': If $\Omega \cup \Lambda \cup \Pi \vDash \beta$, then $\Omega \cup \{\neg\beta\} \cup \Lambda \cup \Pi \vDash \bot$. By claim 8.10, $\Pr(\Omega \cup \{\neg\beta\}|\Lambda \cup \Pi) = 0$, hence by assumption $\Pr(\Omega \cup \{\neg\beta\}|\cup \Pi) = 0$ or $\Pr(\Omega \cup \{\neg\beta\}|\Theta \cup \Pi) = 0$. By claim 8.10, $\Omega \cup \{\neg\beta\} \cup \Pi \vDash \bot$ or $\Omega \cup \{\neg\beta\} \cup \Theta \cup \Pi \vDash \bot$ and thus $\Omega \cup \{\neg\beta\} \cup \Theta \cup \Pi \vDash \bot$ and $\Omega \cup \Theta \cup \Pi \vDash \beta$. □

Thus *making all the probabilistic $\mathscr{B}$-observations* can be considered the probabilistic generalization of *being at least as falsifiable*. It is another justification of definitions 9.1 and 9.2 that they relate to probabilistic and deductive empirical equivalence (definitions 6.38 and 8.7) as expected:

Claim 9.3. *Theory Θ is deductively/probabilistically empirically equivalent to theory Λ relative to Π if and only if Θ makes all the deductive/probabilistic $\mathcal{B}$-assertions of Λ and Λ makes all the deductive/probabilistic $\mathcal{B}$-assertions of Θ relative to Π.*

Proof. Directly from the definitions. □

The absolute conceptions of empirical equivalence are given by

Definition 9.3. Theory Θ is deductively/probabilistically empirically equivalent to theory Λ if and only if Θ is deductively/probabilistically empirically equivalent to theory Λ relative to an honest set Π.

The modified translatability challenge differs fundamentally from the modified translatability challenge in its structure. A theory Θ is not evaluated on its own, but rather compared to another theory Λ, the theory it is meant to replace. Their relation is given by

Claim 9.4. *If Θ makes deductive/probabilistic $\mathcal{B}$-assertions while Λ does not, then Θ makes all the deductive/probabilistic $\mathcal{B}$-assertions of Λ, and Λ does not make all the deductive/probabilistic $\mathcal{B}$-assertions of Θ relative to Π.*

Proof. Immediate from the definitions. □

Since ET makes observational assertions, the modified translatability challenge is harder to meet for ID than the modified falsifiability challenge.

9.3 The testability of Intelligent Design

Since Sober, like Tooley, aims to generalize falsifiability to testability, he modifies the falsifiability challenge. The main difference to Tooley's approach is that Sober gives a definition of 'testing' (definition 8.2).

What role contrastive testability could play in a criticism of ID is somewhat unclear, however, because the definition is symmetric: The definiens is invariant up to logical equivalence under exchange of Θ and Λ. If ID turned out not to be testable against ET, this would therefore be *prima facie* bad for ID if and only if it would be bad for ET. ID may have further properties that distinguish it negatively from ET, but the question would then still be what role contrastive testability can play. Of course, that question only has to be answered if ID is in fact not contrastively testable against ET. As noted, Sober (2007, 3) considers a minimal version of ID to show that ID cannot meet the testability challenge:

> The single thesis of what I will call mini-ID is that the complex adaptations that organisms display (e. g., the vertebrate eye) were crafted by an intelligent designer.

Sober's general strategy is to argue that ID cannot be tested against ET because ID "does not predict much of anything" (Sober 2008, §2.15). I will consider this argument in §9.3.2. Surprisingly, Sober first argues that ID is deductively testable.

9.3.1 The deductive testability of Intelligent Design

Although the designer in ID is not specified and thus has wholly mysterious intentions (Sober 1999, 65; Sober 2007, 6; Sober 2008, 128, n. 14), Hartwig and Meyer (1993, 160) argue that ID is still falsifiable. For "the concept of intelligent design predicts that complex information, such as that encoded in a functioning genome, never arises from purely chemical or physical antecedents". Against this, Sober (2007, 6, 5; 2008, 130) first points out that probabilistic theories are not falsifiable and falsifiability therefore cannot be a good criterion of testability. Second (implicitly assuming that a not purely chemical or physical antecedent is always an intelligent designer), he argues that the argument is invalid because the statement 'Somewhere on the causal chain leading up to "complex information" there is an intelligent designer at work' is not falsifiable. This is because 'somewhere' may refer to a point outside of space and time, and because the intelligent designer may be unobservable, so that ID does not make an *observational* assertion (Sober 2007, 6–7).

The second point is well-taken, and echos Nielsen's remark that every theory is falsifiable if it only has to be incompatible with *some* statement. A falsifiable theory must be incompatible with an *observation* statement, and that something "causes" or (metaphorically) "arises from" something else is not observational. Even if it usually was, the antecedents that cause complex information, or from which complex information arises, can be unobservable; they could be God or some other unobservable designer, for example. And the statement that some two-place relation holds is not generally observational if one of the relation's arguments is not observable. This is especially clear for an otherwise observational relation like 'is standing next to'. However, the point is also a red herring because Sober (2008, 148; cf. 2008, 130; 2007, 4) himself considers it "perfectly clear" that ID deductively entails the observation sentence 'Organisms display complex adaptations' ('ADAP' in the following), so that ID is falsifiable according to claim 6.1. Hartwig and Meyer have only chosen the wrong statement to argue their case.

This is important because Sober's first point, his dismissal of falsifiability, is mistaken according to his own claims: While falsifiability is not a necessary condition for deductive testability, it is a sufficient condition. For Sober (1999, 72, n. 14) states:

> The thesis that testing is contrastive requires that prediction be probabilistic; otherwise, hypotheses could be falsified without any contrastive alternative having to play a role. If $H \wedge A$ deductively entails O, and A is known to be true, then, if we observe $\neg O$, we can conclude

> that H is false.

And in another passage, Sober (2008, 52, n. 29) states:

> There are two exceptions to the thesis that testing is always contrastive. If a true observation statement entails H, there is no need to consider alternatives to H; you can conclude without further ado that H is true; this is just *modus ponens*. And if H entails O and O turns out to be false, you can conclude that H is false, again without needing to contemplate alternatives; this is just *modus tollens*.

I have argued in §8.4 that Sober cannot allow verifiability as a sufficient condition for empirical significance because some verifiable sentences do not make observational assertions. But it is clear that Sober accepts and should accept falsifiability as a sufficient condition for testability. Hence, since ID entails ADAP, it is falsifiable and thus testable simpliciter.

9.3.2 The contrastive testability of Intelligent Design

Sober (2008, §2.12) argues that the "problem with the hypothesis of intelligent design" is that "it doesn't predict much of anything" (Sober 2008, §2.15). (The qualifier 'much of' is necessary because according to Sober, ID does assert ADAP.) And since for "ID to be testable, *it* must make predictions" (Sober 2007, 7), "the design hypothesis is untestable" (Sober 2008, 148; cf. Sober 1999, 66–67). This conclusion is false because ID is falsifiable, so that Sober's claim at least has to be weakened to the claim that ID is not *contrastively* testable against ET. Sober claims that 'The complex adaptations were crafted by an intelligent designer' and 'The complex adaptations are the result of natural selection' both entail ADAP, so that $\Pr(\text{ADAP} \mid \text{ID} \cup \Pi) = 1 = \Pr(\text{ADAP} \mid \text{ET} \cup \Pi)$. More generally, there may be a set **ADAP** of sentences (not very many, since ID "doesn't predict much") whose probabilities are related to that of ADAP, and $\Pr(\mathbf{ADAP}' \mid \text{ID} \cup \Pi) = \Pr(\mathbf{ADAP}' \mid \text{ET} \cup \Pi)$ for each $\mathbf{ADAP}' \subseteq \mathbf{ADAP}$.[6] Therefore ADAP may be the reason for ID being falsifiable, but it cannot be the reason for ID being contrastively testable against ET. I will now argue that ID is nonetheless contrastively testable against ET.

First, note that Sober's argument that ID cannot be tested against ET relies on a hidden premise: Sober claims that ID does not make observational assertions besides those in **ADAP**. Thus $\Pr(\Omega \mid \text{ID} \cup \Pi) = \Pr(\Omega \mid \Pi)$ whenever both probabilities are defined and $\Omega \nsubseteq \mathbf{ADAP}$. Furthermore, if ID cannot be tested against ET, $\Pr(\Omega \mid \text{ET} \cup \Pi) = \Pr(\Omega \mid \text{ID} \cup \Pi)$ whenever both probabilities are defined. Therefore, Sober's argument is valid only if *for any* $\Omega \nsubseteq \mathbf{ADAP}$ *such that* $\Pr(\Omega \mid \text{ET} \cup \Pi) \neq \Pr(\Omega \mid \Pi)$ *and both probabilities are defined,* $\Pr(\Omega \mid \text{ID} \cup \Pi)$ *is undefined*. But ET makes predictions, and, more specifically, there are many Ω'

[6] Sober never spells out precisely which observational assertions ID makes besides ADAP.

with $\Omega' \nsubseteq$ **ADAP** such that $\Pr(\Omega' \mid \mathrm{ET} \cup \Pi) \neq \Pr(\Omega' \mid \Pi)$ and both probabilities are defined.[7] Then, since $\Pr(\Omega' \mid \Pi)$ is defined, by definition 8.9 $\Pr(\Omega' \mid \mathrm{ID} \cup \Pi)$ is defined as well. Thus the hidden premise is false. And since $\Pr(\Omega' \mid \mathrm{ET} \cup \Pi) \neq \Pr(\Omega' \mid A) = \Pr(\Omega' \mid \mathrm{ID} \cup \Pi)$, where all probabilities are defined, ID can be tested against ET.

This inference relies on the possibly contentious component of definition 8.8 according to which, if ID makes no probabilistic assertions, $\Pr(\Omega \mid \mathrm{ID} \cup \Pi)$ is always defined when $\Pr(\Omega \mid \Pi)$ is defined. I will now argue that even under a weaker assumption, ID can be tested against ET. That $\Pr(\Omega \mid \mathrm{ID} \cup \Pi)$ is at least *sometimes* defined when $\Pr(\Omega \mid \Pi)$ is defined follows from Sober's example with Jones's birth control (cf. §8.5). With $\Theta = \{$'God created the eye'$\}$, $\beta =$ 'Jones is pregnant', $\Pi = \{$'Jones is sexually active'$\}$, and $\Lambda = \{$'Jones used birth control'$\}$, Sober (1999, n. 24) states that $\Pr(\beta \mid \Pi) = \Pr(\beta \mid \Theta \cup \Pi) > \Pr(\beta \mid \Lambda \cup \Pi)$. Since for Sober (1999, 62, 65), Θ here stands for ID, $\Pr(\Omega \mid \mathrm{ID} \cup \Pi)$ is defined for some sets Ω of observation sentences with $\Omega \nsubseteq$ **ADAP**, for example $\{$'Jones is pregnant'$\}$. It is thus not clear why exactly for those $\Omega \nsubseteq$ **ADAP** with $\Pr(\Omega \mid \mathrm{ET} \cup \Pi) \neq \Pr(\Omega \mid \Pi)$, ID should render the probability for Ω undefined.

It is thus very plausible that *at least for one* Ω^* with $\Omega^* \nsubseteq$ **ADAP** for which $\Pr(\Omega^* \mid \mathrm{ET} \cup \Pi) \neq \Pr(\Omega^* \mid \Pi)$ and both probabilities are defined, $\Pr(\Omega^* \mid \mathrm{ID} \cup \Pi)$ is defined as well. This very weak assumption is also independently plausible: Take the statement that the eye of some organism has a specific feature. Based on our background knowledge about the ratio of the occurrence of this feature in other organisms' eyes, we might be able to assign a specific probability p to the occurrence of this feature, and based on our background knowledge and ET, we might be able to assign a different probability. But based on our background knowledge, we can also infer that *if* that eye was created by a designer, this designer had the intention and ability to create an eye with that feature with probability p. In conjunction with ID, the probability for the occurrence of the feature then does not become undefined, but rather keeps the value p. Under this very weak assumption, ID can be tested against ET. For then $\Pr(\Omega^* \mid \mathrm{ET} \cup \Pi) \neq \Pr(\Omega^* \mid \Pi) = \Pr(\Omega^* \mid \mathrm{ID} \cup \Pi)$, where all probabilities are defined.

* * *

Sober suggests contrastive testability as an improvement over falsifiability, and argues that ID is not contrastively testable because it does not assert "much of anything". This all sounds as if he claimed that ID fails to meet a modification of the falsifiability challenge. However, when it comes to the observational assertion that organisms display complex adaptations, Sober argues that ET makes the same assertion, and thus ID is still not contrastively testable. This sounds as if Sober

[7] This assumes that **ADAP** does not contain all probabilistic assertions of ET. This is clearly what Sober assumes, for otherwise ID would make all probabilistic empirical assertions of ET, and thus it would hold that ET "doesn't predict much of anything".

claimed that ET fails to meet the modified translatability challenge. A theory that fails to meet the modified falsifiability challenge indeed fails the modified translatability challenge against almost any other theory, but a theory that makes *almost* no observational assertions still makes some.

Sober presumably treats the three challenges as one because he assumes that a theory that does not make observational assertions renders all likelihoods undefined. It is surprising that Sober never argues for this claim. Since it is false, it is unsurprising that it leads to inconsistencies. Given such confusions, the inadequacy of contrastive testability given Sober's own conditions of adequacy, and since it is not clear how a symmetric relation between theories can provide a reason to prefer one theory over another, it is safe to consider it irrelevant whether ID meets the testability challenge against ET.

Whether ID meets the modified falsifiability or translatability challenge depends on whether Sober is right in arguing that ID makes only one observational assertion, and right in his exposition of ID. Sober claims that 'The complex adaptations were crafted by an intelligent designer' and 'The complex adaptations are the result of natural selection' both entail ADAP. But, it seems, so does 'The complex adaptations are the result of circles being round', which gives Sober's inference the air of hocus-pocus. Thus, while the status of the three criteria is reasonably clear, the status of ID still requires investigation.

9.4 What is the theory of Intelligent Design?

To determine whether ID fails either of the two modified challenges, it is necessary to know more or less precisely what the theory of intelligent design is. Sober (2007, 3) considers the minimal version of ID according to which "the complex adaptations that organisms display [. . .] were crafted by an intelligent designer". Since a theory is more likely to fail either of the two challenges the weaker it is, this formulation is unfairly minimal: For instance, while Sober only refers to complex adaptations of organisms, Hartwig and Meyer (1993, 160) refer to any kind of complex information (see §9.3.1). The technical term in the ID literature is 'complex specified information' ('CSI').[8] Thus ID should be taken to be the statement that the objects with CSI were caused by an intelligent designer.

Although according to Sober "it is perfectly clear" that ID entails the existence of objects with CSI (in his version of ID, the existence of complex adaptations), it depends on the interpretation of the definite article whether this is so. Under a suggestion by Sharvy (1980, 615), who follows Russell (1905), Sober's version of ID is to be paraphrased as "There is a collection of objects with CSI such that all collections of objects with CSI are its parts, and this collection of objects with

[8]In this section, it suffices to rely on an intuitive notion of CSI: New York, brains, and computer chips have it, Kazimir Malevich's "Black square", rocks, and sweat do not.

CSI was caused by a designer”. [9] If there are no objects with CSI, there cannot be a collection of them, so that in this paraphrase, ID entails the existence of objects with CSI. The same holds for a suggestion by Brogaard (2007, 164), which leads to the paraphrase “There are objects with CSI that were created by a designer and all objects with CSI are some of them”. Brushing over the subtleties of the natural language formulation, both paraphrases can be formulated as ‘There are objects with CSI and all objects with CSI are caused by an intelligent designer’. Letting, for any x and y, $\ulcorner Dy \urcorner$ stand for $\ulcorner y$ is an intelligent designer$\urcorner$ and $\ulcorner Cyx \urcorner$ stand for $\ulcorner y$ caused $x\urcorner$, this leads to the following paraphrase of Sober’s definition of ID:

ID 1. There are objects with CSI, and for all objects x with CSI,

$$\exists y(Dy \wedge Cyx)\,. \tag{9.3}$$

Russell’s definition of the definite article is not the only plausible one. According to Strawson, the definite article indicates that the existence of the objects in its scope is presupposed, that is, a background assumption (cf. Ludlow 2009, §4.2). Assuming that ‘CSI’ is an observational term, ‘There are objects with CSI’ is justified by observations. The sentence can thus be an honest supplementary sentence, so that ID can be defined as follows:

ID 2. For all x with CSI,

$$\exists y(Dy \wedge Cyx)\,. \tag{9.4}$$

ID 1 entails but is not entailed by ID 2. Hartwig and Meyer (1993, 160) arguably have ID 2 in mind when they claim that ID entails that CSI never arises from purely chemical or physical antecedents *without* claiming that ID entails the existence of objects with CSI. Dembski (2006, emphasis removed) states that “[p]roponents of intelligent design, known as design theorists, [. . .] claim that a type of information, known as specified complexity, is a key sign of intelligence.” If ‘key sign’ means ‘sufficient condition’, then proponents of intelligent design claim that all complex specified information is caused by an intelligent designer, that is, ID 2.

Sober’s version of ID is not the only one suggested. Some definitions differ radically in their logical structure from ID 1 and ID 2: According to the Discovery Institute (Anonymous 2010a),[10]

> [t]he theory of intelligent design holds that certain features of the universe and of living things are best explained by an intelligent cause, not an undirected process such as natural selection.

Two sections below (Anonymous 2010b), the “certain features” of the universe and living things are further specified:

[9] The substitution of ‘caused’ for ‘created’ is meant to align the paraphrase with the definitions of ID proponents discussed below.

[10] I thank Casey Luskin for helpful discussions about this definition.

> Intelligent design begins with the observation that intelligent agents produce complex and specified information (CSI). Design theorists hypothesize that if a natural object was designed, it will contain high levels of CSI. Scientists then perform experimental tests upon natural objects to determine if they contain complex and specified information. One easily testable form of CSI is irreducible complexity, which can be discovered by experimentally reverse-engineering biological structures [...].

Thus it is claimed that, starting from the observed relation between intelligent agents and CSI, ID theorists conjecture that all natural objects caused by an intelligent designer contain CSI and go on to search for natural objects with CSI.[11] Note that the ID theorists only conjecture that in our world, the material implication from an intelligent cause of an object to the occurrence of CSI in that object is true, which is a statement of factual co-occurrence. It is another step to the claim that an intelligent cause *explains* the occurrence of CSI (see the discussion in connection with ID 3 and §9.5.1). The further assumption that undirected processes do *not* explain CSI or are unlikely to do so (critically discussed, for example, by Häggström (2007a,b)) then leads to the Discovery Institute's definition of ID.

Thus the Discovery Institute's definition can be paraphrased as follows:

> The theory of intelligent design holds that those features of the universe and of living things with CSI are best explained by an intelligent cause, not an undirected process.

Since living things are in the universe, their mention is redundant.[12] Since an intelligent cause (a designer) is claimed to explain features with CSI *best*, it is claimed to explain the features better than any process without an intelligent cause. In the following I will call any such process 'undirected'. Finally, the explanation of a feature F by a cause C only makes sense if C is the cause of F. The definition can thus once more be paraphrased as

> The theory of intelligent design holds that for all x with CSI, the statement that x was caused by an intelligent designer explains the CSI of x better than any statement according to which x was caused by an undirected process.

The further discussion of the Discovery Institute's definition will rely on the explication of 'explanation'. This is a notoriously difficult subject, and for

[11] This is somewhat puzzling because the conjecture does not seem to be inductively supported by the observations and, in fact, false: Even many objects *intentionally* produced by intelligent agents, for example the "Black Square", anvils, and (sometimes) sweat, do not show CSI. Note that the conjecture's converse is just ID 2, and thus cannot be assumed to be inductively supported without begging the question (cf. Sober 2008, 176).

[12] Logically redundant, that is. Mentioning living things puts an emphasis on the main application of ID, irreducible complexity in organisms.

this reason, few if any scientific theories rely on the concept as a primitive.[13] For a definition of ID, there are good reasons not to connect explanation to the psychological concept of understanding, for one because then a good explanation for one person is not always a good explanation for another who may, for example, lack specific background knowledge. Furthermore, if explanation is connected to understanding, ID becomes a theory that is at least in part about human psychology rather than objects with CSI.

The classic non-psychological explication of 'explanation' relies on deductive inference (cf. Hempel 1965a), so that, roughly, given our background assumptions Π, Θ explains Ω better than Λ explains Ω if $\Theta \cup \Pi \models \Omega$ and $\Lambda \cup \Pi \not\models \Omega$. Likelihoods provide another way to explicate what it means to be a better explanation, by stipulating that Θ explains Ω better than Λ if $\Pr(\Omega \mid \Theta \cup \Pi) > \Pr(\Omega \mid \Lambda \cup \Pi)$.[14] Using the abbreviations from ID 1 and further letting, for any x, $\ulcorner \text{CSI}_x \urcorner$ stand for a description of the CSI of x, the Discovery Institute's version of ID is captured by

ID 3. Given our background assumptions Π, for all x with CSI and any undirected process u,

$$\exists y(Dy \wedge Cyx) \cup \Pi \models \text{CSI}_x \text{ and } Cux \wedge A \not\models \text{CSI}_x \tag{9.5}$$

or

$$\Pr(\text{CSI}_x \mid \exists y(Dy \wedge Cyx) \cup \Pi) > \Pr(\text{CSI}_x \mid Cux \cup \Pi)\,. \tag{9.6}$$

Background assumptions are almost always needed to assign probabilities to or infer observation statements, and background assumptions should be honest, which I will assume in the following. Since *explanation* is such a difficult concept, one may treat ID 3 simply as one (central) conjunct of the Discovery Institute's natural language definition of ID. For it is arguably a necessary condition for a better explanation of Ω by Θ than by Λ that Θ either entails Ω, while Λ does not, or Θ's likelihood for Ω is higher than Λ's. Thus *explanation* may consist of more than just entailment or higher likelihood. Giving up both of these components would render *explanation* altogether elusive, given that one cannot assume the psychological connection to *understanding*.[15] If the disjunction of (9.5) and (9.6) is taken as a necessary condition for 'explanation', ID 3 becomes a necessary condition for the truth of the full theory of intelligent design: If intelligent design is true, so is ID 3.

ID 3 states a logical or probabilistic relation between design and CSI relative to our background assumptions. This makes ID 3 a rather odd theory: It is not a

[13] Of course, it is claimed of many theories that they explain some phenomenon or other, but that does not mean that the theory itself uses the concept of explanation.

[14] Like Hempel's explication, this explication of *explanation* is extremely implausible if *explanation* is considered as a psychological phenomenon. But taken as a technical concept, it is arguably at least as good as the deductive one.

[15] I will argue below that ID 3 is analytically false, and thus too strong. Hence the use of the disjunction in the definition of ID 3 is a charitable reading, and so is the focus on just one necessary condition.

statement that *in fact* relates to specific observable phenomena as a matter of logic or probability theory, but it *postulates* that some other statement does so relate. One may argue that ID 3 is just the result of a too literal reading of loose language, and that the original quote mixes a description of ID with a meta-theoretic claim *about* ID, namely that ID explains CSI best. In that case, ID boils down to the claim that all things with CSI are caused by an intelligent designer, that is, ID 2.[16]

In an elaborate discussion of ID, Monton (2009, §2) suggests multiple improvements to the Discovery Institute's definition that also avoid the problematic concept of explanation. He conjectures that ID proponents "wouldn't take much solace in the knowledge that their appeal to an intelligent cause is the best explanation, if we've established that their explanation is a false one". Monton (2009, 38, emphasis removed) therefore suggests a preliminary improvement:

> The theory of intelligent design holds that certain features of the universe and of living things are the result of an intelligent cause, not an undirected process such as natural selection.

Keeping in mind that the certain features are those with CSI and that a feature cannot be the result of both an intelligent cause and an undirected process, this paraphrase amounts to ID 2. Since ID 2 differs radically from ID 3 in its logical structure, the justification for this modification is somewhat wanting.

In "a bit of charitable speculation", Monton (2009, 38) suggests that ID proponents speak of an intelligent designer as the best explanation because one of the "key ideas behind intelligent design is that their theory is scientific, and one can get scientific evidence for the existence of the intelligent designer". Therefore "the actions of the intelligent cause [must not be] completely hidden from us". Monton (2009, 38, emphasis removed) thus suggests another preliminary definition ($*$):

> The theory of intelligent design holds that certain features of the universe and of living things provide evidence for the existence of an intelligent cause, and provide evidence against the doctrine that the features are the result of an undirected process such as natural selection.

The exact nature of ID then depends on how 'providing evidence for' is explicated. Monton (2009, 100) assumes that the evidential relationship is to be explicated by probability theory, that is, by by Bayesian confirmation. Definition 8.1 then leads to

ID 4. Given our background assumptions Π, for all x with CSI and any undirected process u,

$$\Pr\left(\exists y[Dy \land Cyx] \,\middle|\, \mathrm{CSI}_x \cup \Pi\right) > \Pr\left(\exists y[Dy \land Cyx] \,\middle|\, \Pi\right) \tag{9.7}$$

[16] Following the principle of charity, one should take ID to be the most plausible of the possible definitions given in this section. Excluding ID 3 from the discussion would thus not help the case of ID.

and

$$\Pr(Cux \mid \mathrm{CSI}_x \cup \Pi) < \Pr(Cux \mid \Pi)\,. \tag{9.8}$$

Of course, there are other conceptions of confirmation. In likelihoodism, for example, Sober's definition of contrastive confirmation would lead to the claim that observation statement $\ulcorner \mathrm{CSI}_x \urcorner$ contrastively confirms $\ulcorner \exists y(Dy \wedge Cyx) \urcorner$ and disconfirms $\ulcorner Cux \urcorner$ against each other if and only if $\Pr(\mathrm{CSI}_x \mid \exists y(Dy \wedge Cyx \wedge A) > \Pr(\mathrm{CSI}_x \mid Cux \cup \Pi)$ (see §10.1). This explication of confirmation and disconfirmation leads to the second disjunct (9.6) of ID 3.

Monton's second modification (∗) of the Discovery Institute's definition brings his own definition again closer to the original (and to ID 3), since (∗) again claims a probabilistic relation between a designer and CSI. His argument for the second modification may indeed provide the reason for the Discovery Institute's definition of ID, but it is an odd one nonetheless: Rather than trying to develop a theory involving a designer that is scientific, that explains something, or that can be supported by evidence, ID theorists, according to Monton's argument, simply *postulate as part of their theory* that for any x with CSI, $\ulcorner \exists y(Dy \wedge Cyx) \urcorner$ is scientific, explains something, or is supported by evidence. I will discuss the implications of this move in §9.5.1.

Monton (2009, 38) charges that both the Discovery Institute's definition of ID and his preliminary definition (∗) are trivially true, because in a crucial point, Monton's interpretation of the Discovery Institute's definition of ID differs from the one assumed in ID 3 and ID 4. This becomes clear when Monton (2009, 72) discusses the assertions of ID:

> I would say that intelligent design proponents are making a prediction: they are claiming that, if one looks, one will find evidence that there is a designer.

I have read 'certain features' in the Discovery Institute's definition as a placeholder for 'instances of CSI'. Monton instead reads 'certain features' as 'some features', so that ID becomes the claim that there are features that are best explained by an intelligent cause, not by an undirected process. Under this reading, it is indeed indisputably and almost trivially true that there are certain features of the universe that provide evidence for a designer and against the claim that they are the result of an undirected process. Monton (2009, 16, 23) gives the Petronas Towers in Kuala Lumpur and Sarah Watson's muscular arms as examples of such features.

I have argued above for my interpretation, and it also fits better with Hartwig and Meyer's claim about the impossibility of CSI arising from purely physical or chemical antecedents, the central role that CSI and irreducible complexity play in the exposition of ID by Dembski and Wells (2008), and specifically Dembski's claim that "if there is a way to detect design, specified complexity is it" (Dembski 2002, 116). The lengths to which Monton has to go to avoid the almost trivial truth of ID provide another argument against his interpretation. His final definition is this (Monton 2009, 39, emphasis changed):

> The theory of intelligent design holds that certain global features of the universe provide evidence for the existence of an intelligent cause, *or* that certain biologically innate features of living things provide evidence for the doctrine that the features are the result of the intentional actions of an intelligent cause which is not biologically related to the living things, and provide evidence against the doctrine that the features are the result of an undirected process such as natural selection.

Except for the change from a conjunction to the disjunction (here emphasized), all the modifications are attempts at avoiding the almost trivial truth of ID (Monton 2009, 17–26). None of the modifications are necessary under the interpretation defended here, since it does not render ID almost trivially true, and renders the mention of living things in the Discovery Institute's definition redundant.[17]

It is furthermore doubtful that Monton avoids triviality or something close to it. If, for example, the hypothesis of a designer asserts a certain feature F in all organisms, while no undirected process does, the existential quantification allows including all and only those organisms with feature F in the evidence. If the hypothesis of a designer, but no undirected process, asserts a specific ratio of F, the existential quantification allows for picking out a set of organisms that has this very ratio as long as there are enough organisms with and enough organisms without the feature to assemble the set. Therefore, as long as the hypothesis of the designer makes any such assertion, ID is true under Monton's definition.[18]

9.5 The two challenges to Intelligent Design

The previous section has resulted in four versions of ID. ID 1 is Sober's own clarification of his definition, using Russell's paraphrase of the definite article. ID 1 entails ID 2, which itself is a clarification of Sober's definition that relies on Strawson's paraphrase of the definite article. The Discovery Institute's definition can be clarified as ID 3. Finally, while Monton's final definition of ID is flawed, his preliminary definition $(*)$ suggests ID 4 if Bayesianism is used as a criterion of confirmation, and the probabilistic disjunct of ID 3 if likelihoodism is used.

[17] Note that neither ID 3 nor ID 4 exclude the alleged fine-tuning of cosmological constants as evidence for an intelligent designer (which is a focus of Monton's discussion): Dembski (2002, xiii) claims that the fine-tuning is an instance of CSI.

[18] Monton (2009, 35, 109) also seems to existentially quantify over all designers, as he rather freely chooses the designer's intentions (to let an atom decay at one specific moment or to have as much intelligent life as possible, for example). Strictly speaking, Monton is thus not dealing with the question whether there could be any evidence $\ulcorner \mathrm{CSI}_x \urcorner$ for $\ulcorner \exists y(Dy \land Cyx) \urcorner$, but with the question whether there could be a y such that there could be any evidence $\ulcorner \mathrm{CSI}_x \urcorner$ for $\ulcorner Dy \land Cyx \urcorner$.

In this section, I will discuss which of these versions of ID meets which of the two remaining challenges. For this discussion, I will assume that 'CSI' is an observational term. This assumption is charitable because either challenge is easier to meet the more terms in a theory are observational, and because CSI is observational only if it is well-defined, which is disputed, for example, by Elsberry and Shallit (2009, §§4–6). In general, I will make the charitable assumption that only sentences in which 'designer' or 'cause' occur are not observational. With this assumption, I thus avoid the problem of determining precisely the set of observational sentences.

9.5.1 The modified falsifiability challenge

According to Swinburne (2004, 7), 'God' is synonymous with 'A necessarily existing person without a body who necessarily is eternal, perfectly free, omnipotent, omniscient, perfectly good, and the creator of all things'. This synonymy is therefore a justified supplementary sentence, which allows inferring from God's existence that all objects, and hence all objects with CSI, are caused by a person (and hence a designer). Any explanation of a phenomenon by this theory is, according to Swinburne (2004, 47–49), a personal explanation, and as such it relies essentially on the intentions and abilities of the designer, in this case God. ID 2 is a personal explanation as well and therefore entails an observation (for example the description of the CSI of some object x) only in conjunction with two specific supplementary sentences. The first states that any designer who caused an object x with CSI has specific intentions i_x and abilities a_x, that is, with $\ulcorner Iyi_x \urcorner$ standing for $\ulcorner y$ has intention $i_x \urcorner$ and $\ulcorner Bya_x \urcorner$ standing for $\ulcorner y$ has ability $a_x \urcorner$:

$$\text{For all } x \text{ with CSI: } \forall y(Dy \wedge Cyx \rightarrow Iyi_x \wedge Bya_x)\,. \tag{9.9}$$

The second states that a designer who caused x, with intentions i_x and abilities a_x, brings about the observation Ω_x:

$$\text{For all } x \text{ with CSI: } \forall y(Dy \wedge Cyx \wedge Iyi_x \wedge Bya_x \rightarrow \Omega_x)\,. \tag{9.10}$$

If some x has CSI, observation Ω_x then follows because for all x with CSI,

$$\exists y(Dy \wedge Cyx) \wedge \forall y(Dy \wedge Cyx \rightarrow Iyi_x \wedge Bya_x) \wedge \\ \forall y(Dy \wedge Cyx \wedge Iyi_x \wedge Bya_x \rightarrow \Omega_x) \vDash \Omega_x\,. \tag{9.11}$$

For probabilistic assertions, the likelihoods of $\ulcorner \exists y(Dy \wedge Cyx) \urcorner$ can be given through generalizations of the deductive premises. Instead of claiming that if x was designed by y, y has intention i_x and ability a_x, it is enough to assert that if x was designed, the probability that it was designed by some y with intention z and ability w has the probability p_{xyzw}:

$$p_{xyzw} = \Pr\bigl(Dy \wedge Cyx \wedge Iyz \wedge Byw \,|\, \exists y(Dy \wedge Cyx)\bigr)\,. \tag{9.12}$$

The second deductive auxiliary assumption (9.10) can be relaxed to the claim that if a designer y of x has intention z and ability w, then an observation Ω has probability $q_{xyzw\Omega}$:

$$q_{xyzw\Omega} = \Pr(\Omega \,|\, Dy \wedge Cyx \wedge Iyz \wedge Byw)\,. \tag{9.13}$$

The probability for Ω for any x with CSI can then be written as[19]

$$\Pr\big(\Omega \,|\, \exists y(Dy \wedge Cyx)\big) = \sum_{yzw} p_{xyzw}\, q_{xyzw\Omega}\,. \tag{9.14}$$

In the theistic theory, the abilities of the designer are known within the precision of natural language: God is omnipotent. The intentions of God are also known within the precision of the word 'good'.[20] ID 2 differs from the theistic theory in this respect: The designer who caused an object with CSI is only specified as being intelligent. Leaving the problem of the level of intelligence aside, Sober argues in a number of ways that the intentions of the designer are unknown. Among them is his point that the designer could be an extraterrestrial, which would give no indication about the intentions. But 'designer' is even less specific than 'extraterrestrial designer', and thus would give even fewer indications of the intentions (Sober 1999, 65–66). And sometimes, the problem of evil is solved by claiming that the intentions of God are unknowable. This answer entails that the intentions of a less specific designer are also unknowable (Sober 2008, §2.17). Sober (1999, 74, n. 25; 2007, 6; 2008, 128, 154, n. 29) also gives lists of proponents of ID who state that the intentions of the designer cannot be known.

In general, for any designer with the intentions and the abilities to do something, one can without inconsistency think of another designer without those intentions or abilities. Thus the first supplementary sentence (9.9) of the deductive assertion of Ω_x (9.11) is unjustified for any intention and ability, with (arguably) one exception. If one observes Ω_x for some x, then one can plausibly infer that if there is a designer of x, the designer had the intention to bring about Ω_x. With the further plausible assumption that having the intention and ability to bring about Ω_x, a designer would bring about Ω_x, one can infer Ω_x. But then, for the supplementary sentence to be honest, they have to contain assumption Ω_x if they contain the first supplementary sentence. Therefore, ID 2 in conjunction with honest supplementary sentences entails Ω_x only if the auxiliary assumptions alone already entail Ω_x, and thus ID 2 makes no deductive observational assertions.

ID 2 makes a probabilistic assertion, that is, assigns a probability to some observation Ω, only if the sum (9.14) is defined. The sum in turn is defined only if

[19] The condition in equation (9.13) for term $q_{xyzw\Omega}$ does not need to contain the conjunct $\ulcorner \exists y(Dy \wedge Cyx) \urcorner$, as this is entailed by $\ulcorner Dy \wedge Cyx \urcorner$ for any y.

[20] This assumes that 'good' has its usual meaning in theistic statements. The contrary assumption led to the original falsifiability challenge (see §9.1).

the conditional probabilities (9.12) and (9.13) are defined, again because ID 2 relies on personal explanations. But analogously to the deductive case, for any designer who would have some intention and ability with one probability, one can, without flouting the laws of probability, think of another designer who has that intention and ability with any other probability. Thus the first auxiliary assumption (9.12) is undefined for almost all intentions and abilities. Similar to the deductive case, the exceptions are those cases in which observation Ω can be assigned a probability s independently of assuming a designer. For some Ω and x, one can then maybe justify the assumption that if a designer caused x, the designer must have had the intention and ability to bring about Ω with probability s. (It is here not obvious what values the probabilities in the two auxiliary assumption have to have, only that the sum of their products has to be s for Ω.) But if $\Pr(\Omega) = s$ is used to justify the auxiliary assumptions, it has to be in the auxiliary assumptions as well if they are honest. Thus, ID 2 only assigns probability s to Ω if the auxiliary assumptions do, too. Therefore ID 2 makes no probabilistic observational assertions either, and thus succumbs to the modified falsifiability challenge.

As already noted, the problem with ID 3 and ID 4 is that, rather than make observational assertions, they postulate deductive or probabilistic relations between the existence of an intelligent designer and observations. Given the results so far, one can see that the postulates are, in fact, false: ID 3 is true only if, given our background assumptions Π, equation (9.5) or equation (9.6) is true for every x with CSI. Equation (9.5) states that for each x, $\ulcorner \exists y(Dy \wedge Cyx) \cup \Pi \urcorner$ entails a description of the CSI of x, while $\ulcorner Ux \cup \Pi \urcorner$ does not. But ID 2 makes no observational assertions for any honest auxiliary assumptions, and for each x with CSI, $\ulcorner \exists y(Dy \wedge Cyx) \urcorner$ is a logically weaker claim than ID 2. Since CSI_x is assumed to be an observation statement, and our background assumptions are specific honest auxiliary assumptions, $\ulcorner \exists y(Dy \wedge Cyx) \urcorner$ therefore asserts no description of the CSI of x. Hence equation (9.5) is false for all x, and ID 3 is false unless equation (9.6) is true for all x. This holds only if for all x with CSI, the probability for the CSI of x given $\ulcorner \exists y(Dy \wedge Cyx) \urcorner$ is higher than its probability given any undirected process u, and thus specifically ET. Sober (2008) has argued extensively that the complex adaptations we observe can be explained by ET, and thus are certainly not *less* probable under the assumption of ET than they are given our background assumptions. Therefore equation (9.6) is true only if the probability for the CSI of x given ID 2 is higher than given our background assumptions. But again, since ID 2 and thus $\ulcorner \exists y(Dy \wedge Cyx) \urcorner$ make no observational assertions, this is false. ID 3 is therefore analytically false, because it postulates an inferential relation that does not exist. And if ID 3 is only taken to be a necessary condition for the truth of the full theory of intelligent design, that theory is false as well by *modus tollens*.

ID 4 is true only if all undirected processes, and thus also ET, are disconfirmed and $\ulcorner \exists y(Dy \wedge Cyx) \urcorner$ is confirmed by each x with CSI given our background assumptions. Sober (2007, 2008) has documented at length that ET has not been disconfirmed by each such x and thus ID 4 is false independently of the status of $\ulcorner \exists y(Dy \wedge Cyx) \urcorner$ because equation (9.8) is false. But ID 4 would be wrong even if ET were disconfirmed by each x with CSI: Purportedly, there are objects for which the designer cannot be directly observed and which show CSI. For those objects, $\ulcorner \exists y(Dy \wedge Cyx) \urcorner$ has to be confirmed by observations other than those of the designer. This holds if and only if equation (9.7) holds, which is equivalent to

$$\Pr\big(\mathrm{CSI}_x \,|\, \exists y(Dy \wedge Cyx) \cup \Pi\big) > \Pr(\mathrm{CSI}_x \,|\, \Pi)\,. \tag{9.15}$$

But this inequality cannot be true for the same reason that equation (9.6) cannot be true, namely because ID 2 makes no observational assertions. This leaves ID 1, Sober's definition, which meets the modified falsifiability challenge.

9.5.2 The modified translatability challenge

ID 1 clearly makes the deductive observational assertion that there are objects with CSI. But ID 1 is equivalent to the conjunction of this existence claim and ID 2, and I have argued in §9.5.1 that also in conjunction with honest auxiliary assumptions, ID 2 makes no observational assertions. Under the assumption that 'CSI' is a well-defined observational concept, it is a justified and hence honest auxiliary assumption that there are objects with CSI.[21] Therefore ID 1 entails only observation statements that are also entailed by this existence claim, and assigns the same probabilities as this existence claim.

ID 1 fails to meet the modified translatability challenge against ET if ET makes all the observational assertions of ID 1, but ID 1 does not make all the observational assertions of ET. Sober (2008, §2.9) argues that if the existence of objects with CSI (or, in his discussion, of complex adaptations of organisms) is to be explicitly included in ID, it should also be included in ET. This is plausible, as the definite article should be interpreted in the same way, no matter whether the statement is 'The objects with CSI were caused by an intelligent designer' or 'The objects with CSI are the result of natural selection'. This is also in keeping with the description of ID by the Discovery Institute that led to ID 3, because it refers to the intelligent cause and undirected processes in the same way.

With both ID 1 and ET entailing the existence of objects with CSI (call this ADAP again), ET makes all the deductive observational assertions of ID 1. Unless there is an observation Ω such that $\Pr(\Omega|\{\mathrm{ADAP}\} \cup \Pi) \neq \Pr(\Omega|\Pi)$ *and* $\Pr(\Omega|\{\mathrm{ADAP}\} \cup \Pi) \neq \Pr(\Omega|\{\mathrm{ADAP}\} \cup \mathrm{ET} \cup \Pi)$, that is, ADAP alone makes a probabilistic observational assertion different from ET, ET also makes all the probabilistic observational assertions of ID 1. Then, since ET makes observational assertions besides ADAP, ID 1 does not meet the modified translatability challenge. But even if ADAP alone makes a probabilistic observational assertion different from ET, ID 1 fails to meet the modified falsifiability challenge against many other

[21] Again because one can observe many objects that, intuitively, have CSI: Miami, eyes, etc.

statements: ADAP itself makes all the observational assertions of ID 1, and, for example, the claim ADAP+ that there are organisms with eyes makes more.

9.6 Intelligent Design as a science

If every version of ID fails one of the modified challenges or is analytically false, what does this mean for the status of ID beyond the obvious that ID is false, makes no observational assertions, or makes fewer observational assertions than ET? Of specific contemporary interest is the question whether ID can still be science.[22] While some consider criteria for the demarcation of science from non-science an important topic of research (Hansson 2008a), others doubt that such a demarcation criterion is possible at all. A number of proponents of the latter view, for example Monton (2009, 49) and Leiter (2010, 6–12), endorse and rely on an influential article by Laudan (1983a). I will argue that the article relies on a non-sequitur and actually suggests the opposite of its purported conclusion.

Laudan places five demands on a criterion for the demarcation of science. First, it would be "a grave drawback for any demarcation criterion" if it did not respect the paradigmatic cases of science and non-science (Laudan 1983a, 117–118). Second, a philosophically significant demarcation criterion should "identify the *epistemic* or *methodological* features which mark off scientific beliefs from unscientific ones" (Laudan 1983a, 118). Third, it should have "sufficient precision that we can tell whether various activities and beliefs whose status we are investigating do or do not satisfy it" (Laudan 1983a, 118). Fourth, it should specify "a set of individually necessary and jointly sufficient conditions" for something to be scientific (Laudan 1983a, 118). Finally, because of the far-reaching practical implications of any demarcation criterion, it should be supported by arguments that are "especially compelling" (Laudan 1983a, 120).

Laudan (1983a, 120–124) argues that none of the demarcation criteria suggested in the 20$^{\text{th}}$ century is a"necessary and sufficient condition for something to count as 'science', at least not as that term is customarily used", that is, the criteria fail to fulfill either his fourth or first demand. [23] The same holds, he argues, for criteria suggested before the 20$^{\text{th}}$ century (while those based on the notion of scientific methodology also fail to fulfill his third demand) (Laudan 1983a, §2). This is the main result of Laudan's paper. Because of the "epistemic heterogeneity of the activities and beliefs customarily regarded as scientific", Laudan (1983a, 124) further suggests that his second demand can never be satisfied.

[22] This section has been published under the title "On an allegedly essential feature of criteria for the demarcation of science" (Lutz 2011d). I thank J. Brian Pitts, Antje Rumberg, and three anonymous referees for helpful comments.

[23] Shortly after publication of his article, Laudan (1986, 120) would criticize Carnap's notion of explication for requiring "in effect, that the methodological theories of the philosopher must, to a very large degree, capture the pre-analytic uses of methodological terminology", thus effectively dropping his first demand (see §3.8.1, n. 60).

In objection to Laudan, a number of proponents of demarcation criteria have contested his fourth demand without, however, addressing the argument that Laudan (1983a, 118–119) provides in support of it (e. g., Thagard 1988, 159; Derksen 1993, 20; Pennock 2011, 183). According to Laudan's fourth demand, a demarcation criterion must provide a set of conditions $\{\chi_i\}_{i\in I}$ for any x to be a science (Sx) that are jointly sufficient, $\forall x\left[\bigwedge_{i\in I}\chi_i(x)\rightarrow Sx\right]$, and individually necessary, $\bigwedge_{i\in I}\forall x[Sx\rightarrow\chi_i(x)]\models\!\!\!\!\dashv\forall x\left[Sx\rightarrow\bigwedge_{i\in I}\chi_i(x)\right]$. This means that there is one necessary and sufficient condition, $\bigwedge_{i\in I}\chi_i(x)$. According to Laudan (1983a, 118), this demand is justified because "it seems unlikely" that "something less ambitious would do the job" of a demarcation criterion: With a criterion that provided only a necessary condition ψ, $\forall x[Sx\rightarrow\psi(x)]$, one could never determine that something *is* a science, and with a criterion that provided only a sufficient condition φ, $\forall x[\varphi(x)\rightarrow Sx]$, one could never determine that something *is not* a science (Laudan 1983a, 118–119). In other words (Laudan 1983a, 119):

> Without conditions which are both necessary and sufficient, we are never in a position to say '*this* is scientific: but *that* is unscientific'.

But Laudan's claim is false: To be able to say that a is scientific (Sa) while b is not ($\neg Sb$), all that is needed is one sufficient condition φ that is fulfilled by a, $\varphi(a)$, and one necessary condition ψ that is not fulfilled by b, $\neg\psi(b)$. That is, the criterion can provide a condition of the form $\sigma\models\!\!\!\!\dashv\forall x[\varphi(x)\rightarrow Sx]\wedge\forall x[Sx\rightarrow\psi(x)]$, where, for σ to be conservative, $\forall x[\varphi(x)\rightarrow\psi(x)]$ must be a logical truth. However, Laudan's demand that the converse must also hold, so that $\forall x[\varphi(x)\leftrightarrow\psi(x)]$ (and $\forall x\left[\psi(x)\leftrightarrow\bigwedge_{i\in I}\chi_i(x)\right]$), is supererogatory. Hedging his claim with "it seems unlikely", Laudan suggests that any condition logically weaker than a necessary and sufficient one is either only necessary or only sufficient. But, if χ is conservative without fulfilling Laudan's supererogatory demand, then χ is weaker than any necessary and sufficient condition and still "does the job" of determining a but not b to be a science.

Without the demand for a single necessary and sufficient condition for scientific theories, Laudan's argument actually suggests the opposite of what he intends to show. He states that, given his first demand, the criteria suggested in the 20th century are implausible as necessary and sufficient conditions, and that "in *most* cases, these are not even plausible as necessary conditions" (Laudan 1983a, 123, my emphasis). But this suggests that *some* criteria do provide plausible necessary conditions $\{\psi_j\}_{j\in J}$ that fulfill his first demand. Assuming that the criteria identify important epistemic or methodological features, the resulting necessary condition $\lambda x\bigwedge_{j\in J}\psi_j(x)$ also fulfills Laudan's second demand. Furthermore, Laudan (1983a, 117–118) states in connection with his first demand that "there is a large measure of agreement at this paradigmatic level" and in fact gives examples of paradigmatic sciences. Thus there is a sufficient condition for scientific theories after all, if only by enumeration of the paradigmatic sciences. If now these paradigmatic sciences have any epistemic or methodologically relevant feature ϑ in common, its conjunction with the necessary condition, $\lambda x \left[\vartheta(x) \wedge \bigwedge_{j\in J} \psi_j(x)\right]$, fulfills Laudan's first and second demand and logically entails the necessary condition.

Whether the resulting criterion with one necessary condition and a different sufficient condition fulfills Laudan's third demand will depend on what is being investigated, but without the fourth demand, a rather weak criterion may already determine some cases and provide a starting point for stronger criteria. For example, the two modified challenges can plausibly be construed as necessary conditions. For a theory should probably not be considered empirical if it fails the modified falsifiability challenge, and even the statement 'There are complex organisms' may not be enough for ID to qualify as a proper theory. And the *pursuit* of ID may not be considered a scientific one if ID fails the modified translatability challenge. Of course, both challenges also have to meet Laudan's fifth demand. But the demand cuts both ways: Given the important practical implications, one should neither adopt nor dismiss a criterion for the demarcation of science based on arguments that are not especially compelling.

9.7 Conclusion

Probably the most important result of the elaboration of ID in §9.4 is that ID is woefully underspecified, leading to different explications with vastly different properties.[24] ID 2 makes no observational assertions and thus fails to meet the modified falsifiability challenge, ID 1 makes fewer observational assertions than ET and in fact the simple claim that there are organisms with eyes, and thus fails to meet the modified translatability challenge. ID 3 and ID 4 are wrong, not because of incorrect observational assertions, but rather because they make incorrect logical or probabilistic claims.

A rather stark example from ID proponents is the difference between two textbook expositions of ID: According to Davis and Kenyon (1993, 99–100) ID "means that various forms of life began abruptly through an intelligent agency, with their distinctive features already intact—fish with fins and scales, birds with feathers, beaks, and wings etc." It seems like more than a mere adjustment of details when in a recent textbook, Dembski and Wells (2008, 109) claim that ID "neither requires nor excludes speciation—even speciation by Darwinian mechanisms. ID is sometimes confused with a static view of species, as though species were designed to be immutable. [...] ID precludes neither significant variation within species nor the evolution of new species from earlier forms". Thus even accepting modifications of theories over time, ID is not plausibly construed as a single theory like quantum mechanics or even evolution. It rather has to be considered a class of

[24]This problem is exacerbated by the fact that ID is also often formulated with so little precision that even charitable interpreters (which I take Monton and me to be) can arrive at fundamentally incompatible conclusions about its content.

related theories with radically different logical structures. The theories may indeed only be related by their common abstraction that there is at least one designer.

As to the abrupt beginning of life, biologists have found overwhelming evidence for common descent, and thus even with the addition of Davis and Kenyon's version to the superabundance of varieties of ID, the conclusion stays the same: ID fails, because it is disconfirmed by the evidence, because it makes false inferential claims, because it fails the modified falsifiability challenge, or because it fails the modified translatability challenge. In all these cases, it is furthermore plausible that either the pursuit of ID given our state of knowledge or ID itself is not scientific.

Chapter 10

Confirmation

Having applied empirical significance directly in the last chapter, I now want to spell out some of the implications for related areas, the most obvious of which may be the concept of confirmation.[1] Contrastive testability, making probabilistic basic assertions, and the Bayesian criterion of empirical significance have all been proposed in the context of theories of confirmation. In this chapter, I will put the cart before the horse and try to arrive at criteria of confirmation starting from probabilistic criteria of empirical significance.

As a first step, I will discuss a criticism of Sober's criterion of testability that will turn out to be rather a criticism of Sober's notion of contrastive confirmation (§10.1). And while this criticism fails, I will give a different criticism of contrastive confirmation, showing that the problem for contrastive testability also applies to contrastive confirmation (§10.2). The two probabilistic criteria of empirical significance found in §8, on the other hand, will lead to viable candidates of confirmation (§10.3). Neither of these candidates, however, provide a complete account of theory acceptance, as I will argue in §10.4.

10.1 Boudry and Leuridan on supplementary sentences

Boudry and Leuridan (2011, §3.6) agree with Sober that testing is contrastive, but they criticize his conception of suitability as too restrictive.[2] Specifically, they argue that restriction IVb of definition 8.3, the demand that supplementary

[1]Parts of this chapter have been presented at the conference of the Eastern Division of the American Philosophical Association on December 30, 2011. I thank my commentator Wesley van Camp and the audience for helpful discussions.

[2]I thank Maarten Boudry and Bert Leuridan for their comments on this section and helpful discussions. Formal notations in this section are changed to fit mine.

sentences be independent of the observation that contrastively tests the theories, does not allow for certain intuitively valid inferences. With respect to ID, they argue that some of these intuitively valid inferences would render ID testable, but disconfirmed.

Remember that Sober (2008, 145) argues that not every set of supplementary sentences is suitable by giving the example of the careless juror who just makes up supplementary sentences to conclude that Smith is a murderer (page 310). I have noted in §8.3.1 that in this example, the question is one about *confirmation*, not testability. The tacit switch relies, first, on the definition of 'testing' via confirmation and disconfirmation: Ω tests Θ relative to suitable supplementary sentences if and only if Ω confirms Θ or disconfirms Θ relative to suitable supplementary sentences. Thus if a stipulated supplementary sentence is not suitable for confirming a hypothesis (such as that Smith is the murderer) or disconfirming it by confirming its competitor, it is also not suitable for testing it. Second, Sober (1999, 48) relies on his postulated relation between testing and testability (see §8.1.2):

> If a set of observations provides a test of a proposition because it bears relation R to that proposition, then a proposition is testable when it is possible for there to be a set of observations that bears relation R to the proposition. Testing is to testability as dissolving is to solubility.

To be precise, he implicitly uses the *converse* of this relation for the tacit switch, so that a stipulated set of supplementary sentences is not suitable for the testability of two hypotheses because it is not suitable for the testing of two hypotheses. Sober thus assumes that Θ is testable if and only there is some Ω that tests Θ relative to suitable supplementary sentences.

Combining the two tacit assumptions, Θ can thus be tested against Λ if and only if there is a possible Ω such that Ω confirms Θ against Λ or Ω disconfirms Θ against Λ relative to suitable supplementary sentences. Sober's definition of testability (page 303) can be summarized as

> Hypothesis Θ can now be tested against hypothesis Λ if and only if there exist true [and suitable] auxiliary assumptions Π and an observation statement Ω such that [...] $\Pr(\Omega \mid \Theta \cup \Pi) \neq \Pr(\Omega \mid \Lambda \cup \Pi)$,

where 'suitable' has to be interpreted by Sober's restrictions III and IV on the supplementary sentences. Through the relation between testability and confirmation, one thus arrives at

Definition 10.1. Ω contrastively confirms hypothesis Θ against hypothesis Λ if and only if there exist true and suitable supplementary sentences Π such that $\Pr(\Omega \mid \Theta \cup \Pi) > \Pr(\Omega \mid \Lambda \cup \Pi)$.

This criterion of confirmation is equivalent to Sober's definition of Ω favoring Θ over Λ (Sober 2008, 32), except that it takes supplementary sentences into account. Sober himself in fact starts out with his definition of confirmation and introduces his definition of testability later. Boudry and Leuridan (2011) tacitly rely on definition 10.1 when they criticize Sober's conception of suitability by arguing that, in some cases, ID is intuitively confirmed but not contrastively confirmed according to definition 10.1. Like Sober and Boudry and Leuridan (2011), I will therefore assume this criterion in this section.[3] The problem with contrastive testability, Boudry and Leuridan (2011) claim, lies with Sober's definition of suitability. To liberalize the conception, they suggest an additional sufficient condition for suitability based on simplicity and unification.

In this section, I will argue that Boudry and Leuridan (2011) misconstrue both Sober's criterion of testability and the version of ID that Sober claims cannot be tested against ET.[4] As a result, the inferences they claim to be intuitively valid either are not, or can be accommodated by Sober's criterion. And while Boudry and Leuridan (2011) do not spell out their sufficient condition for suitability explicitly, in its most plausible interpretations, it trivializes contrastive testability and contrastive confirmation.[5]

10.1.1 Contrastive testability and valid inferences

Boudry and Leuridan (2011) argue against Sober's claim that ID "cannot be tested against evolutionary theory, at least at present" (Sober 1999, 66–67; cf. 2008, 148). More generally, they argue that many examples of intuitive confirmation are not examples of contrastive confirmation. I will discuss each of their four examples in turn.

Excursion or Murder?

Boudry and Leuridan (2011, 562) suggest to change Sober's example of the murder case so that the evidence additionally includes blood stains and broken glass in the bedroom but no body, the supposed victim (a landlord) is a non-smoker who neither owns a gun nor has size twelve feet, and the investigating detective has to decide between

Θ The landlord was murdered.

Λ The landlord is alive and left for an unexpected walk.

[3] I will argue in §10.4 that the criterion and the two tacit assumptions that led to it are incorrect: For confirmation, one must take the maximal set of suitable supplementary sentences into account. Since the following discussion will be about whether specific supplementary sentences *can*, not *should* be included, this will not make a difference.

[4] This is not too surprising, since in contradiction to their conclusion, Sober's definition of suitability is too wide (see §8.3.2).

[5] Boudry and Leuridan's condition for suitability *alone* already trivializes contrastive confirmation. Sober's own trivializing definition is not needed.

Ξ The landlord killed himself and was then dragged away.

Boudry and Leuridan (2011, 562) claim:

> If the detective favors the murder hypothesis, we submit that she is justified in making the additional assumption that the hypothesized murderer, whoever it was, wears a size 12 shoe, smokes cigars, and used a Colt .45 [Π_1]. This would be a matter of sound detective work, not of accusing Smith or Jones without basis.

This is probably correct, but also irrelevant because the question was not whether the detective can infer the anatomy, smoking habit, and possessions of a hypothetical murderer, but whether Θ, Λ, or Ξ is confirmed by the evidence. Sober's definition of suitability does not lead to a problem in this case. For one, given that even unexpected walks tend not to involve gun shells, shattered glass, and profuse bleeding, all of which are more probable in case of a murder, Θ is contrastively confirmed against Λ by the evidence. Thus Sober's notion of suitability is not too exclusive for Θ to be contrastively confirmed against Λ. It is also fairly plausible that Θ is contrastively confirmed against Ξ because it is arguably more probable that someone breaks a window when illegally entering a bedroom than when dragging a body. In any case, it is not obvious that the evidence intuitively confirms Θ against Ξ either, so there is no counterexample.

Of course, Boudry and Leuridan are not interested in the confirmation of Θ, but of Π_1.[6] They give what I take to be meant as a rephrasing of Θ:

Ω A size 12 shoe print, cigar ash, and shells from a Colt .45 revolver were found in the bedroom.

Θ The landlord was murdered by X.

Π_1 X wears a size 12 shoe, smokes cigars, and owns a Colt .45.

Boudry and Leuridan (2011) note that any plausible set of supplementary sentences Π_2 that we may justify independently of the evidence at hand does "not warrant our adopting Π_1. Only the conjunction of Π_2 with Ω and Θ does". But again, the question should not be whether Π_1 is confirmed, but whether Θ is confirmed. Boudry and Leuridan have switched the roles of the hypothesis and the supplementary sentences: In their informal description, Boudry and Leuridan (2011, 562) state that the detective favors Θ, that is, counterfactually assumes Θ to be justified, and can thus infer the anatomy, smoking habit, and possessions of the hypothetical murderer. Thus Θ is the set of supplementary sentences and Π_1 is the theory to be tested. If the competing hypothesis is, for example, 'X does not wear a size 12 shoe, smokes cigars, or owns a Colt .45', one can see that

[6] They assume that the detective "tentatively favors" Θ and want to determine whether Π_1 can be justified (Boudry, personal email from June 18, 2011).

the evidence contrastively confirms Π_1 when we help ourselves to some supplementary sentences that are justified independently of the evidence at hand (e. g. the general frequency of other people's cigar ash, shells, and footprints found in private bedrooms).

Motivated expert or lazy dilettante?

Moving on to ID, Boudry and Leuridan (2011, 564) consider a situation in which "William Paley, reflecting on the origin of the human eye, constructed the following design hypothesis, conjoined with two additional assumptions":

Θ The human camera eye was created by an intelligent designer.

Π_1 The designer is interested in creating camera eyes.

Π_2 The designer is capable of designing something as complex as the camera eye.

Boudry and Leuridan (2011, 565) claim that the "adoption of both Π_1 and Π_2 seems reasonable enough since their negation is completely uninteresting, in the sense of being very unlikely to yield [Ω]", which asserts the existence of camera eyes.[7] Specifically, they state that

> the likelihood of both Θ and $\neg\bigwedge\Pi_1$ and Θ and $\neg\bigwedge\Pi_2$, namely, $\Pr(\Omega\,|\,\Theta\cup\{\neg\bigwedge\Pi_1\})$ and $\Pr(\Omega\,|\,\Theta\cup\{\neg\bigwedge\Pi_2\})$ is extremely low. If we follow Sober's approach, however, this gives us no reason for adopting Π_1 and Π_2, because, in the absence of background knowledge about the designer, the independence rule is violated.

This argument for the intuitive confirmation is somewhat questionable, since $\Pr(\Omega\,|\,\Theta\cup\Pi_1\cup\Pi_2)$ can be low for some Ω, Θ, Π_1, and Π_2 even if $\Pr(\Omega\,|\,\Theta\cup\{\neg\bigwedge\Pi_1\})$ and $\Pr(\Omega\,|\,\Theta\cup\{\neg\bigwedge\Pi_2\})$ are low as well. More importantly, it misconstrues the situation: Paley has developed the hypothesis $\Lambda \mathrel{\text{⊨⫤}} \Theta\cup\Pi_1\cup\Pi_2$, and Boudry and Leuridan claim that, because Ω is extremely improbable under the assumption of $\Theta\cup\{\neg\bigwedge\Pi_1\}$ and $\Theta\cup\{\neg\bigwedge\Pi_2\}$, Λ is intuitively confirmed. Now, since by construction $\Pr(\Omega\,|\,\Lambda)$ is high, Ω also contrastively confirms Λ against $\Xi \mathrel{\text{⊨⫤}} \Theta\cup\{\neg\bigwedge\Pi_1\}\cup\{\neg\bigwedge\Pi_2\}$, because, presumably, $\Pr(\Omega\,|\,\Xi)$ is very low.[8]

[7] Strictly speaking, due to the use of the descriptive phrase, Θ either entails the existence of the human eye (in Russell's paraphrase), or presumes its existence (in Strawson's paraphrase). Θ should rather be 'All camera eyes were created by an intelligent designer' or similar. Similarly, Π_1 and Π_2 should be preceded by "If there is exactly one designer, ...".

[8] In a footnote, Boudry and Leuridan (2011, §3.3, n. 4) further argue against restriction IVb of definition 8.3 on the basis that "there do not seem to be many ways of justifying the introduction of an auxiliary *except* by taking the observations into account which we set out to explain. [...] Take for example: $\Pi_1^* =$ 'Naive set theory suffers from Russell's paradox.' [T]here is no use incorporating it as an auxiliary, because it has no bearing on our observations in any way." This confuses justifying a statement with justifying its inclusion in the set of supplementary sentences that one *considers*. Sober does not restrict the latter.

This response does not show that Sober's claim that ID cannot be tested against ET is mistaken, however, for Sober (2007, 3; cf. 1999, 62; 2008, 132) expressly considers a minimal version of ID:

> The single thesis of what I will call mini-ID is that the complex adaptations that organisms display (e. g., the vertebrate eye) were crafted by an intelligent designer.

Thus Sober considers neither Λ nor Ξ, and in fact repeatedly notes that he assumes that the designer's intentions are unknown because the designer itself is not specified in any way (see §9.3.1).

This also defuses another criticism: Boudry and Leuridan (2011, §3.4, n. 7) claim that in Sober's approach, it is not clear how to "separate the central hypothesis from auxiliary assumptions". Since Sober is considering the contrastive testability of mini-ID, every other assumption must be a supplementary sentence. Otherwise, definition 8.3 (which is not restricted to specifically "central" hypotheses) has been misapplied. Design hypotheses different from mini-ID can of course be contrastively testable against ET (cf. Sober 1999, 61).

This points to a core misunderstanding of Sober's criterion. Boudry and Leuridan (2011, §3.3) take the role of the supplementary sentences to consist in "fleshing out a hypothesis", that is, they consider supplementary sentences to somehow become conjuncts of the hypothesis. They rest this interpretation on Sober's claim that supplementary sentences are "used to bring the hypotheses [...] into contact with the observation" Ω (Sober 2008, 145). But this is mistaken, for Sober (1999, 54; 2007, 5–6; 2008, 144) introduces supplementary sentences to address Duhem's point that hypotheses (whether "fleshed out" or not) rarely make observational assertions on their own, and rely on other hypotheses and individual facts to get "into contact" with observations. As I have noted in §6.8.2, the supplementary sentences can originate from completely different theories, and it would, for example, be problematic to consider the use of sociological theories about color perception to be a "fleshing out" of electrodynamics.

A designer with little choice

Boudry and Leuridan (2011, 567) consider the possibility to "view the goals and abilities of the designer as the adjustable parameters of the model" and imagine a situation in which (Boudry and Leuridan 2011, 569)

> only a few 'parameters' in the design hypothesis [...] provide an elegant explanation for phenomena that resist any conceivable naturalistic explanation [...]. The fact that the choice of auxiliaries about the designer's intentions and attributes ($\pi_1, \ldots, \pi_m$) would depend on the observations we set out to explain ($\beta_1, \ldots, \beta_n$), without the support of independent background knowledge, would then be of little concern.

m is here assumed to be much less than n.[9] But then, assuming that each observation determines one parameter, it is enough to make m observations to determine all parameters, at which point the remaining observations follow from the model. Thus, for example, $\Pr(\{\beta_{m+1},\dots,\beta_n\} \mid \text{ID} \cup \{\beta_1,\dots,\beta_m\}) > \Pr(\{\beta_{m+1},\dots,\beta_n\} \mid \text{ET} \cup \{\beta_1,\dots,\beta_m\})$, because, I assume, ET would be a "naturalistic explanation". Since furthermore $\{\beta_1,\dots,\beta_m\}$ has not been justified by assuming $\{\beta_{m+1},\dots,\beta_n\}$, $\neg(\beta_{m+1} \wedge \dots \wedge \beta_n)$, ET, or ID, it holds that $\{\beta_{m+1},\dots,\beta_n\}$ contrastively confirms ID against ET.[10]

Again it has to be kept in mind that the version of ID that Boudry and Leuridan are considering in the example is not mini-ID. In fact, Sober (2008, §2.19) states explicitly that, since the designer is not specified, mini-ID has enough free parameters (in form of the possible intentions and abilities implicit in the concept of a designer) to accommodate any sequence of observations, that is, $m \geq n$.

A designer who writes on animals

Boudry and Leuridan (2011, 569–570) consider a scenario in which verses of the Hebrew Bible are observed on beetles, and consider the following hypothesis and supplementary sentences:

- Θ Beetles are created by an intelligent designer.
- Π_1 The intelligent designer has the ability to create beetles, is inordinately fond of them, and he has used their bodies to inscribe his Word.

Boudry and Leuridan state that it is doubtful that there is a naturalistic explanation of the observations. However, even if all animals displayed bible verses, there were no maladaptations, and no evidence for ET whatsoever, an "adherent of Sober's approach [...] would remain unmoved [...], because the adoption of auxiliary Π_1 (the properties of the Judeo-Christian God) still depends upon looking at $\beta_1,\dots,\beta_n$ (without independent background knowledge)" (Boudry and Leuridan 2011, 570). Thus, while intuitively the $\mathscr{B}$-sentences clearly confirm ID, they do not confirm ID contrastively.

It is clear that the observation that beetles (or all animals) have bible verses inscribed on them ($\beta_1,\dots,\beta_n$) contrastively confirms the hypothesis $\Lambda \mathrel{\vDash\!\!\dashv} \Theta \cup \Pi_2$ against ET, because even without supplementary sentences, the likelihood of Λ for $\{\beta_1,\dots,\beta_n\}$ is higher than the likelihood of ET. Thus Λ (which is again different from mini-ID) is contrastively confirmed and provides no counterargument to restriction IV.

Boudry and Leuridan (2011, 572) modify their example and consider the scenario in which each bible verse is written on exactly one species, and some bible

[9] Boudry (personal email from June 18, 2011).

[10] It is here important to remember that, unlike logical independence, the independence of justification is not symmetric.

verses are missing from the animal kingdom as catalogued so far. They state that in this case, Λ could predict the existence of the remaining bible verses on some species. Boudry and Leuridan use the example to consider Sober's restriction IV of suitability anew:

> In the (novel) prediction case, the observation Ω that we use to test our competing hypotheses cannot enter into our considerations for choosing auxiliaries $\pi_1, \ldots, \pi_n$, because, by definition, Ω has not been observed yet. In what sense is the "independence" of $\pi_1, \ldots, \pi_n$ to be understood? Is it acceptable if our justification of $\pi_1, \ldots, \pi_n$ depends on other observations that are already known? If so, why does Sober not leave room for such cases of predictive success in setting up his intrinsic argument against design?

The answers are fairly straightforward: First, since restriction IV does not distinguish between observations in the past and the future, 'independence' is to be understood as in the previous cases. Therefore, second, the supplementary sentences may depend on observations already known. The third question can be answered by combining the previous reply and the reply to Boudry and Leuridan's example of ID as a model with intentions as parameters. While some intentions $\pi_1, \ldots, \pi_n$ can be determined by previous observations $\beta_1, \ldots, \beta_n$, the hypothesis that Sober considers, (i. e., mini-ID) contains so many parameters (one intention for each observation) that Ω still cannot be assigned a probability. Thus there is no predictive success.

10.1.2 Boudry and Leuridan's supplementary sentences

Since they consider Sober's definition of suitability to be too restrictive, Boudry and Leuridan (2011, 559) suggest "an alternative and more lenient account of auxiliary assumptions, based on the explanatory virtue of unification". Unfortunately, they leave the notions of unification and simplicity (which also features in their account) and their exact role in a sufficient condition for suitability[11] on an intuitive level. A very strict sufficient condition that plausibly still captures Boudry and Leuridan's intention is given by

Definition 10.2. The set Π of supplementary sentences is suitable for Θ, Λ, and Ω if Π are justified and $\Theta \cup \Pi$ as well as $\Lambda \cup \Pi$ are simple and unifying.

One could further liberalize this condition by (i) dropping the demand that Π be justified, or (ii) changing the 'as well as' into an 'or'. However, even the most restrictive of these sufficient conditions trivializes contrastive testability. To show that any two hypotheses Θ and Λ can be tested against each other, let π^*

[11] Boudry and Leuridan allow other sufficient conditions, e. g. Sober's, as well (Boudry, personal email from Mai 21, 2011).

be any justified sentence (for example some correct hypothesis) that is simple and unifies some true basic sentences Ω, assigning them a high probability. Assume the analogous for $\pi^\dagger$ and Ω'. Let further Θ and Λ be independent from the basic sentences, π^*, and $\pi^\dagger$. Then $\Pi \boxminus \{\neg \bigwedge \Theta \vee \pi^*, \neg \bigwedge \Lambda \vee \pi^\dagger\}$ is justified because it is entailed by $\pi^* \wedge \pi^\dagger$. Now $\Theta \cup \Pi \boxminus \Theta \cup \{\pi^*, \neg \bigwedge \Lambda \vee \pi^\dagger\}$ is almost as simple as $\Theta \cup \{\pi^*\}$ and $\Theta \cup \{\pi^\dagger\}$,[12] and it unifies Ω. Since the analogous holds for $\Lambda \cup \Pi$, Π is a set of suitable supplementary sentences. But similar to Sober's trivialization proof (see §8.3.2), it now holds that $\Pr(\Omega \mid \Theta \cup \Pi) > \Pr(\Omega \mid \Lambda \cup \Pi)$, so that Ω contrastively confirms Θ against Λ. (It also follows that Θ is contrastively testable against Λ.)

If definition 10.2 is liberalized as in (i), the assumption about π^* and $\pi^\dagger$ can be weakened accordingly. For liberalization (ii), Π can be defined as $\{\neg \bigwedge \Theta \vee \pi^*\}$, which puts less strict demands on what can be called 'almost as simple as'. Similarly, if definition 10.1 of contrastive confirmation itself is changed to allow for different supplementary sentences Π_1 for Θ and Π_2 for Λ,[13] then $\Pi_1 \boxminus \{\neg \bigwedge \Theta \vee \pi^*\}$ and $\Pi_2 \boxminus \{\forall x(x = x)\}$ are enough to show that Θ can be confirmed against Λ. In this case, as in (ii), one can even demand that the supplementary sentences must unify the very basic sentences used in their justification. Thus with Boudry and Leuridan's sufficient condition for suitability, contrastive confirmation is trivial.

10.2 Confirmation in likelihoodism

Boudry and Leuridan argue against Sober's definition of suitability under the assumption that it makes it impossible to contrastively confirm or disconfirm ID. Now I want to show that this assumption is false independently of the definition of suitability. In likelihoodism, relative confirmation is defined as follows (Sober 2008, 32):

Definition 10.3. Set Ω of basic sentences contrastively confirms theory Θ against theory Λ relative to Π if and only if

$$\Pr(\Omega \mid \Theta \cup \Pi) > \Pr(\Omega \mid \Lambda \cup \Pi)\,. \tag{10.1}$$

Because of the different logical structure, it is clear that contrastive confirmation is not continuous with Bayesian confirmation. More generally, since likelihoodism defines the confirmation of a theory only in contrast to another theory, likelihoodism on its own requires a fundamental revision of all concepts that rely on confirmation and are commonly not defined contrastively, such as

[12] Similarly, according to Boudry and Leuridan (2011, §3.4) the hypothesis that the disappeared landlord staged the crime scene to fake his own death and go underground is arguably "not far more complex" than the hypothesis that he was murdered. Furthermore, unless there are restrictions on the language, it may be possible to introduce new vocabulary that simplifies Π.

[13] This is in fact what Boudry and Leuridan assume (Boudry, personal email from May 21, 2011).

the notion of probabilistic empirical significance. Likelihoodism therefore also requires the modification of claims that rely on these concepts, including the claim (made by Sober himself) that a theory that does not make basic assertions is not empirically significant. I have argued in §8.5 that the contrastive notion of empirical significance in likelihoodism is inadequate, and argued in §9.3.2 that in likelihoodism, there are theories (for example ID) that do not make basic assertions but are nonetheless contrastively empirically significant. I will now argue that the concept of contrastive confirmation developed in likelihoodism is inadequate as well, and that in likelihoodism, ID and other theories that do not make basic assertions can be confirmed by criticizing a competing theory, absent further assumptions about the theories' relation. This conclusion contradicts Sober's own claims about the confirmation of ID.

Assuming likelihoodism, Sober (2007, 7) criticizes a common argument for the confirmation of ID as follows:

> Defenders of ID often claim to test their position [. . .] by criticizing the theory of evolution. Behe (1996) contends that evolutionary processes cannot produce "irreducibly complex" adaptations; since we observe such traits, evolutionary theory is refuted, leaving ID as the only position standing. [T]his argument does nothing to test ID. For ID to be testable, *it* must make predictions. The fact that a different theory makes a prediction says nothing about whether ID is testable.

Going from bottom to top, Sober's argument amounts to the following: Since ID does not make predictions, it is not testable. Therefore, it is not confirmed by observations of irreducibly complex adaptations (this is a trivial proof given definitions 8.3 and 10.3, assuming the supplementary sentences for contrastive confirmation and testability are the same). Elsewhere, Sober (1999, 66–67) points out that there may be other theories besides ID which predict traits that are improbable according to ET, and continues:

> [T]he defect in this argument that I'm now pointing to is different. [. . .] The worst-case scenario for Darwinism is that the theory, with appropriate auxiliary assumptions, entails that what we observe was very improbable. However, this, by itself, isn't enough to reject Darwinism and opt for the hypothesis of intelligent design. We need to know how probable it is that the features would exist, if they were the result of intelligent design. *Both* hypotheses must make predictions if the observations are to help us choose between them.

Sober here directly infers that because ID does not make predictions, a trait that is improbable according to ET does not test, and therefore also does not confirm, ID. That a low likelihood of ET cannot confirm ID is also a common position outside of likelihoodism (cf. Pennock 2011, 188).

But as in the case of testability, Sober's conclusion is false if there is only one observation Ω^* such that $\Pr(\Omega^* \mid \text{ET} \cup \Pi) \neq \Pr(\Omega^* \mid \Pi)$, $\Pr(\Omega^* \mid \text{ID} \cup \Pi) = \Pr(\Omega^* \mid \Pi)$, where all probabilities are defined. Under this very weak assumption (see §9.3.2), ID can be contrastively confirmed against ET. For let Δ^* express that Ω^* is false.[14] Then $\Pr(\Omega^* \mid \text{ET} \cup \Pi) = 1 - \Pr(\Delta^* \mid \text{ET} \cup \Pi)$, so that either $\Pr(\Delta^* \mid \text{ET} \cup \Pi) < \Pr(\Delta^* \mid \Pi) = \Pr(\Delta^* \mid \text{ID} \cup \Pi)$ or $\Pr(\Delta^* \mid \text{ET} \cup \Pi) > \Pr(\Delta^* \mid \Pi) = \Pr(\Delta^* \mid \text{ID} \cup \Pi)$, where all probabilities are defined. Thus the real worst-case scenario for ET is that it assigns a true observation statement Ω^* a lower probability than the auxiliary assumptions alone. And if Behe were to prove this for some true Ω^*, he would thereby show that ID is contrastively confirmed. Like the argument in the case of testability, this argument can be repeated for any two theories that fulfill the very weak assumption: A tautologous theory can be contrastively confirmed against quantum physics, for instance, and the nonsense theory 'Foo is bar' can be contrastively confirmed against plate tectonics.

Of course, a contrastive *disconfirmation* of ID is more likely, assuming that for most Ω^* with $\Pr(\Omega^* \mid \text{ET} \cup \Pi) \neq \Pr(\Omega^* \mid \Pi)$, it holds that $\Pr(\Delta^* \mid \text{ET} \cup \Pi) < \Pr(\Delta^* \mid \Pi)$. Thus contrary to Boudry and Leuridan, contrastive testability allows the disconfirmation of ID. The problem, however, is that contrastive testability also allows the disconfirmation of 'Foo is bar' against the claim that this section contains a formula.

* * *

It may be possible to amend likelihoodism and avoid the contrastive confirmation of 'Foo is bar'. It is another question whether the resulting definition of contrastive confirmation allows a consistent reformulation of other concepts that rely on confirmation. This question is very much open, for one because even likelihoodism itself (at least in Sober's version) still contains concepts that implicitly rely on non-contrastive confirmation. I have already noted in §8.2.1 that in his informal arguments, Sober sometimes relies on a non-contrastive notion of empirical significance. More importantly, Sober's definition of testability, and thus of confirmation, seems to rely implicitly on a non-contrastive notion of confirmation. For he defines suitable supplementary sentences as being justified (independently of specific observations and theories), but leaves open the definition of 'justification'.[15] Very plausibly, any such definition that would be acceptable to Sober would have to involve confirmation, so that the resulting definition would be recursive (presumably with observation sentences as supplementary sentences in the recursion base). But if confirmation is contrastive, then each step from the recursion's base up to the contrastive confirmation of the two theories under investigation will involve a confirmation of supplementary sentences *relative to some other sentences*. Thus, if the question is the confirmation of Θ against Λ, and

[14] See §8.3.3, n. 16.
[15] See §8.3.1, n. 13.

Θ and Λ are included in the n^{th} iteration of the recursive step, Θ will be confirmed or disconfirmed against Λ relative to $n+1$ different sets of sentences. How such a conception of confirmation can be practically applied, consistent, or even properly spelled out is unknown. Given the other problems of contrastive confirmation, the development of such a conception may not be worth the trouble. It may thus be time to find not only a better criterion of testability, but also a better criterion of confirmation.

10.3 New ways to old criteria

In §10.1, I have shown how Sober relies on the converse of his postulated relation between testing and testability to argue for restrictions on the supplementary sentences. Sober thus has to take the relation between testing and testability as a necessary and sufficient condition. I will follow Sober in this regard, so that taking definition 8.8 as criterion of testability entails that a set Ω tests Θ relative to Π if and only if $\Pr(\Omega\,|\,\Theta \cup \Pi) \neq \Pr(\Omega\,|\,\Pi)$. Further assuming with Sober that testing amounts to confirming or disconfirming, this allows for two definitions of confirmation (one with a 'less-than', one with a 'greater-than' sign), one of which is, unlike contrastive confirmation, continuous with Bayesian confirmation:[16]

Definition 10.4. Set Ω of $\mathscr{B}$-sentences *relevantly confirms Θ relative to Π* if and only if

$$\Pr(\Omega\,|\,\Theta \cup \Pi) > \Pr(\Omega\,|\,\Pi)\,. \tag{10.2}$$

This definition is equivalent to the Bayesian definition 8.1 whenever the theory can be assigned a probability, but, like Sober's definition 10.3, does not presume that this is possible. It has also frequently been defended as a criterion of confirmation (cf. Mackie 1969) under the name 'relevance criterion'.

The question is now what an absolute criterion of confirmation, independent from some not further specified set of supplementary sentences, should be like. As discussed in §10.1, Sober relies on his postulate for relation between testing and testability even for his absolute criterion, which leads to definition 10.1 of confirmation. But the definition states that Ω contrastively confirms Θ against Λ as long as there is *any* honest set of supplementary sentences relative to which Θ's likelihood is greater than Λ's, and analogously for the relevance criterion. And this has at least two problematic implications. First, since different honest sets may lead to different conditional probabilities for an observation, a theory may be both confirmed *and* disconfirmed against another by some observations. Second, the definition allows for ignoring any justified sentence one is inclined to. And this is incompatible with Carnap's principle of sufficient evidence, which "says that you must use all the relevant evidence you have" (Sober 2008, 44), and which Sober (2008, 41–46, §2.10) endorses.

[16] It is also only this one that is the least bit plausible.

The problem, I conjecture, lies with Sober's description of the relation between testing and testability, which should be formulated relative to the supplementary assumptions Π. That is, if Ω tests Θ relative to Π if and only if $R(\Omega,\Theta,\Pi)$, then (and only then) it holds that Θ is testable relative to Π if and only if there is some possible set Ω of basic sentences such that $R(\Omega,\Theta,\Pi)$. This also describes the relation between definition 10.4 and definition 8.8. To arrive at an absolute concept of testability, one has to existentially generalize on the supplementary assumptions as well: If Ω tests Θ relative to Π if and only if $R(\Omega,\Theta,\Pi)$, then (and only then) it holds that Θ is testable if and only if there is some honest set Π of supplementary sentences and some possible set Ω of basic sentences such that $R(\Omega,\Theta,\Pi)$.[17] This relation leads, for example, from $\mathscr{B}$-creativity (see definitions 6.3 and 6.6) to falsifiability (definitions 6.2 and 6.8).

With this amended relation between testing and testability, one can now explicate 'confirmation by Ω'. The principle of total evidence again provides a guide, since the confirmation should be relative to as many justified sentences in Π as possible. Given definition 10.4, this information must not contain Ω or Θ, since then confirmation would be almost trivial: Whenever Ω is justified, $\Omega \subseteq \Pi$, and thus Ω cannot confirm anything relative to Π. Whenever Θ is justified, $\Theta \subseteq \Pi$, and thus nothing can confirm Θ relative to Π. I therefore suggest

Definition 10.5. A set Π is *maximally honest (excluding set Σ)* if and only if Π is honest (and $\Pi \cap \Sigma = \varnothing$) and there is no honest set Γ of sentences such that $\Pi \subset \Gamma$ (and $\Gamma \cap \Sigma = \varnothing$).

Definition 10.5 does not presume that there is a unique maximally honest set of sentences, but if the set of all justified sentences is honest, it is maximally honest. This allows

Definition 10.6. Set Ω of $\mathscr{B}$-sentences *relevantly confirms* Θ if and only if Ω relevantly confirms Θ relative to a maximally honest set excluding $\Omega \cup \Theta$.

Note that, like *making probabilistic $\mathscr{B}$-assertions*, this criterion of confirmation does not presume that theories can be assigned probabilities. If the starting position is the Bayesian notion of probabilistic empirical significance (definition 8.2), definition 10.5 and the relation between empirical significance and confirmation lead to

Definition 10.7. Set Ω of basic sentences *confirms Θ in Bayesianism* if and only if Ω confirms Θ in Bayesianism relative to a maximally honest set excluding $\Omega \cup \Theta$.

This criterion of confirmation presumes that theories can be assigned probabilities. Thus, with two different probabilistic criteria of empirical significance on offer, and assuming the relation between testability and testing postulated by

[17] If 'honest set' turns out not to be a good explicatum for 'suitable set', a new explicatum can be substituted.

Sober, one can arrive at two criteria of confirmation. As is the case for the criteria of empirical significance, the two criteria of confirmation are equivalent whenever all probabilities are defined.

Since definition 10.5 does not ensure the uniqueness of a maximally honest set, it is under both definition 10.6 and definition 10.7 in principle still possible that some observation both confirms and disconfirms a theory. Since the set of maximally honest sets in most cases is a proper subset of the set of honest sets, the danger does not seem overwhelming, however. Furthermore, if there are indeed two incompatible maximally honest sets with this feature, this suggests a more general problem with the system of beliefs under consideration.

* * *

A side note: One may wonder whether the relation between probabilistic empirical significance and confirmation should not be as follows: If Ω tests Θ relative to Π if and only if $R(\Omega, \Theta, \Pi)$, then (and only then) it holds that Θ is testable if and only if there is some *maximally* honest set Π of supplementary sentences and some possible set Ω of basic sentences such that $R(\Omega, \Theta, \Pi)$. This relation, however, has the unwanted effect of rendering every theory Θ untestable that is in a strong sense reducible to another theory Λ (see §11). Assuming that Θ can be derived from Λ, every assertion Θ makes can also be derived from Λ. But Θ is not needed to justify Λ, and thus a maximally honest set excluding Θ would still include Λ (unless Λ is incompatible with other justified sentences). Thus Θ is not probabilistically empirically significant. This is the case for statistical mechanics, for example, if it is reducible to quantum mechanics. It also holds for the theory of selection if, as Matthen and Ariew (2002, 68–69) argue,

> selection is a formally characterized phenomenon, a statistical property of physical substrates that possess certain metrical properties. The causally active physical properties that lie beneath this metric are different from, or even incommensurable with, the properties that form the subject matter of the theory of selection.

If this is a correct interpretation of the theory of selection, then defining probabilistic empirical significance relative to a maximally honest set would put the theory of selection, and indeed any reducible theory, on a par with ID. It would also mean that a theory Θ can cease to be empirically significant, for example if a new theory is accepted to which Θ can be reduced.

10.4 Beyond likelihoods

So far, I have looked only on the *influence* one set of observations can have on the status of a theory. In Carnap's terms, I have analyzed different explications of the *increase in firmness* of a theory (Carnap 1962, xv–xvii). It is clear that this is not enough, for the simple reason that one often needs to now whether, given a maximally honest set of supplementary sentences, one should believe in or accept a theory, that is, in Carnap's terms, whether a theory is *firm*. Such a categorical concept of acceptance is necessary even for the definition of 'honest set' (and Sober's definition of contrastive testability), since the notion of a justified belief is a categorical one. I now want to point out briefly that any definition of firmness that relies solely on likelihoods does not allow ampliative inferences, and is thus of questionable pragmatic value.

Boudry and Leuridan (2011, §§1,3.2) consider Sober's restriction IV a means to avoid gerrymandering of the supplementary sentences to fit the observations (which they themselves try to avoid by demanding that the supplementary sentences be unifying (Boudry and Leuridan 2011, §2.1)). And while their examples do not show that the definition is too restrictive, there is a vital problem with contrastive confirmation that, although it has nothing to do with restriction IV, may lie at the core of their unease. The gerrymandered supplementary sentences can be conjoined with the hypothesis of a designer to form a *new* hypothesis (cf. Sober 2008, 131–133), because Sober's criterion of confirmation contains no restriction whatsoever on the choice of hypotheses to be compared. This poses no problem for contrastive testability, since there seems to be no problem with the testability of a gerrymandered hypothesis against non-gerrymandered hypotheses. A gerrymandered hypothesis may even be confirmed against its competitor by some observation Ω, if *confirmation* is understood as *increase in firmness*. However, the situation is different for the contrastive confirmation of a theory against another if *confirmation* is understood as *firmness*. Following the principle of total evidence, a theory has to be confirmed in this sense in light of *all* observations. But for many a hypothesis Θ and many of its observational assertions β, one can construct another (gerrymandered) hypothesis Λ that differs from Θ only in that it asserts $\neg\beta$ instead.[18] Θ and Λ assign the same probabilities to any set Ω of observations with $\beta \notin \Omega$. Therefore, for any definition of categorical confirmation that relies only on likelihoods, Θ and Λ are either both confirmed or both disconfirmed. This means specifically that neither contrastive confirmation nor the relevance criterion of confirmation as given in definition 10.6 can lead to a criterion of firmness that allows for a distinction between Θ and Λ.

The Bayesian criterion of confirmation given by definition 10.7 by itself also does not obviously solve the problem of gerrymandered theories, since the probability of Λ can be increased just as the probability of Θ. But definition 10.7 *presumes* that theories can be assigned probabilities. Thus, according to Bayesianism, Θ and Λ may differ in their prior probabilities, so that one may be more likely than the other. Proponents of inference to the best explanation may instead

[18] This is often a non-trivial problem to solve formally, as it essentially involves an independence proof. Informally, the above description seems sufficient, however.

opt for one of the theories because it explains better. Boudry and Leuridan might argue that the more simple and unifying theory should be preferred. Contrastive confirmation and relevant confirmation, however, provide no reason to prefer one hypothesis over the other. Thus, considering that β can be 'The world will end tomorrow', contrastive and relevant confirmation may be less discerning than desirable.

* * *

In summary, Sober's postulate about the relation between testing and testability has to be reformulated to take supplementary sentences into account. When this is done, it leads from criteria of testability relative to supplementary sentences to criteria of confirmation relative to supplementary sentences. The principle of total evidence then further suggests absolute criteria of confirmation. These criteria have to be taken as criteria for the increase of firmness of a theory. And if such a criterion relies solely on likelihoods, it cannot support ampliative inference.

Chapter 11

Reduction

The discussions of intelligent design and confirmation relied on the original interpretation of $\mathscr{B}$-terms being empirically easily accessible, and $\mathscr{A}$-sentences being possibly non-empirical. In this chapter and the following, I will discuss applications of the *formalism* of the criteria of empirical significance in new areas. This chapter will cover relations between a basic theory Θ in vocabulary $\mathscr{B}$ and an auxiliary theory Λ that may have terms in common with Θ ($\mathscr{C}$-terms), but also may have some proprietary vocabulary $\mathscr{A}$ with $\mathscr{A} \cap \mathscr{B} = \varnothing$. Thus Λ is formulated in $\mathscr{A} \cup \mathscr{C}$, with $\mathscr{C} \subseteq \mathscr{B}$. The supplementary sentences Π may further connect $\mathscr{B}$-terms to $\mathscr{A}$-terms.

Criteria of empirical significance, both for sentences and terms, can be seen as determining whether there is some connection between $\mathscr{A}$-terms and $\mathscr{A}$-sentences on the one hand, and $\mathscr{B}$-terms and $\mathscr{B}$-sentences on the other hand. Much of the discussion around reduction essentially revolves around the same subject, and thus it is to be expected that criteria of empirical significance relate to criteria of reduction. I will review four kinds of reduction, suggested by Nagel, Kemeny and Oppenheim, Fodor, and Nickels, in this respect. Two more *apropos* discussions, of the status of bridge laws and of structural realism, follow thereafter.

11.1 Nagel

The distinction between a $\mathscr{B}$-vocabulary belonging to one theory and an $\mathscr{A}$-vocabulary proprietary to another is essentially that presumed by Nagel (1951, 330). He sees "the conditions under which one science can be reduced to some other one" as

> the logical and empirical conditions which must be satisfied if the laws and other statements of one discipline can be subsumed under, or explained by, the theories and principles of a second discipline.

Nagel (1951, 330) gives the following conditions for a reduction:

> Omitting details and refinements, the two conditions which seem to be necessary and sufficient for such a reduction are briefly as follows. Let S_1 be [. . .] the "primary discipline," to which a second science, S_2, for example biology, is to be reduced. Then (i) every term which occurs in the statements of S_2 [. . .] must be either explicitly definable with the help of the vocabulary specific to the primary discipline [. . .] or well-established empirical laws must be available with the help of which it is possible to state the sufficient conditions for the application of all expressions in S_2, exclusively in terms of expressions occurring in the explanatory principles of S_1. [. . .] Though the label is not entirely appropriate, this first conditio will be referred to as the condition of definability. (ii) Every statement in the secondary discipline, S_2, and especially those statements which formulate the laws established in S_2, must be derivable logically from some appropriate class of statements that can be established in the primary science, S_1 [. . .]. This second condition will be referred to as the condition of derivability.

It is a matter of considerable discussion what Nagel's final notion of reduction was *precisely* (cf. van Riel 2011). In my rather ahistorical discussion, I will focus on the technical aspects of the notion of reduction suggested by the quote above. I will take science S_1 to be circumscribed by Θ's vocabulary $\mathscr{B}$, and science S_2 by $\mathscr{A} \cup \mathscr{C}$. The definitions or empirical laws referred to in condition (i) I will call Π. Nagel (1951, 330) gives an example for how Π is to "state the sufficient conditions for the application of all expressions in S_2":

> For example, it must be possible to state the truth-conditions of a statement of the form 'x is a cell' by means of sentences constructed exclusively out of the vocabulary belonging to the physico-chemical sciences.

If 'x is a cell' is, plausibly, taken to be a formula, and the truth conditions are stated by giving a $\mathscr{B}$-formula with the same extension in every model of Π, this entails that every $\mathscr{A}$-term must be $\mathscr{B}$-definable (Gupta 2009, §2.3). Thus the above quote plausibly leads to

Definition 11.1. A science phrased in $\mathscr{A} \cup \mathscr{C}$ is N-reducible to a science phrased in $\mathscr{B}$ if and only if there is a set Π of well-established sentences such that (i) every $\mathscr{A}$-term is $\mathscr{B}$-definable given Π and (ii) every well-established theory Λ in $\mathscr{A}$ is verified by a well-established theory Θ in $\mathscr{B}$ given Π.

Note that according to definition 11.1, Λ is an omission from Θ given the background assumptions Π. N-reducibility achieves the two goals of a reduction that Nagel (1951, 330) lists. Condition (i) ensures that all "statements of one discipline can be subsumed under [the] principles of a second discipline", since every $\mathscr{A} \cup \mathscr{C}$-sentence can be translated into a $\mathscr{B}$-sentence by claim 7.1. Condition (ii) ensures that "the laws [...] of one discipline can be [...] explained by the theories [...] of a second discipline " if one takes an explanation to be a derivation from a scientific theory.

Nagel (1949, 330) adds that it is "evident that the second condition cannot be fulfilled unless the first one is, although the realization of the first condition does not entail the satisfaction of the second one". Given claim 7.6, some sentences that contain not $\mathscr{B}$-definable terms can be translated into $\mathscr{B}$-sentences, and thus there are *a fortiori* verifiable sentences with terms that are not $\mathscr{B}$-definable. Thus the first conjunct of this statement is false, and Kemeny and Oppenheim (1952, 10) point out that so is the second conjunct. For if Π contains explicit definitions for all terms in Λ, then Λ can be translated into $\mathscr{B}$ according to claim 7.1. If further Λ and Π are well-established, the translation of Λ into $\mathscr{B}$ is well-established as well, and thus the translation itself is a theory Θ that verifies Λ given Π. Thus the following holds:

Claim 11.1. *If a set of sentences entailed by a set of well-established sentences is itself well-established, then a science phrased in $\mathscr{A}$ is N-reducible to a science phrased in $\mathscr{B}$ if and only if there is a set Π of well-established sentences such that every $\mathscr{A}$-term is $\mathscr{B}$-definable given Π.*

Thus, taken literally, Nagel-reducibility is the strongest deductive criterion for the empirical significance of terms. There is another, more speculative way to read Nagel, however. For if Nagel considers definability to be necessary for derivability, he may have seen the derivability condition as the important one, and condition (i) as a restriction on the possible sets Π: A definition of an $\mathscr{A}$-term in $\mathscr{B}$-terms is $\mathscr{B}$-conservative, which is important to not trivialize the notion of reduction. For if Π could be any set, it could be chosen to be identical to Λ, so that Λ is trivially derivable from $\Theta \cup \Pi$. But condition (i) could be weakened from the demand that all terms must be definable in $\mathscr{B}$-terms given Π to the demand that Π be $\mathscr{B}$-conservative. This would lead to the following

Definition 11.2. A science phrased in $\mathscr{A} \cup \mathscr{C}$ is *idiosyncratically N-reducible* to a science phrased in $\mathscr{B}$ if and only if there is a set Π of well-established, $\mathscr{B}$-conservative sentences such that every well-established theory Λ in $\mathscr{A} \cup \mathscr{C}$ is verified relative to Π by a well-established theory Θ in $\mathscr{B}$.

Definition 11.2 is weaker than definition 11.1, but achieves that the laws of one discipline can be derived from the laws of the other. While it has the disadvantage of not taking account of Nagel's demand that all statements of one discipline be subsumed under the other, it avoids the irrelevance of the derivability condition. Definition 11.2 could be strengthened to demand that Π must be such that every well-established theory in $\mathscr{A} \cup \mathscr{C}$ is *translatable* into a well-established theory in $\mathscr{B}$. Since according to claim 7.6, some sentences that can be translated into $\mathscr{B}$ contain terms that are not $\mathscr{B}$-definable, this strengthening would still be weaker than N-reducibility.

11.2 Kemeny and Oppenheim

Kemeny and Oppenheim (1952, 13–15) consider claim 11.1 a reason to develop a new account of reduction, $\mathrm{Red}(\Theta, \Lambda)$, that they use to proof the following (Kemeny and Oppenheim 1952, 15, my notation):

> $\mathrm{Red}(\Theta, \Lambda)$ if and only if (1) $\mathscr{A} \cup \mathscr{C}$ contains terms not in $\mathscr{B}$, (2) every observational statement implied by Λ is also implied by Θ, and (3) Θ is at least as well systematized as Λ.

Let observational statements be $\mathscr{C}$-sentences. Assuming that the implication of a $\mathscr{C}$-sentences may rely on justified supplementary assumptions Π, every observational statement implied by Λ is implied by Θ if and only if Θ makes all the deductive $\mathscr{C}$-assertions of Λ relative to Π according to definition 9.1. By claim 9.1, this leads to

Definition 11.3. Λ is K-O-reducible to Θ if and only if (1) $\mathscr{A} \neq \emptyset$, (2) Θ is at least as syntactically falsifiable as Λ (with $\mathscr{C}$ as basic terms), and (3) Θ is at least as well systematized as Λ.

Note that 'at least as falsifiable as' here does not take into account implications in higher order logic with infinite additional $\mathscr{C}$-sentences as premises (see page 224). Note also that Λ is reducible to Θ only if Λ's abstraction in $\mathscr{C}$-terms omits from Θ's abstraction in $\mathscr{C}$-terms. Apart from the demand on the vocabularies and the systematization of the theories, K-O-reducibility is thus a comparative criterion of empirical significance. Claims 9.1 and 9.2 furthermore suggest that substituting definition 9.2 of *making at least as many probabilistic observational assertions* for *being at least as falsifiable as* leads to a probabilistic version of K-O-reducibility. Finally, it is trivial to show that if a theory Λ is deductively or probabilistically K-O-reducible to some other well-supported theory Θ, then abandoning Θ in favor of Λ is not scientific according to the modified translatability criterion.

11.3 Fodor

Fodor (1974, 97–98) attempts to develop a weaker kind of intertheoretic relation than that of reduction:

> [T]he assumption that the subject-matter of psychology is part of the subject-matter of physics is taken to imply that psychological theories

> must reduce to physical theories, and it is this latter principle that makes the trouble. I want to avoid the trouble by challenging the inference.

Thus Fodor intends to develop an account of $\ulcorner x$ is part of subject matter $y \urcorner$ that is weaker than $\ulcorner x$ is reducible to $y \urcorner$. His first step is to introduce the notion of a "natural kind" into the discussion by way of a definition (Fodor 1974, 102, my notation):

> If I knew what a law is, and if I believed that scientific theories consist just of bodies of laws, then I could say that P is a natural kind predicate relative to [theory] S iff S contains proper laws of the form $Px \rightsquigarrow \alpha(x)$ or $\alpha(x) \rightsquigarrow Px$;

'$S_1 \rightsquigarrow S_2$' is a law and "intended to be read as something like 'all S_1 situations bring about S_2 situations'" (Fodor 1974, 98). Since the universal quantification is implied (Fodor 1974, 111), '$Px \rightsquigarrow \alpha(x)$' probably has to be read as 'For all x, Px brings about $\alpha(x)$'. This can be written more generally as $\Theta(P)$, that is, a theory Θ that contains the predicate P. In short, then, a natural kind predicate is one that occurs in a theory, probably non-vacuously.[1]

Now Fodor (1974, 98) characterizes the notion of reduction as follows: He assumes "that a science is individuated largely by reference to its typical predicates", so that $\mathscr{B}$ can be the vocabulary of the reducing and $\mathscr{A}$ the proprietary vocabulary of the reduced science. A necessary and sufficient condition of the reduction of a science in $\mathscr{A} \cup \mathscr{C}$ to a science in $\mathscr{B}$ is that for all the $\mathscr{A} \cup \mathscr{C}$-laws $\Lambda(V)$, there is a $\mathscr{B}$-law $\Theta(B)$ such that $Vx \leftrightsquigarrow Bx$, where $Vx \leftrightsquigarrow Bx$ is called a 'bridge law' (Fodor 1974, 98–99). Since Fodor (1974, 99) interprets bridge laws as identity statements, one can generalize his account of reducibility to

Definition 11.4. A science in $\mathscr{A} \cup \mathscr{C}$-terms is *F-reducible* to a science in $\mathscr{B}$-terms if and only if for all $\mathscr{A} \cup \mathscr{C}$-laws $\Lambda(\bar{V})$ there are a $\mathscr{B}$-law $\Theta(\bar{B})$ and a law $\Pi = \{\bar{V} = \bar{B}\}$.

Note that every F-reducible science is also N-reducible, but not vice versa, since an $\mathscr{A}$-term may be definable in $\mathscr{B}$-terms without being identical to any of them.

Fodor (1974, 103–107) argues that there can be no natural kind terms in physics that are identical to natural kind terms in the special sciences, say, psychology or economics, so that there are no F-reducible special sciences. Instead of bridge laws, Fodor (1974, 108) suggests, one should allow true empirical generalizations of the form $Sx \leftrightsquigarrow P_1x \vee P_2x \vee \cdots \vee P_nx$, which are enough for a (liberalized) concept of

[1] There is much hidden in the phrase 'brings about'. Fodor would presumably not say that Pa brings about $Pa \vee Qa$, but it is not clear how to spell out those subtleties explicitly.

reduction.[2] In a slight generalization, one can thus give

Definition 11.5. A science in $\mathscr{A} \cup \mathscr{C}$-terms is *liberalized F-reducible* to a science in $\mathscr{B}$-terms if and only if for each law $\Lambda(P_1,\ldots,P_n)$ in $\mathscr{A} \cup \mathscr{C}$ and for each $i, 1 \leq i \leq n$, there are k_i $\mathscr{B}$-terms $B_{i_1},\ldots,B_{i_{k_i}}$ that occur non-vacuously in $\mathscr{B}$-laws, and there are true empirical generalizations

$$\left\{\forall \bar{x}\left(P_i\bar{x} \leftrightarrow \bigvee_{j=1}^{k_i} B_{i_j}\bar{x}\right) \,\middle|\, 1 \leq i \leq n\right\}. \tag{11.1}$$

Fodor (1974, 109) himself points out the central problem with liberalized F-reducibility: From $\Lambda(P_1,\ldots,P_n)$ and the empirical generalizations (11.1) (which, like a bridge law, is taken to be an identity), one can infer $\Lambda\big(\lambda\bar{x}[B_{1_1}\bar{x} \vee \ldots \vee B_{1_{k_1}}\bar{x}],$ $\ldots, \lambda\bar{x}[B_{n_1}\bar{x} \vee \ldots \vee B_{n_{k_n}}\bar{x}]\big)$, which is a law. This is analogous to Kemeny and Oppenheim's observation with respect to N-reducibility. Since Fodor phrases his discussion in terms of natural kind terms, he further infers that for each $1 \leq i \leq n$, $\lambda\bar{x}(B_{i_1}\bar{x} \vee \cdots \vee B_{i_{k_i}}\bar{x})$ is a natural kind term and thus in $\mathscr{B}$, which makes the empirical generalizations (11.1) bridge laws in the sense of F-reducibility, and thus liberalized F-reducibility is the same as F-reducibility. Incidentally, using the newly defined terms, one has found an abstraction of the $\mathscr{B}$-theory $\Lambda\big(\lambda\bar{x}[B_{1_1}\bar{x} \vee \ldots \vee B_{1_{k_1}}\bar{x}],\ldots,\lambda\bar{x}[B_{n_1}\bar{x} \vee \ldots \vee B_{n_{k_n}}\bar{x}]\big)$ in terms of $P_1,\ldots,P_n$.

In response, Fodor (1974, 109–110) argues that the inference is invalid because 'it's a law that __' is not truth-functional. For otherwise,

> one gives up the possibility of identifying the natural kind predicates of a science with those predicates which appear as the antecedents or the consequents of its proper laws. [...] One thus inherits the need for an alternative construal of the notion of a natural kind, and I don't know what that alternative might be like.

However, if this the only problem with treating 'it's a law that __' as truth-functional, there seems to be no good reason to give up truth-functionality. For one, the price of giving up truth-functionality is steep: It means disallowing equivalent reformulations of laws, so that the ideal gas law $PV = nRT$ (R is the gas constant, P the volume, V the pressure, T the temperature, and n the molar number of the gas) is a law, but its equivalent reformulation $p = nRT/V$ may not be. Or, if the latter formulation is a law, the original formulation may not be one.

[2]This is indeed Fodor's terminology: He supposes "that it is enough [...] that every law of the special sciences should be *reducible* to physics by bridge statements which express true empirical generalizations" (Fodor 1974, 108, my emphasis). Thus according to Fodor, the following definition 11.5 is a case of reduction.

It is furthermore not clear whether antecedents of laws would be in general acceptable as natural kind terms to Fodor himself. For P, V, and T may be intuitively satisfying natural kinds, but when phrased as an implication, the ideal gas law looks like this:

$$\forall xy\,[Ix \land Px \cdot Vx = y \rightarrow n(x) \cdot R \cdot T(x) = y]\ , \tag{11.2}$$

where I stands for 'is an ideal gas'. Thus the antecedent relation is $\lambda xy(Ix \land Px \cdot Vx = y)$, which makes for an odd natural kind term. If the equivalent reformulation of the ideal gas law given above is the real gas law, another natural kind term results. Or one could introduce uncountably many natural kind terms $\lambda x(Ix \land Px \cdot Vx = c)$, one for each $c \in \mathbb{R}^{\geq 0}$.

It is thus doubtful that Fodor succeeds in avoiding the collapse of liberalized F-reducibility into F-reducibility. But even if he did, liberalized F-reducibility is still much stronger than N-reducibility, since N-reducibility allows *any* explicit definition of $\mathscr{A}$-terms, while liberalized F-reducibility demands an explicit definition of $\mathscr{A}$-terms with disjunctive definientia. Fodor thus has by no means weakened the standard notion of reduction, but rather introduced his own, stronger notion, and weakened *that*—but not enough. If all that was needed was a notion of reducibility weaker than F-reducibility, N-reducibility would be adequate.

While Fodor is unsure about the nature of laws, the concept of a natural kind is similarly unclear. Given the above considerations, it is by no means obvious how one would go about identifying one. But it may be possible to use Nagel's comparatively weak notion of reduction to save Fodor's intuition that "the subject-matter of psychology is part of the subject-matter of physics" without thereby having "psychological theories [...] reduce to physical theories" (Fodor 1974, 97–98). Relying on Lewis's notion of aboutness (definition 6.9), one can state that psychological laws are about the subject matter determined by $\mathscr{B}$, where $\mathscr{B}$ is the vocabulary of physics. By claim 7.5, it is enough for the terms of psychology to be definable *in every model* for all sentences of psychology to be about the subject matter of physics. The models here could be those of the theories of physics Θ and the bridge laws Π. Since definability in every model is strictly weaker than definability, this results in psychology being about the physical world, but not N-reducible to physics.

If a relation P_i in first order logic is definable in every model of $\Theta \cup \Pi$, it is piecewise $\mathscr{B}$-definable by claim 7.4:

$$\Theta \cup \Pi \vDash \left\{ \bigvee_{j=1}^{k_i} \forall \bar{x}\,(P_i\bar{x} \leftrightarrow \beta_{i_j}\bar{x}) \;\middle|\; 1 \leq i \leq n \right\}, \tag{11.3}$$

where each β_{i_j} is a $\mathscr{B}$-formula. The crucial difference to liberalized F-reducibility is that there is no explicit definition of P_i by a disjunction of $\mathscr{B}$-formulas, but a disjunction of explicit definitions of P_i by $\mathscr{B}$-formulas. For one-place predicates, Humberstone (1998) shows that piecewise $\mathscr{B}$-definability is equivalent to the *supervenience* of P_i over $\mathscr{B}$, defined as

$$\Theta \cup \Pi \vDash \left\{ \forall xy \left[\bigwedge_{j=1}^{k_i} (B_{i_j} x \leftrightarrow B_{i_j} y) \rightarrow (P_i x \leftrightarrow P_i y) \right] \middle| \, 1 \leq i \leq n \right\} \tag{11.4}$$

for some $\{B_{i_j}\}_{j \leq k_i} \subseteq \mathscr{B}$. Since Lewis considers his definition of aboutness to be an explication of supervenience, non-reductive reduction can be defined as the supervenience of sentences or the supervenience of terms on $\mathscr{B}$. Therefore Fodor's intuition about the relation of psychology and physics can be described as follows: The sentences and terms of psychology supervene on the language of physics without being N-reducible to physics.

Through the relation of supervenience and definability in each model, supervenience without N-reducibility entails that a complete description of the physical world (a $\mathscr{B}$-structure) provides definitions for all supervening concepts. On the other hand, the laws of physics themselves may not, for they do not necessarily determine $\mathscr{B}$-structures up to isomorphism. In Fodor's terms (Fodor 1974, 104):

> [A]n immortal econophysicist might, when the whole show is over, find a predicate in physics that was, in brute fact, coextensive with 'is a monetary exchange'. If physics is general—if the ontological biases of reductivism are true—then there must be such a predicate. But (a) [. . .] nothing but brute enumeration could convince us of this brute co-extensivity, and (b) there would seem to be no chance at all that the physical predicate employed in stating the coextensivity is a natural kind term, and (c) there is still less chance that the coextension would be lawful (i.e., that it would hold not only for the nomologically possible world that turned out to be real, but for any nomologically possible world at all).

Given the above discussion about natural kind terms and laws, claim (b) seems false. But it is also irrelevant since N-reducibility can be weakened without recourse to natural kinds. Through the relation between definability in a model and Lewis's notion of aboutness, supervenience can be seen as one way to explicate the idea that "the natural facts include all the facts that a complete science will acknowledge", which itself is one way to explicate the position of substantive naturalism (Feldman 2008, §4).[3]

[3]See §2.5.

11.4 Nickles

In his analysis of different types of reduction of a theory Λ to a theory Θ, Sklar (1967, 111) notes that often, "[w]hat can be derived from the reducing theory is an *approximation* to the reduced". Typically, approximations will be of a certain degree ε, and in a specific domain (described by some one-place formula φ), so that one can give

Definition 11.6. Theory Λ is *approximately reducible* to theory Θ in domain φ to degree ε if and only if there is a theory Λ' such that $\Lambda' \approx_\varepsilon \Lambda^{(\varphi)}$ and Λ' is reducible to $\Theta^{(\varphi)}$.

$\Lambda^{(\varphi)}$ is here again the set of relativizations of all elements of Λ to φ. One can then use verifiability, K-O-reducibility, supervenience, or even translatability as an explicatum for the reducibility of Λ' to $\Theta^{(\varphi)}$. How two theories Λ and Λ' can be identical to degree ε, $\Lambda \approx_\varepsilon \Lambda'$, can be further elucidated with the help of an account by Nickles (1973), which purportedly describes a new type of reduction.

Nickles defines this new type of reduction as pertaining when a theory Θ is transformed by some operation into another theory Λ. The only operation that he spells out, though, is the limit operation $\lim_{\kappa\to 0}\Theta(\kappa) = \Lambda$. In the case of special relativity and Newtonian mechanics, κ could be $v_{\max}/c$, that is, the maximal velocity $v_{\max}$ becomes arbitrary small compared to the speed of light c. In this case, Nickles speaks of a reduction of special relativity to Newtonian mechanics. This gives the following

Definition 11.7. Θ L-reduces to Λ if and only if for some κ,

$$\lim_{\kappa\to 0}\Theta(\kappa) = \Lambda\,. \tag{11.5}$$

But $\lim_{\kappa\to 0}\Theta(x) = \Lambda$ is not a very precise way of describing the situation. At least in the case of special relativity, Nickles's type of reduction amounts to a restriction of the maximal absolute speed Vx of all objects x. Special relativity and Newtonian physics are thus restricted to the domain $\lambda x(Vx \leq v_{\max})$ and identified in the limit of a maximal speed infinitely close to 0.[4] The limit procedure is well-defined only if there is a space of theories with a distance measure d(. , .), namely by the contextual definition

$$\lim_{\kappa\to 0}\Theta(\kappa) = \Lambda \text{ if and only if}$$
$$\forall\varepsilon > 0\, \exists\delta\, \forall\kappa \leq \delta : \mathrm{d}\Big(\Theta^{(\lambda x[Fx\leq\kappa])}, \Lambda^{(\lambda x[Fx\leq\kappa])}\Big) \leq \varepsilon\,, \tag{11.6}$$

[4] Mathematically, this identification is similar to identifying the two theories in the limit of an infinite speed of light. But while the latter is usually more tractable, it is of course not at all what is happening physically.

where F may in principle be a symbol for any function to the real numbers.

The existence of a distance measure between sets of sentences is thus presupposed by Nickles, and leads to the possibility of giving a precise definition of 'approximation of Λ by Λ' to degree ε':

Definition 11.8. Λ' *approximates* Λ to the degree ε, that is, $\Lambda' \approx_\varepsilon \Lambda$, iff

$$\mathrm{d}(\Lambda', \Lambda) \leq \varepsilon \; . \tag{11.7}$$

I will also speak of Λ' being *approximately identical* to Λ to the degree ε. With this definition, the following holds:

Claim 11.2. *Θ L-reduces to Λ only if for each ε there is a domain φ such that Λ is approximately reducible to Θ in domain φ to degree ε, with 'reducible' interpreted as 'identical'.*

Proof. Choose $\Lambda' = \Theta^{(\varphi)}$ and $\varphi = \lambda x(Fx \leq \kappa)$. The claim follows from formula (11.6) for $\kappa = \delta$. □

Given that Batterman (2002, 5) claims (without proof) that L-reduction is a necessary condition for every "philosopher's reduction", it is noteworthy that Nickles's reduction is not the most general type of approximate reduction. For one, identity is a stronger demand than even translatability. But even independently from the interpretation of 'reduction', Nickles's criterion also relies on a strengthening of 'approximation'. For even if for some ε, there are a domain φ and a theory $\Lambda' \approx_\varepsilon \Lambda^{(\varphi)}$, this does not entail that there are such a domain and a theory for every degree ε. Thus to prove that there cannot be a reduction between two theories, it is not enough to show that L-reduction fails. This holds all the more true for generalizations of definition 11.8 of approximation. For instance, one could introduce a total order $\precsim$ over pairs of sets of sentences, and let one specific pair $\langle \Delta, \Sigma \rangle$ play the role of ε in definition 11.8. This leads to

Definition 11.9. Λ' *generally approximates* Λ to the degree $\langle \Delta, \Sigma \rangle$ if and only if

$$\langle \Lambda', \Lambda \rangle \precsim \langle \Delta, \Sigma \rangle \; . \tag{11.8}$$

Of course, not every order is appropriate. The specific order may very well depend on the context (for an example, see Lutz 2011b).

11.5 On the status of bridge laws

I now want to take a look at the status of Π for the special case of N-reducibility. One could argue that, since the reduced theory Λ already consists to a great extent of empirically established statements, it is no qualitatively new step to allow further empirical research to establish additional connections between $\mathscr{A}$-terms and $\mathscr{B}$-formulas. This is plausible, but brushes over an important point: The empirical research that is necessary to connect the two theories consists in research within Θ and Λ, not some overarching theory containing both. In short, this is because each theory is considered sufficiently interpreted by itself, without the need for the respective other theory. The individual interpretations determine the sets of possible models for each theory, and these sets of intended models can be compared *without further empirical research*. On the basis of this comparison, one may then additionally introduce meaning postulates, which are wholly non-empirical.

It is clear that there has to be some comparison of the possible structures of two theories when considering their common vocabulary $\mathscr{C}$. For to ensure that not only the $\mathscr{C}$-terms but also their possible structures are identical, the set of the possible $\mathscr{C}$-structures of Λ has to be compared to the set of possible $\mathscr{C}$-structures of Θ. Let then $\mathscr{C}$ contain all the terms that the two theories share and whose interpretation $\mathbf{M}_{\mathscr{C}}$ can thus be determined independently of each of the two theories individually. The sets of $\mathscr{B}$ and $\mathscr{A} \cup \mathscr{C}$-structures that are possible given the two theories are then given by

$$\mathbf{M}_{\mathscr{B}} = \{\mathfrak{A}_{\mathscr{B}} \mid \mathfrak{A}_{\mathscr{B}}|_{\mathscr{C}} \in \mathbf{M}_{\mathscr{C}} \text{ and } \mathfrak{A}_{\mathscr{B}} \vDash \Theta\} \tag{11.9}$$

and

$$\mathbf{M}_{\mathscr{A}\cup\mathscr{C}} = \{\mathfrak{B}_{\mathscr{A}\cup\mathscr{C}} \mid \mathfrak{B}_{\mathscr{A}\cup\mathscr{C}}|_{\mathscr{C}} \in \mathbf{M}_{\mathscr{C}} \text{ and } \mathfrak{B}_{\mathscr{A}\cup\mathscr{C}} \vDash \Lambda\}\,. \tag{11.10}$$

It may be, of course, that neither Θ nor Λ determines $\mathbf{M}_{\mathscr{C}}$ up to isomorphism, and thus there can be empirical research that determines the two theories' sets $\mathbf{N}_{\mathscr{B}}$ and $\mathbf{N}_{\mathscr{A}\cup\mathscr{C}}$ of intended structures more precisely. Such empirical research, however, can again be pursued without ever considering the conjunction of the two theories.

Barring additional empirical information, there are now two relevant possible relations of an $\mathscr{A}$-term A to a $\mathscr{B}$-formula β. For one, the possible interpretations of the term and the formula may be identical in every possible structure. In that case, $\Theta \cup \Lambda \vDash A = \beta$, and an N-reduction has been achieved. Or the sets of possible structures may be such that the set of possible interpretations of the $\mathscr{A}$-term is not disjoint from the set of possible interpretations of the $\mathscr{B}$-formula. Expressed with the help of definition 7.10, this means that for each $\mathscr{C}$-structure $\mathfrak{A}_{\mathscr{C}}$ that is possible given Θ and Λ, the sets of $\mathscr{C}$-possible extensions of β given Θ and of A given Λ are not disjoint: $\mathbf{A}^{\mathfrak{A}_{\mathscr{C}},\Theta} \cap \beta^{\mathfrak{A}_{\mathscr{C}},\Lambda} \neq \varnothing$. In this second case, there are then two ways that A and β may become identified. First, further empirical research may restrict the set of intended $\mathscr{C}$-structures $\mathbf{N}_{\mathscr{C}}$ more than both Θ and Λ. This is simply empirical research that can be undertaken independently from both Θ and Λ and leads to empirical laws expressed by a set Γ of $\mathscr{C}$-sentences.

Second, one can introduce $\mathscr{C}$-conservative bridge laws Π, $\mathscr{V}$-sentences which restrict the expansions from $\mathscr{C}$ to $\mathscr{V}$. It is only at this point that both vocabularies are needed at once, and Π only describes language conventions. Thus bridge laws, connections between $\mathscr{B}$-terms and $\mathscr{A}$-terms that do not already follow from the two theories separately or from empirical research independently of either theory, are analytic.

This view, it seems, is contrary to much of the discussion in the philosophy of science. Witness Sklar (1967, 118), who says about the classic discussion by Nagel (1961, ch. 11):

> He assumes, as I shall, that the concepts appearing in the reduced theory but absent from the reducing theory are not eliminable in terms of the concepts of the reducing theory by linguistic investigation alone. If such were the case, [. . .] then the reduction would constitute a matter of linguistic insight and perhaps clever logical inference, but it would hardly require experimental justification and observational confirmation. But the reduction of, say, physical optics to electromagnetic theory is not of this sort, but is instead an important factual discovery of empirical science.

It is indeed true that the reduction between two theories is often an important empirical discovery, but the discovery stems from empirical investigation within each science individually: The interpretation of each science's terms are determined with higher accuracy, for example by determining under what experimental setups specific phenomena occur, that is, arriving at $\mathscr{C}$-sentences Γ. The increased accuracy of the interpretation then allows a reduction of one theory's term to another theory's formula without further empirical research. It is true that scientific practice does not clearly distinguish between research within a science and research done to reduce one theory to another, but from this, one cannot infer that, logically, both theories are needed to do the empirical research. Questions of reduction may motivate specific experiments, but the results can be phrased independently within each theory.

It is another matter that empirical results Γ may *suggest* the identification of some $\mathscr{A}$-term and a $\mathscr{B}$-formula. But the choice to indeed identify the two is then again a matter of convention, given in a set Π of meaning postulates. Note also that this result does not rely on semantic empiricism: If one assumes that each theory's possible structures can be determined directly for its whole vocabulary, then the comparison of the interpretations of the theories' terms and formulas determines the bridge laws directly. And again, all that is needed is research into each theory individually, but not in the relation of the two theories. For the comparison of the theories' possible structures is a purely formal exercise once the possible structures are determined.

11.6 A note on structural realism

Going back to a suggestion by Maxwell (1970), it has often been suggested that, first, the factual content of a theory is given by its structure, and second, this structure is expressed by the theory's Ramsey sentence. I want to argue briefly that the second claim is false and that the first claim, called structural realism, is best expressed in terms of reduction.

The idea of using the Ramsey sentence $\mathsf{R}_{\mathscr{B}}(\Theta)$ to express the structure of theory $\Theta(\bar{B},\bar{A})$ is intuitively appealing. For the resulting sentence $\exists\bar{X}\Theta(\bar{B},\bar{X})$ contains no references to auxiliary terms, but still seems to contain the same logical structure that the auxiliary terms had. I have already argued in §2.9 that the problem with this idea is that a theory's Ramsey sentence does not say anything over and above the theory's basic implications. Allowing sentences of higher order logic as basic sentences, this is illustrated nicely by claim 6.7: A theory that is not semantically falsifiable also has no empirical content according to the Ramsey sentence. Another problem is that if the structural content of a theory is identified with the relations between the existentially generalized variables $\bar{X}$, the structural content is not invariant under equivalence transformations of the theory. Trotting out once more the fate of 'pain' in Papineau's toy theory of pain (2.14), an equivalent transformation (2.15a) can make its associated variable and all the relations that come with it disappear.

Looking at the standard semantics of the situation, these problems are unsurprising. In §2.9, I already noted for a special case that the existentially generalized variables do not generally correspond to elements of the models of the Ramsey sentences. For the models of Θ are structures $\mathfrak{A}$ of the type $\langle|\mathfrak{A}|,B_1^{\mathfrak{A}},\ldots,B_m^{\mathfrak{A}},A_1^{\mathfrak{A}},\ldots,A_n^{\mathfrak{A}}\rangle$. The models of $\mathsf{R}_{\mathscr{B}}(\Theta)$, on the other hand, are structures $\mathfrak{A}_{\mathscr{B}}$ of the type $\langle|\mathfrak{A}|,B_1^{\mathfrak{A}},\ldots,B_m^{\mathfrak{A}}\rangle$. The referents of all auxiliary terms, and with them their structural relations, have disappeared.[5] It is thus rather puzzling to claim (Ladyman 2009, §3.2) that

> it is a mistake to think that the Ramsey sentence allows us to eliminate theoretical entities, for it still states that these exist. It is just that they are referred to not directly, by means of theoretical terms, but by description, that is via variables, connectives, quantifiers and predicate terms whose direct referents are (allegedly) known by acquaintance.

Assuming that equivalent transformation is allowed, the Ramsey sentence states that the referents of the auxiliary terms exist only insofar $\exists x(x=x)$ states that there are at least two of such referents. For $\exists x(x=x) \models\!\!\!\dashv \exists X_1X_2(\exists xX_1x\wedge\neg\exists xX_2x)$.

It seems that the Ramsey sentence approach looks for structure in all the wrong

[5]If $\mathfrak{A}$ is a model of Θ, it is also a model of $\mathsf{R}_{\mathscr{B}}(\Theta)$, but only trivially so. So is $\langle|\mathfrak{A}|,B_1^{\mathfrak{A}},\ldots,B_m^{\mathfrak{A}},C_1^{\mathfrak{A}},\ldots,C_k^{\mathfrak{A}}\rangle$ for any k and $C_i, 1\leq i\leq k$.

places.[6] The core idea of structural realism is that a theory's structure is real, and that the objects that occur in the structures are, in one way or another, less so. In other words, the factual content of a theory abstracts from the objects of theories. This suggests that when one (approximately true) theory Θ replaces another theory Λ (which may be approximately true to a lesser degree), Λ's structure can be recovered from Θ (possibly with the help of supplementary sentences Π). To achieve that, the reducts of the models of $\Theta \cup \Pi$ to the vocabulary $\mathscr{A} \cup \mathscr{C}$ of Λ must be the models of Λ or at least a subset thereof, since Θ may be logically stronger than Λ, and especially make stronger assertions in their common vocabulary $\mathscr{C}$. Thus Θ verifies Λ relative to Π, and Λ either is the abstraction of Θ in $\mathscr{A} \cup \mathscr{C}$, or omits from it.

This explication of structural realism may seem just like realism, since some of the original structures of Λ remain, with their original domains. But this impression is mistaken because verification and abstraction are defined for higher order logic as well; thus Θ may contain the objects of the domains of Λ as relations. For instance, it may be that Λ, say, rigid body mechanics, asserts that two objects in its domain (rigid bodies) cannot be co-located, $\Lambda \vDash \forall xy(Rx \wedge Ry \wedge x \neq y \rightarrow \neg Cxy)$. Θ, say, condensed matter physics, then may contain a higher order relation R that identifies specific first order relations between the objects of its domain as rigid body relations, and assert that non-identical rigid body relations cannot have a property that corresponds to co-location, $\Theta \vDash \forall XY(RX \wedge RY \wedge X \neq Y \rightarrow \neg CXY)$. Thus what is an object in Λ is a relation in Θ, and the objects of Θ do not occur in Λ.

In principle, the relation variables X and Y can range over the superset of the Cartesian product of any domain whatsoever, as long as the abstraction of Θ to the vocabulary $\{R, C\}$ results in Λ or a stronger theory. Since in Θ, R and C may also apply to *sets* of relations (and thus be at least of third order), $X \neq Y \rightarrow \neg CXY$ may also state that there are no relations *of the same equivalence class* that are co-located. This fits nicely with the discussion of the bridges of Königsberg in §2.12.4, where the bridges can also vary widely in their microstructure without ceasing to be the same bridges, and without the problem of how to cross them losing its structure. The bridges can be seen as equivalence classes of microstructures, and the problem be phrased in terms of relations between the equivalence classes.

The objects in the domain of Λ's models are thus indeed irrelevant, since they can always turn out to be nothing but relations of an underlying theory with different objects in its domain. This, I take it, is also Carnap's idea in the *Aufbau*, §30, where he considers different types of higher order logic to be about different object spheres (see also §3.3). Especially Carnap's discussion of the difference between 'containing' (in the sense of a part-whole relationship) and 'being constructed out of' (in the way some definienda have only lower types in their definientia) should fit with structural realism (Carnap 1928a, §35–§38).

[6]But not for the reasons that Cei and French (2006) are considering.

11.7 Conclusion

There are two main results to this chapter. The first one is that many criteria of empirical significance are closely connected to different definitions of reduction. Even a comparably recent definition like that of Nickles has at its core a strengthening of the concept of translatability, and in fact can be used to generalize older concepts like those of Nagel or Kemeny and Oppenheim to include approximation.

The second main result is that the criteria of empirical significance used in most definitions of reduction are not always the strongest ones. Specifically, idiosyncratic N-reducibility relies on verifiability and K-O-reducibility relies on the comparative criterion of falsifiability. It is thus not obvious why so many discussions about reduction revolve around much stronger concepts, namely definability and supervenience. It would be interesting to investigate what goals exactly definability and supervenience achieve that verifiability fails to achieve. The same question also arises, of course, for (liberalized) F-reducibility and L-reducibility.

The prominent role of weaker criteria of empirical significance also suggests looking not at the strongest possible end result of scientific research, but rather at the beginning: If there is some—possibly very weak—law in psychology, how would one go about verifying it starting from physics or neurobiology? Or how would one determine whether a law in physics is more falsifiable than a law in psychology? These questions are arguably also the conceptually more important ones, since they require providing a first relation between the proprietary $\mathscr{A}$-terms of the to-be-reduced theory and the $\mathscr{B}$-terms of the to-be-reducing theory. It is at this point that the weakest criteria of deductive empirical significance for sentences and for terms become of interest.[7] Once this first step of a conceptual relation is made, the move to the verification of *all* psychological laws or the *complete* definition of all psychological terms may be more of the same.

[7] It remains to be seen how probabilistic relations between sentences can contribute to the meaning of the sentences' terms.

Chapter 12

Concept formation

In this chapter, my discussion will come full circle. I have started with a chapter outlining and defending artificial language philosophy, and on this basis have analyzed and developed criteria of empirical significance. Now I want to use criteria of empirical significance to further analyze and develop a core assumption of artificial language philosophy, the need for and possibility of conventional concept formation.

The core idea of artificial language philosophy is that the analytic component of a set of postulates is conventional, and that the remaining component is empirical. However, given an arbitrary postulate ϑ, it is not always obvious what the empirical (synthetic) component of ϑ is, and what the analytic component is. The most famous and most analyzed technical solution to this problem is that of the Carnap sentence, mentioned in §2.8.2.[1] The Carnap sentence approach defines the analytic component of ϑ as its Carnap sentence

$$\mathsf{C}_{\mathscr{B}}(\vartheta) = \mathsf{R}_{\mathscr{B}}(\vartheta) \rightarrow \vartheta \; . \tag{12.1}$$

Since ϑ's Ramsey sentence can be considered its synthetic content $\mathrm{Syn}(\vartheta)$, and $\mathsf{R}_{\mathscr{B}}(\vartheta) \wedge \mathsf{C}_{\mathscr{B}}(\vartheta) \Dashv\vDash \vartheta$, Carnap (1963c, §24.D) suggests ϑ's Carnap sentence as its analytic content $\mathrm{An}(\vartheta)$.

The relation between semantic falsifiability and Ramsey sentences suggests that the criteria of empirical significance discussed here are indeed criteria of *empirical* meaningfulness, not meaningfulness simpliciter. For the Carnap sentence approach can be applied to any sentence and never classifies a component as meaningless. Rather, a semantically non-falsifiable sentence is wholly analytic. Since this also holds for ostensibly metaphysical sentences, it suggest a view of metaphysics not as meaningless but as engaged in language choice, just as I argued

[1] Since I will focus on the Carnap sentence approach, which applies only to single sentences, I will not assume a set Θ of postulates but a single postulate ϑ.

in §2.7: Assuming that metaphysical claims are not falsifiable, they can be *chosen* to be true. Thus there is no problem with non-synthetic sentences *per se*; the problem rather lies in the treatment of non-synthetic sentences as synthetic. *If used* as synthetic sentences, non-synthetic sentences are *pseudo-synthetic* (Diamond 1975a, 16–20), but in the same vein, non-analytic sentences that are used as if they were analytic are *pseudo-analytic*. Thus in the dispute over the role of a criterion of empirical significance between Popper, who intended to demarcate empirical sentences (see §5.2) and Carnap, who intended to demarcate meaningful sentences, (see §8.1), Carnap's own account of analyticity speaks in favor of Popper.

However, the notion of analyticity may have to be refined, for a non-falsifiable sentence may still be verifiable. Thus while one can choose it to be true without the possibility of being mistaken, one may not be able to choose it to be false without the possibility of being mistaken. Therefore the Carnap sentence of a sentence, and other non-falsifiable sentences as well, may be verified at some point. Indeed, the Carnap sentence $\mathsf{C}_{\mathscr{B}}(\vartheta)$ is verified simply by $\neg\mathsf{R}_{\mathscr{B}}(\vartheta)$. If, in the terminology of Peacocke (1986, 47), the Ramsey sentence $\mathsf{R}_{\mathscr{B}}(\vartheta)$ is akin to the "canonical commitment of the content that" ϑ, then the negation of the reverse Ramsey sentence $\neg\mathsf{R}_{\mathscr{B}}(\neg\vartheta)$ (see §6.3) is akin to the "canonical ground for the content that" ϑ. Only a sentences that has neither commitment nor ground, that is, is neither falsifiable nor verifiable, and thus not weakly $\mathscr{B}$-determined, allows the choice of either truth value. I will not explore this line of thought further, however. Instead, I will assume that the empirical content of a sentence is given by its Ramsey sentence and focus on possible liberalizations of the Carnap sentence.

12.1 The analytic component of sentences

$\mathsf{R}_{\mathscr{B}}(\vartheta)$ and $\mathsf{C}_{\mathscr{B}}(\vartheta)$ fulfill three conditions of adequacy that Carnap (1963c, 965) suggests for any split of a sentence ϑ into an analytic component $\mathrm{An}(\vartheta)$ and a synthetic component $\mathrm{Syn}(\vartheta)$. In my terminology, they are the following:

1. $\mathrm{An}(\vartheta) \wedge \mathrm{Syn}(\vartheta) \mathrel{\models\!\!\!\dashv} \vartheta$.

2. $\mathrm{Syn}(\vartheta)$ and ϑ are deductively empirically equivalent relative to the empty set.

3. Only $\mathrm{An}(\vartheta)$ contains $\mathscr{A}$-terms; $\mathrm{An}(\vartheta)$ has no $\mathscr{B}$-content relative to the empty set.

Winnie (1970, theorem 4) shows that for a consistent sentence ϑ in a first order language without identity, only tautological $\mathscr{A}$-sentences are analytic if $\mathsf{C}_{\mathscr{B}}(\vartheta)$ is taken as the analytic component of ϑ (cf. Williams 1973, theorem 5). Under the same condition, Winnie (1970, theorem 5) shows further that beyond $\mathsf{C}_{\mathscr{B}}(\vartheta)$, $\mathsf{C}_{\mathscr{B}}(\vartheta) \wedge \mathsf{R}_{\mathscr{A}}(\vartheta)$ also fulfills the conditions for $\mathrm{An}(\vartheta)$ (cf. Williams 1973, 404–408).

In other words, instead of treating only tautological $\mathscr{A}$-sentences as analytic, one can also treat all of ϑ's $\mathscr{A}$-consequences as analytic.

Winnie (1970, 294–296) and Demopoulos (2007, V) consider this result something of a confirmation of the Quinean charge that the distinction between analytic and synthetic sentences is arbitrary (cf. Quine 1951). As a defense of Carnap's approach, Winnie (1970, 296–297) suggests, with agreement of Demopoulos (2007, V), an additional condition of adequacy on $\mathrm{An}(\vartheta)$, based on

Definition 12.1. σ is *observationally vacuous* in ϑ if and only if $\vartheta \vDash \sigma$ and for any $\mathscr{B}$-sentence β and $\mathscr{V}$-sentence τ with $\vartheta \vDash \tau$, $\tau \wedge \sigma \vDash \beta$ only if $\tau \vDash \beta$.

Winnie (1970, 296–297) and Demopoulos (2008, 371) justify definition 12.1 as similar to the notion of $\mathscr{B}$-conservativeness relative to an empty set in first order logic,[2] and point out that an observationally vacuous sentence can never contribute to the inference of a $\mathscr{B}$-sentence (Winnie 1970, 297; Demopoulos 2007, 259). This, of course, is shorthand for the claim that an observationally vacuous sentence can never contribute to the inference of a $\mathscr{B}$-sentence *from a sentence entailed by* ϑ. In fact, σ is observationally vacuous in ϑ if and only if, first, it is entailed by ϑ, and second, it is $\mathscr{B}$-conservative relative to *every sentence* entailed by ϑ. This is quite obviously a much stronger condition than, for example, $\mathscr{B}$-conservativeness relative to ϑ. Indeed, it is so strong that one might suspect that no sentence at all is observationally vacuous in ϑ, since for any $\mathscr{B}$-sentence β and $\mathscr{V}$-sentence σ entailed by ϑ, $\vartheta \vDash \sigma \rightarrow \beta$. Since $\sigma \wedge (\sigma \rightarrow \beta) \vDash \beta$, σ therefore fails to be observationally vacuous unless $\sigma \rightarrow \beta \vDash \beta$. And this holds if and only if $\vDash (\sigma \rightarrow \beta) \rightarrow \beta$, that is, $\vDash \neg\beta \rightarrow \sigma$ and thus $\neg\beta \vDash \sigma$. In other words, a sentence is observationally vacuous only if it is entailed by ϑ and by every $\mathscr{B}$-sentence that falsifies ϑ. Since a $\mathscr{B}$-sentence falsifies ϑ if and only if it is incompatible with $\mathsf{R}_{\mathscr{B}}(\vartheta)$, σ is observationally vacuous only if $\neg\mathsf{R}_{\mathscr{B}}(\vartheta) \vDash \sigma$ and $\vartheta \vDash \sigma$, that is, $\neg\mathsf{R}_{\mathscr{B}}(\vartheta) \vee \vartheta \vDash \sigma$, or simply $\mathsf{C}_{\mathscr{B}}(\vartheta) \vDash \sigma$. Winnie (1970, corollary 12) also proves the converse, so that the following holds:

Claim 12.1. $\mathsf{C}_{\mathscr{B}}(\vartheta) \vDash \sigma$ *if and only if* σ *is observationally vacuous in* ϑ.

Since Winnie demands that $\mathrm{An}(\vartheta)$ be observationally vacuous in ϑ, he thus shows that only the Carnap sentence is an adequate explication of the analytic component of ϑ. Winnie's and Demopoulos's justification for observational vacuity shows at best that the criterion is not too wide. Given the comparative strength of the criterion, however, it is much more interesting whether, first, it is too narrow and second, whether it is needed at all. Winnie and Demopoulos claim that without the demand for observational vacuity, the analytic-synthetic dichotomy is arbitrary. But this is a tendentious formulation, since $\mathrm{Syn}(\vartheta)$ is uniquely determined, and $\mathrm{An}(\vartheta)$ is somewhat, but not completely vague (all $\mathscr{B}$-creative sentences

[2]Demopoulos (2007, n. 12) also calls observational vacuity a "special case" of $\mathscr{B}$-conservativeness; but this is misleading, since the former is stronger than the latter.

are in its negative extension, and all sentences that are observationally vacuous are in its positive extension). On the other hand, the set of analytic sentences is even more vague than Winnie and Demopoulos let on, since from a formal point of view, any theory ϑ is as good as any other theory τ as long as $\mathsf{R}_{\mathscr{B}}(\vartheta) \mathrel{\models\!\!\!\!\dashv} \mathsf{R}_{\mathscr{B}}(\tau)$, so that even if $\mathrm{An}(\vartheta)$ were uniquely determined by ϑ, the analytic sentences could nonetheless be different, namely $\mathrm{An}(\tau)$. And once one accepts that the truth of $\mathscr{B}$-conservative sentences is a matter of choice, then there is no obvious reason to demand that if a theory ϑ is given, there must be no more choice with respect to $\mathrm{An}(\vartheta)$, for example between $\mathsf{C}_{\mathscr{B}}(\vartheta)$ and stronger $\mathscr{B}$-conservative implications of ϑ. But *if* one were to demand that $\mathrm{An}(\vartheta)$ has to be fixed by ϑ, $\mathsf{C}_{\mathscr{B}}(\vartheta)$ seems like the wrong criterion to determine the dichotomy, because for every sentence ϑ, some of its analytic implications will become non-analytic if our empirical knowledge increases in any way:

Claim 12.2. *For two compatible theories ϑ and τ, $\mathsf{C}_{\mathscr{B}}(\vartheta \wedge \tau) \vDash \mathsf{C}_{\mathscr{B}}(\vartheta)$ if and only if τ has no empirical content relative to ϑ.*

Proof. $\mathsf{C}_{\mathscr{B}}(\vartheta \wedge \tau) \vDash \mathsf{C}_{\mathscr{B}}(\vartheta)$ iff $\mathsf{R}_{\mathscr{B}}(\vartheta \wedge \tau) \to \vartheta \wedge \tau \vDash \mathsf{R}_{\mathscr{B}}(\vartheta) \to \vartheta$. This holds iff $\neg\mathsf{R}_{\mathscr{B}}(\vartheta \wedge \tau) \vDash \mathsf{R}_{\mathscr{B}}(\vartheta) \to \vartheta$ and $\vartheta \wedge \tau \vDash \mathsf{R}_{\mathscr{B}}(\vartheta) \to \vartheta$, where the latter is logically true. The former holds iff $\neg\mathsf{R}_{\mathscr{B}}(\vartheta \wedge \tau) \vDash \neg\mathsf{R}_{\mathscr{B}}(\vartheta)$ or $\neg\mathsf{R}_{\mathscr{B}}(\vartheta \wedge \tau) \vDash \vartheta$. The first disjunct is true iff $\mathsf{R}_{\mathscr{B}}(\vartheta) \vDash \mathsf{R}_{\mathscr{B}}(\vartheta \wedge \tau)$, that is, iff τ has no empirical content relative to ϑ. The second disjunct is true iff $\mathsf{R}_{\mathscr{B}}(\neg\vartheta) \vDash \mathsf{R}_{\mathscr{B}}(\vartheta \wedge \tau)$, which is logically false. $\square$

On the Carnap sentence approach, the introduction of any sentence into our system of beliefs that contains new empirical information thus renders some previously analytic sentence non-analytic. Mathematics, for instance, cannot be analytic for long because mathematical theories are axiomatized by auxiliary sentences. And once an auxiliary sentence α is conjoined with an empirical theory ϑ, α ceases to be analytic, since $\mathsf{C}_{\mathscr{B}}(\alpha \wedge \vartheta) \mathrel{\models\!\!\!\!\dashv} \mathsf{R}_{\mathscr{B}}(\alpha \wedge \vartheta) \to \alpha \wedge \vartheta$; if $\mathsf{R}_{\mathscr{B}}(\alpha)$ is a tautology and contains no terms of ϑ, $\mathsf{C}_{\mathscr{B}}(\alpha \wedge \vartheta) \mathrel{\models\!\!\!\!\dashv} \mathsf{R}_{\mathscr{B}}(\vartheta) \to \alpha \wedge \vartheta$. In conclusion, *if* for every sentence ϑ, there has to be a unique analytic component $\mathrm{An}(\vartheta)$, it will have to be $\mathsf{C}_{\mathscr{B}}(\vartheta) \wedge \mathsf{R}_{\mathscr{A}}(\vartheta)$. For otherwise, one could always introduce $\mathsf{R}_{\mathscr{A}}(\vartheta)$ as lone auxiliary sentence, which would become non-analytic with the addition of any empirical theory. In a sense then, the Carnap sentence approach leads to contradictory results when the gain of empirical knowledge is taken into consideration.

Since there is no obvious reason to treat all auxiliary sentences of all theories as analytic, there cannot be a unique set $\mathrm{An}(\sigma)$ for every sentence σ. Therefore, as the conventionality of the choice between ϑ and τ with $\mathsf{R}_{\mathscr{B}}(\vartheta) \mathrel{\models\!\!\!\!\dashv} \mathsf{R}_{\mathscr{B}}(\tau)$ suggests, and as Przełęcki and Wójcicki (1969, 386) note, it is another conventional choice which sentence σ that fulfills Carnap's conditions of adequacy one chooses to be analytic. I now want to look at a historically important case that will turn out to be still relevant.

12.2 Reduction pairs

In definition 7.5 of reducibility, Carnap (1936, §8) introduces the concept of a reduction sentence, which is simply a necessary or sufficient condition for relations. The problem with pairs of reduction sentences is that they are not generally $\mathscr{B}$-conservative. The conjunction of a general *reduction pair*

$$\forall x[\varphi(x) \to Px] \tag{12.2a}$$

$$\forall x[\psi(x) \to \neg Px]\ , \tag{12.2b}$$

where $P \in \mathscr{A}$ and φ and ψ are $\mathscr{B}$-formulas, entails what Carnap (1936, 451) calls the "*representative sentence*"

$$\forall x \neg[\varphi(x) \wedge \psi(x)]\ . \tag{12.3}$$

That this is also the only $\mathscr{B}$-consequence of the reduction pair can be seen by comparison with Papineau's toy theory of pain (2.14). The reduction pair's representative sentence is thus its synthetic component. As argued above, there will be different possibilities for the analytic component.

Unsurprisingly, Carnap (1963c, 964–966) suggests the Carnap sentence of the general reduction pair as its analytic component,[3] that is,

$$\forall x[\varphi(x) \to \neg\psi(x)] \to \forall x[\varphi(x) \to Px] \wedge \forall x[\psi(x) \to \neg Px]\ . \tag{12.4}$$

This is equivalent to the *Carnap reduction pair*

$$\forall x\big(\forall y[\varphi(y) \to \neg\psi(y)] \wedge \varphi(x) \to Px\big) \tag{12.5a}$$

$$\forall x\big(\forall y[\varphi(y) \to \neg\psi(y)] \wedge \psi(x) \to \neg Px\big)\ . \tag{12.5b}$$

The same result can be achieved with the Ramsey and the reverse Ramsey constant, although not immediately.

Claim 12.3. *For* $\vartheta \mathrel{\vert\!\!\!=\!\!\!\vert} \forall x[\varphi(x) \to Px] \wedge \forall x[\psi(x) \to \neg Px]$,

$$\forall x\left[\neg Я^{\vartheta}_{\mathscr{B},P} x \leftrightarrow \big(\forall y[\varphi(y) \to \neg\psi(y)] \to \varphi(x)\big)\right] \tag{12.6}$$

$$\forall x\left[\neg R^{\vartheta}_{\mathscr{B},P} x \leftrightarrow \big(\forall y[\varphi(y) \to \neg\psi(y)] \to \psi(x)\big)\right]\ . \tag{12.7}$$

Proof. I will prove the first conjunct; the proof of the second is similar. Assume for any x that $\neg Я^{\vartheta}_{\mathscr{B},P} x$ and hence $\forall X\big(\forall y[\varphi(y) \to Xy] \wedge \forall y[\psi(y) \to \neg Xy] \to Xx\big)$. When substituting $\lambda y\big(\forall z[\varphi(z) \to \neg\psi(z)] \wedge \varphi(y)\big)x$ for X, the antecedent of the implication is a logical truth, so that $\forall y[\varphi(y) \to \neg\psi(y)] \to \varphi(x) \wedge \neg\psi(x)$

[3]Maybe somewhat surprisingly, since Carnap suggested this solution before he found the general formalism of the Carnap sentence.

holds. Assuming for the converse that $\forall y[\varphi(y) \to \neg\psi(y)] \to \varphi(x) \wedge \neg\psi(x)$ and $\forall y[\varphi(y) \to Xy] \wedge \forall y[\psi(y) \to \neg Xy]$ for an arbitrary X entails $\forall y[\varphi(y) \to \neg\psi(y)]$, hence $\varphi(x) \wedge \neg\psi(x)$, and finally Xx. Since X is arbitrary, this entails $\forall X\left(\forall y[\varphi(y) \to Xy] \wedge \forall y[\psi(y) \to \neg Xy] \to Xx\right) \mathrel{\Dashv\vDash} \neg Я^{\vartheta}_{\mathcal{B},P}x$. Since x is arbitrary as well, the two formulas are equivalent. Finally, $\forall y[\varphi(y) \to \neg\psi(y)] \to \varphi(x) \wedge \neg\psi(x) \mathrel{\Dashv\vDash} \forall y[\varphi(y) \to \neg\psi(y)] \to \varphi(x)$. □

Thus, if $\neg Я^{\vartheta}_{\mathcal{B},P}x$ was used as a sufficient condition for P and $\neg R^{\vartheta}_{\mathcal{B},P}x$ was used as a sufficient condition for $\neg P$, their conjunction would be $\mathcal{B}$-creative, since both $\neg Я^{\vartheta}_{\mathcal{B},P}x$ and $\neg R^{\vartheta}_{\mathcal{B},P}x$ apply to all objects in the domain whenever $\mathsf{R}_{\mathcal{B}}(\vartheta)$ is false. To arrive at the Carnap reduction pair, the case in which $\mathsf{R}_{\mathcal{B}}(\vartheta)$ is false has to be excluded:

Corollary 12.4. *For* $\vartheta \mathrel{\Dashv\vDash} \forall x[\varphi(x) \to Px] \wedge \forall x[\psi(x) \to \neg Px]$,

$$\forall x[\mathsf{R}_{\mathcal{B}}(\vartheta) \wedge \neg Я^{\vartheta}_{\mathcal{B},P}x \to Px] \tag{12.8}$$

$$\forall x[\mathsf{R}_{\mathcal{B}}(\vartheta) \wedge \neg R^{\vartheta}_{\mathcal{B},P}x \to \neg Px] \tag{12.9}$$

are equivalent to the Carnap reduction pair for ϑ.

Proof. Immediately from claim 12.3. □

These results hold in general; as in claim 12.3, the Ramsey and reverse Ramsey constant are in general conditional on the Ramsey sentence:

Claim 12.5. *For any sentence* ϑ *and any auxiliary relation* P_i,

$$\vDash \forall \bar{x}\left(\neg R^{\vartheta}_{\mathcal{B},P_i}\bar{x} \leftrightarrow \left[\mathsf{R}_{\mathcal{B}}(\vartheta) \to \neg R^{\vartheta}_{\mathcal{B},P_i}\bar{x}\right]\right) \tag{12.10}$$

and

$$\vDash \forall \bar{x}\left(\neg Я^{\vartheta}_{\mathcal{B},P_i}\bar{x} \leftrightarrow \left[\mathsf{R}_{\mathcal{B}}(\vartheta) \to \neg Я^{\vartheta}_{\mathcal{B},P_i}\bar{x}\right]\right)\,. \tag{12.11}$$

Proof.

$$\neg R^{\vartheta}_{\mathcal{B},P_i}\bar{x} \leftrightarrow \neg\exists\bar{X}\left[\vartheta(\bar{B},\bar{X}) \wedge \neg X_i\bar{x}\right] \tag{12.12a}$$

$$\leftrightarrow \neg\left(\exists\bar{X}\,\vartheta(\bar{B},\bar{X}) \wedge \exists\bar{X}[\vartheta(\bar{B},\bar{X}) \wedge \neg X_i\bar{x}]\right) \tag{12.12b}$$

$$\leftrightarrow \neg\mathsf{R}_{\mathcal{B}}(\vartheta) \vee \neg R^{\vartheta}_{\mathcal{B},P_i}\bar{x} \tag{12.12c}$$

$$\leftrightarrow \left[\mathsf{R}_{\mathcal{B}}(\vartheta) \to \neg R^{\vartheta}_{\mathcal{B},P_i}\bar{x}\right] \tag{12.12d}$$

The proof for $\neg Я^{\vartheta}_{\mathcal{B},P_i}\bar{x}$ is analogous. □

To arrive at a $\mathscr{B}$-conservative reduction pair, it thus has to be assumed that the Ramsey sentence of the theory is true. Call

$$\forall \bar{x}\left[\mathsf{R}_{\mathscr{B}}(\vartheta) \wedge \neg \mathcal{A}^{\vartheta}_{\mathscr{B},P_i}\bar{x} \rightarrow P_i\bar{x}\right] \tag{12.13}$$

$$\forall \bar{x}\left[\mathsf{R}_{\mathscr{B}}(\vartheta) \wedge \neg R^{\vartheta}_{\mathscr{B},P_i}\bar{x} \rightarrow \neg P_i\bar{x}\right] \tag{12.14}$$

the *Martin reduction pairs* for ϑ and P_i, since Martin (1966) has suggested the notion of the Ramsey constant. Martin reduction pairs are entailed by a theory's Carnap sentence:

Claim 12.6. *For any sentence ϑ and any auxiliary relation P_i, $\mathsf{C}_{\mathscr{B}}(\vartheta)$ entails the Martin reduction pair for every auxiliary relation P_i.*

Proof. Trivially, $\forall \bar{x}(\neg \mathcal{A}^{\vartheta}_{\mathscr{B},P_i}\bar{x} \rightarrow \neg \mathcal{A}^{\vartheta}_{\mathscr{B},P_i}\bar{x})$, and thus by lemma 7.10, $\vartheta \vDash \forall \bar{x}(\neg \mathcal{A}^{\vartheta}_{\mathscr{B},P_i}\bar{x} \rightarrow P_i\bar{x})$. Therefore $\mathsf{R}_{\mathscr{B}}(\vartheta) \rightarrow \vartheta \vDash \mathsf{R}_{\mathscr{B}}(\vartheta) \rightarrow \mathsf{R}_{\mathscr{B}}(\vartheta) \wedge \forall \bar{x}(\neg \mathcal{A}^{\vartheta}_{\mathscr{B},P_i}\bar{x} \rightarrow P_i\bar{x}) \vDash \forall \bar{x}(\mathsf{R}_{\mathscr{B}}(\vartheta) \rightarrow [\mathsf{R}_{\mathscr{B}}(\vartheta) \wedge \neg \mathcal{A}^{\vartheta}_{\mathscr{B},P_i}\bar{x} \rightarrow P_i\bar{x}]) \vDash \forall \bar{x}[\mathsf{R}_{\mathscr{B}}(\vartheta) \wedge \neg \mathcal{A}^{\vartheta}_{\mathscr{B},P_i}\bar{x} \rightarrow P_i\bar{x}]$. The proof for $R^{\vartheta}_{\mathscr{B},P_i}\bar{x}$ is analogous. □

When all $\mathscr{A}$-terms of a sentence ϑ are relations, the conjunction of all their Martin reduction pairs with $\mathsf{R}_{\mathscr{B}}(\vartheta)$ is not always equivalent to ϑ, and thus the Martin reduction pairs cannot always be the *only* analytic components of a sentence. But this is, of course, no argument against taking the Martin reduction pairs as analytic.

Instead of Carnap reduction pairs, Przełęcki (1961b, 136) suggests the stronger *Przełęcki reduction pair*

$$\forall x([\varphi(x) \wedge \neg \psi(x)] \rightarrow Px) \tag{12.15a}$$

$$\forall x([\psi(x) \wedge \neg \varphi(x)] \rightarrow \neg Px)\ , \tag{12.15b}$$

because together with ϑ's Ramsey sentence, they meet all of Carnap's conditions of adequacy for the analytic component of a sentence and, in contradistinction to Carnap's solution, allow the relation P to stay reducible even if the representative sentence of the general reduction pair (12.2) turns out false.[4] Intuitively, φ becomes and indicator for P that is defeasible by ψ, and ψ becomes and indicator for $\neg P$ that is defeasible by φ. Przełęcki reduction pairs can thus be useful in cases where a theory turns out false, but its concepts are still considered to be at least in part helpful.

A particularly intuitive way of arriving at the Przełęcki reduction pair is to think of it as a restriction of the concepts of a theory to the domain in which the theory is true. Even if $\forall x \neg[\varphi(x) \wedge \psi(x)]$ is false, there may be some objects a

[4]Since Przełęcki reduction pairs are logically stronger than Carnap reduction pairs, they cannot fulfill Winnie's condition of adequacy.

in the domain for which $\neg[\varphi(a) \wedge \psi(a)]$ is true. The restriction to these objects, that is, the relativization of the general reduction pair (12.2) to $\lambda x \neg[\varphi(x) \wedge \psi(x)]$ results in

$$\forall x \big([\neg\varphi(x) \vee \neg\psi(x)] \rightarrow [\varphi(x) \rightarrow Px] \big) \tag{12.16a}$$

and

$$\forall x \big([\neg\varphi(x) \vee \neg\psi(x) \rightarrow [\psi(x) \rightarrow \neg Px] \big) \,, \tag{12.16b}$$

which is equivalent to the Przełęcki reduction pair. In some contexts, Przełęcki reduction pairs are a significant improvement over Carnap reduction pairs, as the following case study from ethics shows.

12.3 Przełęcki reduction pairs in ethics

Mark Alfano (2009) argues that the response-dependence theory of Prinz and others and the fitting-attitudes theory first articulated by Brentano are false because they imply empirically false statements. He further concludes that these statements cannot be avoided by revising the definitions of the terms 'good' and 'bad' used in the two theories. In this section, I strengthen Alfano's first conclusion by arguing that the two theories are false even if they imply empirically true but analytically contingent statements, and show how, contrary to his second conclusion, the theories can avoid both empirically false and analytically contingent implications.[5]

12.3.1 The case against response-dependence and fitting-attitude semantics

An empirical inconsistency

Response-dependence and fitting-attitude theory contain explicit definitions for 'good' and 'bad'. In response-dependence theory, something is good (bad) if and only if someone is disposed to have a positive (negative) attitude towards it upon careful reflection (Alfano 2009, 3); in fitting-attitude theory, something is good (bad) if and only if it would be fitting to take an approbative (disapprobative) attitude towards it (Alfano 2009, 8). Since 'good' and 'bad' are polar predicates, they are contraries, that is, nothing is both good and bad.

The two theories share a structure with all theories that contain explicit definitions of contraries. It can be expressed as the simple theory η (Alfano 2009, 1):

$$\forall x [Fx \leftrightarrow \beta(x)] \tag{12.17}$$

$$\forall x [\text{un-}Fx \leftrightarrow \omega(x)] \tag{12.18}$$

$$\neg \exists x (Fx \wedge \text{un-}Fx) \,. \tag{12.19}$$

[5] A slightly different version of this section has been published in the *Journal of Ethics and Social Philosophy* (Lutz 2010). I thank an anonymous referee for the journal for helpful comments.

Postulates (12.17) and (12.18) are simply the explicit definitions of 'good' (F) and 'bad' (un-F). Postulate (12.19) is the claim that the two defined terms are contraries.

If the definientia of the two terms are not themselves contraries, then the world can turn out to be such that they are co-instantiated. The core of Alfano's article (Alfano 2009, 4–7, 9–10) is a defense of the claim that the definientia in both theories are indeed co-instantiated, that is, for both theories it holds

$$\exists x[\beta(x) \wedge \omega(x)] . \tag{12.20}$$

This claim is incompatible with η; η is therefore empirically false.

Alfano further argues that none of the six possible responses to the discovery of the inconsistency is tenable:

(I) Dialetheism is too high a price to pay (Alfano 2009, 2).

(II) Giving up postulate (12.19) means that some things are both good and bad, which is not better for an ethical theory than outright dialetheism (Alfano 2009, 7).

(III) Giving up postulate (12.17) or (12.18) results in an ethical theory that has nothing to say about both 'good' and 'bad' (Alfano 2009, 7).

(IV) Changing the definitions so that the empirical claim (12.20) seems false leaves the ethical theory at least in principle vulnerable to empirical refutations (Alfano 2009, 2).

(V) One could make one of the two defined terms into a trouser-word, that is, introduce a new definition

$$\forall x\big(Fx \longleftrightarrow [\beta(x) \wedge \neg\text{un-}Fx]\big) \tag{12.21}$$

or a new definition

$$\forall x\big(\text{un-}Fx \longleftrightarrow [\omega(x) \wedge \neg Fx]\big) , \tag{12.22}$$

but the choice between them is arbitrary and therefore *ad hoc*: There is no plausible argument for 'good' being prior to 'bad' or vice versa (Alfano 2009, 8).

(VI) Changing one of the two definitions to yield contradictories, that is, change postulate (12.17) into $\forall x(Fx \longleftrightarrow \neg\text{un-}Fx)$ or postulate (12.18) into $\forall x(\text{un-}Fx \longleftrightarrow \neg Fx)$ is even less plausible than the introduction of a trouser-word (Alfano 2009, 7).

Thus Alfano can conclude that neither theory can be saved.

A conceptual inconsistency

Alfano's rebuttal IV suggests a way of strengthening his argument that the two theories are false. Specifically, it is not necessary to establish that (12.20) is true, only that its contradictory,

$$\neg\exists x[\beta(x) \wedge \omega(x)]\,, \tag{12.23}$$

is analytically contingent. It may be only an empirical (but not an analytic) truth, for example, that the disposition to have a positive sentiment is contrary to a disposition to have a negative sentiment. Similarly, it might not be an analytic truth that a fitting approbative attitude is contrary to a fitting disapprobative attitude.[6] So while definitions (12.17) and (12.18) ensure that (12.23) entails (12.19), one could argue analogously to a consideration by Rabinowicz (2008, 40) that these definitions are not satisfactory as complete reductions of F and un-F because they would reduce the analytic truth (12.19) to the non-analytic truth (12.23).[7] But the problem is more severe. For assume that (12.23) is not analytically true. Response-dependence and fitting-attitudes theories are meant to capture the concepts *good* and *bad*; in other words, definitions (12.17) and (12.18) are analytically true. Since postulate (12.19) is analytically true and, in connection with (12.17) and (12.18), entails (12.23), (12.23) is analytically true as well, which contradicts the assumption.[8]

* * *

Alfano's argument against response-dependence and fitting-attitude theory consists of the identification of an empirical implication of η, the argument that this implication is false, and the claim that the concepts F and un-F cannot be changed to avoid empirical implications. On this abstract level, his argument can be connected to the preceding discussion of reduction pairs.

[6] Note that Alfano (2009, §3) argues that it is sometimes fitting to have both an approbative and a disapprobative attitude, even though for his argument, he only needs to establish that it is sometimes both fitting to have an approbative attitude and fitting to have a disapprobative attitude. The latter claim is the correct paraphrase of Alfano's formula (20) (Alfano 2009, 9); the former is Alfano's paraphrase.

[7] I thank an anonymous referee for *Journal of Ethics and Social Philosophy* for pointing out this problem and its discussion by Rabinowicz.

[8] Since all and only analytic truths are analytically necessary, the argument can be expressed (using "$\Box$" for "it is analytically necessary that") as follows: $\{\Box\forall x[Fx \leftrightarrow \beta(x)], \Box\forall x[\text{un-}Fx \leftrightarrow \omega(x)], \Box\neg\exists x(Fx \wedge \text{un-}Fx)\} \vDash \Box\neg\exists x[\beta(x) \wedge \omega(x)]$, which contradicts $\neg\Box\neg\exists x[\beta(x) \wedge \omega(x)]$, i. e., the assumption that (12.23) is not analytically necessary. Since for analytic necessity $\exists x[\beta(x) \wedge \omega(x)] \vDash \neg\Box\neg\exists x[\beta(x) \wedge \omega(x)]$ and $\neg\Box\neg\exists x[\beta(x) \wedge \omega(x)] \nvDash \exists x[\beta(x) \wedge \omega(x)]$ hold, the assumption is weaker than what Alfano needs to establish for his argument.

12.3.2 Isolating the empirical content of ethical theories

η can be equivalently rewritten as a necessary and a sufficient condition for F

$$\forall x[\beta(x) \to Fx] \tag{12.24a}$$

$$\forall x\big([\neg\beta(x) \lor \omega(x)] \to \neg Fx\big)\ , \tag{12.24b}$$

and a necessary and a sufficient condition for un-F:

$$\forall x[\omega(x) \to \text{un-}Fx] \tag{12.25a}$$

$$\forall x\big([\neg\omega(x) \lor \beta(x)] \to \neg\text{un-}Fx\big)\ . \tag{12.25b}$$

As Alfano's objection to response V shows, he assumes that β and ω do not contain F or un-F; the conditions (12.24) and (12.25) are therefore reduction pairs. Both reduction pairs' representative sentence is just the negation of Alfano's empirical claim (12.20). If the analytic component of η *had* to be its Carnap sentence, the story would end here, since (assuming Alfano is right) η is false, and thus F and un-F are completely undetermined by their respective Carnap reduction pairs. Przełęcki reduction pairs, however, still determine F and un-F up to a point. Applied to the two reduction pairs (12.24) and (12.25), the Przełęcki reduction pairs are

$$\forall x\big([\beta(x) \land \neg\omega(x)] \to Fx\big) \tag{12.26a}$$

$$\forall x[\neg\beta(x) \to \neg Fx] \tag{12.26b}$$

and

$$\forall x\big([\omega(x) \land \neg\beta(x)] \to \text{un-}Fx\big) \tag{12.27a}$$

$$\forall x[\neg\omega(x) \to \neg\text{un-}Fx]\ . \tag{12.27b}$$

By design of Przełęcki's general solution, neither of the new reduction pairs has empirical implications, and together with the representative sentence (12.23) of the original reduction pairs (12.24) and (12.25), these new reduction pairs are equivalent to (12.24) and (12.25), and therefore to η.

In the discussion above, the new postulates (12.26) and (12.27) for F and un-F were obtained by first equivalently reformulating η so that un-F does not appear in the postulates for F and vice versa, and then applying Przełęcki's solution to both polar predicates. But there are different ways to equivalently reformulate the theory, and another one leads to the explicit definitions

$$\forall x\big(Fx \leftrightarrow [\beta(x) \land \neg\omega(x)]\big) \tag{12.28}$$

and

$$\forall x\big(\text{un-}Fx \leftrightarrow [\omega(x) \land \neg\beta(x)]\big) \tag{12.29}$$

conjoint with η's empirical claim (12.23). Since explicit definitions are $\mathscr{B}$-conservative (Belnap 1993), they do not have empirical implications. They also entail (12.19) (the postulate that F and un-F are contraries) and the reduction pairs (12.26) and (12.27), so the conjunctions of (12.19) with the new definitions and (12.19) with the new reduction pairs do not have empirical implications either.

Note that this solution is only possible because η already entails explicit definitions for F and un-F, and that, for instance, the original explicit definitions (12.17) and (12.18) cannot be used in the same way. This is a very good example of how the choice of the analytic sentences within a theory (η in this case) has to be dependent on factors outside the pure formalism of the theory, and a nice illustration of van Fraassen's point (van Fraassen 1980, §3.5) that equivalent axiomatizations of a theory can be very different when it comes to the generalization of concepts (see §4.1.1).

12.3.3 New postulates for 'good' and 'bad'

The new reduction pairs (12.26), (12.27) and the new definitions (12.28), (12.29) avoid all of Alfano's criticisms: First and foremost, the postulates do not have any empirical implications, that is, neither are they incompatible with Alfano's core claim (12.20), nor can they be shown false by any other empirical result (response IV). The postulates do not assume dialetheism (response I) because they are compatible with each other and with the claim (12.19) that F and un-F are contraries. Specifically, the postulates do not give up postulate (12.19) (response II). They do not force an arbitrary choice between taking 'good' to be prior to 'bad' or vice versa (response V) because the postulates for F do not contain un-F and vice versa; furthermore, the changes to the original definitions are analogous for the two predicates. The postulates also do not make the polar predicates into contradictories (response VI).

One might criticize the new reduction pairs (12.26), (12.27) for failing to fully address rebuttal III. This is because for some objects, it is not defined whether they are F or not F (un-F or not un-F)—or in this case, good or not good (bad or not bad). There are two responses. The first is to bite the bullet and accept that 'good' and 'bad' are vague; just as some things are good, some are bad, and some are neither, some things are clearly good, some things are clearly not good, and for some things, it is not clear whether they are good or not good. This may just be a fact about the predicates, but it may also be a lack of knowledge: Alfano objects to a simple switch to definientia β and ω for which postulate (12.20) is false (response IV) because such a change may lead to empirical objections in the future. In light of this worry, it may simply be cautious to keep the predicates undefined for cases in which not enough is known.

For response-dependence theory, this response means that it is either a fact of language or of our knowledge that whenever someone is disposed to have a positive attitude towards x and someone is disposed to have a negative attitude towards x, it is not clear whether x is good, bad, not good, or not bad. For fitting-attitude theory, this situation occurs if it would both be fitting to take an approbative and be fitting to take a disapprobative attitude towards x. In this response, β and ω become defeasible indicators of goodness and badness, respectively. When the indicators for goodness and badness both apply to the same instance, all bets are off.

The first response is unsatisfactory if fitting-attitude and response- dependence theories are intended to provide complete reductions of 'good' and 'bad' because the two reduction pairs do not entail that 'good' and 'bad' are contraries, and therefore the purported reductions fail to reproduce all of the predicates' properties. This problem is solved by the second response, which consists in using the explicit definitions (12.28) and (12.29) instead of (12.26) and (12.27). This response leads to a very exclusive notion of 'good' and 'bad'. In response-dependence theory, it means that if anyone is disposed to have a negative sentiment toward x, it is not good, and if anyone is disposed to have a positive sentiment toward x, it is not bad. The situation in fitting-attitudes theory is analogous.

The new postulates also avoid the conceptual inconsistency to which the original definitions (12.17) and (12.18) give rise if claim (12.23) is not an analytic truth. The new definitions (12.28) and (12.29) logically entail postulate (12.19) independently of the status of (12.23); this makes (12.19) an analytic truth, as intended. The reduction pairs (12.26) and (12.27) do not entail (12.19), but are consistent with (12.19) being analytically true while (12.23) is analytically contingent, because the conjunction of (12.19), (12.26) and (12.27) is conservative. On the other hand, claim (12.23) entails $\forall x\big([\beta(x) \wedge \neg\omega(x)] \leftrightarrow \beta(x)\big)$ and $\forall x\big([\omega(x) \wedge \neg\beta(x)] \leftrightarrow \omega(x)\big)$. So if (12.23) is analytically true, then for all x, $\beta(x) \wedge \neg\omega(x)$ is equivalent to $\beta(x)$ and $\omega(x) \wedge \neg\beta(x)$ is analytically equivalent to $\omega(x)$. Therefore substituting the new postulates for the original definitions (12.17) and (12.18) does not lead to a change of the analytic truths.

* * *

Without the assumption of Alfano's empirical claim (12.20), the argument of this section works as follows: Response-dependence and fitting-attitudes theories have a structure η that entails (12.23), which either is or is not an analytic truth. If (12.23) is *not* an analytic truth, then the original definitions (12.17) and (12.18) cannot both be analytic truths. Since response-dependence and fitting-attitudes theories claim the analytic truth of (12.17) and (12.18), they are false. The new reduction pairs (12.26) and (12.27) and the new definitions (12.28) and (12.29) can be analytically true even if (12.23) is not, and insofar as they avoid all of Alfano's rebuttals I–VI, they are acceptable substitutes for the original definitions (12.17) and (12.18). If (12.23) is an analytic truth, substituting the new postulates for the original definitions does not amount to a change of the meaning postulates. In either case, then, substitution of the new postulates is acceptable. In some cases, it may even save response-dependence and fitting-attitudes theories from contradiction.

Alfano concludes that response-dependence and fitting-attitudes theories are false because they have false empirical implications, and that neither theory can be salvaged. I have argued instead, first, that both theories are false if these implications are analytically contingent (even if they are in fact true), and second, that the two theories can be saved. Specifically, by adopting new postulates for the polar predicates 'good' and 'bad' that either entail or are consistent with the postulate that the two predicates are contraries, the theories can avoid any non-analytic implications. Therefore, even if the theories are false in their current formulation, the conclusion should not be that they are completely misguided, but only that their postulates for 'good' and 'bad' require modification.

The specific case discussed here points to some general strategies for developing postulates for value notions. If all postulates can be expressed as reduction pairs, empirical and analytic inconsistencies can be precluded by ensuring that the representative sentences are tautologies. In general, postulates for value notions must be $\mathscr{B}$-conservative if the $\mathscr{B}$-terms are empirical. If value notions are to be completely reduced, then the analytic truths holding between them must be entailed by the reducing postulates (whether they are explicit definitions, reduction pairs or of some other form).

12.4 Concluding remarks on language choice

The discussion of the two ethical theories also points to a meta-philosophical conclusion: The success of the methods I have employed to arrive at new postulates for 'good' and 'bad' shows that some results from the theory of concept formation are applicable outside of their original domain. This is unsurprising, for the theory of definition has long been considered to be analogous to the theory of inference (see, for example Belnap 1993), and so methods to improve concepts, just as methods to improve arguments, should be expected to be useful in a wide variety of cases.

In particular, the discussion in this chapter has shown how valuable criteria of empirical significance are when they are understood as tools for concept formation. Here, even the criteria for terms can have applications, in spite of their lack of justification as demarcation criteria for sentences with empirical import. The discussion has also shown once more that synthetic claims do not determine analytic claims, but that they can *suggest* them. Whether, for example, general reduction pairs are made $\mathscr{B}$-conservative using Carnap or Przełęcki reduction pairs is a matter of choice, but will be influenced by the truth or falsity of the Ramsey sentence of the general reduction pair. The choice between the different reduction pairs is only a special case of the more general choice for postulates for auxiliary terms.

However, the choice may not be quite as unrestricted as I have made it out to be in this chapter. For the Carnap sentence approach to analyticity is built on the assumption that there is no empirical difference between theories with the same Ramsey sentence. But I have also shown that the Ramsey sentence, first, only determines what a theory asserts, not what can verify the theory. Second, the Ramsey sentence only determines what a theory asserts *deductively*, not probabilistically. Therefore sentences that do not differ in their Ramsey sentences can differ in their deductive verifiability or in their probabilistic assertions or probabilistic verifiability. I have mentioned in §9.2 Hempel's point that non-observational statements are necessary for "inductive explanatory and predictive use and [. . .] systematic economy and heuristic fertility" (Hempel 1965g, 222). And while systematic advantages of one set of meaning postulates over another may already show in deductive settings (I have given possible examples in §2.9), the analysis of the inductive advantage of one set of sentences over another will in all probability require a probabilistic approach. A truly analytic sentence may thus have to lack falsifiability, verifiability, probabilistic relevance, and Bayesian confirmability.[9] I hope that the analysis of the different criteria of empirical significance presented here will contribute to the development of corresponding criteria of analyticity.

However, even though I do expect the set of analytic sentence to be smaller than the set of non-falsifiable ones, I doubt that all sentences in theories will turn out empirically significant. For one, a first connection has to be made between $\mathscr{B}$ and $\mathscr{A}$ terms, and this connection will at least be in part free of assertions and also be unverifiable. Thus I expect the search for a criterion that distinguishes between metaphysics and science to never recover. But I also think that there would be fewer debates about pseudo-synthetic sentences if the distinction between synthetic and analytic sentences was considered more thoroughly (and accepted in the first place).

I want to end with a by now hopefully obvious note on maybe the most perseverant criticism of all criteria of empirical significance, that they show their own meaninglessness. There are two conjuncts to the reply. The first is that the criteria I have discussed here are formulated in one language, and a criterion for sentences about that object language will be a *new* criterion in a metalanguage. The second conjunct is that a criterion in the metalanguage should show that the criteria in the object language are not empirically significant, since as explications, they are suggestions for a language choice. This is essentially the answer given by Popper (cf. Horgan 1996, 38–39),[10] and Carnap (Stein 1992, 278–279; cf. Reichenbach 1951, 48–49).

[9] Raatikainen (2011) provides a somewhat more detailed argument that the Carnap sentence cannot be taken as a theory's analytic content when the context allows inductive inferences.

[10] Horgan's account is neither entirely clear nor entirely trustworthy, however (cf. Hoffman 1998).

Epilogue

Criteria of empirical significance have always been linked to the notion of analyticity, for example by the definition of meaningless sentences as being neither empirically significant nor analytic. I have done my best to strengthen this link by arguing for the definition of analytic sentences as empirically non-significant. My discussion was restricted almost entirely to languages of predicate logic and probability theory, and while generalizations to other formal languages will be straightforward in some cases, the step to natural language will be, without a doubt, difficult. Nonetheless, just as the study of formal inference can improve one's precision in natural language arguments (many of the informal proofs in the preceding chapters in fact grew out of more cumbersome formal proofs), it may be possible to improve one's ability to spot non-empirical claims by studying formal criteria of empirical significance, and even arrive at some good explicit indicators (cf. Bohnert 1963, 421–422). Unclear cases will have to be resolved by conscious language choice, which may sometimes involve straightforward formalization of the respective concepts. In the end, then, there does not seem to be a principled problem with empirical significance, analyticity, or artificial language philosophy in general. Says Mates (1951, 533–534):

> I do not find, in the considerations set forth by Quine [(1951)] and White [(1950)], any basis for pessimism about the explicability of 'analytic'. The task is difficult, but progress has certainly been made by Carnap and others; in any case, a difficult task is not necessarily an impossible one. There is great value in the searching criticism which White and Quine have devoted to the notion of analyticity; all distinctions and terminology should undergo this treatment regularly. Experience shows that our favorite philosophical concepts usually do not stand up very well under such scrutiny. At the same time, it is wise, even in the face of adverse results, to be hesitant about exchanging our old notions and distinctions for new ones which may be even less satisfactory. In the present case, despite all the unclarity connected with the term 'analytic' and its various associates, namely 'true', 'valid', 'necessary', 'synonymous', and others, it is difficult

> to see what is to be gained by exchanging these for the businessman vocabulary of pragmatism. How, for example, are we to decide whether a belief "pays" or is "expedient" or is "fruitful"? What does it mean to say that experience is "recalcitrant" to our system of beliefs, or that this latter is "a man-made fabric which impinges on experience only along the edges"?

Of course, artificial language philosophy also relies on the notions of expedience and fruitfulness, but only for statements whose truth or falsity cannot be decided in any other way. It could be said to be an attempt at arriving at objective, or at least intersubjective, decision procedures for as many questions as possible (empirical and formal ones), while acknowledging a remainder of questions that can only be answered by *fiat*. Those answers, however, are not ascribed to any ineffable sense of truth or intuition, but to simple choice.

Quine in essence argued that there was something fundamentally wrong with this view, which would justify his introduction of comparably hazy concepts like man-made fabrics impinging on experience—a hazy concept is still better than a fundamentally wrong one, or one based on a fundamentally wrong idea. Luckily, Stein (1992, 278–279) reports a discussion between Carnap and Quine on exactly this topic:[11]

> Carnap's summary of the issue between Quine and himself was on the following lines: "Quine", he said (I am not quoting verbatim, but giving the gist as I remember it), "and I really differ, not concerning any matter of fact, nor any question with cognitive content, but rather in our respective estimates of the most fruitful course for science to follow. Quine is impressed by the continuity between scientific thought and that of daily life—between scientific language and the language of ordinary discourse—and sees no philosophical gain, no gain either in clarity or in fruitfulness, in the construction of distinct formalized languages for science. I concede the continuity, but, on the contrary, believe that very important gains in clarity and fruitfulness are to be had from the introduction of such formally constructed languages. This is a difference of opinion which, despite the fact that it does not concern (in my own terms) a matter with cognitive content, is nonetheless in principle susceptible of a kind of rational resolution. In my view, both programs—mine of formalized languages, Quine's of a more free-flowing and casual use of language—ought to be pursued; and I think that if Quine and I could live, say, for two hundred years, it would be possible at the end of that time for us to agree on which of the two programs had proved more successful".

[11]The occasion was Quine's presentation of his unpublished paper "Ontology and Analyticity" at the University of Chicago in February 1951 (Creath 1991, 364).

> [A]s I recall, Quine happily assented to Carnap's diagnosis. [...] I have never understood why Quine continued to argue his case against Carnap with no suggestion that the issue concerned the fruitfulness of a program, and not the tenability—or intelligibility—of a doctrine.

Quine's seeming reversion to an out-and-out critical stance towards artificial language philosophy suggests that this discussion was another instance of the "persuasive and holistically critical" style of philosophical argument that Rozeboom identified (§3.10). Once the possibility of *choosing* between two programs is recognized, however, the possibility for collaboration becomes obvious.

I have already argued that traditional and ordinary language philosophy can be captured in artificial language philosophy, and that naturalized and experimental philosophy can provide suggestions for language choice (§2.7). I now want to point out that Quine's metaphor of the "totality of our so-called knowledge or beliefs" as a "a man-made fabric which impinges on experience only along the edges" can be explicated (or even precisified) by taking the experiences to be expressed in basic terms or, more generally, by basic sentences. Then 'impinging on experience' may be precisified as empirical significance, and 'man-made fabric' as the postulates expressing our beliefs. Quine (1960, 35–36) would later suggest, roughly, that experiences are expressed in *occasion sentences*. On this basis, Quine (1960, 64) argues:

> The significant trait of [non-occasion] sentences is that experience is relevant to them largely in indirect ways, through the mediation of associated sentences. Alternatives emerge: experiences call for changing a theory, but do not indicate just where and how.

As a feature of non-basic sentences, this can be read directly off the Carnap sentence: If a theory's basic assertions are false, its auxiliary vocabulary is completely undetermined. And whether the analytic component of a theory is to be taken as stronger than its Carnap sentence is not determined by the theory itself. Arguably, then, Quine's stance is that of artificial language philosophy.

It is thus unsurprising that the two effects that according to Quine (1951, 20) come from abandoning the dogma of an analytic-synthetic distinction, the "blurring of the supposed boundary between speculative metaphysics and natural science" and the "shift toward pragmatism" are also effects of embracing the distinction: I have argued at length for the importance of pragmatic considerations in concept formation (§2.3). And if the sciences rely on the same methodology as artificial language philosophy (§2.10) and artificial language philosophy can capture the practices of traditional philosophy (§2.7), the distinction between traditional philosophy and science becomes blurred. Specifically, the position of ontological naturalism becomes ill-defined (§2.13).

In artificial language philosophy, the final arbiter between different empirically equivalent methodologies is their fruitfulness. I have found artificial language philosophy as suggested by Carnap and the formalism suggested by Przełęcki to be a flexible and enlightening method of thinking about philosophical problems, analyzing other people's solutions, and coming up with my own. I hope the preceding results show this, and thus provide a cumulative argument for artificial language philosophy, syntactic approaches, the Received View, and the different criteria of empirical significance.

Bibliography

Concordance		
Short title	Title	Reference
Aufbau	*Der logische Aufbau der Welt*	Carnap (1928a)
"Aufgabe"	'Über die Aufgabe der Physik'	Carnap (1923)
"Autobiography"	'Intellectual autobiography'	Carnap (1963a)
Begriffsbildung	*Phyikalische Begriffsbildung*	Carnap (1926)
"Beobachtungssprache"	'Beobachtungssprache und theoretische Sprache'	Carnap (1958)
Concept Formation	*Fundamentals of Concept Formation in Empirical Sciences*	Hempel (1952)
Der Raum	*Der Raum. Ein Beitrag zur Wissenschaftslehre*	Carnap (1922)
"Dilemma"	'The theoretician's dilemma'	Hempel (1958)
"Dreidimensionalität"	'Dreidimensionalität des Raumes und Kausalität'	Carnap (1924)
Introduction	*Philosophical Foundations of Physics: An Introduction to the Philosophy of Science*	Carnap (1966)
L&M	*Foundations of Logic and Mathematics*	Carnap (1939)
Probability	*Logical Foundations of Probability*	Carnap (1962)
"Testability"	'Testability and meaning'	Carnap (1936)
"Theoretical concepts"	'The methodological character of theoretical concepts'	Carnap (1956b)

Works are cited on the pages indicated between square brackets.

Achinstein, P. (1963–1964). Theoretical terms and partial interpretation. *The British Journal for the Philosophy of Science*, 14:89–105. [217, 271]

Achinstein, P. (1968). *Concepts of Science*. Johns Hopkins Press, Baltimore. [217, 218]

Achinstein, P., Barbour, I. G., Brodbeck, M., Buck, R., Cornman, J. W., Craig, W., Feigl, H., Feyerabend, P., Grünbaum, A., Hanson, N. R., Hempel, C. G., Hesse, M., Hill, E. L., McMullin, E., Maxwell, G., Rozeboom, W. W., Salmon, W., Van Vliet, C. M., and Williams, F. M. (1970). Discussion at the conference on correspondence rules. In Radner and Winokur (1970). [151]

Achinstein, P., Bromberger, S., Causey, R. L., Hempel, C. G., Putnam, H., and Suppes, P. C. (1974). Discussion of "Formulation and formalization of scientific theories". In Suppe (1974b), pages 255–265. Discussion of the presentation by Hempel (1974). [134, 146, 158]

Alfano, M. (2009). A danger of definition: Polar predicates in metaethics. *Journal of Ethics & Social Philosophy*, 3(3). [400, 401, 402]

Andreas, H. (2010). Semantic holism in scientific language. *Philosophy of Science*, 77(4):524–543. [57, 58, 59, 157]

Anellis, I. H. (1996). Reply to query: How old is first-order logic? *Modern Logic*, 6(3):313–314. [102]

Angelides, A. (2004). The last collapse? An essay review of Hilary Putnam's *The Collapse of the Fact/Value Dichotomy and Other Essays*. *Philosophy of Science*, 71(3):402–411. [182]

Anonymous (2010a). Definition of intelligent design. http://www.intelligentdesign.org/whatisid.php. Archived at http://www.webcitation.org/5tZWKxwwV. [347]

Anonymous (2010b). Is intelligent design a scientific theory? http://www.intelligentdesign.org/whatisid.php. Archived at http://www.webcitation.org/5tZWKxwwV. [347]

Ayer, A. J. (1936). *Language, Truth and Logic*. Victor Gollanz, London, 1st edition. References are to the second edition (Ayer 1946). [223, 225, 227, 247, 265, 317, 329, 339]

Ayer, A. J. (1946). *Language, Truth and Logic*. Victor Gollanz, London, 2nd edition. [226, 259, 414]

Bailer-Jones, D. M. (2003). When scientific models represent. *International Studies in the Philosophy of Science*, 17(1). [133]

Barker-Plummer, D. (2011). Turing machines. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, Stanford, spring 2011 edition. [252]

Batterman, R. W. (2002). *The Devil in the Details. Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford Studies in Philosophy of Science. Oxford University Press, Oxford and New York. [73, 74, 386]

Beatty, J. (1980). What's wrong with the received view of evolutionary theory? In *Proceedings of the Biennial Meeting of the Philosophy of Science Association. Volume Two: Symposia and Invited Papers*, pages 397–426. Philosophy of Science Association, University of Chicago Press. [101, 107, 141, 208]

Beauchamp, T. L. and Childress, J. F. (2008). *Principles of Biomedical Ethics*. Oxford University Press, New York, NY, 6th edition. [72, 76]

Behe, M. J. (1996). *Darwin's Black Box: The Biochemical Challenge to Evolution*. Free Press, New York. [370]

Bell, J. L. and Slomson, A. B. (1974). *Models and Ultraproducts: An Introduction*. North-Holland, Amsterdam, 3rd edition. [169, 173, 174, 175, 183]

Belnap, N. (1993). On rigorous definitions. *Philosophical Studies*, 72:115–146. [224, 225, 252, 274, 276, 304, 404, 406]

Belot, G. (2005). Whose devil? Which details? *Philosophy of Science*, 72(1):128–153. [87]

Benacerraf, P. (1965). What numbers could not be. *The Philosophical Review*, 74(1):47–73. [80]

Bergmann, G. (1949). Two criteria for an ideal language. *Philosophy of Science*, 16(1):71–74. [9, 23, 35]

Bergmann, G. (1957). The revolt against logical atomism—I. *The Philosophical Quarterly*, 7(29):323–339. [18]

Berkowitz, L. J. (1979). Discussion: Achinstein on empirical significance: A matter of principle. *Philosophy of Science*, 46:459–465. [217, 218, 271]

Betti, A. and de Jong, W. R. (2008). Introduction. *Synthese*, 174(2):181–183. Guest editors' introduction for the issue "The Classical Model of Science I: A Millennia-Old Model of Scientific Rationality". [86, 110]

Betti, A., de Jong, W. R., and Martijn, M. (2009). The axiomatic method, the order of concepts and the hierarchy of sciences: An introduction. *Synthese*, 183(1):1–5. Guest editors' introduction for the issue "The Classical Model of Science II: The Axiomatic Method, the Order of Concepts and the Hierarchy of Sciences". [86, 110]

Bishop, M. A. (1992). The possibility of conceptual clarity in philosophy. *American Philosophical Quarterly*, 29(3):267–277. [141]

Bohnert, H. (1971a). Review of 'Theoretical terms and partial interpretation' by Peter Achinstein. *Journal of Symbolic Logic*, 36(2):312–322. [50]

Bohnert, H. G. (1963). Carnap's theory of definition and analyticity. In Schilpp (1963), pages 407–430. [16, 23, 127, 409]

Bohnert, H. G. (1971b). Review [untitled]. *The Journal of Symbolic Logic*, 36(1):178. Review of (Martin 1966). [277]

Boudry, M. and Leuridan, B. (2011). Where the design argument goes wrong: Auxiliary assumptions and unification. *Philosophy of Science*, 78(4):558–578. [361, 363, 364, 365, 366, 367, 368, 369, 371, 375, 376]

Brogaard, B. (2007). Sharvy's theory of definite descriptions revisited. *Pacific Philosophical Quarterly*, 88:160–180. [347]

Buck, R. C. and Cohen, R. S., editors (1970). *In Memory of Rudolf Carnap: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, volume VIII of *Boston Studies in the Philosophy of Science*, Dordrecht. Philosophy of Science Association, D. Reidel Publishing Company. [423, 446]

Bueno, O. (1997). Empirical adequacy: A partial structures approach. *Studies in the History and Philosophy of Science*, 28(4):585–610. [199, 200, 279]

Bueno, O. (2000). Empiricism, scientific change and mathematical change. *Studies in the History and Philosophy of Science*, 31(2):269–296. [202]

Bueno, O., French, S., and Ladyman, J. (2002). On representing the relationship between the mathematical and the empirical. *Philosophy of Science*, 69:497–518. [196, 200]

Burgos, J. E. (2007). The theory debate in psychology. *Behavior and Philosophy*, 35:149–183. [87, 101]

Campbell, N. R. (1920). *Physics. The Elements*. Cambridge University Press, Cambridge. [94]

Carnap, R. (1922). *Der Raum. Ein Beitrag zur Wissenschaftslehre*, volume 56 of *„Kant-Studien". Ergänzungshefte im Auftrag der Kant-Gesellschaft*. Verlag von Reuther & Reichard, Berlin. [95, 100, 413]

Carnap, R. (1923). Über die Aufgabe der Physik und die Anwendung des Grundsatzes der Einfachstheit. *Kant-Studien*, 28:91–107. [94, 95, 96, 97, 98, 99, 100, 102, 111, 113, 115, 116, 117, 118, 119, 123, 126, 127, 128, 129, 132, 149, 155, 157, 158, 413]

Carnap, R. (1924). Dreidimensionalität des Raumes und Kausalität. Eine Untersuchung über den logischen Zusammenhang zweier Fiktionen. *Annalen der Philosophie und philosophischen Kritik*, 4(1):105–130. [94, 99, 100, 111, 112, 113, 116, 117, 121, 413]

Carnap, R. (1926). *Physikalische Begriffsbildung*, volume 39 of *Wissen und Wirken. Einzelschriften zu den Grundfragen des Erkennens und Schaffens*. Verlag G. Braun, Karlsruhe. [94, 95, 98, 99, 100, 112, 113, 117, 118, 123, 124, 128, 129, 131, 132, 157, 413]

Carnap, R. (1928a). *Der logische Aufbau der Welt*. Weltkreis-Verlag, Berlin-Schlachtensee. References are to the translation (Carnap 1967a). [18, 27, 88, 94, 99, 100, 102, 104, 110, 113, 115, 118, 119, 124, 128, 129, 223, 271, 272, 273, 274, 390, 413]

Carnap, R. (1928b). *Scheinprobleme in der Philosophie: Das Fremdpsychische und der Realismusstreit*. Weltkreis-Verlag, Berlin-Schlachtensee. References are to the translation (Carnap 1967a). [232, 247, 329]

Carnap, R. (1931a). Die physikalische Sprache als Universalsprache der Wissenschaft. *Erkenntnis*, 2(5/6):432–465. Nominal publication date incorrect. Published in 1932. [122, 148]

Carnap, R. (1931b). Überwindung der Metaphysik durch logische Analyse der Sprache. *Erkenntnis*, 2(4). Nominal publication date incorrect. Published in 1932. [419]

Carnap, R. (1932). Über Protokollsätze. *Erkenntnis*, 3(1):215–228. [123, 185]

Carnap, R. (1934a). *Logische Syntax der Sprache*, volume 8 of *Schriften zur wissenschaftlichen Weltauffassung*. Springer-Verlag, Wien. References are to the corrected translation (Carnap 1967b). [10, 18, 102, 104, 142]

Carnap, R. (1934b). On the character of philosophic problems. *Philosophy of Science*, 1(1):5–19. [36]

Carnap, R. (1935a). Formalwissenschaft und Realwissenschaft. *Erkenntnis*, 5(1):30–37. [123, 276, 287, 290, 294]

Carnap, R. (1935b). *Philosophy and Logical Syntax*. Number 70 in Psyche Miniatures, General Series. Kegan Paul, Trench, Trubner & Co., London. References are to the corrected reprint (Carnap 1967b). [262]

Carnap, R. (1936). Testability and meaning. *Philosophy of Science*, 3(4):420–468. [88, 89, 91, 93, 117, 123, 148, 157, 245, 246, 247, 262, 271, 276, 294, 397, 413]

Carnap, R. (1937). Testability and meaning—continued. *Philosophy of Science*, 4(1):2–35. [88, 104, 148, 246, 312]

Carnap, R. (1938). Empiricism and the language of science. *Synthese*, 3(1). [89, 104, 106]

Carnap, R. (1939). *Foundations of Logic and Mathematics*, volume I,3 of *Foundations of the Unity of Science. Toward an International Encyclopedia of Unified Science*. University of Chicago Press, Chicago and London. References are to the two-volume edition. [17, 40, 74, 75, 85, 88, 89, 90, 92, 93, 98, 104, 105, 106, 109, 113, 124, 129, 133, 135, 137, 149, 157, 216, 413]

Carnap, R. (1948). *Introduction to Semantics*, volume 1 of *Studies in Semantics*. Harvard University Press, Cambridge, Massachusetts. [26]

Carnap, R. (1950a). Empiricism, semantics, and ontology. *Revue Internationale de Philosophie*, 4:20–40. References are to the slightly modified reprint (Carnap 1956a, appendix A). [21, 54]

Carnap, R. (1950b). *Logical Foundations of Probability*. University of Chicago Press, Chicago. References are to the 2nd edition (Carnap 1962). [9, 107, 109, 110, 140, 142, 143, 250, 252, 305]

Carnap, R. (1952). Meaning postulates. *Philosophical Studies*, 3(5):65–73. [40, 71, 86, 206]

Carnap, R. (1954). *Einführung in die symbolische Logik*. Julius Springer, Wien. References are to the translation (Carnap 1958). [104, 136]

Carnap, R. (1955). Meaning and synonymy in natural languages. *Philosophical Studies*, 6(3):33–47. [22, 25, 35]

Carnap, R. (1956a). *Meaning and Necessity*. University of Chicago Press, Chicago, 2nd edition. [418]

Carnap, R. (1956b). The methodological character of theoretical concepts. In Feigl and Scriven (1956). [91, 92, 95, 97, 101, 102, 103, 104, 105, 106, 113, 133, 135, 148, 150, 153, 157, 168, 222, 271, 282, 283, 284, 289, 293, 294, 312, 413]

Carnap, R. (1958). Beobachtungssprache und theoretische Sprache. *Dialectica*, 12:236–248. [43, 86, 91, 92, 101, 103, 113, 117, 174, 275, 413, 418]

Carnap, R. (1962). *Logical Foundations of Probability*. University of Chicago Press, Chicago, 2nd edition. [18, 19, 20, 53, 91, 116, 117, 375, 413, 418]

Carnap, R. (1963a). Intellectual autobiography. In Schilpp (1963), pages 1–84. [88, 95, 96, 112, 116, 117, 118, 119, 413]

Carnap, R. (1963b). The physical language as the universal language of science. In Alston, W. P. and Nakhnikian, G., editors, *Readings in Twentieth-Century Philosophy*, pages 393–424. Free Press of Glencoe, London. Revised translation of (Carnap 1931b) with author's introduction and addenda. [104, 148]

Carnap, R. (1963c). Replies and systematic expositions. In Schilpp (1963), pages 859–1016. [9, 13, 22, 23, 25, 101, 103, 127, 133, 142, 145, 148, 216, 273, 300, 329, 393, 394, 397]

Carnap, R. (1966). *Philosophical Foundations of Physics: An Introduction to the Philosophy of Science*. Basic Books, Inc., New York and London. Edited by Martin Gardner. [56, 66, 92, 93, 95, 96, 97, 99, 101, 103, 105, 115, 117, 122, 130, 131, 138, 155, 195, 223, 413]

Carnap, R. (1967a). *The Logical Structure of the World. Pseudoproblems of Philosophy*. University of California Press, Berkeley and Los Angeles. Translation by Rolf A. George. [34, 273, 417]

Carnap, R. (1967b). *The Logical Syntax of Language*. Routledge & Kegan Paul Ltd, London. Reprinted with corrections. Translation by Amethe Smeaton (Countess von Zeppelin). [417, 418]

Cartwright, N. (1989). *Nature's Capacities and their Measurement*. Oxford University Press, Oxford. [77]

Carus, A. W. (2007). *Carnap and Twentieth-Century Thought. Explication as Enlightenment*. Cambridge University Press, Cambridge. [10, 13, 86, 87, 96, 102, 141, 142, 145]

Cat, J. (2011). Otto Neurath. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, Stanford, winter 2011 edition. [128]

Cei, A. and French, S. (2006). Looking for structure in all the wrong places: Ramsey sentences, multiple realisability, and structure. *Studies In History and Philosophy of Science Part A*, 37:633–655. [390]

Chakravartty, A. (2001). The semantic or model-theoretic view of theories and scientific realism. *Synthese*, 127:325–345. [183, 208]

Chang, C. C. and Keisler, H. J. (1990). *Model Theory*, volume 73 of *Studies in Logic and the Foundations of Mathematics*. North Holland, Amsterdam, 3rd edition. 3rd impression 1992. [4, 173, 175]

Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress*. Oxford Studies in the Philosophy of Science. Oxford University Press, New York. [54, 132]

Chang, H. (2005). A case for old-fashioned observability, and a reconstructed constructive empiricism. *Philosophy of Science*, 72(5):876–887. [18]

Chomsky, N. (1957). *Syntactic Structures*. Mouton and Co., The Hague, 2nd edition. [140, 142]

Church, A. (1939). Review: Rudolf Carnap, *Foundations of Logic and Mathematics*. *Bulletin of the American Mathematical Society*, 45(11):821–822. [52]

Church, A. (1949). Review of Ayer's *Language, Truth, and Logic*, 2nd edition. *The Journal of Symbolic Logic*, 14(1):52–53. [214, 227, 252]

Cohen, R. S., editor (1983). *Physics, Philosophy and Psychoanalysis: Essays in Honour of Adolf Grünbaum*. Boston Studies in the Philosophy of Science. Kluwer, Dordrecht. [427, 430]

Cohnitz, D. and Häggqvist, S. (2009). The role of intuitions in philosophy. *Studia Philosophica Estonica*, 2.2:1–14. Special issue: *The Role of Intuitions in Philosophical Methodology*, edited by Daniel Cohnitz and Sören Häggqvist. http://www.spe.ut.ee/ojs-2.2.2/index.php/spe/article/view/84/53. [10]

Copeland, B. J. (2008). The Church-Turing thesis. In Zalta (2008a). [252]

Corkett, C. J. (2002). Fish stock assessment as a non-falsifiable science: Replacing an inductive and instrumental view with a critical rational one. *Fisheries Research*, 56:117–123. [214]

Craig, W. (1953). On axiomatizability within a system. *The Journal of Symbolic Logic*, 18(1):30–32. [86]

Creath, R. (1976). On Kaplan on Carnap on significance. *Philosophical Studies*, 30:393–400. [271, 290, 444]

Creath, R. (1991). Every dogma has its day. *Erkenntnis*, 35(1–3):347–389. Special issue: *Special Volume in Honor of Rudolf Carnap and Hans Reichenbach*. [xii, 68, 410]

da Costa, N. and French, S. (1990). The model-theoretic approach in the philosophy of science. *Philosophy of Science*, 57:248–265. [133, 158, 173, 199, 200] da Costa, N. and French, S. (2000). Models, theories, and structures: Thirty years on. *Philosophy of Science*, 67 (Proceedings):S116–S127. [183, 199]

da Costa, N. C. A., Bueno, O., and French, S. (1998). The logic of pragmatic truth. *Journal of Philosophical Logic*, 27(6):603–620. [200]

Daniels, N. (2008). Reflective equilibrium. In Zalta (2008a). [16]

Davidson, D. (1966). Emeroses by other names. *The Journal of Philosophy*, 63(24):778–780. [181]

Davis, P. and Kenyon, D. H. (1993). *Of Pandas and People: The Central Question of Biological Origins*. Foundation for Thought and Ethics, Richardson, TX, 2nd edition. [359, 360, 425]

de Regt, H. W. (2009). The epistemic value of understanding. *Philosophy of Science*, 76(5):585–597. [339]

Dembski, W. (2006). In defense of intelligent design. In Clayton, P. and Simpson, Z., editors, *The Oxford Handbook of Religion and Science*, Oxford Handbooks in Religion and Theology. Oxford University Press, Oxford. [347]

Dembski, W. A. (2002). *No Free Lunch: Why Specified Complexity Cannot be Purchased Without Intelligence*. Rowman & Littlefield, Lanham, MD. [351, 352]

Dembski, W. A. and Wells, J. (2008). *The Design of Life: Discovering Signs of Intelligence in Biological Systems*. Foundation for Thought and Ethics, Dallas, TX. [351, 359]

Demopoulos, W. (2003). On the rational reconstruction of our theoretical knowledge. *The British Journal for the Philosophy of Science*, 54:371–403. [61, 62, 296]

Demopoulos, W. (2007). Carnap on the rational reconstruction of scientific theories. In Friedman, M. and Creath, R., editors, *The Cambridge Compation to Carnap*, pages 248–272. Cambridge University Press, Cambridge. [61, 395]

Demopoulos, W. (2008). Some remarks on the bearing of model theory on the theory of theories. *Synthese*, 164(3):359–383. [61, 62, 63, 296, 395]

Demopoulos, W. (2011). Three views of theoretical knowledge. *The British Journal for the Philosophy of Science*, pages 177–205. [62, 270]

Derksen, A. A. (1993). The seven sins of pseudo-science. *Journal for General Philosophy of Science*, 24(1):17–42. [358]

Diamond, M. L. (1975a). The challenge of contemporary empiricism. In Diamond and Litzenburg (1975), pages 3–62. [335, 394]

Diamond, M. L. (1975b). Verificationism: Difficulties and proposals. In Diamond and Litzenburg (1975), pages 435–445. [335, 338]

Diamond, M. L. and Litzenburg, T. V., editors (1975). *The Logic of God: Theology and Verification*. Bobbs-Merill, Indianapolis, IN. [214, 422, 424, 444]

Dizadji-Bahmani, F., Frigg, R., and Hartmann, S. (2010). Who's afraid of Nagelian reduction? *Erkenntnis*, 73(3):393–412. Special issue: *Reduction and the Special Sciences*, edited by Mark Colyvan and Stephan Hartmann. [87]

Dorr, C. (2010). Review of Ladyman and Ross (2007). In Gutting, G. and Gutting, A. F., editors, *Notre Dame Philosophical Reviews*. Philosophy Department, University of Notre Dame, Notre Dame, IN. [36, 230]

Duhem, P. (1914). *La théorie physique, son objet et sa structure*. Marcel Rivière, Paris, 2nd edition. References are to the translation (Duhem 1954). [257]

Duhem, P. (1954). *The Aim and Structure of Physical Theory*. Princeton University Press, Princeton, NJ. [422]

Earman, J. (1993). Carnap, Kuhn, and the philosophy of scientific methodology. In Horwich (1993), pages 9–36. [156]

Eberle, R. (1990). Classification by comparison with paradigms. *American Philosophical Quarterly*, 27(4):295–304. [182]

Einstein, A. (1934). On the method of theoretical physics. *Philosophy of Science*, 1(2):163–169. The Herbert Spencer Lecture, delivered at Oxford, June 10, 1933. [52]

Elsberry, W. and Shallit, J. (2009). Information theory, evolutionary computation, and Dembski's "complex specified information". *Synthese*, 178(2):237–270. Special issue: *Evolution and its Rivals*, edited by Glenn Branch and James H. Fetzer. [353]

Enderton, H. B. (2009). Second-order and higher-order logic. In Zalta (2009a). [170]

Ereshefsky, M. (1991). The semantic approach to evolutionary theory. *Biology and Philosophy*, 6:59–80. [87, 101, 107]

Eshleman, A. (2004). Moral responsibility. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, Stanford, fall 2004 edition. [38]

Essler, W. K. (1982). *Wissenschaftstheorie I. Definition und Reduktion*. Kolleg Philosophie. Verlag Karl Alber, Freiburg and München, 2nd edition. [258, 273]

Fedyk, M. (2009). Philosophical intuitions. *Studia Philosophica Estonica*, 2.2:54–80. Special issue: *The Role of Intuitions in Philosophical Methodology*, edited by Daniel Cohnitz and Sören Häggqvist. http://www.spe.ut.ee/ojs-2.2.2/index.php/spe/article/view/57. [17, 29]

Feigl, H. (1950). Existential hypotheses: Realistic versus phenomenalistic interpretations. *Philosophy of Science*, 17:35–62. [189, 216]

Feigl, H. (1956). Some major issues and developments in the philosophy of science of logical empiricism. In Feigl and Scriven (1956). [152, 217]

Feigl, H. (1958). Critique of intuition according to scientific empiricism. *Philosophy East and West*, 8(1/2):1–16. [14, 15, 16, 29, 30, 148]

Feigl, H. (1963). Physicalism, unity of science, and the foundations of psychology. In Schilpp (1963), pages 227–268. [148]

Feigl, H. (1970). The "orthodox" view of theories: Remarks in defense as well as critique. In Radner and Winokur (1970), pages 3–16. [85, 86, 94, 105, 108, 137, 145, 146, 151, 152]

Feigl, H., Hempel, C. G., Jeffrey, R. C., Quine, W. V., Shimony, A., Bar-Hillel, Y., Bohnert, H. G., Cohen, R. S., Hartshorne, C., Kaplan, D., Morris, C., Reichenbach, M., and Stegmüller, W. (1970). Homage to Rudolf Carnap. In Buck and Cohen (1970). [150]

Feigl, H. and Scriven, M., editors (1956). *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, volume 1 of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, Minneapolis, MN. [418, 423]

Feldman, R. (2008). Naturalized epistemology. In Zalta (2008a). [11, 26, 384]

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, volume 2. John Wiley & Sons, Inc., New York, N.Y., 2nd edition. [316]

Fine, K. (1975). Vagueness, truth and logic. *Synthese*, 30(3–4):265–300. References are to the corrected reprint (Fine 1997). [40, 45]

Fine, K. (1997). Vagueness, truth and logic. In Keefe, R. and Smith, P., editors, *Vagueness. A Reader*, pages 119–150. The MIT Press, Cambridge, MA, and London. [423]

Fitelson, B. (2002). Putting the irrelevance back into the problem of irrelevant conjunction. *Philosophy of Science*, 69(4):611–622. [168]

Flew, A. (1950). Theology and falsification. *University*, 1:1–8. References are to the reprint (Diamond and Litzenburg 1975, 257–259). [215, 266, 337]

Flew, A. (1975). "Theology and falsification" in retrospect. In Diamond and Litzenburg (1975), pages 269–283. [215, 337]

Fodor, J. A. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese*, 28:97–115. [230, 380, 381, 382, 383, 384]

Frege, G. (1918). Der Gedanke: eine logische Untersuchung. *Beiträge zur Philosophie des deutschen Idealismus*, 1:58–77. References are to the translation (Frege 1956). [305]

Frege, G. (1956). The thought: A logical inquiry. *Mind*, 65(259):289–311. [424]

French, S. and Ladyman, J. (1999). Reinflating the semantic approach. *International Studies in the Philosophy of Science*, 13(2):103–121. [94, 101, 114, 171, 173, 178, 184, 185, 187, 199, 279]

Friedman, M. (1982). Review. *The Journal of Philosophy*, 79(5):274–283. Review of (van Fraassen 1980). [194]

Friedman, M. (1987). Carnap's *Aufbau* reconsidered. *Noûs*, 21(4):521–545. [156]

Friedman, M. (1992). Epistemology in the *Aufbau*. *Synthese*, 93(1–2):15–57. [119]

Friedman, M. (2003). Hempel and the Vienna Circle. In Hardcastle, G. L. and Richardson, A. W., editors, *Logical Empiricism in North America*, volume 18 of *Minnesota Studies in the Philosophy of Science*, pages 94–114. University of Minnesota Press, Minneapolis, MN. [156]

Frigg, R. and Hartmann, S. (2008). Models in science. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, Stanford, spring 2008 edition. [101, 133, 136]

Gaines, B. (2010). Human rationality challenges universal logic. *Logica Universalis*, 4:163–205. [141]

Gemes, K. (1998a). Hypothetico-deductivism: The current state of play; the criterion of empirical significance: Endgame. *Erkenntnis*, 49:1–20. [218]

Gemes, K. (1998b). Logical content and empirical significance. In Weingartner, P., Schurz, G., and Dorn, G., editors, *Proceedings of the 20th International Wittgenstein Symposium*, Vienna. Hölder-Pichler-Tempski. [104, 213, 216, 227, 246, 257]

George, A. (2000). On washing the fur without wetting it: Quine, Carnap, and analyticity. *Mind*, 109(433):1–24. [68]

Giere, R. N. (1985). Philosophy of science naturalized. *Philosophy of Science*, 52(3):331–356. [11]

Gödel, K. (1930). Die Vollständigkeit der Axiome des logischen Funktionenkalküls. *Monatshefte für Mathematik*, 37(1):349–360. [102]

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik*, 38(1):173–198. [102]

Goodman, N. (1951). *The Structure of Appearance*. Harvard University Press, Cambridge. [142]

Goodman, N. (1963). The significance of *Der logische Aufbau der Welt*. In Schilpp (1963), pages 545–558. [274]

Goodman, N. (1965). *Fact, Fiction, and Forecast*. Bobbs-Merrill, Indianapolis, 2nd edition. [180]

Grice, H. P. (1975). Logic and conversation. In Cole, P. and Morgan, J. L., editors, *Speech Acts*, volume 3 of *Syntax and Semantics*, pages 41–58. Academic Press, New York. [260]

Gupta, A. (2009). Definitions. In Zalta (2009a). [224, 273, 378]

Häggström, O. (2007a). Intelligent design and the NFL theorems. *Biology and Philosophy*, 22(2):217–230. [348]

Häggström, O. (2007b). Uniform distribution is a model assumption. http://www.math.chalmers.se/~olleh/reply_to_Dembski.pdf. Archived at http://www.webcitation.org/5tsNDuhBm. [348]

Halvorson, H. (2012). What scientific theories could not be. *Philosophy of Sciences*, 79(2):183–206. [110]

Hansson, S. O. (2008a). Philosophy and other disciplines. *Metaphilosophy*, 39:472–483. [65, 66, 67, 68, 357]

Hansson, S. O. (2008b). Science and pseudo-science. In Zalta (2008a). [265]

Hare, R. M. (1960). Philosophical discoveries. *Mind*, 69(274):145–162. [22]

Hare, R. M. (1975). Abortion and the golden rule. *Philosophy and Public Affairs*, 4(3):201–222. [16]

Hartwig, M. D. and Meyer, S. C. (1993). A note to teachers. In Davis and Kenyon (1993), pages 153–164. [343, 346, 347, 351]

Hempel, C. G. (1939). Vagueness and logic. *Philosophy of Science*, 6(2):163–180. [133]

Hempel, C. G. (1950). Problems and changes in the empiricist criterion of meaning. *Revue Internationale de Philosophie*, 11:41–63. [143, 217, 248, 250]

Hempel, C. G. (1951). The concept of cognitive significance: A reconsideration. *Proceedings of the American Academy of Arts and Sciences*, 80(1):61–77. [145, 217]

Hempel, C. G. (1952). *Fundamentals of Concept Formation in Empirical Sciences*, volume II,7 of *Foundations of the Unity of Science. Toward an International Encyclopedia of Unified Science*. University of Chicago Press, Chicago and London. References are to the two-volume edition. [53, 54, 93, 94, 95, 99, 105, 108, 131, 132, 142, 143, 144, 156, 157, 249, 250, 305, 413]

Hempel, C. G. (1953). Reflections on Nelson Goodman's *The Structure of Appearance*. *The Philosophical Review*, 62(1):108–116. [143]

Hempel, C. G. (1958). The theoretician's dilemma. In Feigl, H., Scriven, M., and Maxwell, G., editors, *Concepts, Theories, and the Mind-Body Problem*, volume 2 of *Minnesota Studies in the Philosophy of Science*, pages 173–226. University of Minnesota Press, Minneapolis, MN. [109, 155, 157, 413]

Hempel, C. G. (1963). Implications of Carnap's work for the philosophy of science. In Schilpp (1963), pages 685–709. [101, 103, 104, 106, 135, 157]

Hempel, C. G. (1965a). Aspects of scientific explanation. In Hempel (1965b), pages 331–496. [135, 136, 137, 144, 349]

Hempel, C. G. (1965b). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. The Free Press, New York. [133, 217, 264, 426]

Hempel, C. G. (1965c). Empiricist criteria of cognitive significance: Problems and changes. In Hempel (1965b), pages 81–119. [86, 145, 217, 222, 223, 225, 233, 234, 249, 252, 255, 264, 272, 294, 306]

Hempel, C. G. (1965d). Postscript (1964) on cognitive significance. In Hempel (1965b), pages 120–122. [235, 248, 264]

Hempel, C. G. (1965e). Preface. In Hempel (1965b), pages vii–viii. [144]

Hempel, C. G. (1965f). Studies in the logic of confirmation. In Hempel (1965b), pages 3–51. [106, 168, 227]

Hempel, C. G. (1965g). The theoretician's dilemma. In Hempel (1965b), pages 173–226. [101, 104, 105, 106, 339, 340, 407]

Hempel, C. G. (1970). On the “standard conception” of scientific theories. In Radner and Winokur (1970), pages 142–163. [86, 94, 105, 108, 144, 147, 153, 155]

Hempel, C. G. (1973). Rudolf Carnap, logical empiricist. *Synthese*, 25(3–4):256–268. [104, 106]

Hempel, C. G. (1974). Formulation and formalization of scientific theories—A summary-abstract. In Suppe (1974b), pages 244–265. [94, 108, 134, 151, 208, 414]

Hempel, C. G. (1983). Valuation and objectivity in science. In Cohen (1983), pages 73–100. [146]

Hempel, C. G. (1993). Thomas Kuhn, colleague and friend. In Horwich (1993), pages 7–8. [144, 156]

Hempel, C. G. and Oppenheim, P. (1936). *Der Typusbegriff im Lichte der neuen Logik. Wissenschaftstheoretische Untersuchungen zur Konstitutionsforschung und Psychologie*. Sijthoff, Leiden. Nagel (1936) gives a summary in English. [143, 249]

Hempel, C. G. and Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, 15(2):135–175. [86]

Hendry, R. F. and Psillos, S. (2007). How to do things with theories: An interactive view of language and models in science. In Brzeziński, J., Klawiter, A., Kuipers, T. A., Łastowski, K., Paprzycka, K., and Przybysz, P., editors, *The Courage of Doing Philosophy: Essays Dedicated to Leszek Nowak*, pages 59–115. Rodopi, Amsterdam/New York. [180]

Henkin, L. (1950). Completeness in the theory of types. *The Journal of Symbolic Logic*, 15(2):81–91. [102, 105]

Hintikka, J. (1999). The emperor’s new intutions. *The Journal of Philosophy*, 96(3):127–147. [30]

Hodges, W. (1993). *Model Theory*, volume 42 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge. Digitally printed in 2008. [4, 40, 41, 58, 62, 168, 170, 172, 175, 176, 180, 188, 191, 192, 193, 196, 197, 206, 239, 240, 269, 273, 274, 275]

Hodges, W. (2001). Elementary predicate logic. In *Handbook of Philosophical Logic*, volume 1, pages 1–130. Kluwer Academic Publishers, Dordrecht, Boston, London. [102]

Hoffman, D. (1998). Book review—*The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age*. *Notices of the American Mathematical Society*, 45. Review of (Horgan 1996). [407]

Hookway, C. (2010). Pragmatism. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, Stanford, spring 2010 edition. [295]

Horgan, J. (1996). *The End of Science: Facint the Limits of Knowledge in the Twilight of the Scientific Age*. Helix Books, Addison-Wesley Publishing, Reading, MA. [407, 428]

Horwich, P., editor (1993). *World Changes: Thomas Kuhn and the Nature of Science*. MIT Press, Cambridge, Massachusetts, and London, England. [422, 427, 430]

Howson, C. and Urbach, P. (1993). *Scientific Reasoning: The Bayesian Approach*. Open Court, Chicago and La Salle, IL. Second edition. Second printing 1996. [302]

Hughes, R. I. G. (1989). *The Structure and Interpretation of Quantum Mechanics*. Harvard University Press, Cambridge, MA. [87]

Humberstone, L. (1998). Note on supervenience and definability. *Notre Dame Journal of Formal Logic*, 39(2):243–252. [384]

Irzik, G. and Grünberg, T. (1995). Carnap and Kuhn: Arch enemies or close allies? *The British Journal for the Philosophy of Science*, 46(3):285–307. [156]

Jones, M. R. (2005). Idealization and abstraction: A framework. In Jones, M. R. and Cartwright, N., editors, *Idealization XII: Correcting the Model. Idealization and Abstraction in the Sciences*, volume 86 of *Poznań Studies in the Philosophy of the Sciences and the Humanitites*, pages 173–217. Rodopi, Amsterdam/New York, NY. [70, 74, 75]

Jones, N. (2008). Is all abstracting idealizing? *The Reasoner*, 2(4):4–5. [74]

Justus, J. (2010). The search for a formal criterion of empirical significance: A reconsideration. Unpublished talk presented at the *Philosophy of Science Association Biennial Meeting*, Montréal, November 2010. [217]

Justus, J. (2011). Carnap on concept determination: Methodology for philosophy of science. *European Journal for Philosophy of Science*. Forthcoming. [55, 87]

Kaplan, D. (1975). Significance and analyticity: a comment on some recent proposals of Carnap. In Hintikka, J., editor, *Rudolph Carnap, Logical Empiricist: Materials and Perspectives*. D. Reidel, Dordrecht, The Netherlands, Boston, MA. [271, 290]

Kauppinen, A. (2007). The rise and fall of experimental philosophy. *Philosophical Explorations*, 10(2):95–118. [9, 12, 29, 33, 35]

Kemeny, J. G. (1963). Analyticity versus fuzziness. *Synthese*, 15(1):57–80. [21, 22, 68, 230]

Kemeny, J. G. and Oppenheim, P. (1952). On reduction. *Philosophical Studies*, 7(1–2):6–19. [379, 380, 382]

Ketland, J. (2004). Empirical adequacy and ramsification. *British Journal for the Philosophy of Science*, 55:287–300. [270]

Kim, J. (1988). What is "naturalized epistemology"? *Philosophical Perspectives*, 2:381–405. [13, 35]

Kim, J. (1999). Hempel, explanation, metaphysics. *Philosophical Studies*, 94(1–2):1–20. [133]

Kim, J. (2006). *Philosophy of Mind*. Westview Press, Cambridge, MA, 2nd edition. [50]

Kitcher, P. (2001). Carl g. hempel, 1905–1997. In Martinich, A. P. and Sosa, D., editors, *Blackwell Companion to Philosophy: A Companion to Analytic Philosophy*, pages 148–159. Blackwell, Malden, MA. [101]

Kitcher, P. and Salmon, W., editors (1989). *Scientific Explanation*, volume 13 of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, Minneapolis, MN. [170]

Kitts, D. B. (1977). Karl popper, verifiability, and systematic zoology. *Systematic Zoology*, 26(2):185–194. [235]

Klein, C. (2011). Multiple realizability and the semantic view of theories. *Philosophical Studies*. Forthcoming. [101]

Knobe, J. (2007). Experimental philosophy and philosophical significance. *Philosophical Explorations*, 10(2):119–121. [12]

Kokoszyńska, M. (1964). Review:"some considerations concerning 'interpretative systems'." by harry v. stopes-roe. *The Journal of Symbolic Logic*, 29(4):195–197. Review of (Stopes-Roe 1958). [83]

Kornblith, H. (1994). What is naturalistic epistemology? In Kornblith, H., editor, *Naturalizing Epistemology*, pages 1–14. MIT Press, Cambridge, MA, 2nd edition. [28]

Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A. (1971). *Additive and Polynomial Representations*, volume 1 of *Foundations of Measurement*. Academic Press, New York, NY, and London. [131, 132]

Kuhn, T. S. (1993). Afterword. In Horwich (1993), pages 311–341. [156]

Kuipers, T. A. F. (2007). Introduction. Explication in philosophy of science. In Kuipers, T. A. F., editor, *General Philosophy of Science—Focal Issues*, volume 1 of *Handbook of the Philosophy of Science*, pages vii–xxiii. Elsevier, Amsterdam. [12, 20, 36, 82]

Kyburg, Jr., H. E. (1990). Theories as mere conventions. In Savage (1990), pages 158–174. [40]

Ladyman, J. (2009). Structural realism. In Zalta (2009a). [389]

Ladyman, J. and Ross, D. (2007). *Every Thing Must Go: Metaphysics Naturalized*. Oxford University Press, Oxford. With David Spurrett and John Collier. [36, 422]

Lakatos, I. (1974). Falsification and the methodology of scientific research programmes. In Lakatos, I. and Musgrave, A., editors, *Criticism and the Growth of Knowledge*, volume 4 of *Proceedings of the International Colloquium in the Philosophy of Science*, pages 91–196. Cambridge University Press, Cambridge. [262]

Laudan, L. (1982). Commentary: Science at the bar—causes for concern. *Science, Technology, & Human Values*, 7(41):16–19. [339]

Laudan, L. (1983a). The demise of the demarcation problem. In Cohen (1983), pages 111–127. [357, 358]

Laudan, L. (1983b). More on creationism. *Science, Technology, & Human Values*, 8(1):36–38. [339]

Laudan, L. (1986). Some problems facing intuitionist meta-methodologies. *Synthese*, 67(1):115–129. [20, 143, 357]

Le Bihan, S. (2011). Defending the semantic view: What it takes. *European Journal for Philosophy of Science*. Forthcoming. doi: 10.1007/s13194-011-0026-6. [199]

Leiter, B. (2010). The demarcation problem in jurisprudence: a new case for skepticism. Technical Report 319, University of Chicago. To be published as (Leiter 2011). [357, 431]

Leiter, B. (2011). The demarcation problem in jurisprudence: a new case for skepticism. In Ferrer, J. and Moreso, J., editors, *Neutrality and the Theory of Law*. Marcial Pons, Madrid. References are to the online version (Leiter 2010). [430]

Leitgeb, H. (2009). Logic in general philosophy of science: Old things and new things. *Synthese*. [86, 110]

Leivant, D. (1994). Higher order logic. In Gabbay, D. M., Hogger, C., and Robinson, J., editors, *Deduction Methodologies*, volume 2 of *Handbook of Logic in Artificial Intelligence and Logic Programming*, pages 229–321. Oxford University Press, Oxford. [40, 105, 107, 117, 170, 258, 274]

Lenzen, V. F. (1938). *Procedures of Empirical Science*, volume I,5 of *Foundations of the Unity of Science. Toward an International Encyclopedia of Unified Science*. University of Chicago Press, Chicago and London. References are to the two-volume edition. [132]

Lewis, D. (1970). How to define theoretical terms. *The Journal of Philosophy*, 67(13):427–446. [18, 49, 50]

Lewis, D. (1988a). Ayer's first empiricist criterion of meaning: Why does it fail? *Analysis*, 48(1):1–3. [226, 254, 259, 265]

Lewis, D. (1988b). Statements partly about observation. *Philosophical Papers*, 17:1–31. References are to the reprint (Lewis 1998, 125–155). [xi, 213, 226, 230, 231, 243, 252, 263, 312]

Lewis, D. (1998). *Papers in Philosophical Logic*. Cambridge Studies in Philosophy. Cambridge University Press, Cambridge. [431]

Liao, S. M. (2008). A defense of intuitions. *Philosophical Studies*, 140(2):247–262. [31]

Loomis, E. J. (2006). Empirical equivalence in the Quine-Carnap debate. *Pacific Philosophical Quarterly*, 87:499–508. [68]

Łoś, J. (1955). On the extending of models I. *Fundamenta Mathematicae*, 42. [269]

Luce, R. D. (1978). Dimensionally invariant numerical laws correspond to meaningful qualitative relations. *Philosophy of Science*, 45(1):1–16. [271, 274]

Ludlow, P. (2009). Descriptions. In Zalta (2009a). [347]

Lutz, S. (2009). Ideal language philosophy and experiments on intuitions. *Studia Philosophica Estonica*, 2.2:117–139. Special issue: *The Role of Intuitions in Philosophical Methodology*, edited by Daniel Cohnitz and Sören Häggqvist. http://www.spe.ut.ee/ojs-2.2.2/index.php/spe/article/view/65. [9, 87, 145]

Lutz, S. (2010). Discussion note—Concept formation in ethical theories: Dealing with polar predicates. *Journal of Ethics & Social Philosophy*. http://jesp.org/articles/download/DiscussionNote_PolarPredicates.pdf. [400]

Lutz, S. (2011a). Artificial language philosophy of science. *European Journal for Philosophy of Science*. Forthcoming. [9, 87]

Lutz, S. (2011b). Generalizing empirical adequacy I: Multiplicity and approximation. In preparation. Preprint: http://philsci-archive.pitt.edu/id/eprint/8744. [199, 386]

Lutz, S. (2011c). Generalizing empirical adequacy II: Partial structures. In preparation. Preprint: http://philsci-archive.pitt.edu/id/eprint/8743. [199]

Lutz, S. (2011d). On an allegedly essential feature of criteria for the demarcation of science. *The Reasoner*, 5(8):125–126. http://www.kent.ac.uk/secl/philosophy/jw/TheReasoner/vol5/TheReasoner-5(8).pdf. [357]

Lutz, S. (2012). On a straw man in the philosophy of science: A defense of the Received View. *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, 2(1). [85]

MacFarlane, J. (2009). Logical constants. In Zalta (2009b). [280]

MacGibbon, D. and Ross, T. (1887). *The Castellated and Domestic Architecture of Scotland From the Twelfth to the Eighteenth Century*. David Douglas, Edinburgh. [166]

Mackie, J. L. (1969). The relevance criterion of confirmation. *The British Journal for the Philosophy of Science*, pages 27–40. [372]

Mackie, J. L. (1982). *The Miracle of Theism. Arguments for and against the Existence of God*. Clarendon Press, Oxford. [335, 336, 340]

Maher, P. (2007). Explication defended. *Studia Logica*, 86(2):331–341. [87]

Malzkorn, W. (2001). Defining disposition concepts: A brief history of the problem. *Studies In History and Philosophy of Science Part A*, 32(2):335–353. [276]

Marhenke, P. (1949–1950). The criterion of significance. *Proceedings and Addresses of the American Philosophical Association*, 23:1–21. [217]

Martin, R. M. (1952). On 'analytic'. *Philosophical Studies*, 3:42–47. [68, 215]

Martin, R. M. (1966). On theoretical constructs and Ramsey constants. *Philosophy of Science*, 33(1/2):1–13. [277, 399, 416]

Mates, B. (1951). Analytic sentences. *Philosophical Review*, 60:525–534. [68, 409]

Mates, B. (1958). On the verification of statements about ordinary language. *Inquiry*, 1:161–171. [23, 24, 29]

Matthen, M. and Ariew, A. (2002). Two ways of thinking about fitness and natural selection. *The Journal of Philosophy*, 99(2):55–83. [374]

Maxwell, G. (1970). Structural realism and the meaning of theoretical terms. In Radner and Winokur (1970), pages 181–192. [389]

Maxwell, G. and Feigl, H. (1961). Why ordinary language needs reforming. *The Journal of Philosophy*, 58(18):488–498. [9, 13, 18, 23, 24, 26, 31, 35, 36, 38, 145]

Mikenberg, I., da Costa, N. C. A., and Chuaqui, R. (1986). Pragmatic truth and approximation to truth. *The Journal of Symbolic Logic*, 51(1):201–221. [199, 200]

Misak, C. J. (1995). *Verificationism: Its History and Prospects*. Routledge, London and New York, NY. References are to the republication (Misak 2005). [217]

Misak, C. J. (2005). *Verificationism: Its History and Prospects*. Routledge, London and New York, NY. [433]

Montague, R. (1962). Deterministic theories. In Washburne, N. F., editor, *Decisions, values and groups*, volume 2, pages 325–370, New York. Macmillian Co. [108]

Monton, B. (2009). *Seeking God in Science: An Atheist Defends Intelligent Design*. Broadview Press, Peterborough, ON. [350, 351, 352, 357]

Monton, B. and Mohler, C. (2008). Constructive empiricism. In Zalta (2008b). [191]

Mormann, T. (1991). Husserl's philosophy of science and the semantic approach. *Philosophy of Science*, 58(1):61–83. [87]

Mormann, T. (2007). The structure of scientific theories in logical empiricism. In Richardson and Uebel (2007), pages 136–162. [94]

Morrison, M. and Morgan, M. S. (1999). Introduction. In Morgan, M. S. and Morrison, M., editors, *Models as Mediators. Perspectives on Natural and Social Science*. Cambridge University Press, Cambridge. [87, 101, 133]

Moschovakis, Y. N. (1974). *Elementary Induction on Abstract Structures*, volume 77 of *Studies in Logic and the Foundations of Mathematics*. Elsevier, New York, NY. [227, 258]

Muller, F. A. (2010). Reflections on the revolution at Stanford. *Synthese*, 183(1):87–114. Special issue: *The Classical Model of Science II: The Axiomatic Method, the Order of Concepts and the Hierarchy of Sciences*, edited by Arianne Betti, Willem de Jong and Marije Martijn. [87, 101, 107, 108, 133, 139, 175, 184, 195]

Nadelhoffer, T. and Nahmias, E. (2007). The past and future of experimental philosophy. *Philosophical Explorations*, 10(2):124–149. [12, 27, 28, 29, 30, 32, 33, 35, 39]

Naess, A. (1938). *"Truth" as Conceived by those who are not Professional Philosophers*. Number 4 in Skrifter Norske Videnskaps-Akademi, Oslo, II. Hist.-Filos. Klasse 1938. Jacob Dybwad, Oslo. [26]

Naess, A. (1953). *Interpretation and Preciseness: A Contribution to the Theory of Communication*. Number 1 in Skrifter Norske Videnskaps-Akademi, Oslo, II. Hist.-Filos. Klasse 1953. Jacob Dybwad, Oslo. [25]

Nagel, E. (1936). Review of *Der Typusbegriff im Lichte der neuen Logik. Wissenschaftstheoretische Untersuchungen zur Konstitutionsforschung und Psychologie*. *The Journal of Philosophy*, 33(22):611. [427]

Nagel, E. (1949). The meaning of reduction in the natural sciences. In Stauffer, R. C., editor, *Science and Civilization*. University of Wisconsin Press, Madison. [86, 379]

Nagel, E. (1951). Mechanistic explanation and organismic biology. *Philosophy and Phenomenological Research*, 11(3):327–338. [86, 230, 377, 378, 379]

Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation*. Harcourt, Brace & World, New York and Burlingame. [388]

Nagel, E., Suppes, P., and Tarski, A., editors (1962). *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress*. Stanford University Press, Stanford. [437, 443]

Neurath, O. (1932). Protokollsätze. *Erkenntnis*, 3(1):204–214. [23, 185]

Nichols, S. and Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs*, 41:663–685. [37]

Nickles, T. (1973). Two concepts of intertheoretic reduction. *The Journal of Philosophy*, 70(7):181–201. [385]

Nielsen, K. (1966). On fixing the reference range of 'God'. *Religious Studies*, 2(1):13–36. [215, 335, 337]

Niiniluoto, I. (1972). Inductive systematization: Definition and a critical survey. *Synthese*, 25(1–2). [233, 340]

Nott, K. (1959). Review of Karl Popper: *The Logic of Scientific Discovery*. *Times Literary Supplement*. Published on the 27th of February. Republished online on Januay 8, 2009 at http://entertainment.timesonline.co.uk/tol/arts_and_entertainment/the_tls/article5472346.ece. Archived at http://www.webcitation.org/60ayEdi27. [263]

Nowak, L. (2000). The idealizational approach to science: A new survey. In Nowakowa, I. and Nowak, L., editors, *Idealization X: The Richness of Idealization*, volume 69 of *Poznań Studies in the Philosophy of the Sciences and the Humanitites*, pages 109–184. Rodopi, Amsterdam/Atlanta, GA. [74, 75, 77]

Oberdan, T. (1990). Positivism and the pragmatic theory of observation. In *Proceedings of the Biennial Meeting of the Philosophy of Science Association. Volume One: Contributed Papers*, pages 25–37. Philosophy of Science Association, University of Chicago Press. [124]

O'Neill, O. (1988). Ethical reasoning and ideological pluralism. *Ethics*, 98(4):705–722. [74, 75]

Paley, W. (1802). *Natural Theology: Or, Evidences of the Existence and Attributes of the Deity. Collected From the Appearances of Nature*. R. Foulder, London. References are to the reprint (Paley 1830). [336]

Paley, W. (1830). *Natural Theology: Or, Evidences of the Existence and Attributes of the Deity. Collected From the Appearances of Nature*. Hilliard and Brown, Cambridge. [435]

Papineau, D. (1996). Theory-dependent terms. *Philosophy of Science*, 63(1):1–20. [50]

Papineau, D. (2009a). Naturalism. In Zalta (2009a). [26, 46, 48, 52]

Papineau, D. (2009b). The poverty of analysis. *Aristotelian Society Supplementary Volume*, 83(1):1–30. [13, 14, 15, 46, 47, 48, 64, 87]

Peacocke, C. (1986). *Thoughts: an Essay on Content*, volume 4 of *Aristotelian Society Series*. Basil Blackwell, New York, NY. [394]

Pearce, D. (1981). Is there any theoretical justification for a nonstatement view of theories? *Synthese*, 46:1–39. [158]

Pearce, D. and Rantala, V. (1985). Approximative explanation is deductive-nomological. *Philosophy of Science*, 52(1):126–140. [87, 101]

Peirce, C. S. (1878). How to make our ideas clear. *Popular Science Monthly*, 12:286–302. [294]

Pennock, R. T. (2011). Can't philosophers tell the difference between science and religion?: Demarcation revisited. *Synthese*, 178(2):177–206. Special issue: *Evolution and its Rivals*, edited by Glenn Branch and James H. Fetzer. [358, 370]

Philipse, H. (2009). Can philosophy be a rigorous science? *Royal Institute of Philosophy Supplement*, 84:155–176. [64]

Pincock, C. (2007). A role for mathematics in the physical sciences. *Noûs*, 41:253–275. [78, 79, 80, 82]

Pokriefka, M. L. (1983). Ayer's definition of empirical significance revisited. *Analysis*, 43(4):166–170. [213, 227]

Pokriefka, M. L. (1984). More on empirical significance. *Analysis*, 44(2):92–93. [213, 265]

Popper, K. (1963). The demarcation between science and metaphysics. In Schilpp (1963), pages 183–226. [26, 264]

Popper, K. R. (1935). *Logik der Forschung: Zur Erkenntnistheorie der modernen Naturwissenschaft*, volume 9 of *Schriften zur wissenschaftlichen Weltauffassung*. Julius Springer, Wien. Published in 1934, nominal publication date incorrect. References are to the extended and revised translation (Popper 2000). [216, 223, 225, 251, 255, 262, 264, 265, 300, 436]

Popper, K. R. (2000). *The Logic of Scientific Discovery*. Routledge, London and New York, NY. Translated and extended from the German original (Popper 1935) by Karl Popper, Julius Freed, and Lan Freed. [436]

Przełęcki, M. (1961a). Pojęcia teoretyczne a doświadczenie. *Studia Logica*, 11(1):91–129. [436]

Przełęcki, M. (1961b). Theoretical concepts and experience. *Studia Logica*, 11(1):135–138. Summary of (Przełęcki 1961a). [399]

Przełęcki, M. (1969). *The Logic of Empirical Theories*. Monographs in Modern Logic Series. Routledge & Kegan Paul/Humanities Press, London/New York. [xi, 2, 40, 41, 43, 44, 46, 83, 157, 158, 182, 198, 207, 216, 224, 236, 269, 459]

Przełęcki, M. (1974a). Empirical meaningfulness of quantitative statements. *Synthese*, 26:344–355. [xi, 78, 222, 228, 236, 238, 244, 253]

Przełęcki, M. (1974b). On model theoretic approach to empirical interpretation of scientific theories. *Synthese*, 26:401–406. [157, 216]

Przełęcki, M. (1974c). A set theoretic versus a model theoretic approach to the logical structure of physical theories. *Studia Logica*, 33(1):91–105. [41]

Przełęcki, M. (1975). Review of Ryszard Wójcicki: *Metodologia formalna nauk empirycznych. Podstawowe pojęcia i zagadnienia* (Formal methodology of empirical sciences. Basic concepts and problems), Ossolineum, Wrocław, 1974. *Studia Logica*, 34(3):275–284. [157]

Przełęcki, M. (1976). Fuzziness as multiplicity. *Erkenntnis*, 10:371–380. [44, 45, 202]

Przełęcki, M. and Wójcicki, R. (1969). The problem of analyticity. *Synthese*, 19(3–4):374–399. [43, 44, 396]

Przełęcki, M. and Wójcicki, R. (1971). Inessential parts of extensions of first-order theories. *Studia Logica*, 28(1):83–95. [269]

Psillos, S. (2000). Rudolf Carnap's 'Theoretical concepts in science'. *Studies in History and Philosophy of Science Part A*, 31:151–172. [106, 135, 223, 232, 312]

Putnam, H. (1962). What theories are not. In Nagel et al. (1962). [86, 121, 125]

Quine, W. V. O. (1951). Two dogmas of empiricism. *The Philosophical Review*, 60(1):20–43. [57, 265, 395, 409, 411]

Quine, W. V. O. (1960). *Word and Object*. MIT Press, Cambridge, MA. [411]

Quine, W. V. O. (1963). Two dogmas of empiricism. In *From a Logical Point of View. Logico-Philosophical Essays*, pages 20–46. Harper & Row, New York et al., 2nd edition. [68]

Quine, W. V. O. (1969a). Epistemology naturalized. In Quine (1969b), pages 69–90. [11]

Quine, W. V. O. (1969b). *Ontological Relativity and Other Essays*. Number 1 in The John Dewey Essays in Philosophy. Columbia University Press, New York and London. [27, 56, 119, 437]

Quine, W. V. O. (1970). *Philosophy of Logic*. Prentice-Hall, Englewood-Cliffs, NJ, 2nd edition. [107]

Raatikainen, P. (2010). Ramsification and inductive inference. *Synthese*. Print publication forthcoming. [233]

Raatikainen, P. (2011). On Carnap sentences. *Analysis*, 71(2):245–246. [407]

Rabinowicz, W. (2008). Value relations. *Theoria*, 74:18–49. [402]

Radner, M. and Winokur, S., editors (1970). *Analyses of Theories and Methods of Physics and Psychology*, volume 4 of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, Minneapolis, MN. [414, 423, 427, 433, 439]

Ramsey, F. P. (1929). Theories. In *The Foundations of Mathematics and other Logical Essays*, chapter IX, pages 212–236. Routledge & Kegan Paul, London. Edited by R. B. Braithwaite. [86]

Rantala, V. (1978). The old and the new logic of metascience. *Synthese*, 39:233–247. [186]

Ratzsch, D. (2009). Teleological arguments for God's existence. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, Stanford, winter 2009 edition. [69]

Reck, E. H. (2007). Carnap and modern logic. In Creath, R. and Friedman, M., editors, *The Cambridge Companion to Carnap*. Cambridge University Press, Cambridge. [102]

Redhead, M. (1980). Models in physics. *The British Journal for the Philosophy of Science*, 31(2):145–163. [73, 74]

Reichenbach, H. (1938). *Experience and Prediction; an Analysis of the Foundations and the Structure of Knowledge*. University of Chicago Press, Chicago. [64, 65, 330]

Reichenbach, H. (1951). The verifiability theory of meaning. *Proceedings of the American Academy of Arts and Sciences*, 80(1):46–60. [18, 140, 150, 407]

Richardson, A. (2007). "That sort of everyday image of logical positivism"—Thomas Kuhn and the decline of logical empiricist philosophy of science. In Richardson and Uebel (2007), pages 346–370. [12, 86]

Richardson, A. and Uebel, T., editors (2007). *The Cambridge Companion to Logical Empiricism*. Cambridge University Press, Cambridge. [433, 438, 441]

Roberts, F. S. and Franke, C. H. (1976). On the theory of uniqueness in measurement. *Journal of Mathematical Psychology*, 14:211–218. [238, 240]

Rorty, R. (1967a). Introduction: Metaphilosophical difficulties of linguistic philosophy. In Rorty (1967b), pages 1–39. [9, 11, 13, 16, 145]

Rorty, R., editor (1967b). *The Linguistic Turn: Recent Essays in Philosophical Method*. University of Chicago Press, Chicago and London. [13, 22, 23, 438]

Rosen, G. (2009). Abstract objects. In Zalta (2009b). [75]

Rothbart, D. (1982). Demarcating genuine science from pseudoscience. In Grim, P., editor, *Philosophy of Science and the Occult*, SUNY Series in Philosophy. State University of New York Press, Albany. [339, 340]

Rozeboom, W. W. (1962). The factual content of theoretical concepts. In Feigl, H., Scriven, M., and Maxwell, G., editors, *Scientific Explanation, Space, and Time*, volume 3 of *Minnesota Studies in the Philosophy of Science*, pages 273–357. University of Minnesota Press, Minneapolis, MN. [43, 56, 135, 216, 271, 282, 285]

Rozeboom, W. W. (1970). The crisis in philosophical semantics. In Radner and Winokur (1970), pages 196–219. [135, 151, 158, 159, 216]

Ruja, H. (1961). The present status of the verifiability criterion. *Philosophy and Phenomenological Research*, 22(2):216–222. [216]

Russell, B. (1905). On denoting. *Mind*, 14:479–493. [346]

Russell, B. (1946/1961). *History of Western Philosophy and its Connection with Political and Social Circumstances from the Earliest Times to the Present Day*. Routledge, London. [158]

Rynasiewicz, R. A. (1983). Falsifiability and the semantic eliminability of theoretical languages. *The British Journal for the Philosophy of Science*, 34(3):225–241. [269]

Rynin, D. (1956–1957). Vindication of l*g*c*l p*s*t*v*sm. *Proceedings and Addresses of the American Philosophical Association*, 30:45–67. [216, 235, 248, 255, 264]

Salmon, W. C. (1967). *The Foundations of Scientific Inference*. University of Pittsburgh Press, Pittsburgh, PA. [263, 320]

Salmon, W. C. (1971). *Statistical Explanation and Statistical Relevance*. University of Pittsburgh Press, Pittsburgh, PA. With contributions by Richard C. Jeffrey and James G. Greeno. [321]

Savage, C. W., editor (1990). *Scientific Theories*, volume 14 of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, Minneapolis, MN. [430, 441]

Schaffner, K. F. (1969). Correspondence rules. *Philosophy of Science*, 36(3):280–290. [127]

Scheffler, I. (1968). Reflections on the ramsey method. *The Journal of Philosophy*, 65(10):269–274. Double issue: *On Scientific Hypotheses*. [233, 277]

Schilpp, P. A., editor (1963). *The Philosophy of Rudolf Carnap*, volume 11 of *The Library of Living Philosophers*. Open Court Publishing Company, Chicago and LaSalle, IL. [416, 419, 423, 425, 426, 436, 442]

Schurz, G. (1991). Relevant deduction. *Erkenntnis*, 35:391–437. [260]

Schwabl, F. (1995). *Quantum Mechanics*. Springer, New York et al., 2nd, rev. edition. [73]

Shackel, N. (2005). The vacuity of postmodernist methodology. *Metaphilosophy*, 36:295–320. [165]

Shapiro, S. (2000). *Foundations without Foundationalism: A Case for Second-order Logic*. Oxford University Press, Oxford. [274]

Sharvy, R. (1980). A more general theory of definite descriptions. *The Philosophical Review*, 89(4):607–624. [346]

Shoenfield, J. R. (1967). *Mathematical Logic*. Addison-Wesley, Reading, MA. References are to the reprint (Shoenfield 2000). [62]

Shoenfield, J. R. (2000). *Mathematical Logic*. Association for Symbolic Logic, Natick, MA. [440]

Simon, H. A. (1947). The axioms of Newtonian mechanics. *Philosophical Magazine*, 38(7). [50]

Simon, H. A. (1970). The axiomatization of physical theories. *Philosophy of Science*, 37:16–26. [43, 50]

Simpson, J. A. (1981). Foundations of metrology. *Journal of Research of the National Bureau of Standards*, 86(3):281–292. [131]

Sklar, L. (1967). Types of inter-theoretic reduction. *The British Journal for the Philosophy of Science*, 18:109–124. [385, 388]

Skyrms, B. (1984). *Prediction and Experience*. Yale University Press, New Haven, CT. [329, 330]

Smith, P. (2008). Introducing Wilfrid Hodges, *A Shorter Model Theory*. Typescript. [173]

Smolin, L. (2006). *The Trouble With Physics: The Rise of String Theory, the Fall of a Science, and What Comes Next*. Houghton Mifflin, Boston, MA. [214]

Sneed, J. D. (1971). *The Logical Structure of Mathematical Physics*. D. Reidel Publishing Co., Dordrecht, The Netherlands. [269]

Sneed, J. D. (1979). *The Logical Structure of Mathematical Physics*. D. Reidel Publishing Co., Dordrecht, The Netherlands, 2nd edition. [209]

Sobel, J. H. (2004). *Logic and Theism. Arguments For and Against Beliefs in God*. Cambridge University Press, Cambridge. [335, 336, 340]

Sober, E. (1990). Contrastive empiricism. In Savage (1990), pages 392–410. [57, 195, 214, 299, 318, 323]

Sober, E. (1999). Testability. *Proceedings and Addresses of the American Philosophical Association*, 73(2):47–76. [214, 218, 299, 302, 303, 305, 308, 310, 312, 317, 318, 321, 336, 343, 344, 345, 354, 362, 363, 366, 370]

Sober, E. (2002). Intelligent design and probability reasoning. *International Journal for Philosophy of Religion*, 52(2):65–80. [300, 301, 319]

Sober, E. (2007). What is wrong with intelligent design? *The Quarterly Review of Biology*, 82(1):3–8. [214, 257, 299, 300, 308, 310, 336, 342, 343, 344, 346, 354, 355, 366, 370]

Sober, E. (2008). *Evidence and Evolution: The Logic Behind the Science*. Cambridge University Press, Cambridge. [214, 216, 218, 219, 259, 264, 265, 299, 300, 301, 302, 303, 304, 305, 308, 309, 310, 317, 318, 319, 326, 330, 336, 343, 344, 348, 354, 355, 356, 362, 363, 366, 367, 369, 372, 375]

Sober, E. (2010). Comments on Sebastian Lutz's "On Sober's criterion of contrastive testability". Unpublished typescript. [300, 305, 308, 322]

Sosa, E. (2007). Experimental philosophy and philosophical intuition. *Philosophical Studies*, 132(1):99–107. [14, 32, 37, 38]

Steglich-Petersen (2008). The philosophical lexicon. http://www.philosophicallexicon.com/. Retrieved 2012-02-03. [xi]

Stegmüller, W. (1973). Der sogenannte Zirkel des Verstehens. In Hübner, K. and Menne, A., editors, *Natur und Geschichte. X. Deutscher Kongreßfür Philosophie, Kiel 8.–12. Oktober 1972*, pages 21–46, Hamburg. Felix Meiner. [272]

Stegmüller, W. (1978). A combined approach to the dynamics of theories. *Theory and Decision*, 9(1). [101]

Stegmüller, W. (1979). *The Structuralist View of Theories*. Springer Verlag, New York. [108, 187]

Stein, H. (1992). Was Carnap entirely wrong, after all? *Synthese*, 93(1–2):275–295. Special issue: *Carnap: A Centenary Reappraisal*, edited by Sahotra Sarkar. [68, 140, 407, 410]

Stern, D. (2007). Wittgenstein, the Vienna circle, and physicalism: A reassessment. In Richardson and Uebel (2007), pages 305–331. [87]

Stopes-Roe, H. V. (1958). Some considerations concerning "interpretative systems". *Philosophy of Science*, 25(3):143–156. [83, 429]

Stotz, K. (2009). Experimental philosophy of biology: Notes from the field. *Studies In History and Philosophy of Science Part A*, 40:233–237. [29, 54, 55]

Stotz, K., Griffiths, P. E., and Knight, R. (2004). How scientists conceptualize genes: An empirical study. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 35(4):647–673. [64]

Strawson, P. F. (1963). Carnap's views on constructed systems versus natural languages in analytic philosophy. In Schilpp (1963), pages 503–518. [22, 23, 35]

Suarez, A. G. (2000). The verification principle: Another puncture. *Analysis*, 60(3):293–295. [217]

Suárez, M. (2005). The semantic view, empirical adequacy, and application. *Crítica, Revista Hispanoamericana de Filosofia*, 37(109):29–63. [191, 198]

Suppe, F. (1971). On partial interpretation. *The Journal of Philosophy*, 68(3):57–76. [122]

Suppe, F. (1972). What's wrong with the received view on the structure of scientific theories? *Philosophy of Science*, 39(1):1–19. [94, 122, 123, 125, 126, 128]

Suppe, F. (1974a). The search for philosophic understanding of scientific theories. In Suppe (1974b), pages 3–241. [87, 88, 94, 100, 101, 107, 113, 122, 124, 125, 127, 134, 140, 141, 148, 158, 164, 178, 179, 189, 208]

Suppe, F., editor (1974b). *The Structure of Scientific Theories*. University of Illinois Press, Urbana, IL. [146, 167, 414, 427, 442]

Suppe, F. (1989). *The Semantic Conception of Theories and Scientific Realism*. University of Illinois Press, Urbana, IL. [94, 171]

Suppe, F. (2000). Understanding scientific theories: An assessment of developments, 1969–1998. *Philosophy of Science*, 67:S102–S115. Supplement. Proceedings of the 1998 Biennial Meetings of the Philosophy of Science Association. Part II: Symposia Papers. [85, 86, 101, 133, 140, 151, 158, 167, 169]

Suppes, P. (1951). A set of independent axioms for extensive quantities. *Portugaliae Mathematica*, 10:163–172. [131]

Suppes, P. (1954). Some remarks on problems and methods in the philosophy of science. *Philosophy of Science*, 21(3):242–248. [107]

Suppes, P. (1959). Measurement, empirical meaningfulness, and three-valued logic. In Churchman, C. W. and Ratoosh, P., editors, *Measurement: Definitions and Theories*, pages 129–143. Wiley, New York. [77, 78, 237, 239, 240]

Suppes, P. (1960). A comparison of the meaning and uses of models in mathematics and the empirical sciences. *Synthese*, 12:287–301. [139]

Suppes, P. (1962). Models of data. In Nagel et al. (1962), pages 252–261. [127, 128, 150, 189, 196]

Suppes, P. (1967). What is a scientific theory? In Morgenbesser, S., editor, *Philosophy of Science Today*, pages 55–67. Basic Books, New York. [101, 102, 114, 127]

Suppes, P. (1968). The desirability of formalization in science. *The Journal of Philosophy*, 65(20):651–664. Sixty-Fifth Annual Meeting of the American Philosophical Association Eastern Division, (Oct. 24, 1968). [86, 110, 158, 163, 208]

Suppes, P. (1992). Axiomatic methods in science. In Carvallo, M. E., editor, *Nature, Cognition and System II: On Complementarity and Beyond*, volume 10 of *Theory and Decision Library D*, pages 205–232. Springer, Heidelberg. [94, 107, 110, 187]

Suppes, P. and Zinnes, J. (1963). Basic measurement theory. In Luce, R. D., Bush, R. R., and Galanter, E., editors, *Handbook of Mathematical Psychology*, volume I, pages l–76. Wiley, New York. [238, 239, 240]

Swinburne, R. (2004). *The Existence of God*. Clarendon Press, Oxford, 2nd edition. [335, 353]

Symons, J. (2008). Intuition and philosophical methodology. *Axiomathes*, 18(1):67–89. [30, 32]

Sytsma, J. (2010). The proper province of philosophy: Conceptual analysis and empirical investigation. *Review of Philosophy and Psychology*, 1(3):427–445. Special issue: *Psychology and Experimental Philosophy (Part II)*, edited by Edouard Machery, Tania Lombrozo, and Joshua Knobe. [29]

Tarski, A. (1935). Einige methodologische Untersuchungen über die Definierbarkeit der Begriffe. *Erkenntnis*, 5(1). [117, 274]

Tarski, A. (1944). The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research*, 4(3):341–376. [20]

Thagard, P. (1988). *Computational Philosophy of Science*. MIT Press, Cambridge, MA. [358]

Thompson, P. (1988). Explanation in the semantic conception of theory structure. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association. Volume Two: Symposia and Invited Papers*, pages 286–296. Philosophy of Science Association, University of Chicago Press. [127]

Tooley, M. (1975). Theological statements and the question of an empiricist criterion of cognitive significance. In Diamond and Litzenburg (1975), pages 481–524. [335, 337, 338, 339, 340]

Tuomela, R. (1973). *Theoretical Concepts*. Number 10 in Library of Exact Philosophy. Springer, Wien. [62, 233, 340]

Turney, P. (1990). Embeddability, syntax, and semantics in accounts of scientific theories. *Journal of Philosophical Logic*, 19:429–451. [189, 191, 197, 209]

Uebel, T. (2008). Vienna circle. In Zalta (2008a). [87]

Väänänen, J. (2001). Second-order logic and foundations of mathematics. *The Bulletin of Symbolic Logic*, 7(4):504–520. [107, 170, 185, 187]

van Benthem, J. (1978). Ramsey eliminability. *Studia Logica*, 37:321–336. [269]

van Benthem, J. (1982). The logical study of science. *Synthese*, 51:431–472. [207, 269]

van Benthem, J. (2011). The logic of empirical theories revisited. *Synthese*. Forthcoming. [207]

Van Cleve, J. (1971). Empiricist criteria of meaningfulness. Unpublished, cited by Creath (1976). [271, 290]

van Fraassen, B. C. (1970). On the extension of Beth's semantics of physical theories. *Philosophy of Science*, 37(3):325–339. [86, 113, 197]

van Fraassen, B. C. (1972). A formal approach to the philosophy of science. In Colodny, R. G., editor, *Paradigms and Paradoxes. The Philosophical Challenge of the Quantum Domain*, volume 5 of *University of Pittsburgh Series in Philosophy of Science*, pages 303–367. University of Pittsburgh Press, Pittsburgh, PA. [107, 108]

van Fraassen, B. C. (1980). *The Scientific Image*. The Clarendon Library of Logic and Philosophy. Clarendon Press, Oxford. [57, 94, 107, 158, 165, 167, 168, 171, 190, 191, 195, 197, 207, 404, 424]

van Fraassen, B. C. (1989). *Laws and Symmetry*. The Clarendon Library of Logic and Philosophy. Clarendon Press, Oxford. [101, 107, 114, 158, 159, 171, 179, 182, 197, 198]

van Fraassen, B. C. (1991). *Quantum Mechanics: An Empiricist View*. Clarendon Press, Oxford. [190, 198] van Fraassen, B. C. (2000). The semantic approach to scientific theories. In Sklar, L., editor, *The Nature of Scientific Theory*, volume 2 of *Philosophy of Science*, pages 175–194. Garland Publishing, Inc., New York. [101]

van Fraassen, B. C. (2002). *The Empirical Stance*. The Terry Lectures. Yale University Press, New Haven, CT. [190]

van Fraassen, B. C. (2006). Representation: The problem for structuralism. *Philosophy of Science*, 73:536–547. [186, 196]

van Fraassen, B. C. (2008). *Scientific Representation: Paradoxes of Perspective*. Clarendon Press, Oxford. [190]

van Riel, R. (2011). Nagelian reduction beyond the Nagel model. *Philosophy of Science*, 78(3):353–375. [378]

Volpe, G. (1995). A semantic approach to comparative verisimilitude. *The British Journal for the Philosophy of Science*, 46(4):563–581. [158]

Wagner, P. (2009). Carnap's rational reconstruction of theories and Newman's argument. Unpublished talk presented at the *EPSA 09*, Amsterdam, October 2009. [61]

Wassermann, G. D. (1978). Testability of the role of natural selection within theories of population genetics and evolution. *British Journal for the Philosophy of Science*, 29:223–242. [214]

Waters, C. K. (1994). Genes made molecular. *Philosophy of Science*, 61(2):163–185. [69]

Waters, C. K. (2004). What concept analysis in philosophy of science should be (and why competing philosophical analyses of gene concepts cannot be tested by polling scientists). *History and Philosophy of the Life Sciences*, 26(1):29–58. [64, 65, 70]

Weinberg, J. M. and Crowley, S. (2009). The x-phi(les): Unusual insights into the nature of inquiry. *Studies in History and Philosophy of Science*, 40:227–232. [27, 28]

Weisberg, M. (2007). Who is a modeler? *British Journal for the Philosophy of Science*, 58:207–233. [136]

Wessels, L. (1974). PSA 1974: Proceedings of the 1974 biennial meeting of the philosophy of science association. In Cohen, R. S., Hooker, C. A., Michalos, A. C., and van Evra, J. W., editors, *PSA 1974: Proceedings of the 1974 Biennial Meeting of the Philosophy of Science Association*, volume XXXII of *Boston Studies in the Philosophy of Science*, pages 215–234, Dordrecht, The Netherlands. Philosophy of Science Association, D. Reidel Publishing Company. [86]

White, M. G. (1950). The analytic and the synthetic: An untenable dualism. In *John Dewey: Philosopher of Science and Freedom. A Symposium*, pages 316–330. The Dial Press, New York, NY. [409]

Whittle, A. (2008). A functionalist theory of properties. *Philosophy and Phenomenological Research*, 77:59–82. [50]

Williams, P. M. (1973). On the conservative extensions of semantical systems: A contribution to the problem of analyticity. *Synthese*, 25(3–4):398–416. [394]

Williamson, T. (2007). *The Philosophy of Philosophy*, volume 2 of *The Blackwell/Brown Lectures in Philosophy*. Blackwell Publishing, Malden, MA. [10, 12, 13, 15, 16, 29, 64]

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, 104(19):7780–7785. [257]

Winnie, J. A. (1967). The implicit definition of theoretical terms. *British Journal for the Philosophy of Science*, 18:223–229. [83]

Winnie, J. A. (1970). Theoretical analyticity. In Buck and Cohen (1970), pages 289–305. [63, 168, 394, 395]

Woit, P. (2006). *Not Even Wrong: The Failure of String Theory and the Search for Unity in Physical Law*. Basic Books, New York. [214]

Wójcicki, R. (1966). Semantical criteria of empirical significance. *Studia Logica*, 19:75–102. [271, 276, 277, 280, 282, 291, 292]

Woodger, J. H. (1939). *The Technique of Theory Construction*, volume II,5 of *Foundations of the Unity of Science. Toward an International Encyclopedia of Unified Science*. University of Chicago Press, Chicago and London. References are to the two-volume edition. [105, 187]

Woodward, J. (2008). Causation and manipulability. In Zalta (2008b). [20]

Wright, C. (1986). Scientific realism, observation and the verification principle. In Macdonald, G. and Wright, C., editors, *Fact, Science and Morality: Essays on A. J. Ayer's Language, Truth and Logic*, pages 247–274. Basil Blackwell, Oxford. [213]

Wright, C. (1989). The verification principle: Another puncture—another patch. *Mind*, 98(392):611–622. [213]

Yi, B.-U. (2001). Compact entailment and Wright's verification principle. *Mind*, 110(437):414–421. [213]

Zalta, E. N., editor (2008a). *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, Stanford, fall 2008 edition. [420, 421, 423, 425, 444]

Zalta, E. N., editor (2008b). *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, Stanford, winter 2008 edition. [433, 446]

Zalta, E. N., editor (2009a). *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, spring 2009 edition. [422, 425, 430, 431, 435]

Zalta, E. N., editor (2009b). *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, fall 2009 edition. [432, 438]

List of symbols

Symbol	Interpretation	Page
$\mathscr{V}$	vocabulary	4
$\mathscr{B}$	set of basic terms	18
$\mathscr{A}$	set of auxiliary terms	18
P_i	relation symbol	4
F_j	function symbol	4
c_k	constant symbol	4
$\bar{x}$	string or tuple	5
$\bar{x}_{-s}$	string or tuple without 'x_s'	5
$\sigma^{(\varphi)}$	relativization of sentence σ to $\varphi = \lambda x \varphi(x)$	192
$\Sigma^{(\varphi)}$	$\{\sigma^{(\varphi)} \mid \sigma \in \Sigma\}$	192
$\mathfrak{A}$	model theoretic structure: $\mathfrak{A} = \langle \lvert\mathfrak{A}\rvert, \mathscr{I} \rangle$	4
$\lvert\mathfrak{A}\rvert$	domain of $\mathfrak{A}$	4
$\mathscr{I}$	interpretation: $\mathscr{I} : \mathscr{V} \longrightarrow \mathscr{I}(\mathscr{V})$	4
$\mathfrak{A}_{\mathscr{I}}$	structure in vocabulary $\mathscr{I}$	5
$\mathfrak{A}\rvert_{\mathscr{I}}$	reduct of $\mathfrak{A}$ to $\mathscr{I}$	41
$\mathfrak{A}\lvert E_{\mathscr{I}}$	relativized reduct of $\mathfrak{A}$ to vocabulary $\mathscr{I}$ and domain $E \subseteq \lvert\mathfrak{A}\rvert$	58
$\hat{\mathfrak{A}}$	pure structure	174
$\langle A, a, \prec \rangle$	representative of a pure structure	174
$\tilde{\mathfrak{A}}$	partial structure	199
$\left\langle P_i^{\tilde{\mathfrak{A}},+}, P_i^{\tilde{\mathfrak{A}},-}, P_i^{\tilde{\mathfrak{A}},\circ} \right\rangle$	partial relation	199
$\simeq$	isomorphism	4
$\equiv$	syntactical equivalence between structures	216
$\mathrm{Th}(\mathfrak{A})$	theory of $\mathfrak{A}$: $\{\sigma \mid \mathfrak{A} \vDash \sigma\}$	216
$\vdash$	syntactic inference (deduction)	5
$\vDash$	semantic inference (entailment)	5
$\top$	tautology	
$\bot$	contradiction	

Symbol	Interpretation	Page
$\Theta\|_{\mathscr{S}}$	$\mathscr{S}$-consequences of Θ	75
$\circ$	function concatenation	5
$\complement$	complement	75
$\mathrm{card}(A)$	cardinality of A	181
$\mathbf{\Phi}$	$\mathbf{\Phi}(\Sigma) := \{\mathfrak{S} \mid \mathfrak{S} \models \Sigma\}$	168
$\mathbf{\Psi}$	$\mathbf{\Psi}(\mathbf{S}) := \{\varphi \mid \mathfrak{S} \models \varphi$ for all $\mathfrak{S} \in \mathbf{S}\}$	169
Π	set of meaning postulates (analytic sentences)	21
Π	set of supplementary sentences	257
$\Pi_{\mathscr{B}}$	set of meaning postulates in $\mathscr{B}$ (analytic $\mathscr{B}$-sentences)	40
$\mathbf{M}$	class of possible $\mathscr{V}$-structures	41
$\mathbf{M}_{\mathscr{B}}$	class of possible $\mathscr{B}$-structures	40
$\mathbf{N}$	class of intended structures	44
$\mathbf{N}_{\mathscr{B}}$	class of intended $\mathscr{B}$-structures	41
$\mathsf{R}_{\mathscr{I}}(\sigma)$	Ramsey sentence generalizing on all terms in $\mathscr{V} - \mathscr{I}$	43
$\mathsf{R}_{\mathscr{I}}(\Sigma)$	$\mathsf{R}_{\mathscr{I}}(\bigwedge \Sigma)$	43
$\mathsf{C}_{\mathscr{I}}(\sigma)$	Carnap sentence: $\mathsf{C}_{\mathscr{I}}(\sigma) \to \sigma$	43
$\mathsf{C}_{\mathscr{I}}(\Sigma)$	$\mathsf{C}_{\mathscr{I}}(\bigwedge \Sigma)$	43
$\Theta(B_1,\ldots,B_m,A_1,\ldots,A_n)$	theory Θ with terms $\{B_1,\ldots,B_m,A_1,\ldots,A_n\}$	43
$\Theta(\bar{B},\bar{A})$	$\Theta(B_1,\ldots,B_m,A_1,\ldots,A_n)$	43
$R^{\Theta}_{\mathscr{B},P_i}$	Ramsey constant: $\exists\bar{X}\,[\Theta(\bar{B},\bar{X}) \wedge X_i\bar{x}]$	277
$Я^{\Theta}_{\mathscr{B},P_i}$	reverse Ramsey constant: $\exists\bar{X}\,[\Theta(\bar{B},\bar{X}) \wedge \neg X_i\bar{x}]$	278
$\mathrm{Syn}(\vartheta)$	synthetic component of ϑ	393
$\mathrm{An}(\vartheta)$	analytic component of ϑ	393
P_i^{+}	positive extension of P_i	44
P_i^{-}	negative extension of P_i	44
P_i°	neutral extension of P_i	44
$F_j^{+\circ}$	non-negative extension of F_j	45
$c_k^{+\circ}$	non-negative extension of c_k	45
$P_i^{\mathfrak{A}_{\mathscr{B}},\Theta,+}$	positive extension of P_i relative to $\mathscr{B}$ and Θ in $\mathfrak{A}_{\mathscr{B}}$	277
$P_i^{\mathfrak{A}_{\mathscr{B}},\Theta,-}$	negative extension of P_i relative to $\mathscr{B}$ and Θ in $\mathfrak{A}_{\mathscr{B}}$	277
$W(P_i,F_j,c_k)_{i\in I,j\in J,k\in K}$	penumbral connections	45
$\mathbf{A}$	vagueness set	45

Symbol	Interpretation	Page
$\{\mathfrak{T}_n\}_{n\in N}$	theory (as defined by van Fraassen)	190
$\mathbf{E}_n$	set of empirical substructures of $\mathfrak{T}_n \in \{\mathfrak{T}_n\}_{n\in N}$	190
$\mathbf{P}$	set of appearances	190
$\mathbf{C}_{\Sigma}$	content of Σ	231
$\mathscr{E}\mathscr{F}f$	$\mathscr{E} \cup \mathscr{F} \cup \{f\}$	240
EF	$\lambda x(Ex \vee Fx)$	240
$\mathfrak{A}\|EF_{\mathscr{E}\mathscr{F}f}$	relativized reduct to $\mathscr{E}\mathscr{F}f$ and EF	240
$\mathbf{\Omega}$	set of observation sentences	255
$\mathbf{\Omega}$	set of sets of observational structures	255
$\mathbf{V_i}^{\mathfrak{A}_{\mathscr{B}},\Theta}$	set of $\mathscr{B}$-possible extensions of V_i	282
$\boldsymbol{\mathscr{I}}^{\mathfrak{A}_{\mathscr{B}},\Theta}$	set of $\mathscr{B}$-possible $\mathscr{I}$-structures	290
$\rightsquigarrow$	bringing about by law	381
ET	evolutionary theory	304
ID	intelligent design	3
ADAP	‘Organisms display complex adaptations’	343
CSI	complex specific information	346
“. . . ”	quotation	3
‘. . . ’	name of phrase between the marks	4
$\ulcorner \ldots \urcorner$	quasi-quotation	4

Index

Curriculum vitae

Sebastian Lutz

Ph. D., Theoretical Philosophy, Utrecht University, The Netherlands, 2012.

Dissertation:	*Criteria of Empirical Significance: Foundations, Relations, Applications*
Supervisors:	Thomas Müller (Theoretical Philosophy, Utrecht University) Janneke van Lith (Theoretical Philosophy, Utrecht University) Albert Visser (Theoretical Philosophy, Utrecht University)
Committee:	Otávio Bueno (Philosophy, University of Miami) Dennis Dieks (Institute for History and Foundations of Science, Utrecht University) Hannes Leitgeb (Munich Center for Mathematical Philosophy, Ludwig-Maximilians-Universität München) F. A. Muller (Faculty of Philosophy, Erasmus University Rotterdam) Paul Ziche (History of Philosophy, Utrecht University)

Visiting Fellow, Tilburg Center for Logic and Philosophy of Science, Tilburg University, The Netherlands, 2007–2008.

Graduate studies, Philosophy, University of Western Ontario, Canada, 2005–2008

Diplom, Physics, Universität Hamburg, Germany, 2004.

Thesis:	*Dekohärenz komplexer Quantengatter* [Decoherence of Complex Quantum Gates]
Supervisor:	Sigmund Kohler (Theoretische Physik I, Universität Augsburg)
Primary referee:	Peter Hänggi (Theoretische Physik I, Universität Augsburg)
Secondary referee:	Daniela Pfannkuche (I. Institut für theoretische Physik, Universität Hamburg)

Visiting Researcher, Theoretical Physics I, Universität Augsburg, Germany, 2003–2005.

Visiting Student Fellow, Minnesota Center for Philosophy of Science, University of Minnesota–Twin Cities, USA, 2001–2002.

Vordiplom, Physics, Universität Bayreuth, Germany, 2000.

Zwischenprüfung, Philosophy, Universität Bayreuth, Germany, 2000.

Samenvatting

(Summary in Dutch)

Deze dissertatie bevat drie delen.[12] Deel I is een verdediging van kunsttalige filosofie (een methodologie van filosofie op basis van kunstmatige taal), en een historische en systematische verdediging van de toepassing van kunsttalige filosofie op wetenschappelijke theorieën in het logisch empirisme. Deze verdedigingen steunen de vooronderstellingen van een groot aantal criteria voor empirische significantie die ik in deel II analyseer, vergelijk en ontwikkel. Op basis van deze analyse hanteer ik in deel III verschillende criteria om de wetenschappelijke status van Intelligent Design te beoordelen en verder noties van confirmatie, reductie en begripsvorming te bediscussiëren.

I. Grondslagen

Grondslagen van de wijsbegeerte

Ik beargumenteer in §2.4 en §2.6 dat kunsttalige filosofie (ook 'ideaaltalige filosofie' genoemd) voordelen heeft in vergelijking met de concurrerende methodologieën van filosofie gebaseerd op intuïties (§2.1), natuurtalige filosofie (§2.2), experimentele filosofie en genaturaliseerde filosofie (§2.5). Kunsttalige filosofie gebruikt de keuze van taal om filosofische problemen op te lossen; deze keuze determineert welke zinnen analytisch waar zijn, maar heeft zelf geen feitelijke implicaties, zodat de resultaten van kunsttalige filosofie analytisch zijn. Alhoewel de keuze van de taal ook de keuze van de logica met zich mee kan brengen is het voor mijn discussie voldoende een vaste logica en een vaste basis-woordenschat te veronderstellen, en te veronderstellen dat alleen de betekenispostulaten van een supplementair-woordenschat kunnen worden gekozen. In dit geval komt de keuze van een taal neer op de vorming van begrippen. Ik bediscussieer in §2.8 de semantiek van begripsvorming zoals ontwikkeld door Przełęcki (1969, §§4–6) en beargumenteer

[12] Ik dank Janneke van Lith voor het verbeteren van mijn Nederlands.

dat kunsttalige filosofie methodologisch naturalistisch is, omdat de wetenschappen ook uitgebreid op de keuze van taal berusten (§2.10). Daarom steunt kunsttalige filosofie ook veel waargenomen vruchtbare interacties tussen wetenschap en wetenschapsfilosofie (§2.11). Verder maakt kunsttalige filosofie ook de herinterpretatie en productieve gebruik van resultaten van de andere vijf reeds vermelde methodologieën (§2.7). Als voorbeeld van de toepassing van kunsttalige filosofie expliciteer ik de begrippen 'abstractie' en 'idealisatie' (§2.12).

Grondslagen van wetenschapsfilosofie

De "Received View" op wetenschappelijke theorieën van Carnap, Hempel en Feigl is een toepassing van kunsttalige filosofie op wetenschappelijke theorieën. Ik beargumenteer dat de sterke kritiek die tot de ondergang van de Received View leidde op valse veronderstellingen rust. In het bijzonder is de Received View niet gebaseerd op uitputtende axiomatisering (§3.4) in logica van de eerste orde (§3.3) en heeft hij in feite vaak dezelfde formalisering van wetenschappelijke theorieën gebruikt als zijn critici (§3.5). De Received View vereenvoudigt de daadwerkelijke relatie tussen theorieën en observaties ook niet te sterk, noch introduceert hij in dit opzicht onnodige complicaties (§3.6). Hij is ook niet vijandig tegen het gebruik van modellen in wetenschap (§3.7). Voorts heeft de Received View niet in zijn poging gefaald om het begrip van een theorie nauwkeuriger te maken, omdat hij niet bedoeld is dit te doen. In tegendeel: Ten eerste was hij bedoeld als generaliseerbaar kader waarin men *specifieke* theorieën kan expliciteren (§3.8.2). Ten tweede verschilt explicatie van precisering (§3.8.1).

In aanvulling op deze historische verdediging beargumenteer ik dat de syntactische benaderingen in het algemeen net zo krachtig zijn als concurrerende semantische benaderingen (§4.1). In het bijzonder kan de Received View van Fraassens begrippen van een theorie en van empirische adequaatheid vangen en veralgemeniseren (§4.2). Hij kan verder de kernbegrippen van de partial structures benadering op twee verschillende manieren vangen (§4.3).

II. Relaties

Het doel van criteria van empirische significantie is te onderscheiden tussen zinnen of begrippen die met empirische eisen of toestanden zijn verbonden en diegene die dat niet zijn. Dergelijke criteria kunnen onderverdeeld worden in criteria voor de empirische significantie van zinnen en criteria voor de empirische significantie van begrippen. Langs een andere kant kan men onderscheiden tussen deductieve en probabilistische criteria, die deductieve respectievelijk probabilistische inferenties veronderstellen.

Ik laat zien dat alle belangrijke criteria voor zinnen equivalent zijn met falsifieerbaarheid (§6.2), verifieerbaarheid (§6.3), hun disjunctie (§6.6) of superveniëntie

(§6.5). Op die equivalenties voortbouwend geef ik vergelijkende criteria van empirische significantie (§6.7.1) en geef ik een zeer algemene notie van empirische eis en empirische feit (§6.8.1), en ontwikkel ik een notie van supplementaire vooronderstellingen die de problemen van Ayers criterium en zijn opvolgers vermijdt (§6.8.2). Verder laat ik zien dat de meest belangrijke deductieve criteria voor begrippen equivalent of gelijkaardig zijn aan expliciete definieerbaarheid, superveniëntie van begrippen (§7.2), reduceerbaarheid (§7.3), of een criterium dat onafhankelijk door William Rozeboom en door Carnap werd voorgesteld (§7.4). Gebaseerd op deze resultaten veralgemeniseer ik sommige van deze criteria tot criteria voor verzamelingen van begrippen (§7.5). Ik beargumenteer dat die criteria voor begrippen die niet slechts een middel zijn om criteria voor zinnen opnieuw te formuleren überhaupt niet kunnen worden gebruikt om empirisch significante zinnen te identificeren. Nochtans kunnen sommige criteria voor begrippen gebruikt worden om theoretische concepten van wiskundige concepten te onderscheiden (§7.6). Ik toon verder aan dat Sobers (probabilistische) criterium van testbaarheid ontoereikend is volgens zijn eigen voorwaarden van adequaatheid en beveel in plaats daarvan twee verschillende criteria aan: het relevantie-criterium van probabilistische significantie, dat het probabilistische analogon van falsifieerbaarheid is, en het Bayesiaanse criterium, het analogon van de disjunctie van falsifieerbaarheid en verifieerbaarheid (§8).

III. Toepassingen

Ik pas de resultaten van delen I en II op vier verschillende gebieden toe. In een directe toepassing van deductieve en probabilistische criteria voor zinnen analyseer ik de structuur van de theorie van Intelligent Design (ID) en laat zien dat, in zijn meest aannemelijke formuleringen, ID noch falsifieerbaar is noch probabilistisch relevant. Dit resultaat is een indicatie dat ID geen wetenschap is (§9).

Door de relatie tussen probabilistische criteria van empirische significantie en het begrip van testbaarheid leidt het falen van likelihoodisme om een criterium van empirische significantie te geven tot een kritiek van likelihoodisme als een explicatie van confirmatie. Op analoge wijze leidt het succes van het relevantie-criterium en het Bayesiaanse criterium van empirische significantie tot het relevantie-criterium van confirmatie en, respectievelijk, het Bayesianisme zelf (§10). Verder kunnen de deductieve criteria voor zinnen en sommige criteria voor begrippen gebruikt worden in de explicatie van verschillende begrippen van reductie (§11).

Tot slot kunnen de criteria van empirische significantie zelf worden gebruikt om het formalisme van de kunsttalige filosofie te analyseren, te veralgemeniseren en te verbeteren. Ik geef een overzicht van een dergelijke ontwikkeling en pas sommige van de resultaten toe op een kwestie van conceptvorming in twee ethische theorieën (§12).

Quaestiones Infinitae

Publications of the Z∃NO Institute of Philosophy

VOLUME 20. H. H. A. VAN DEN BRINK, *The Tragedy of Liberalism* (dissertation), 1997.
VOLUME 21. D. VAN DALEN, *Torens en Fundamenten* (valedictory lecture), 1997.
VOLUME 22. J. A. BERGSTRA, W. J. FOKKINK, W. M. T. MENNEN, S. F. M. VAN VLIJMEN, *Spoorweglogica via EURIS*, 1997.
VOLUME 23. I. M. CROESE, *Simplicius on Continuous and Instantaneous Change* (dissertation), 1998.
VOLUME 24. M. J. HOLLENBERG, *Logic and Bisimulation* (dissertation), 1998.
VOLUME 25. C. H. LEIJENHORST, *Hobbes and the Aristotelians* (dissertation), 1998.
VOLUME 26. S. F. M. VAN VLIJMEN, *Algebraic Specification in Action* (dissertation), 1998.
VOLUME 27. M. F. VERWEIJ, *Preventive Medicine Between Obligation and Aspiration* (dissertation), 1998.
VOLUME 28. J. A. BERGSTRA, S. F. M. VAN VLIJMEN, *Theoretische Software-Engineering: kenmerken, faseringen en classificaties*, 1998.
VOLUME 29. A. G. WOUTERS, *Explanation Without A Cause* (dissertation), 1999.
VOLUME 30. M. M. S. K. SIE, *Responsibility, Blameworthy Action & Normative Disagreements* (dissertation), 1999.
VOLUME 31. M. S. P. R. VAN ATTEN, *Phenomenology of choice sequences* (dissertation), 1999.
VOLUME 32. V. N. STEBLETSOVA, *Algebra, Relations and Geometries (an equational perspective)* (dissertation), 2000.
VOLUME 33. A. VISSER, *Het Tekst Continuüm* (inaugural lecture), 2000.
VOLUME 34. H. ISHIGURO, *Can we speak about what cannot be said?* (public lecture), 2000.
VOLUME 35. W. HAAS, *Haltlosigkeit; Zwischen Sprache und Erfahrung* (dissertation), 2001.
VOLUME 36. R. POLI, *ALWIS: Ontology for knowledge engineers* (dissertation), 2001.
VOLUME 37. J. MANSFELD, *Platonische Briefschrijverij* (valedictory lecture), 2001.
VOLUME 37a. E. J. BOS, *The Correspondence between Descartes and Henricus Regius* (dissertation), 2002.
VOLUME 38. M. VAN OTEGEM, *A Bibliography of the Works of Descartes (1637-1704)* (dissertation), 2002.
VOLUME 39. B. E. K. J. GOOSSENS, *Edmund Husserl: Einleitung in die Philosophie: Vorlesungen 1922/23* (dissertation), 2003.
VOLUME 40. H. J. M. BROEKHUIJSE, *Het einde van de sociaaldemocratie* (dissertation), 2002.
VOLUME 41. P. RAVALLI, *Husserls Phänomenologie der Intersubjektivität in den Göttinger Jahren: Eine kritisch-historische Darstellung* (dissertation), 2003.
VOLUME 42. B. ALMOND, *The Midas Touch: Ethics, Science and our Human Future* (inaugural lecture), 2003.
VOLUME 43. M. DÜWELL, *Morele kennis: over de mogelijkheden van toegepaste ethiek* (inaugural lecture), 2003.
VOLUME 44. R. D. A. HENDRIKS, *Metamathematics in Coq* (dissertation), 2003.

VOLUME 45. TH. VERBEEK, E. J. BOS, J. M. M. VAN DE VEN, *The Correspondence of René Descartes: 1643*, 2003.
VOLUME 46. J. J. C. KUIPER, *Ideas and Explorations: Brouwer's Road to Intuitionism* (dissertation), 2004.
VOLUME 47. C. M. BEKKER, *Rechtvaardigheid, Onpartijdigheid, Gender en Sociale Diversiteit; Feministische filosofen over recht doen aan vrouwen en hun onderlinge verschillen* (dissertation), 2004.
VOLUME 48. A. A. LONG, *Epictetus on understanding and managing emotions* (public lecture), 2004.
VOLUME 49. J. J. JOOSTEN, *Interpretability formalized* (dissertation), 2004.
VOLUME 50. J. G. SIJMONS, *Phänomenologie und Idealismus: Analyse der Struktur und Methode der Philosophie Rudolf Steiners* (dissertation), 2005.
VOLUME 51. J. H. HOOGSTAD, *Time tracks* (dissertation), 2005.
VOLUME 52. M. A. VAN DEN HOVEN, *A Claim for Reasonable Morality* (dissertation), 2006.
VOLUME 53. C. VERMEULEN, *René Descartes, Specimina philosophiae: Introduction and Critical Edition* (dissertation), 2007.
VOLUME 54. R. G. MILLIKAN, *Learning Language without having a theory of mind* (inaugural lecture), 2007.
VOLUME 55. R. J. G. CLAASSEN, *The Market's Place in the Provision of Goods* (dissertation), 2008.
VOLUME 56. H. J. S. BRUGGINK, *Equivalence of Reductions in Higher-Order Rewriting* (dissertation), 2008.
VOLUME 57. A. KALIS, *Failures of agency* (dissertation), 2009.
VOLUME 58. S. GRAUMANN, *Assistierte Freiheit* (dissertation), 2009.
VOLUME 59. M. AALDERINK, *Philosophy, Scientific Knowledge, and Concept Formation in Geulincx and Descartes* (dissertation), 2010.
VOLUME 60. I. M. CONRADIE, *Seneca in his cultural and literary context: Selected moral letters on the body* (dissertation), 2010.
VOLUME 61. C. VAN SIJL, *Stoic Philosophy and the Exegesis of Myth* (dissertation), 2010.
VOLUME 62. J. M. I. M. LEO, *The Logical Structure of Relations* (dissertation), 2010.
VOLUME 63. M. S. A. VAN HOUTE, *Seneca's theology in its philosophical context* (dissertation), 2010.
VOLUME 64. F. A. BAKKER, *Three Studies in Epicurean Cosmology* (dissertation), 2010.
VOLUME 65. T. FOSSEN, *Political legitimacy and the pragmatic turn* (dissertation), 2011.
VOLUME 66. T. VISAK, *Killing happy animals. Explorations in utilitarian ethics.* (dissertation), 2011.
VOLUME 67. A. JOOSSE, *Why we need others: Platonic and Stoic models of friendship and self-understanding* (dissertation), 2011.
VOLUME 68. N. M. NIJSINGH, *Expanding newborn screening programmes and strengthening informed consent* (dissertation), 2012.
VOLUME 69. R. PEELS, *Believing Responsibly: Intellectual Obligations and Doxastic Excuses* (dissertation), 2012.
VOLUME 70. S. LUTZ, *Criteria of Empirical Significance: Foundations, Relations, Applications* (dissertation), 2012.

Typeset with pdfLaTeX 2_ε in 10pt Garamond. LaTeX 2_ε is maintained by the LaTeX Team and Leslie Lamport, pdfTeX is maintained by the pdfTeX Team and Hàn Thế Thành. The Garamond font was designed by URW++, the package is maintained by Walter Schmidt.

Packages used	
Package	Maintainer
amsmath	The American Mathematical Society
amsthm	The American Mathematical Society
babel	Johannes L. Braams
booktabs	Simon Fear and Danie Els
calc	Frank Jensen and Kresten Krab Thorup
caption	Axel Sommerfeldt
ccicons	Michael Ummels
color	David Carlisle
emptypage	Karl Wette
enumerate	David Carlisle
fancyhdr	Piet van Oostrum
geometry	Hideo Umeki
graphicx	David Carlisle
hyperref	Sebastian Rahtz and Heiko Oberdiek
inputenc	Frank Mittelbach and Alan Jeffrey
longtable	David Kastrup and David Carlisle
ltxtable	David Carlisle
makeidx	The LaTeX Team
mathdesign	Paul Pichaureau
mdwtab	Mark Wooding
microtype	R. Schlicht
natbib	Arthur Ogawa and Patrick W. Daly
tabularx	David Carlisle
tikz	Christian Feuersänger and Till Tantau
tocloft	Peter R. Wilson and Will Robertson
varwidth	Donald Arseneau
xspace	Morten Høgholm and David Carlisle