Criteria of Empirical Significance: A Success Story

Sebastian Lutz

Preprint: 2013-03-27

Abstract

The sheer multitude of criteria of empirical significance has been taken as evidence that the pre-analytic notion being explicated is too vague to be useful. I show instead that a significant number of these criteria—by Ayer, Popper, Przełęcki, Suppes, and David Lewis, among others—not only form a coherent whole, but also connect directly to the theory of definition, the notion of empirical content as explicated by Ramsey sentences, and the theory of measurement; a criterion by Carnap is trivial, but can be saved and connected to the other criteria by slight modifications. A corollary is that the ordinary language defense of Lewis, the conceptual arguments by Ayer and Popper, the theoretical considerations by Przełęcki, and the practical considerations by Suppes all apply to the same criterion or closely related criteria. The equivalences of some criteria allows for their individual justifications to be taken cumulatively and suggest a variety of further lines of inquiry, for instance into analyticity and empirical equivalence. The inferential relations between the non-equivalent criteria suggest comparative notions of empirical significance. In a short case study, I discuss the debate about realism, structural realism, and antirealism.

Keywords: empirical significance; cognitive significance; testability; meaningfulness; empirical content; definition; measurement; falsifiability; verifiability; aboutness; supervenience; empirical equivalence; realism; structural realism; antirealism

Contents

1 Introduction **2**

2 Preliminaries **3**
2.1 Methodological presumptions 3
2.2 On the explicandum 7

3 Falsifiability **8**
3.1 Syntactic criteria 8
3.2 Semantic criteria 12

4 **Verifiability** 17

5 **Falsifiability and verifiability** 18

6 **Strong $\mathscr{B}$-determinacy** 19

7 **Weak $\mathscr{B}$-determinacy** 28

8 **Import of the relations** 33
8.1 Consolidating the justifications 35
8.2 The non-arbitrariness of the classes of empirically significant sentences 37

9 **The fruitfulness of the criteria** 39
9.1 Six comparative concepts of empirical significance 40
9.2 The relation of the criteria to other concepts 42
9.3 A short application: Empiricism and realism, structural realism, and antirealism 44

10 **Conclusion** 47

1 Introduction

Criteria of empirical significance are meant to demarcate those statements or terms that have some connection to empirical statements from those that do not. An early criterion suggested by Ayer was quickly shown to be trivial, and proposed amendments did not fair better. This history has "done a lot to discredit the very idea of delineating a class of statements as empirical" (Lewis 1988a, §I) and, as Lewis (1988a, 127, footnote removed) further notes, has

> 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 charge of arbitrariness 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.

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 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. Specifically, several criteria of empirical significance are equivalent (§3.1) or nearly equivalent (§3.2) to falsifiability, which is also the non-trivial core of Ayer's criteria. Falsifiability in turn is closely connected to verifiability (§4). Falsifiability and verifiability are more inclusive than the (universally panned) criterion demanding both (§5), which is itself more inclusive than the criterion of strong $\mathscr{B}$-determinacy, suggested independently by Patrick Suppes, Marian Przełęcki, and David Lewis (§6). 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 (§7). Falsifiability, verifiability, their disjunction, and strong $\mathscr{B}$-determinacy thus make up the four major criteria of empirical significance. 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 (§8).

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 (reviewed in §2), 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 the discussions of the empirical significance of claims about a designer of life whose intentions and abilities are unknown (Sober 2008, Lutz 2012c). In other cases, a generally accepted endeavor is put under scrutiny, like string theory (Smolin 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.

2 Preliminaries

2.1 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* (cf. Kuipers 2007). According to Carnap (1950, 7), the criterion of empirical significance (the *explicatum*) should be similar to the intuitive concept (the *explicandum*), and furthermore precise, fruitful, and as simple as possible given the preceding desiderata.

An explicatum does not have to agree with its explicandum within the bounds of the latter's vagueness (Lutz 2012d, §5.1),[1] and thus the demand for similarity cannot entail the demand that the explicatum 'empirical significance' must apply to all intuitively empirically significant sets of sentences or must exclude all intuitively non-significant sets of sentences.[2] The demand for similarity to the explicandum can rather be captured by conditions of adequacy, which are suggested by the pre-theoretic use of the explicandum and identify in what contexts the explicatum should be applicable (cf. Tarski 1944, §4; Kuipers 2007, §2; Lutz 2012a, §3). Carnap explicates 'fruitful' as 'useful for the formulation of many universal statements'[3] and Kemeny (1963, 76) adds that a fruitful concept should also suggest many new research questions. The idea underlying Carnap's and Kemeny's desiderata is expressed by Hempel (1952, 663), who demands, with reference to Carnap, that "it should be possible to develop, in terms of the reconstructed concepts, a comprehensive [...] and sound theoretical system". The fruitfulness of an explicatum will depend on the goals of the explication, and similarly the conditions of adequacy will be chosen according to the intended uses of the explicata. Thus explications are not true or false claims, but more or less expedient suggestions (Popper 1935, 37–38; Hempel 1952, 663).

Concepts are typically explicated in a restricted domain. For instance, Tarski 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, 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. Especially opponents of the syntactic view on theories (as developed by the logical empiricists) will consider this an extreme restriction. It is part of philosophical folklore that the syntactic view failed because of its reliance on predicate logic, and has now been completely superseded by the semantic view, which relies only on set- or model theory. This is less of a problem than it might seem, for, first, the equivalences discussed here will suggest immediate generalizations beyond predicate logic. Second, not all major criticisms of the syntactic view are in fact justified (Lutz 2012d). Third, it is doubtful that the use of (higher order) predicate logic poses more restrictions on the formalization of theories than the use of set- or model theory (Lutz 2012b, §4.1). In fact, 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 $\mathscr{V}$ of the language into basic terms $\mathscr{B}$ and auxil-

[1] In other words, not every explication is a precisification.

[2] Hence I will also not try to determine the criteria's adequacy by applying them to examples.

[3] Unqualified, Carnap's explication is a bit silly. I think it is clear that he meant 'useful for the formulation of many *important* universal statements', or something similar.

iary terms $\mathscr{A}$,[4] with sentences containing $\mathscr{V}$-terms called $\mathscr{V}$-sentences, sentences containing only $\mathscr{B}$-terms called basic or $\mathscr{B}$-sentences, and sentences containing only $\mathscr{A}$-terms called auxiliary or $\mathscr{A}$-sentences.[5] This assumption is implausible for ordinary languages, and has been criticized in this regard (Putnam 1962). But the explicata assume an artificial language that is designed for a specific purpose, which in this case is the analysis of the relation between theories and observations. And there is no reason to assume that it is impossible to develop such a language (Suppe 1972, §I), where it is encapsulated in the vocabulary what is or is not observable (cf. Przełęcki 1974a, §III). Przełęcki (1969, §10.II) suggests to achieve this result by simply taking all terms in the sciences as auxiliary and introducing an artificial basic language. A similar strategy can be used to capture the notion of an empirical substructure with the help of a bipartition of the vocabulary (Lutz 2012b, §4.2). (This is of specific interest because of van Fraassen's influential conjecture that this is impossible (van Fraassen 1980, §3.6).) Furthermore, the bipartition need not stay fixed, but may change depending on the context (Rozeboom 1970, 201–203, Lewis 1970, 428). Reichenbach (1951, 49) suggests that the basic sentences should be assumed to have "primitive meaning, i. e., a meaning which is not under investigation during the analysis to be performed". Under this suggestion, basic sentences do not have to be about observations in any sense of the word, but must only be unproblematic for the purposes at hand. Specifically, it must be uncontroversial for any object or tuple of objects whether it is in the extension of a basic term, and it must be uncontroversial under what circumstances a $\mathscr{B}$-sentence is true and under what circumstances it is false.[6]

All of the criteria in the following also refer to a consistent set of *analytic sentences* or *meaning postulates* Π, sometimes bipartitioned into meaning postulates $\Pi_{\mathscr{B}}$ for $\mathscr{B}$-terms (cf. Carnap 1952), and meaning postulates $\Pi_{\mathscr{A}}$ for $\mathscr{A}$-terms. The meaning postulates for $\mathscr{B}$-terms are $\mathscr{B}$-sentences, while those for $\mathscr{A}$-terms are $\mathscr{V}$-sentences, because they give the $\mathscr{A}$-terms' relations to each other and to $\mathscr{B}$-terms. Przełęcki (1974a, 345) argues that $\Pi_{\mathscr{A}}$ should be $\mathscr{B}$-conservative with respect to $\Pi_{\mathscr{B}}$, 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 do assume that Π is closed under entailment, so that any set Λ of sentences with $\Pi \vDash \Lambda$ is *analytic* (*analytically true*). Any set of sentences incompatible with Π is *analytically false*; analytically true sets and analytically false sets are *analytically determined*. Note that under this definition, logically determined sets are also analytically determined. A set not analytically determined is *analytically contingent*. Finally, Γ *analytically entails* Λ if and only if $\Gamma \cup \Pi \vDash \Lambda$. Here and in the following, a definition for sets of sentences holds for a single sentence if and only if they hold for the

[4]I will follow the tradition in the philosophy of science and call all non-logical constants 'terms'. It is unfortunate that this is not the common usage in symbolic logic.

[5]Of course, both $\mathscr{B}$ and $\mathscr{A}$-sentences can also contain logical constants.

[6]This entails that all $\mathscr{B}$-terms must be perfectly precise, an assumption that again is plausible only if the language is artificial.

sentence's singleton set. Once again, the assumption of a clearly delineated set of analytic sentences is plausible for artificial languages (cf. Mates 1951, Martin 1952, Kemeny 1963), and it is important to keep in mind that for an artificial language, analytic sentences are not found to be true, but *chosen* to be true (Lutz 2012a, §3). The assumption of a set of analytic sentences is also not obviously a restriction, since Π may be empty. On the other hand, letting $\Pi = \varnothing$ severely restricts the inferences that are possible, excluding, for example, the inference from 'function f is linear' to 'function f is continuous'.

A non-trivial example of analytic sentences with a scientific flavor is given by the introduction of a mass-function based on comparative measurements. Assume that, for instance by using a balance scale, one has decided that the binary $\mathscr{B}$-predicate 'at least as heavy as' (written as '$\precsim$') and the binary $\mathscr{B}$-function 'the physical combination of' (written as '$*$') should obey the axioms for extensive quantities. This would mean that the sentences '$\forall x \forall y \forall z (x \precsim y \land y \precsim z \rightarrow x \precsim z)$', '$\forall x \forall y \forall z [(x * y) * z \precsim x * (y * z)]$', and others are in $\Pi_{\mathscr{B}}$ (cf. Przełęcki 1974a, 348).[7] Based on these choices, one may further decide to introduce a numerical quantity 'mass' (written as 'm') to the $\mathscr{T}$-vocabulary and add the analytic $\mathscr{V}$-sentences '$\forall x \forall y (x \precsim y \leftrightarrow mx \leq my)$' and '$\forall x \forall y (m(x * y) = mx + my)$' to Π.[8] A scientific theory described by a set Σ of sentences (e. g., Newtonian mechanics) may accordingly refer only to the mass of objects, without the need to state its claims in the comparably cumbersome basic terms. The connection to basic (observational) claims is rather given by Π.[9]

In general, analytic sentences Π determine which $\mathscr{B}$-sentences can be true within the chosen language, which suggests

Definition 1. A *set of $\mathscr{V}$-sentences Γ* is *possible* if and only if $\Gamma \cup \Pi$ has a model.

In other words, a set of $\mathscr{V}$-sentences is possible if and only if it is compatible with Π. Specifically, Π may restrict the sets of observation sentences that are compatible with the rules of the language.

Unless it is tautologous, Π also puts restrictions on the possible interpretations of terms, so that, say, every function in the extension of 'linear' must also be in the extension of 'continuous'. Π therefore may restrict how the interpretations of basic terms relate. 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). Any $\mathscr{B}$-structure that does not have an expansion to a model of Π is then impossible. Since a $\mathscr{V}$-structure is its own expansion, one can give

[7] The interplay of statistics, convention, and errors of observations is subtle when introducing analytic $\mathscr{B}$-sentences. Kyburg (1984, ch. 2–5, §5.3) and Carnap (1926, 16–18) provide analyses of random and, respectively, systematic errors that fit with the discussion here.

[8] The status of '$\leq$' and '$+$' will be discussed in §6.

[9] Carnap (1926, II.E; 1966, ch. 5–7) and Hempel (1952, ch. 12) give early discussions of the determination of mass along these lines. Simpson (1981) gives a more recent discussion.

Definition 2. A *structure* is *possible* if and only if it can be expanded to a model of Π.

It is clear that a set of $\mathscr{V}$-sentences is possible if and only if it is true in a possible structure.

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 ($\mathrm{Th}(\mathfrak{A}) \models\!\!\!\dashv \mathrm{Th}(\mathfrak{B})$), that is, for all sentences φ, $\mathfrak{A} \models \varphi$ if and only if $\mathfrak{B} \models \varphi$. In first order logic, syntactic equivalence is called 'elementary equivalence' (and is not equivalent to isomorphism).

2.2 On the explicandum

For now, I will only give a circumscription of the explicandum precise enough to counter some common criticisms. A somewhat more thorough discussion follows in §8.2.

First, 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 (1956) considers criteria for terms possible and perhaps even preferable to criteria for sentences (see also Hempel 1965b, §3), Przełęcki (1974a, 345–346), for example, considers such criteria misguided. And if 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.[10]

The criteria are also not meant to determine the meaning of sentences as Ruja (1961), for instance, assumes in his critique. Rynin (1957, 51–53) and Gemes (1998, §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 can be empirically significant or not), while a criterion of meaning has to define a relation between sentences and meanings.

Pace Rynin (1957, 51), 'empirical significance' does not explicate 'meaningfulness', either, because the meaning of a sentence is generally accepted to be determined by both the sentence's empirical import and the rules that govern its use with other sentences (Carnap 1939, §25). Thus even a sentence not connected in the slightest to observation can be meaningful (cf. Sober 2008, 149–150). Whether there is more to the meaning of sentences beyond their empirical import and relation to other sentences depends on the status of semantic empiricism,

[10] Arguably, most criteria for the empirical significance of terms are meant to provide the basis for criteria for the empirical significance of sentences; Carnap (1956, 60, D3) for example, uses his criterion for terms in this way. However, the logical structure of the resulting criteria for sentences is typically so different from that of the criteria discussed here that the analysis of the criteria for terms would lead too far afield (Lutz 2012b, §7).

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 his criterion as a demarcation criterion between empirical and non-empirical sentences (Popper 1935, §4, §9; cf. Carnap 1963, §6.A). Incidentally, if a sentence can be meaningful without being empirically significant, most of the criticisms by Hempel (1965b) are invalid (Hempel 1965c; Sober 2008, 149–150).

Gemes (1998, §1.4) argues that a criterion of empirical significance does not have to be a criterion of inductive confirmability. In connection with claim 8 below, 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 probabilistic inferences are possible. In this article, I will assume that the criteria are meant to be applied in contexts that allow only deductive inferences, and thus are specifically not meant 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 1965b, §2) is the best example of this. In its stead, I will 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).

3 Falsifiability

3.1 Syntactic criteria

Hempel's formulation of the "requirement of complete falsifiability in principle" (Hempel 1965b, 106) can serve as a good starting point of my discussion:

> 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. 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.[11] Finally, I will generalize the crite-

[11]This focus on the empirical significance of sentences with $\mathscr{A}$-terms is not as ahistorical as it may seem: Carnap (1928b, 325) already motivated his "meaning criterion" with the need to evaluate the meaningfulness of sentences containing new concepts, and, more specifically, Carnap (1928a, §§61–67) 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 Carnap's and Reichenbach's

rion 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 and not analytically false.[12]

Definition 3. A *set Ω of sentences falsifies* a set Σ of sentences if and only if $\Omega \cup \Sigma \cup \Pi \vDash \bot$.

Definition 4. 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.

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.

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 5. 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.[13]

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 is equivalent to $\Sigma \cup \Delta \vDash \bigwedge \Omega' \rightarrow \beta$. Hence for

criticism of Driesch's term 'entelechy' in 1934 (Carnap 1966, 14–15) and Schlick's and Ayer's example sentence 'The Absolute enters into, but is itself incapable of, evolution and progress', whose 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 to this focus on an auxiliary vocabulary.

[12] As noted in §2.1, the qualifier 'syntactic' here does not refer to the use of syntactic deduction ('$\vdash$'), but to the syntactic description of empirical states (by sentences).

[13] Note again that 'syntactic' refers to the syntactic description of the observations. This terminology is essentially that of Przełęcki (1969, 52).

first order logic, and if the set of basic sentences is closed under truth functional composition,[14] 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 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 $\Omega \cup \{\neg\beta\} \cup \Sigma \cup \Pi \vDash \bot$. $\square$

The relation between falsifiability and $\mathscr{B}$-creativity provides a justification for Reichenbach's claim that the $\mathscr{B}$-sentences only need to be unproblematic, not observational: 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 (1965b, 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, so that for Ayer, it is a condition of adequacy for any criterion of empirical significance that it distinguish those sets of sentences that assert the occurrence of experiences from those that do not. Under this assumption, he arrives at his first criterion of empirical significance, namely that "the

[14]This is always the case if the $\mathscr{B}$-sentences are defined as all those containing only $\mathscr{B}$-terms.

mark of a genuine factual proposition [is] that some 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'[15] and the criterion is meant as a necessary and sufficient condition, this can be paraphrased as

Definition 6. 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 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.

Unfortunately, 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.

[15]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.

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 Ayer develops his criterion under the assumption 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 Hempel (1965d, §7). Gemes (1998, §1.4) discusses its historical importance in the search for criteria of empirical significance and argues that the failure of Ayer's criterion is inherited from the failure of the prediction criterion of confirmation.

3.2 Semantic criteria

Syntactic $\mathscr{B}$-conservativeness has a semantic counterpart:

Definition 7. 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 $\mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$.

Definition 7 is slightly more general than that given, for example, by Przełęcki (1974a, 345), so that it allows for any $\mathscr{V}$-sentence in Δ. Note that, like the other semantic definitions discussed here, this definition relies essentially on a bipartition of the vocabulary.

As announced in §2.1, 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 ('$\equiv$'):[16]

Claim 3. *A set Σ of $\mathscr{V}$-sentences is syntactically $\mathscr{B}$-conservative with respect to Δ 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 $\mathfrak{B}|_{\mathscr{B}} \equiv \mathfrak{A}_{\mathscr{B}}$, 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 $\mathfrak{C} \vDash \Omega \cup \{\neg\beta\} \cup \Sigma \cup \Delta \vDash\!\!\!\!\dashv \mathrm{Th}(\mathfrak{A}_{\mathscr{B}}) \cup \Sigma \cup \Delta$. Thus there is a $\mathfrak{C} \vDash \Sigma \cup \Delta$ such that $\mathfrak{C}|_{\mathscr{B}} \equiv \mathfrak{A}_{\mathscr{B}}$.

'$\Leftarrow$': Let $\Omega \cup \Delta \nvDash \beta$. Choose $\mathfrak{A} \vDash \Omega \cup \Delta \cup \{\neg\beta\}$; by assumption, there is a $\mathfrak{C} \vDash \Sigma \cup \Delta$ with $\mathfrak{C}|_{\mathscr{B}} \equiv \mathfrak{A}|_{\mathscr{B}}$ and thus $\mathfrak{C} \vDash \Omega \cup \Sigma \cup \Delta \cup \{\neg\beta\}$, so that $\Omega \cup \Sigma \cup \Delta \nvDash \beta$. □

[16] It is fairly easy to show that in definition 7, '$=$' can be substituted by '$\simeq$'.

This suggests

Claim 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 3 because $\mathfrak{A}|_{\mathscr{B}} = \mathfrak{B}|_{\mathscr{B}}$ only if $\mathfrak{A}|_{\mathscr{B}} \equiv \mathfrak{B}|_{\mathscr{B}}$.
'$\not\Leftarrow$': Van Benthem (1978, 324). □

Of course, the two criteria are equivalent in all languages in which syntactic equivalence amounts to isomorphism.

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 8. 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 9. 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 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 if and only if 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 $\mathfrak{C} \not\models \Sigma$. □

The relation between syntactic and semantic falsifiability is then given by claims 5, 4 and 1.

To arrive at a workable criterion of empirical significance, Lewis (1988b, 127–128) suggests going back to Ayer's failed criterion and taking it to be an attempt at identifying sentences that are partly about observation. To elucidate what that means, Lewis first explicates aboutness and thereafter the modifier 'partly'. 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". There are different ways to explicate the notion of a possible world, but it is clear from Lewis's definition of supervenience that if the language under investigation is that of predicate logic with its usual semantics, the worlds have to be structures in the sense used here. As Kemeny (1956, 9–10; cf. 1963, §IV) further points out, a structure may be identified with a possible world under the assumption that all and only those worlds in which the analytic sentences are true are possible. 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. This approach is fairly typical. In his discussion of reduction, for instance, Fodor (1974, 98) explicitly assumes "that a science is individuated largely by reference to its typical predicates", so that two worlds would be exactly alike with respect to biology if and only if they do not differ in their interpretation of the biological terms. Nagel (1951, 330) similarly assumes in his notion of reduction that different disciplines typically rely on different vocabularies. Possible worlds are thus exactly alike with respect to subject matter $\mathscr{B}$ if and only if they have the same $\mathscr{B}$-structures. This leads to

Definition 10. A set Σ of $\mathscr{V}$-sentences is *about subject matter* $\mathscr{B}$ if and only if for any $\mathscr{V}$-structures $\mathfrak{A} \vDash \Pi, \mathfrak{B} \vDash \Pi$ with $\mathfrak{A}|_{\mathscr{B}} = \mathfrak{B}|_{\mathscr{B}}$ it holds that $\mathfrak{A} \vDash \Sigma$ if and only if $\mathfrak{B} \vDash \Sigma$.

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 10 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 §7 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 $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{C}|_{\mathscr{B}}$, $\mathfrak{B} \in \mathbf{C}_\Sigma$ if and only if $\mathfrak{C} \in \mathbf{C}_\Sigma$. The parts of the content of Σ are then defined as the subsets of $\mathbf{C}_\Sigma$, which leads to

Definition 11. *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 $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{C}|_{\mathscr{B}}$, $\mathfrak{B} \in \mathbf{F}$ if and only if $\mathfrak{C} \in \mathbf{F}$.

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 $\mathscr{B}$ if and only if the sentence is syntactically falsifiable.[17] But Lewis's definition 11 is better compared to *semantic* falsifiability:

Claim 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. '⇒': 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 11 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 $\mathbf{C}_\Sigma$, $\mathfrak{B} \nvDash \Sigma$, and the possible structure $\mathfrak{A}_{\mathscr{B}}$ falsifies Σ.

'⇐': Assume Σ is semantically falsified by $\mathfrak{A}_{\mathscr{B}}$. Define $\mathbf{F} := \{\mathfrak{B} \mid \mathfrak{B} \vDash \Pi \text{ and } \mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}\}$. Since $\varnothing \neq \mathbf{F} \subseteq \mathbf{C}_\Sigma$, part of Σ's content is about subject matter $\mathscr{B}$. □

Because of claims 1, 5, and 6, the relation between syntactic falsifiability and Lewis's definition 11 is the same as that between syntactic and semantic $\mathscr{B}$-creativity, which is given in claim 3.

[17] 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 10 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.

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 32). 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 (Psillos 2000)

$$\mathsf{R}_{\mathscr{B}}(\sigma) := \exists \bar{X} \sigma(\bar{B}, \bar{X}) , \tag{1}$$

which results from σ by existentially generalizing on all $\mathscr{A}$-terms in σ. $\mathsf{R}_{\mathscr{B}}(\sigma)$ entails the same $\mathscr{B}$-sentences as σ (Rozeboom 1962, 291–293), which makes the choice of the Ramsey sentence as a sentence's $\mathscr{B}$-content plausible. If Σ is a finite set of sentences, I will sometimes write $\ulcorner \mathsf{R}_{\mathscr{B}}(\Sigma) \urcorner$ instead of $\ulcorner \mathsf{R}_{\mathscr{B}}(\bigwedge \Sigma) \urcorner$ and speak of the $\mathscr{B}$-content of Σ. Now, a criterion of the meaning of a set of sentences cannot be a criterion of empirical significance (see §2.2). Analogously, a description of the $\mathscr{B}$-content of a set of sentences cannot be a criterion of empirical significance either. 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 12. 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 $\tilde{\Pi} \nvDash \mathsf{R}_{\mathscr{B}}(\sigma \wedge \bigwedge \tilde{\Pi})$.

Under this definition, Carnap's later notion of $\mathscr{B}$-content squares well with the notion of falsifiability, as can be seen from

Lemma 7. *$\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ can be expanded to a model of $\mathscr{V}$-sentence σ if and only if $\mathfrak{A}_{\mathscr{B}} \vDash \mathsf{R}_{\mathscr{B}}(\sigma)$.*

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 $\exists X_1 \dots X_n \sigma[A_1/X_1, \dots, A_n/X_n]$. 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 $\mathfrak{A}_{\mathscr{B}} \vDash \mathsf{R}_{\mathscr{B}}(\sigma)$. Then there is 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 $\mathfrak{A}_{\mathscr{B}}, \nu \vDash \gamma[A_1/X_1, \dots, A_n/X_n]$. Induction on the complexity of formulas then shows that any extension f of $\nu|_{\{X_1,\dots,X_n\}} \circ g$ to all $\mathscr{A}$-terms can be used to expand $\mathfrak{A}_{\mathscr{B}}$ to a model of σ.

'$\Rightarrow$': Similar. □

Claim 8. *If Π can be finitely axiomatized, then a $\mathscr{V}$-sentence σ has $\mathscr{B}$-content if and only if σ is semantically falsifiable.*

Proof. Since $\mathsf{R}_{\mathscr{B}}(\sigma \wedge \bigwedge \tilde{\Pi})$ is a $\mathscr{B}$-sentence, $\tilde{\Pi} \nvDash \mathsf{R}_{\mathscr{B}}(\sigma \wedge \bigwedge \tilde{\Pi})$ if and only if $\mathsf{R}_{\mathscr{B}}(\tilde{\Pi}) \nvDash \mathsf{R}_{\mathscr{B}}(\sigma \wedge \bigwedge \tilde{\Pi})$, that is for some $\mathfrak{A}_{\mathscr{B}}$, $\mathfrak{A}_{\mathscr{B}} \vDash \tilde{\Pi}$ and $\mathfrak{A}_{\mathscr{B}} \nvDash \sigma \wedge \bigwedge \tilde{\Pi}$. By lemma 7, this holds if and only if there is a possible $\mathscr{B}$-structure that cannot be expanded to a model of $\sigma \wedge \bigwedge \tilde{\Pi}$ or, in other words, if and only if σ is falsified by a possible $\mathscr{B}$-structure. □

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 probabilistic inferences.

4 Verifiability

Another criterion of empirical significance that has been proposed very early on is that of syntactic verifiability (Hempel 1965b, 104). Modifying Hempel's formulation in a way analogous to his formulation of falsifiability leads to

Definition 13. 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 14. 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.

Hempel (1965b, 106) points out the following straightforward

Claim 9. *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 15. A *$\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ verifies* a set Σ of $\mathscr{V}$-sentences if and only if for all $\mathfrak{C} \vDash \Pi$ with $\mathfrak{C}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$, $\mathfrak{C} \vDash \Sigma$.

Definition 16. 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 10. *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 3 and 4 therefore transfer to the verifiability of sentences.

For sets of sentences, a relation analogous to claim 3 holds as well:

Claim 11. *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 $\mathfrak{B} \vDash \Omega \cup \Pi$, $\mathfrak{B} \vDash \Sigma$. 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 $\mathfrak{C} \vDash \Omega$. 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 $\Omega \mathrel{\vDash\!\dashv} \mathrm{Th}(\mathfrak{A}_{\mathscr{B}})$. Since $\mathfrak{A}_{\mathscr{B}}$ is possible, $\Omega \cup \Pi$ has a model, and thus Ω is possible. By assumption, $\mathfrak{B} \vDash \Omega \cup \Pi$ only if $\mathfrak{B} \vDash \Sigma$, 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 11 'verifiable' by 'falsifiable' and 'verified' by 'falsified' results in a simple paraphrase of claim 3 that makes this analogy obvious.

Furthermore, claims 10 and 8 entail the following:

Claim 12. *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, $\mathsf{R}_{\mathscr{B}}(\neg\sigma)$ is not the basic content of σ for $\Pi = \varnothing$. Rather, it can be shown similarly to the proof of claim 8 that the possible $\mathscr{B}$-structures that verify σ are the models of $\neg\mathsf{R}_{\mathscr{B}}(\neg\sigma)$. 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 12 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, $\vDash \mathsf{R}_{\mathscr{B}}(\neg\sigma)$. If, in the terminology of Peacocke (1986, 47), $\mathsf{R}_{\mathscr{B}}(\Sigma)$ is akin to the "canonical commitment of the content that" Σ, then $\neg\mathsf{R}_{\mathscr{B}}(\neg \bigwedge \Sigma)$ is akin to the "canonical ground for the content that" Σ.

5 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 (1965c, 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 seems to lack any ground.

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. A good dialectical 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). But there is also a non-pragmatic reason for this criterion. As argued above, the criteria discussed so far are applicable only to deductive inferences. And Hempel (1958, §3), for example, argues that the reason to employ non-basic sentences is the inference of basic sentences from other basic sentences. But whenever only deductive inferences can be employed, a set Σ of $\mathscr{V}$-sentences can contribute non-trivially to such an inference if and only if, first, it is entailed by possible basic sentences, that is, is syntactically verifiable. For otherwise, Σ can never be established and thus cannot be relied upon for inferring any other sentences. Second, together with some possible basic sentences Ω, Σ must entail another basic sentence not entailed by Ω alone, that is, Σ must be syntactically falsifiable. Thus, to be useful in a deductive inference of basic sentences from other basic sentences, a set Σ of $\mathscr{V}$-sentences must be both verifiable and falsifiable. (One could make an analogous argument for the inference of one possible $\mathscr{O}$-structure from a set of other possible $\mathscr{O}$-structures.)

A straightforward response to such an argument is that deductive inferences are possible in some but not all situations in the sciences, and that Σ may very well have been established by inductive inference, or may be one theory of many that might be considered for inductive testing. But even an inductively established theory may allow deductive inferences, so that Σ, while established or to be established by a type of inference that is outside the scope of this article, may still meet or fail to meet the falsifiability criterion. And no matter how a theory is established, it is of interest whether it deductively entails basic sentences or not. (An analogous argument could be made for verifiability of Σ and inductive inferences from Σ.)

6 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:[18]

Definition 17. 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 $\mathfrak{B} \vDash \Sigma$ if and only if $\mathfrak{C} \vDash \Sigma$.

Definition 18. 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 13. *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.

For single sentences, this can be phrased in terms of Ramsey sentences.

Claim 14. *If* Π *can be finitely axiomatized, let* $\tilde{\Pi}$ *be this axiomatization. Then a* $\mathscr{V}$*-sentence* σ *is strongly semantically* $\mathscr{B}$*-determined if and only if*

$$\tilde{\Pi} \vDash \neg\left[\mathsf{R}_{\mathscr{B}}(\sigma \wedge \bigwedge \tilde{\Pi}) \wedge \mathsf{R}_{\mathscr{B}}(\neg\sigma \wedge \bigwedge \tilde{\Pi})\right] . \tag{2}$$

Proof. '$\Rightarrow$': Assume $\mathfrak{A} \vDash \tilde{\Pi}$. Then $\mathfrak{A}|_{\mathscr{B}}$ has an expansion to a model of $\tilde{\Pi}$, and by assumption and claim 13, all possible expansions of $\mathfrak{A}|_{\mathscr{B}}$ are models of $\neg\sigma$ or all possible expansions of $\mathfrak{A}|_{\mathscr{B}}$ are models of σ. In other words, there is no expansion of $\mathfrak{A}|_{\mathscr{B}}$ that is a model of $\tilde{\Pi}$ and σ or there is no expansion of $\mathfrak{A}|_{\mathscr{B}}$ that is a model of $\tilde{\Pi}$ and $\neg\sigma$. By lemma 7, then, $\mathfrak{A}|_{\mathscr{B}} \nvDash \mathsf{R}_{\mathscr{B}}(\sigma \wedge \tilde{\Pi})$ or $\mathfrak{A}|_{\mathscr{B}} \nvDash \mathsf{R}_{\mathscr{B}}(\neg\sigma \wedge \tilde{\Pi})$ and hence $\mathfrak{A}|_{\mathscr{B}} \vDash \neg\mathsf{R}_{\mathscr{B}}(\sigma \wedge \tilde{\Pi}) \vee \neg\mathsf{R}_{\mathscr{B}}(\neg\sigma \wedge \tilde{\Pi})$. Thus $\mathfrak{A} \vDash \neg\left[\mathsf{R}_{\mathscr{B}}(\sigma \wedge \bigwedge \tilde{\Pi}) \wedge \mathsf{R}_{\mathscr{B}}(\neg\sigma \wedge \bigwedge \tilde{\Pi})\right]$.

'$\Leftarrow$': Similar. □

In higher order logic, strong semantic $\mathscr{B}$-determinacy is equivalent to translatability, another central concept in the theory of definition (Gupta 2009, §2.3; Belnap 1993).[19]

Claim 15. *If* Π *can be finitely axiomatized, let* $\tilde{\Pi}$ *be this axiomatization. Then* $\mathscr{V}$*-sentence* σ *is strongly semantically* $\mathscr{B}$*-determined if and only if* $\Pi \vDash \sigma \leftrightarrow \mathsf{R}_{\mathscr{B}}(\sigma \wedge \bigwedge \tilde{\Pi})$.

[18]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.

[19]It is a necessary condition for an explicit definition that every sentence in which its definiendum occurs can be translated by the definition into a sentence in which it does not occur.

Proof. '$\Rightarrow$': Obviously, $\sigma \rightarrow \mathsf{R}_{\mathscr{B}}(\sigma \wedge \bigwedge \tilde{\Pi})$ in every model of $\tilde{\Pi}$. Now assume that $\mathsf{R}_{\mathscr{B}}(\sigma \wedge \bigwedge \tilde{\Pi})$ is true in some model of $\tilde{\Pi}$. Then, if $\neg\sigma$ held, $\neg\sigma \wedge \bigwedge \tilde{\Pi}$ would hold as well, and thus $\mathsf{R}_{\mathscr{B}}(\neg\sigma \wedge \bigwedge \tilde{\Pi})$. But this is impossible because by claim 14, $\mathsf{R}_{\mathscr{B}}(\neg\sigma \wedge \bigwedge \tilde{\Pi})$ cannot hold in the same model of $\tilde{\Pi}$ as $\mathsf{R}_{\mathscr{B}}(\sigma \wedge \bigwedge \tilde{\Pi})$. Hence σ, so that $\mathsf{R}_{\mathscr{B}}(\sigma \wedge \bigwedge \tilde{\Pi}) \rightarrow \sigma$ in all models of $\tilde{\Pi}$.

'$\Leftarrow$': Immediate. □

In other words, σ is strongly semantically $\mathscr{B}$-determined if and only if it can be translated into a higher order $\mathscr{B}$-sentence by Π.

As Przełęcki (1974a, 346–347) already notes, definition 18 is very exclusive even in first order logic. 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 16. *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, $\mathfrak{B} \vDash \Sigma$ if and only if $\mathfrak{C} \vDash \Sigma$.

'$\Leftarrow$': 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, $\mathfrak{B} \vDash \Sigma$ if and only if $\mathfrak{C} \vDash \Sigma$. □

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 the justification of his criterion of "empirical meaningfulness" 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

$$\begin{aligned} \Pi_{\text{mass}} := \{&\forall x \forall y\, (x \precsim y \leftrightarrow mx \leq my), \\ &\forall x \forall y\, (m(x * y) = mx + my)\}\,, \end{aligned} \tag{3}$$

where, as in §2.1, '$\precsim$' stands for 'is at most as heavy as' and '$*$' stands for physical combination. $\precsim$ and $*$ thus 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. x and y are silently understood to range over physical objects.[20] 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, 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:[21]

Definition 19 (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 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 17. *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 19 if and only if σ is about subject matter $\mathscr{B}$.*

Proof. Przełęcki (1974a, 349). □

In claim 17, 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. In effect, '+' and '$\leq$' are thus taken to be basic terms. One could also ensure the standard interpretation with the usual axioms in second order logic, with '+' and '$\leq$' as auxiliary terms.

[20] Together, the two works by Przełęcki (1969, 1974a) provide a model-theoretic semantics for this exact situation that fits with the discussions by Carnap (1926, 1966) and Hempel (1952) and his own criteria of empirical significance.

[21] Przełęcki's paraphrase is slightly different, for one because he aims to prove its equivalence with definition 18, not definition 10, but also because his definition of $\mathscr{B}$-conservativeness is slightly less general.

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 $\mathfrak{E} = \langle |\mathfrak{E}|, \{\langle P_1, P_1^{\mathfrak{E}}\rangle, \dots, \langle P_p, P_p^{\mathfrak{E}}\rangle\}, \{\langle \circ_1, \circ_1^{\mathfrak{E}}\rangle, \dots, \langle \circ_q, \circ_q^{\mathfrak{E}}\rangle\}\rangle$, sometimes called '*empirical*', to another relational system $\mathfrak{F} = \langle |\mathfrak{F}|, \{\langle Q_1, Q_1^{\mathfrak{F}}\rangle, \dots, \langle Q_p, Q_p^{\mathfrak{F}}\rangle\}, \{\langle *_1, *_1^{\mathfrak{F}}\rangle, \dots, \langle *_q, *_q^{\mathfrak{F}}\rangle\}\rangle$, 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 $a_1^{\mathfrak{E}}, \dots, a_{k_i}^{\mathfrak{E}}, a^{\mathfrak{E}}, b^{\mathfrak{E}} \in |\mathfrak{E}|$ with $i = 1, \dots, p$ and for all $j = 1, \dots, q$ it holds that

$$P_i^{\mathfrak{E}}(a_1^{\mathfrak{E}}, \dots, a_{k_i}^{\mathfrak{E}}) \text{ if and only if } Q_i^{\mathfrak{F}}(\mu(a_1^{\mathfrak{E}}), \dots, \mu(a_{k_i}^{\mathfrak{E}}))\ , \tag{4a}$$

$$\mu(a^{\mathfrak{E}} \circ_j^{\mathfrak{E}} b^{\mathfrak{E}}) = \mu(a^{\mathfrak{E}}) *_j^{\mathfrak{F}} \mu(b^{\mathfrak{E}})\ . \tag{4b}$$

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 (Hodges 1993, 202). 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 $\mathfrak{A}|E_{\mathscr{E}}$ 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).[22]

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 (4). This is the case if and only if $\mathfrak{A} \vDash \Pi_{\text{adeq}}$ with

$$\begin{aligned}\Pi_{\text{adeq}} := \{\forall a(Ea \to Ffa)\} \cup \\ &\bigcup_{i=1}^{p} \left\{\forall a_1 \ldots \forall a_{k_i} \left[Ea_1 \wedge \cdots \wedge Ea_{k_i} \to \left(P_i a_1 \ldots a_{k_i} \leftrightarrow Q_i f a_1 \ldots f a_{k_i}\right)\right]\right\} \\ &\cup \bigcup_{j=1}^{q} \left\{\forall a \forall b \left(Ea \wedge Eb \to f(a \circ_j b) = fa *_j fb\right)\right\} .\end{aligned} \tag{5}$$

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 $\mathfrak{A} \vDash \Pi_{\text{scale}}$ 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, $\mathfrak{A}|EF_{\mathscr{E}\mathscr{F}f} \vDash \Sigma$ if and only if $\mathfrak{A} \vDash \Sigma^{(EF)}$ for any $\mathscr{E}\mathscr{F}f$-sentence Σ. 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

[22] The relativization $\sigma^{(E)}$ of a sentence σ consists of the restriction of all quantifiers in σ to E. For any set Σ of formulas, $\Sigma^{(E)}$ is the set of the relativization of the elements of Σ.

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, $\mathfrak{A} \vDash \Sigma^{(EF)}$ 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 20. If $\mathfrak{A} \vDash \Pi_{\text{scale}}$, 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 $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \vDash \Pi_{\text{adeq}}$.

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 21. A set Σ of $\mathscr{E}\mathscr{F}f$-sentences is *strongly invariant* if and only if for any $\mathfrak{A} \vDash \Pi_{\text{scale}}$ and any admissible transformation φ relative to $\mathfrak{A}$, it holds that $\mathfrak{A} \vDash \Sigma$ if and only if $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \vDash \Sigma$.

The restriction of definition 21 to relativizations of $\mathscr{E}\mathscr{F}f$-sentences to EF is indeed equivalent to the original definition by Suppes and Zinnes (1963):

Claim 18. *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 $\mathfrak{A} \vDash \Pi_{\text{scale}}$, $\mathfrak{A}|E_{\mathscr{E}} = \mathfrak{E}$, $\mathfrak{A}|F_{\mathscr{F}} = \mathfrak{F}$, 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 construction of $\mathfrak{A}$, $\mathfrak{A} \vDash \Sigma^{(EF)}$. 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 5. 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 $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \vDash \Sigma^{(EF)}$. Now $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \vDash \Pi_{\text{scale}}$ 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 $\mathfrak{A} \vDash \Sigma^{(EF)}$, and that φ is an admissible transformation relative to $\mathfrak{A}$. By construction, if $\mathfrak{A} \vDash \Sigma^{(EF)}$, 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 5 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 $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \vDash \Pi_{\text{scale}}$ 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

$f^{\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]}|_{|\mathfrak{C}|} = \varphi(f^{\mathfrak{A}})|_{|\mathfrak{C}|} = \psi(f^{\mathfrak{A}}|_{|\mathfrak{C}|}) = \psi(\mu)$, so that $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \vDash \Sigma^{(EF)}$. By analogous reasoning, $\mathfrak{A} \nvDash \Sigma^{(EF)}$ only if $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \nvDash \Sigma^{(EF)}$, 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 19. *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. '⇐': First, note that for any $\mathfrak{A} \vDash \Pi_{\text{scale}}$ and any admissible transformation φ relative to $\mathfrak{A}$, there is some $\mathfrak{B} \vDash \Pi_{\text{scale}}$ 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 20, $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \vDash \Pi_{\text{scale}}$, 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 $\mathfrak{A}, \mathfrak{B} \vDash \Pi_{\text{scale}}$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$, $\mathfrak{A} \vDash \Sigma$ if and only if $\mathfrak{B} \vDash \Sigma$. Let $\mathfrak{C} \vDash \Pi_{\text{scale}}$ and φ be admissible relative to $\mathfrak{C}$. Then, because of (∗), there is some $\mathfrak{D} \vDash \Pi_{\text{scale}}$ with $\mathfrak{D}|_{\mathscr{B}} = \mathfrak{C}|_{\mathscr{B}}$ and $\mathfrak{D} = \mathfrak{C}[f/\varphi(f^{\mathfrak{C}})]$. Therefore, by assumption, $\mathfrak{C}[f/\varphi(f^{\mathfrak{C}})] \vDash \Sigma$ if and only if $\mathfrak{C} \vDash \Sigma$.

'⇒': First, note that for any $\mathfrak{A} \vDash \Pi_{\text{scale}}, \mathfrak{B} \vDash \Pi_{\text{scale}}$ 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 $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] = \mathfrak{B} \vDash \Pi_{\text{adeq}}$, 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 $\mathfrak{B} \vDash \Pi_{\text{scale}}$. 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, $\mathfrak{B} \vDash \Sigma$ if and only if $\mathfrak{A} \vDash \Sigma$. □

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 10. 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 1 and 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 3 for falsifiability and claim 11 for verifiability. To arrive at an analogous relation for strong $\mathscr{B}$-determinacy, I thus suggest

Definition 22. A *set Γ of $\mathscr{V}$-sentences determines* a set Σ of $\mathscr{V}$-sentences if and only if $\Gamma \cup \Pi \vDash \Sigma$ or $\Gamma \cup \Sigma \cup \Pi \vDash \bot$.

Definition 23. A set Ω of $\mathscr{B}$-sentences is *maximal* if and only if for every $\mathscr{B}$-sentence β, $\Omega \cup \Pi \vDash \beta$ or $\Omega \cup \Pi \vDash \neg\beta$.

Then one can formulate

Definition 24. 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 20. *A set Σ of $\mathscr{V}$-sentences is strongly syntactically $\mathscr{B}$-determined if and only if for any $\mathscr{V}$-structures $\mathfrak{A}, \mathfrak{B} \vDash \Pi$ with $\mathfrak{A}|_{\mathscr{B}} \equiv \mathfrak{B}|_{\mathscr{B}}$, it holds that $\mathfrak{A} \vDash \Sigma$ if and only if $\mathfrak{B} \vDash \Sigma$.*

Proof. '$\Rightarrow$': Let $\mathfrak{A}, \mathfrak{B} \vDash \Pi$ and $\mathfrak{A}|_{\mathscr{B}} \equiv \mathfrak{B}|_{\mathscr{B}}$. Then $\mathfrak{A}, \mathfrak{B} \vDash \mathrm{Th}(\mathfrak{B}|_{\mathscr{B}}) =: \Omega$. It is straightforward to show that Ω is maximal and possible. Thus, by assumption, $\Omega \cup \Pi \vDash \Sigma$ or $\Omega \cup \Sigma \cup \Pi \vDash \bot$. Thus $\mathfrak{A} \vDash \Sigma$ if and only if $\mathfrak{B} \vDash \Sigma$.

'$\Leftarrow$': Assume Ω is possible and maximal. Then for any $\mathfrak{A}, \mathfrak{B} \vDash \Omega \cup \Pi$, $\mathfrak{A}|_{\mathscr{B}} \equiv \mathfrak{B}|_{\mathscr{B}}$. Therefore, by assumption, $\mathfrak{A} \vDash \Sigma$ if and only if $\mathfrak{B} \vDash \Sigma$ and thus either all $\mathfrak{A} \vDash \Omega \cup \Pi$ are models of Σ or none is. Thus $\Omega \cup \Pi \vDash \Sigma$ or $\Omega \cup \Sigma \cup \Pi \vDash \bot$. □

This entails

Claim 21. *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 16 and 20 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[23]

Claim 22. *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 $\Pi \nvDash \Sigma$. Then for some $\mathfrak{A}$, $\mathfrak{A} \vDash \Pi$ and $\mathfrak{A} \nvDash \Sigma$. If Σ is syntactically $\mathscr{B}$-determined, then $\mathrm{Th}(\mathfrak{A}|_{\mathscr{B}}) \cup \Sigma \cup \Pi \vDash \bot$ because $\mathrm{Th}(\mathfrak{A}|_{\mathscr{B}}) \cup \Pi \nvDash \Sigma$. Thus $\mathrm{Th}(\mathfrak{A}|_{\mathscr{B}})$ falsifies Σ. If Σ is semantically $\mathscr{B}$-determined, then for all $\mathfrak{B} \vDash \Pi$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$, $\mathfrak{B} \nvDash \Sigma$. Thus $\mathfrak{A}|_{\mathscr{B}}$ falsifies Σ.

The proofs for verifiability are analogous. □

[23] Here and in the following definitions and claims, choosing uniformly the first, second, etc. of the n phrases connected by slashes '/' leads to one of n conjuncts of the definition or claim. In this claim, for example, Λ is falsifiable if and only if it is not analytically true, and Λ is verifiable if and only if it is not analytically false. Analogously, uniformly substituting a phrase by the parenthetical one following it leads to another conjunct of the definition or claim. Claim 22 therefore has four conjuncts.

7 Weak $\mathscr{B}$-determinacy

Since Przełęcki considers strong semantic $\mathscr{B}$-determinacy too exclusive, he suggests a straightforward weakening of definition 18:

Definition 25. 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 $\Pr(X\,|\,\sigma) > \frac{1}{2}$ and $\Pr(X\,|\,\neg\sigma) > \frac{1}{2}$. Under some additional assumptions, the notion of partial supervenience that results is equivalent to the standard probabilistic criterion that σ is empirically significant if and only if $\Pr(\sigma\,|\,\beta) \neq \Pr(\sigma)$ for some basic sentence β.

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 worlds, 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 26. 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 $\mathfrak{B} \models \Pi$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}}$ are in X, and for any $\mathfrak{A}, \mathfrak{B} \in X$ with $\mathfrak{A}|_{\mathscr{B}} = \mathfrak{B}|_{\mathscr{B}}$, $\mathfrak{B} \models \Sigma$ if and only if $\mathfrak{A} \models \Sigma$.

As announced, this is the same as weak semantic $\mathscr{B}$-determinacy:

Claim 23. *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. $\square$

Przełęcki (1974a, 347) points out that under his assumption that $\Pi_{\mathscr{A}}$ is $\mathscr{B}$-conservative with respect to $\Pi_{\mathscr{B}}$, definition 25 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 observation can be phrased in a more general (and more conspicuous) way:

Claim 24. *A set Σ of $\mathscr{V}$-sentences is weakly semantically $\mathscr{B}$-determined if and only if Σ is semantically falsifiable or semantically verifiable.*

Proof. Σ is semantically falsifiable or semantically verifiable if and only if there is a possible $\mathfrak{A}_{\mathscr{B}}$ such that Σ is false in all structures $\mathfrak{B} \models \Pi$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$ or true 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. $\square$

With claims 5 and 9, 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 essentially a reformulation of Przełęcki's claim. Claim 24 can also be phrased in terms of Ramsey sentences

Claim 25. *If Π can be finitely axiomatized, let $\tilde{\Pi}$ be this axiomatization. Then a $\mathscr{V}$-sentence σ is weakly semantically $\mathscr{B}$-determined if and only if*

$$\tilde{\Pi} \not\models \mathsf{R}_{\mathscr{B}}(\sigma \wedge \bigwedge \tilde{\Pi}) \vee \mathsf{R}_{\mathscr{B}}(\neg\sigma \wedge \bigwedge \tilde{\Pi}) \,. \tag{6}$$

Proof. By claim 24, σ is weakly semantically $\mathscr{B}$-determined if and only if σ is semantically falsifiable or verifiable. This holds if and only if, according to claim 8, $\tilde{\Pi} \not\models A$ or, according to claim 12, $\tilde{\Pi} \not\models B$, where A and B are the respective Ramsey sentences. This in turn holds if and only if there is some $\mathfrak{A}$ such that $\mathfrak{A} \vDash \tilde{\Pi}$ and $\mathfrak{A} \nvDash A$, or some $\mathfrak{A}$ such that $\mathfrak{A} \vDash \tilde{\Pi}$ and $\mathfrak{A} \nvDash B$. It is straightforward to show that this in turn holds if and only if there is some $\mathfrak{A}$ such that $\mathfrak{A} \vDash \tilde{\Pi}$ and $\mathfrak{A} \nvDash A \vee B$, that is, $\tilde{\Pi} \nvDash A \vee B$. □

Considerations analogous to those leading to the definition of strong syntactic $\mathscr{B}$-determinacy lead to

Definition 27. 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 26. *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 $\mathfrak{B}, \mathfrak{C} \vDash \Pi$ with $\mathfrak{B}|_{\mathscr{B}} \equiv \mathfrak{C}|_{\mathscr{B}} \equiv \mathfrak{A}_{\mathscr{B}}$, it holds that $\mathfrak{B} \vDash \Sigma$ if and only if $\mathfrak{C} \vDash \Sigma$.*

Proof. '⇒': Choose $\Omega := \mathrm{Th}(\mathfrak{A}_{\mathscr{B}})$ and proceed as in the proof of claim 20.
'⇐': Choose some $\mathfrak{A}_{\mathscr{B}} \vDash \Omega$ and proceed as in the proof of claim 20 □

And analogously to the semantic case, the following holds:

Claim 27. *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 took place with the help of the prediction criterion of confirmation discussed in connection with Ayer's trivial definition of indirect verifiability (§ 3.1). 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 1998, §1.4).

Following the chain of definitions is somewhat tedious, 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 28. *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. $\square$

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, Carnap's 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). Because of claim 28, 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 $\{c_i \mid i \in I\} \cup \{\lambda x(\text{It is cold at } x)\} \subseteq \mathscr{B}$.) 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) brackets the problem of how to capture induction 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, inductively confirmable or inductively 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.

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. Thus weak $\mathscr{B}$-determinacy turns out to be a very popular criterion of empirical significance, as it has been proposed by Przełęcki, Carnap, Ayer, and Rynin, and is suggested by Lewis's defense of partial supervenience.

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 (1965c, 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. 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).

8 Import of the relations

Using the definitions above, one arrives at the notable number of equivalences and entailment relations shown in figures 1 and 2, with strong and weak $\mathscr{B}$-determinacy, falsifiability, and verifiability as the four major criteria of empirical significance. Many of the circumscriptions of the explicandum of the criteria and the defense of the methodological presumptions in §1 provide arguments for the feasibility of a criterion of empirical significance. In the sequel, I will argue that the equivalences show that the extant criteria are already adequate.

The importance of the equivalence results is suggested by the analogous (though more spectacular) case of different explicata of 'computable function', whose equivalence is often cited as evidence for their adequacy and sometimes even for the truth of the Church-Turing thesis (Barker-Plummer 2011, Copeland 2008). To see the strength of the analogy, it is helpful to first look at the disanalogous case of Tarski's definition of 'truth'. Tarski (1944, §4) demands as a condition of (material) adequacy for any definition of truth that all and only true sentences p fulfill the T-schema $\ulcorner\ulcorner p\urcorner$ is true if and only if $p\urcorner$. Tarski (1944, §12) is in the enviable position

> that the conditions for the material adequacy of the definition determine uniquely the extension of the term "*true*". Therefore, every definition of truth which is materially adequate would necessarily be equivalent to that actually constructed [by Tarski].

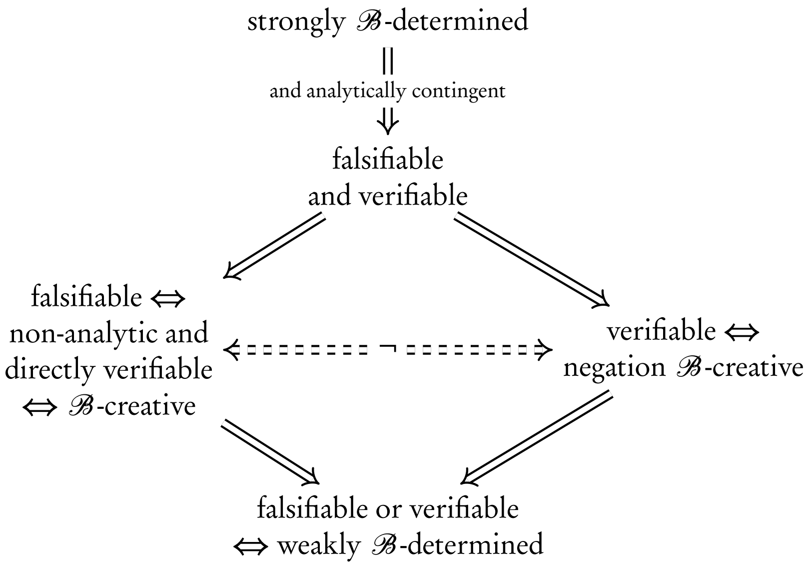

Figure 1: Relations between the syntactic definitions. The equivalence holds for direct verifiability and the negation of sets of sentences whenever the concepts are defined. A strongly $\mathscr{B}$-determined set of sentences is also weakly $\mathscr{B}$-determined even if not analytically contingent. Criteria of empirical significance typically also require that a set of sentences be analytically contingent.

But even for this nice state of affairs, it is at least not obvious how the extensional equivalence of all adequate definitions of 'truth' show that Tarski's is the correct one of the lot—after all the definitions are not all identical.[24] Luckily, even if Tarski's definition does not capture the real meaning of 'truth' (assuming there is one), it does provide an adequate *criterion* for truth, since it identifies exactly those sentences that are true. The extensional equivalence of the different definitions of computability accordingly suggest the definitions' adequacy as *criteria* of computable functions, that is, as a means for determining whether a specific function is computable. And the equivalences between the criteria of significance suggest that they, too, are adequate criteria, that is, adequate means for determining whether a specific sentence is empirically significant.[25] Thus there is a good chance that Lewis's demand for "the right class" of sentences (not "the right definition of 'empirical significance'") has been met. Of course, while Tarski can rely on a single very strong condition of adequacy, things are not as clear-cut in the case of empirical significance.

[24] I thank an anonymous referee for this point.

[25] Here and in the following, I will often for the sake of readability speak of single sentences rather than sets thereof. But all my claims hold for sets of sentences unless specifically indicated.

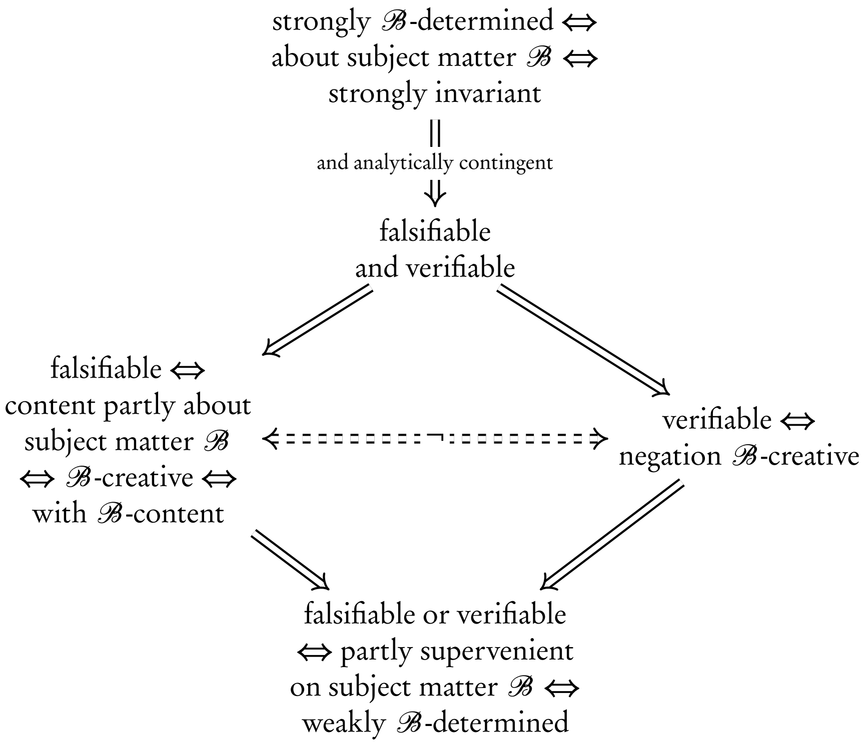

Figure 2: Relations between the semantic definitions. The equivalence holds for strong invariance, empirical content, and the negation of sets of sentences whenever the concepts are defined. A strongly $\mathscr{B}$-determined set of sentences is also weakly $\mathscr{B}$-determined even if not analytically contingent. Each of the nodes is entailed by its syntactic counterpart from figure 1. Criteria of empirical significance typically also demand that a set of sentences be analytically contingent.

8.1 Consolidating the justifications

The first step in defending the criteria of empirical significance discussed here consists in addressing extant criticisms, the multitude of different criteria being the most pressing. For if there are so many different criteria, determining the class of empirically significant sentences by any specific criterion must seem arbitrary. A partial reply is that, as has been shown, there are in fact only four major classes of significant sentences (with some vagueness resulting from the comparably small differences between the semantic and syntactic versions of the criteria in logics in which syntactic equivalence and isomorphism come apart). Lewis's analysis of partial aboutness completes the reply: The four classes are different because they are extensions of different but closely related notions. Strong semantic $\mathscr{B}$-determinacy is equivalent to Lewis's notion of aboutness, semantic falsifiability is equivalent to his notion of partial aboutness of content, and weak semantic $\mathscr{B}$-determinacy is arguably equivalent to partial supervenience. 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 27). It is thus at least the link connecting two different notions of partial aboutness. Thus the elements of the four classes have arguably all been called empirically significant because of a confusion of aboutness with partial aboutness and the ambiguity of partial aboutness itself.

Lewis's analysis also provides a reply to his own charge that many criteria of empirical significance have strayed too far from the intuitive explicandum, since each of his criteria is meant to capture some notion from ordinary language. The equivalences then show that the class of sentences picked out by the respective notion is robust under a change of formalism from predicate logic to model theory to set theory, and a change of formulation within each formalism.

Alas, the close connection of each class of sentences to ordinary language may prompt another criticism: that the classes are of little use in the sciences, the way the extension of the ordinary language notion of 'fish', which includes the likes of whales and dolphins, is of little use in biology (cf. Carnap 1950, §3). If the sciences are taken to include mathematics, then the equivalence of falsifiability to $\mathscr{B}$-creativity and of strong semantic $\mathscr{B}$-determinacy to translatability already provide rebuttals, for definitions are essential in mathematics, and every definition has to be $\mathscr{B}$-conservative and, in the standard case, ensure translatability (cf. Belnap 1993). Claim 10 shows the relevance of the verifiability of single sentences, since their negation is $\mathscr{B}$-creative, and claim 27 shows the relevance of weak $\mathscr{B}$-determinate sentences, whose class is just the union of the classes of falsifiable and verifiable sentences. Lest one argue that only the translatability and falsifiability of sentences is relevant, 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 is too inclusive if it includes every sentence or its negation among the empirically significant sentences. Thus, because of claims 10 and 27, falsifiability would already be too inclusive if every sentence was falsifiable or verifiable, that is, weakly $\mathscr{B}$-determined. Hence the class of weakly $\mathscr{B}$-determined sentences is considered so closely related to the class of falsifiable sentences 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 17 and claims 18 and 19 show that at least strongly $\mathscr{B}$-determined sentences are important for measurements because in this context, strongly $\mathscr{B}$-determined sentences are strongly invariant. The close relation of strong 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. More generally, the relation between strong invariance and strong $\mathscr{B}$-determinacy makes obvious the role of symmetry (i. e., invariance under transformations) in all semantic criteria: Σ is strongly $\mathscr{B}$-determined if and only if for every $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$, its truth value is invariant under all transformations allowed by Π and $\mathfrak{A}_{\mathscr{B}}$. Σ is verifiable if and only if its truth is so invariant for at least one possible $\mathscr{B}$-structure. Analogous relations hold for weak $\mathscr{B}$-determinacy and falsifiability. Given the enormous role of symmetry considerations in the sciences (and even philosophy), the criteria of significance should play a role as well.

Conversely, the equivalence of strong invariance 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 to be misled by the special conditions of a context, in this case the features of measurements, but the equivalence shows that a strongly invariant sentence is also empirically significant according to a much more generally motivated criterion.

The equivalences also provide a positive justification rather than a defense, because now the arguments in favor of each individual criterion turn out to be arguments for the delineation of the same class of sentences as empirically significant. Thus Przełęcki's general argument, Suppes's measure theory-specific argument, and arguments in favor of translatability already lead to the class of strongly $\mathscr{B}$-determined sentences, and Lewis's argument for aboutness adds additional support. The class of falsifiable sentences 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 class. The verifiability of a sentence, while historically not often defended by itself, again receives some justification because it is a sufficient condition for the sentence being weakly $\mathscr{B}$-determined.

8.2 The non-arbitrariness of the classes of empirically significant sentences

For those with realist tendencies about formal concepts, the equivalences between the criteria should arguably be proof enough for the correctness of the classes they determine. Like the different ways of determining Avogadro's number convinced scientists of the reality of atoms (cf. Salmon 1984, 214–220), the different ways of determining the class of empirically significant sentences should convince the realist of the existence of some natural property delineating that class. More concretely, one could argue that there is no better explanation for the equivalence of the criteria than the existence of some corresponding natural property and make an inference to the best explanation.

Without recourse to realism, one can fall back on Carnap's notion of explication and argue that the criteria fulfill the desiderata of explicata, that is, they are similar to the explicandum, precise, fruitful, and simple. The equivalences to Lewis's ordinary language notions[26] 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" (Carnap 1950, 7). Note, however, that this formulation suggests that there is something sacred about the current usage of a term, as Laudan (1986, 120) has criticized, since strictly speaking more than half of current uses must be captured. As noted in §2.1, I consider this demand problematic and would suggest that the current use of a term only provides some conditions of adequacy for its explicatum. Lewis's ordinary language intuitions could then only serve 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.

To avoid recourse to current usage, Carnap's demand for similarity could be dropped completely. In this case, the criteria could not be plausibly considered explications of intuitive concepts, but would simply be newly formed concepts. There is nothing wrong with this approach, since many new concepts are very helpful. But I want to suggest some conditions of adequacy that the different criteria fulfill (without, however, spending much time on their defense). The most important condition of adequacy is non-triviality. For if a criterion were to identify all contingent sentences or even all sentences as significant, or were to identify only observation sentences as significant, nothing would have been gained by the new criterion. Effectively, one would have only introduced a new name for a known concept. That none of the criteria given in figures 1 and 2 is trivial can be easily shown by exploiting their inferential relations: Assume $\mathscr{B} = \{B, b\}$, $\mathscr{A} = \{A_1, A_2\}$, and $\Pi = \{\forall x[Bx \leftrightarrow \neg A_1 x]\}$. Then the sentence $A_1 b$ 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.

Besides not being trivial, all criteria meet different conditions of adequacy. This is was to be expected given Lewis's claim of an ambiguity in the explicandum: If the explicandum is used differently in different contexts, each of these contexts leads to different conditions of adequacy. A condition of adequacy that determines falsifiability up to equivalence is the one suggested by Popper and Ayer. They both demand that an empirically significant sentence σ be such that it allows the deduction (or, in Ayer's formulation, the "prediction") of a sentence that is not deducible without σ. This is in effect syntactic $\mathscr{B}$-creativity relative to the set of analytic sentences, and thus claim 1 establishes that up to equivalence, syntactic falsifiability is the only adequate criterion of empirical significance. The semantic analogue of this condition of adequacy is the demand that all and only empirically significant sentences, when conjoined with the analytic sentences, ex-

[26]Lewis's purported ordinary language analyses are very likely themselves already explications (cf. Lutz 2012a, §4), although Lewis's explicata are arguably fairly similar to their explicanda.

clude $\mathscr{B}$-structures that are not excluded by the analytic sentences alone. This is semantic $\mathscr{B}$-creativity, and claim 5 establishes that up to equivalence, only semantic falsifiability is an adequate criterion.

While syntactic falsifiability is adequate if significant sentences have to allow for the inference of basic sentences, verifiability is adequate if the converse inference is required. Indeed, syntactic verifiability is an expression of the condition of adequacy. Semantic verifiability is accordingly adequate if it has to be possible for a sentence to be established as true by a basic structure. A related condition of adequacy that, by claim 24, leads to weak semantic $\mathscr{B}$-determinacy is the demand that for an empirically significant sentence it has to be possible to establish that it is true or establish that it is false. In other words, an empirically significant sentence has to be testable. Since the truth value of a strongly semantically $\mathscr{B}$-determined sentence is fixed for any possible $\mathscr{B}$-structure (claim 13), one can accordingly consider it adequate given the demand that an empirically significant sentence be tested (rather than testable) no matter the state of the world (as given by a $\mathscr{B}$-structure). Under the assumption that it is not the state of the world that counts, but rather our most precise description, claim 20 leads to strong syntactic $\mathscr{B}$-determinacy. Analogously, claim 27 leads to weak syntactic $\mathscr{B}$-determinacy if the condition of adequacy is that it be possible to establish the truth or possible to establish the falsity of an empirically significant sentence from basic sentences.

In summary, then, each of the criteria of empirical significance is uniquely adequate up to logical equivalence given some condition of adequacy plausible in some contexts. For establishing the success of the explication, it therefore remains to be shown that the criteria are precise, simple, and fruitful. The criteria are precise, or at least more precise than the phrase 'empirically meaningful', and some of the formulations are fairly simple—at the least, they are not "page-long", as Lewis (1988a, 127) feared. In fact, the equivalences allow for the application of different formulations according to expedience. Which leaves the central desideratum of fruitfulness.

9 The fruitfulness of the criteria

The criteria are clearly fruitful according to Carnap's own elucidation of his desideratum, since they are useful for establishing all the universal sentences given in this article. In the following, I will first show that the criteria more generally allow for the development of a comprehensive and sound theoretical system as demanded by Hempel and that they suggest new research question as demanded by Kemeny. Second, I will outline a possible application of the criteria to the debate about realism, structuralism realism, and antirealism.[27]

[27] I thank an anonymous referee for this suggestion.

9.1 Six comparative concepts of empirical significance

One way in which the criteria lead to a sound and comprehensive theoretical system and new research questions is by suggesting new, related concepts. For instance, 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. This is in agreement with Hempel's claim that "cognitive significance in a system is a matter of degree" Hempel (1965b, 117). However, 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 temperature than' illustrates (Carnap 1950, §4, Hempel 1952, §10).[28]

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. Therefore I suggest

Definition 28. 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/verifiable/$\mathscr{B}$-determined', in both its syntactic and its semantic guise. '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).

The semantic comparative notion of falsifiability connects directly to the notion of $\mathscr{B}$-content as explicated by the Ramsey sentence:

Claim 29. *Let σ and γ be $\mathscr{V}$-sentences and $\tilde{\Pi}$ a finite axiomatization of Π. Then σ is at least as semantically falsifiable as γ if and only if* $\mathsf{R}_{\mathscr{B}}(\sigma \wedge \bigwedge \tilde{\Pi}) \vDash \mathsf{R}_{\mathscr{B}}(\gamma \wedge \bigwedge \tilde{\Pi})$.

Proof. Immediately from lemma 7. □

[28]Indeed, Hempel (1950, 211) seems to take just this stance towards comparative criteria of empirical significance in an earlier work.

Thus Σ is at least as semantically falsifiable as Γ if and only if the $\mathscr{B}$-content of Σ is logically stronger than the $\mathscr{B}$-content of Γ.

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 28 has natural greatest and, more importantly, least elements.

Claim 30. *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 $\mathfrak{A} \vDash \Pi \cup \Sigma$ and a structure $\mathfrak{B} \vDash \Pi$ with $\mathfrak{B} \nvDash \Sigma$. 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 31. *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 32. *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.

9.2 The relation of the criteria to other concepts

That the criteria of empirical significance suggest new comparative criteria shows that the former allow for the development of a comprehensive and sound theoretical system. Next, I want to briefly point out some new direction of research.

Analyticity. The relation between semantic falsifiability and Ramsey sentences suggests that the criteria discussed here are indeed criteria of *empirical* meaningfulness, not meaningfulness simpliciter. For Carnap, when he suggested the Ramsey sentence $\mathsf{R}_{\mathscr{B}}(\Sigma)$ as the empirical content of Σ, did not claim that only the Ramsey sentence is meaningful. Rather, he suggested that the Carnap sentence $\mathsf{C}_{\mathscr{B}}(\Sigma) := \mathsf{R}_{\mathscr{B}}(\Sigma) \rightarrow \bigwedge \Sigma$ be considered the sentence's analytic component (since its Ramsey sentence is a tautology) and thus meaningful as well. Since the conjunction of Σ's Carnap sentence with its Ramsey sentence is equivalent to Σ, Σ has no meaningless component, no matter what the sentences in Σ are. Hence there are no well-formed meaningless sentences; a non-falsifiable sentence is rather wholly analytic. And since this also holds for ostensibly metaphysical sentences, it suggests a view of metaphysics not as meaningless but as engaged in language choice: Assuming that metaphysical claims are not falsifiable, they can be *chosen* to be true (cf. Lutz 2012a, §4).

At the same time, 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 in general many non-falsifiable sentences, may be verified at some point. Only sentences that are neither falsifiable nor verifiable, and thus not weakly $\mathscr{B}$-determined, allow the choice of either truth value.

Empirical equivalence. Reichenbach (1951, 45–55) considers it necessary for the development of a criterion of empirical significance that one also develop a criterion of empirical equivalence. This is easy given the comparative notions developed in §9.1:

Definition 29. Two sets of $\mathscr{V}$-sentences *are syntactically (semantically) verification-/falsification-/determinacy-equivalent* if and only if each set is at least as syntactically (semantically) falsifiable/verifiable/$\mathscr{B}$-determined as the other.

It will be convenient for the discussion in the following section to have a paraphrase of falsification-equivalence in terms of Ramsey sentences.

Claim 33. *Let $\tilde{\Pi}$ be a finite axiomatization of Π. $\mathscr{V}$-sentences σ and γ are then semantically falsification-equivalent if and only if* $\mathsf{R}_{\mathscr{B}}(\sigma \wedge \bigwedge \tilde{\Pi}) \mathrel{\vDash\!\!\!\dashv} \mathsf{R}_{\mathscr{B}}(\gamma \wedge \bigwedge \tilde{\Pi})$.

Proof. Immediately from claim 29. □

Reichenbach sees the need for a criterion of empirical equivalence because he assumes that a sentence should be empirically non-significant if and only if it is empirically equivalent to a tautology. And while he discusses both concepts in probabilistic terms, his demand is also fulfilled for falsifiability and falsification-equivalence:

Claim 34. *A set Σ of $\mathscr{V}$-sentences is syntactically (semantically) verifiable/falsifiable if and only if Σ is not syntactically (semantically) verification-/falsification-equivalent to $\perp/\top$.*

Proof. Immediate. □

As claim 34 makes clear, Reichenbach's reasoning depends essentially on the kind of criterion of significance under consideration, as it does not hold for verifiability. The relation between strong $\mathscr{B}$-determinacy and determinacy-equivalence is also markedly different:

Claim 35. *A set Σ of $\mathscr{V}$-sentences is syntactically (semantically) strongly $\mathscr{B}$-determined if and only if Σ is syntactically (semantically) determinacy-equivalent to $\perp$ or $\top$.*

Proof. Immediate. □

Thus when empirical significance is defined as strong $\mathscr{B}$-determinacy and empirical equivalence accordingly as determinacy-equivalence, the *contradictory* of Reichenbach's demand is fulfilled. The class of weakly $\mathscr{B}$-determined sentences is not co-extensive with any of the classes of sentences that are determinacy-equivalent to $\perp$, $\top$, or a truth-functional combination of the two. Rather, the following holds.

Claim 36. *Let $A_1, A_2 \in \mathscr{A}$ be such that they do not occur in Π. Then Σ is syntactically (semantically) weakly $\mathscr{B}$-determined if and only if Σ is not syntactically (semantically) determinacy-equivalent to $\forall x(A_1x \leftrightarrow A_2x)$.*

Proof. Since A_1 and A_2 do not occur in Π, every possible $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ has an expansion in which $\forall x(A_1x \leftrightarrow A_2x)$ is true, namely any one with $A_1^{\mathfrak{A}_{\mathscr{B}}} = A_2^{\mathfrak{A}_{\mathscr{B}}} = |\mathfrak{A}_{\mathscr{B}}|$, and an expansion in which it is false, namely any one with $A_1^{\mathfrak{A}_{\mathscr{B}}} = \varnothing$ and $A_2^{\mathfrak{A}_{\mathscr{B}}} = |\mathfrak{A}_{\mathscr{B}}|$. The claim follows immediately. □

Claim 36 provides a first glimpse into the logic that one is actually dealing with when analyzing criteria of empirical significance. While $\top$ is true in every structure that interprets only the terms in Π, and $\perp$ is always false, $\forall x(A_1x \leftrightarrow A_2x)$ is always neither true nor false. For A_1 and A_2 are, in a sense, maximally vague for every structure that interprets only the terms in Π. This suggests that a natural logic for the analysis of empirical significance is one that can deal with vague terms, as, for example, Przełęcki (1958) has maintained (cf. Wójcicki

1966, 87). For instance, a set Σ of sentences is syntactically (semantically) verifiable/falsifiable if and only if it is supertrue/superfalse in all possible models of some possible set of $\mathscr{B}$-sentences (in all possible expansions of some possible $\mathscr{B}$-structure) (cf. Fine 1975).

9.3 A short application: Empiricism and realism, structural realism, and antirealism[29]

Given the preceding, it is clear that the criteria of empirical significance allow the derivation of many universal claims about them, lead to a comprehensive theoretical system, and suggest new directions of research. I now want to show that they are also fruitful in a different sense; they are helpful in clarifying philosophical problems and positions. Although it might seem heroic, I will discuss a specific debate about realism, structural realism, and anti-realism about our best scientific theories as an example. To make the task manageable, I will restrict the discussion to those best theories that have true observational content, rather than approximately true observational content. A generalization of the discussion to theories with approximately true observational content would not be feasible without a thorough analysis of the notion of approximation, which is beyond the scope of this discussion. In the following, 'best theories' will thus always refer to our best theories with only true observational content. For obvious reasons, I will also assume versions of realism, structural realism, and antirealism that are simple enough to be explicated in predicate logic. For most of the discussion, I will further assume that there are no analytic sentences ($\Pi = \varnothing$). Finally, I will assume that the observational content of a theory is given by its Ramsey sentence. This presupposes that the observational vocabulary of the theory's language can be treated as basic ($\mathscr{O} = \mathscr{B}$) and the theoretical vocabulary can be treated as auxiliary ($\mathscr{T} = \mathscr{A}$). Within these constraints, I will discuss Elliott Sober's claim that traditional empiricism should be replaced by what he calls 'contrastive empiricism', which does not evaluate theories on their own, but rather *compares* theories (Sober 1990).

A realist stance towards a best theory ϑ is the simple claim that ϑ is literally true, that is, realism claims that ϑ. The structural realist position that I want to discuss asserts the truth of ϑ's Ramsey sentence, where the existential generalization on the theory's theoretical terms is restricted to the class of somehow privileged properties, for instance "founded" ones or natural kinds (cf. Ainsworth 2009, §§6.1, 6.3). Thus structural realism claims that $\exists \bar{X} \vartheta(\bar{O}, \bar{X}) \wedge \bigwedge_{i=1}^{m} P X_i$, where the extension of P contains all and only privileged properties and $\bar{O}$ is the tuple of observational terms contained in ϑ. While structural realism claims that the theoretical terms in our best theories are interpreted by privileged properties, antirealism claims that there are no privileged properties for theoretical terms at all, and thus specifically that the theoretical terms in our best theories cannot re-

[29] I thank Seamus Bradley, Lorenzo Casini, and Thomas Meier for comments on this section.

fer to any. Antirealism thus claims that $\neg\exists X(PX \wedge UX)$, where U refers to the class of theoretical concepts. However, antirealism does not claim that the observational contents of our best theories are false. Rather, the contents are true by assumption, so that an antirealist stance towards a best theory is the claim that $\mathsf{R}_{\mathcal{O}}(\vartheta) \wedge \neg\exists X(PX \wedge UX)$.

In his argument for contrastive empiricism, Sober (1990, 397–398) argues in a probabilistic framework that for our best theories, realism and antirealism are empirically equivalent. Staying within the confines of deductive inference, Sober's claim is easily shown to be true if empirical equivalence is defined as semantic falsification-equivalence (in keeping with the definition of empirical content of a theory as its Ramsey sentence) and it is assumed that the concept of a privileged property is itself theoretical (Ainsworth 2009, §5.3,§6.3). Then it is straightforward to prove that the realistic, structurally realistic, and antirealistic stances towards any best theory ϑ are semantically falsification-equivalent. For the empirical content of the realist stance is $\mathsf{R}_{\mathcal{O}}(\vartheta)$, that of the structural realist stance is

$$\mathsf{R}_{\mathcal{O}}\left[\exists\bar{X}\vartheta(\bar{O},\bar{X}) \wedge \bigwedge_{i=1}^{n} PX_i\right] \mathrel{\vDash\!\dashv} \exists Y\left[\exists\bar{X}\vartheta(\bar{O},\bar{X}) \wedge \bigwedge_{i=1}^{n} YX_i\right] \mathrel{\vDash\!\dashv} \mathsf{R}_{\mathcal{O}}(\vartheta)\,, \tag{7}$$

and that of the antirealist stance is

$$\mathsf{R}_{\mathcal{O}}[\mathsf{R}_{\mathcal{O}}(\vartheta) \wedge \neg\exists X(PX \wedge UX)] \mathrel{\vDash\!\dashv} \exists Y \exists Z[\mathsf{R}_{\mathcal{O}}(\vartheta) \wedge \neg\exists X(YX \wedge ZX)] \mathrel{\vDash\!\dashv} \mathsf{R}_{\mathcal{O}}(\vartheta)\,.\text{[30]} \tag{8}$$

The semantic falsification-equivalence of the three positions then follows from claim 33.

Sober (1990, 398) now suggests adopting contrastive empiricism, which is the view that there is no genuine disagreement between two empirically equivalent positions. In this way contrastive empiricism dissolves the philosophical problem of deciding between realism, structural realism, and antirealism by showing that they are empirically indistinguishable: any of the positions towards scientific theories can be adopted because "science is not in the business of discriminating between empirically equivalent theories" (Sober 1990, 404). He further claims that this dissolution of the problem shows the superiority of contrastive empiricism over traditional empiricism, because traditional empiricists were unsuccessful in their attempts at dissolving the problem by establishing that realism and antirealism are empty doctrines.

I now want to show that if contrastive empiricism is a viable position, so is non-contrastive, traditional empiricism. Specifically, the assumptions that allow proving the three positions empirically equivalent also allow proving each position non-falsifiable. For neither realism, nor structural realism, nor antirealism

[30]For antirealism, the first equivalence of the derivation 8 assumes that the name for the class of theoretical concepts is theoretical itself. However, the second equivalence would also hold if the Ramseyfication had not existentially generalized on 'U'.

make a claim about the empirical content of a theory. Realism only makes the claim that a best theory is also true. This claim is logically weaker than the claim that a theory with true observational content is true itself, that is,

$$\mathsf{R}_{\mathscr{O}}(\vartheta) \rightarrow \vartheta \ . \tag{9}$$

Structural realism makes the claim that there are privileged properties to which the theoretical terms of a best theory refer, which is logically weaker than the claim that there are such privileged properties for a theory with true observational content:

$$\mathsf{R}_{\mathscr{O}}(\vartheta) \rightarrow \exists \bar{X} \vartheta(\bar{O}, \bar{X}) \wedge \bigwedge_{i=1}^{n} PX_i \ . \tag{10}$$

Antirealism, finally, makes the simple claim that there are no natural kinds as referents for theoretical terms:

$$\neg \exists X (PX \wedge UX) \ . \tag{11}$$

Showing that the Ramsey sentences of all three claims are tautologies is straightforward. And since claims (9) and (10) are logically stronger than realism and structural realism, respectively, and claim (11) expresses antirealism, none of these positions has observational content. As one instance of claim 8, it is now clear that the positions are not semantically falsifiable, so that according to empiricism, each can be chosen to be true with impunity, as they are conventions. Thus there is no need to adopt contrastive empiricism to dissolve the problem. Indeed, in marked distinction to contrastive empiricism, traditional empiricism can establish that realism, structural realism, and antirealism *each on their own* are non-factual positions.

It is clear that within empiricism, this result is obtainable independently of the specific criterion of empirical significance, as long as it is equivalent to semantic falsifiability. For example, it is but one instance of claim 5 that because of their non-falsifiability, all of the positions are semantically $\mathscr{O}$-conservative as well, so that their adoption adds nothing to our empirical knowledge. That the positions are not falsifiable further means that not even a part of the content of the positions is about observation. For the content of each position is given by the set $\mathbf{C}$ of structures that make the position false, and part of the content is about observation if and only if there is a non-empty subset $\mathbf{F} \subseteq \mathbf{C}$ that contains at least for one $\mathscr{O}$-structure all the expansions that make the position false. But since the position is not falsifiable, there are no such subsets. This result is just an instance of claim 6, and it shows that each of the positions individually are not even partly about observation. Syntactically, it is simple to show that none of the positions is creative, that is, that none of them makes an observational assertion, and that none of them can be falsified. Again, traditional empiricism on its own is then sufficient to dissolve the problem.

Finally, it is interesting to relate at least the realist stance to invariance as it is used in Suppes's criterion. For this, ϑ has to be a theory involving measurement, and specifically contain a name $f \in \mathscr{A}$ for a homomorphism μ. It is an indication of the restricted applicability of Suppes's notion that neither structural realism nor antirealism can be evaluated by his criterion. This is because Suppes's criterion cannot be applied to sentences with any terms that are not observational, formal, or names of homomorphisms, and N is none of these ($N \notin \mathscr{F}$ because N has no observational counterpart whose extension could be mapped to the extension of N by the homomorphism μ.) As noted in §8.1, one can define an analogue of strong invariance as a special case of semantic falsifiability:

Definition 30. A set Σ of $\mathscr{E}\mathscr{F}f$-sentences is *falsification-invariant* if and only if there is some $\mathfrak{A} \vDash \Pi_{\text{scale}}$ such that $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \nvDash \Sigma$ for any admissible transformation φ.

To evaluate the falsification-invariance of realism, one has to assume a set Π_{scale} of analytic sentences for scales. Realism (about $\mathscr{E}\mathscr{F}f$-sentences) is then again logically weaker than the claim that given these analytic sentences, if the observational content of $\mathscr{E}\mathscr{F}f$-sentence ϑ is true, then ϑ is true as well, where $\mathscr{E} = \mathscr{O}$, $\mathscr{F} \subseteq \mathscr{T}$, and $f \in \mathscr{T}$:

$$\varrho := \mathsf{R}_{\mathscr{O}}(\vartheta \wedge \bigwedge \Pi_{\text{scale}}) \wedge \bigwedge \Pi_{\text{scale}} \rightarrow \vartheta \ . \tag{12}$$

Since $\vDash \mathsf{R}_{\mathscr{O}}(\varrho)$, ϱ is clearly not semantically falsifiable and hence should also not be falsification-invariant. This can be shown directly.

Claim 37. *ϱ is not falsification-invariant.*

Proof. Assume any $\mathfrak{A} \vDash \Pi_{\text{scale}}$. It has to be shown that not for every φ with $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \vDash \Pi_{\text{scale}}$ it holds that $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \nvDash \varrho$. If for all φ, $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \nvDash \mathsf{R}_{\mathscr{O}}(\vartheta \wedge \bigwedge \Pi_{\text{scale}}) \wedge \bigwedge \Pi_{\text{scale}}$, ϱ is vacuously true for any φ, and the claim follows. Now assume that for some φ, $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \vDash \exists \bar{X} \vartheta(\bar{O}, \bar{X}) \wedge \bigwedge \Pi_{\text{scale}}) \wedge \bigwedge \Pi_{\text{scale}}$. As in the proof of lemma 7, there is then a satisfaction function that can be concatenated with the bijection between theoretical terms and variables that results from Ramseyfication to provide a mapping ζ from theoretical terms to extensions such that $\mathfrak{A}[f/\zeta(f), \bar{T}/\zeta(\bar{T})] \vDash \vartheta(\bar{O}, f, \bar{T}) \wedge \bigwedge \Pi_{\text{scale}}$, where $\bar{T}$ is a sequence of theoretical terms other than f. Since Π_{scale} determines $\mathfrak{A}|_{\mathscr{F}}$ up to isomorphism, $\mathfrak{A}[f/\zeta(f), \bar{T}/\zeta(\bar{T})] = \mathfrak{A}[f/\zeta(f)]$ up to isomorphism. Now choose φ^* such that $\varphi^*(\mu) = \zeta(f)$ for every function μ. Then $\mathfrak{A}[f/\varphi^*(f^{\mathfrak{A}})] \vDash \vartheta(\bar{O}, f, \bar{T})$ and thus $\mathfrak{A}[f/\varphi^*(f^{\mathfrak{A}})] \vDash \varrho$. Again, the claim follows. □

Thus if empirical significance is defined as falsification-invariance, realism towards ϑ is not empirically significant either, and again traditional empiricism is sufficient to dissolve the problem of realism.

10 Conclusion

The belief that the search for a criterion of empirical significance has been a failure is usually based on the string of trivialization proofs for purported criteria. I have given an alternative view on this search, based on criteria that are provably non-trivial. I have also argued that the criteria successfully explicate their explicandum: For one, they fruitfully connect to each other, to comparative criteria of empirical significance, Ramsey sentences, and concepts from measurement and definition theory. The criteria also suggest new lines of research, for instance in connection with the notions of analyticity, empirical equivalence, and vagueness. Furthermore, the criteria and their relations can be fruitfully applied to problems in philosophy, for example the debate about realism, structural realism, and anti-realism. That the criteria are successful explications counters the charge of arbitrariness.

Because of the inferential relations between the criteria, the motivations of the individual criteria lead to a cumulative defense of the criteria's material adequacy. Furthermore, the relations allow for a more informed search for generalizations of the criteria; falsifiability, for example, is easily transferred to any logic in which the notion of creativity can be explicated. My hope is that the relations also provide guidance in the evaluation of criteria for the empirical significance of terms and in the search for criteria that can be applied in probabilistic contexts.

Thus a host of criteria for the empirical significance of sentences, justified in a variety of ways, stand in strong inferential relations to each other, fulfill the desiderata for explications, and provide paths towards their generalizations. For this reason, I consider the search for criteria of empirical significance a success.

References

Ainsworth, P. M. (2009). Newman's objection. *The British Journal for the Philosophy of Science*, 60:135–171. 44, 45

Ayer, A. J. (1936). *Language, Truth and Logic*. Victor Gollanz, London, 1st edition. References are to the second edition (Ayer 1946). 9, 10, 12, 31, 32

Ayer, A. J. (1946). *Language, Truth and Logic*. Victor Gollanz, London, 2nd edition. 11, 48

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. 33

Belnap, N. (1993). On rigorous definitions. *Philosophical Studies*, 72:115–146. 9, 10, 20, 36

Carnap, R. (1926). *Physikalische Begriffsbildung*, volume 39 of *Wissen und Wirken. Einzelschriften zu den Grundfragen des Erkennens und Schaffens*. Verlag G. Braun, Karlsruhe. 6, 22

Carnap, R. (1928a). *Der logische Aufbau der Welt*. Weltkreis-Verlag, Berlin-Schlachtensee. References are to the translation (Carnap 1967). 8

Carnap, R. (1928b). *Scheinprobleme in der Philosophie: Das Fremdpsychische und der Realismusstreit*. Weltkreis-Verlag, Berlin-Schlachtensee. References are to the translation (Carnap 1967). 8, 16, 32

Carnap, R. (1936). Testability and meaning. *Philosophy of Science*, 3(4):420–468. 30, 31

Carnap, R. (1937). Testability and meaning—continued. *Philosophy of Science*, 4(1):2–35. 30

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. 7

Carnap, R. (1950). *Logical Foundations of Probability*. University of Chicago Press, Chicago. References are to the 2nd edition (Carnap 1962). 3, 36, 38, 40

Carnap, R. (1952). Meaning postulates. *Philosophical Studies*, 3(5):65–73. 5

Carnap, R. (1956). The methodological character of theoretical concepts. In Feigl, H. and Scriven, M., editors, *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. 7

Carnap, R. (1962). *Logical Foundations of Probability*. University of Chicago Press, Chicago, 2nd edition. 49

Carnap, R. (1963). Replies and systematic expositions. In Schilpp, P. A., editor, *The Philosophy of Rudolf Carnap*, volume 11 of *The Library of Living Philosophers*, pages 859–1016. Open Court Publishing Company, Chicago and LaSalle, IL. 8

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. 6, 8, 9, 22

Carnap, R. (1967). *The Logical Structure of the World. Pseudoproblems of Philosophy*. University of California Press, Berkeley and Los Angeles. Translation by Rolf A. George. 49

Church, A. (1949). Review of Ayer's *Language, Truth, and Logic*, 2nd edition. *The Journal of Symbolic Logic*, 14(1):52–53. 2, 11, 36

Copeland, B. J. (2008). The Church-Turing thesis. 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 2008 edition. 33

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. 3

Diamond, M. L. and Litzenburg, T. V., editors (1975). *The Logic of God: Theology and Verification*. Bobbs-Merill, Indianapolis, IN. 3

Fine, K. (1975). Vagueness, truth and logic. *Synthese*, 30(3–4):265–300. References are to the corrected reprint (Fine 1997). 44

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. 50

Fodor, J. A. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese*, 28:97–115. 14

Gemes, K. (1998). Logical content and empirical significance. In Weingartner, P., Schurz, G., and Dorn, G., editors, *Proceedings of the 20^th^ International Wittgenstein Symposium*, Vienna. H¨older-Pichler-Tempski. 7, 8, 12, 30

Gupta, A. (2009). Definitions. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, spring 2009 edition. 9, 20

Hempel, C. G. (1950). Problems and changes in the empiricist criterion of meaning. *Revue Internationale de Philosophie*, 11:41–63. 32, 40

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. 4, 6, 22, 40

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. 19

Hempel, C. G. (1965a). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. The Free Press, New York. 50, 51

Hempel, C. G. (1965b). Empiricist criteria of cognitive significance: Problems and changes. In Hempel (1965a), pages 81–119. 7, 8, 10, 17, 40

Hempel, C. G. (1965c). Postscript (1964) on cognitive significance. In Hempel (1965a), pages 120–122. 8, 18, 33

Hempel, C. G. (1965d). Studies in the logic of confirmation. In Hempel (1965a), pages 3–51. 12

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. 40

Hodges, W. (1993). *Model Theory*, volume 42 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge. Digitally printed in 2008. 6, 23, 24

Kemeny, J. G. (1956). A new approach to semantics—part I. *The Journal of Symbolic Logic*, 21(1):1–27. 14

Kemeny, J. G. (1963). Analyticity versus fuzziness. *Synthese*, 15(1):57–80. 4, 6, 14

Kitts, D. B. (1977). Karl Popper, verifiability, and systematic zoology. *Systematic Zoology*, 26(2):185–194. 19

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. 3, 4

Kyburg, H. E. (1984). *Theory and Measurement*. Cambridge Studies in Philosophy. Cambrigde University Press, Cambridge. 6

Laudan, L. (1986). Some problems facing intuitionist meta-methodologies. *Synthese*, 67(1):115–129. 38

Lewis, D. (1970). How to define theoretical terms. *The Journal of Philosophy*, 67(13):427–446. 5

Lewis, D. (1988a). Ayer's first empiricist criterion of meaning: Why does it fail? *Analysis*, 48(1):1–3. 2, 11, 39

Lewis, D. (1988b). Statements partly about observation. *Philosophical Papers*, 17:1–31. References are to the reprint (Lewis 1998, 125–155). 11, 13, 14, 15, 28

Lewis, D. (1998). *Papers in Philosophical Logic*. Cambridge Studies in Philosophy. Cambridge University Press, Cambridge. 51

Lutz, S. (2012a). Artificial language philosophy of science. *European Journal for Philosophy of Science*, 2(2):181–203. Special issue: *The Future of Philosophy of Science*, edited by Stephan Hartmann and Jan Sprenger. 4, 6, 38, 42

Lutz, S. (2012b). *Criteria of Empirical Significance: Foundations, Relations, Applications*. PhD thesis, Utrecht University. http://philsci-archive.pitt.edu/id/eprint/9117. 4, 5, 7

Lutz, S. (2012c). Empiricism and Intelligent Design II: Analyzing Intelligent Design. *Erkenntnis*. Forthcoming. DOI: 10.1007/s10670-012-9392-5. 3

Lutz, S. (2012d). 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):77–120. 4

Martin, R. M. (1952). On 'analytic'. *Philosophical Studies*, 3:42–47. 4, 6

Mates, B. (1951). Analytic sentences. *Philosophical Review*, 60:525–534. 6

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. 51

Nagel, E. (1951). Mechanistic explanation and organismic biology. *Philosophy and Phenomenological Research*, 11(3):327–338. 14

Niiniluoto, I. (1972). Inductive systematization: Definition and a critical survey. *Synthese*, 25(1–2). 17

Peacocke, C. (1986). *Thoughts: an Essay on Content*, volume 4 of *Aristotelian Society Series*. Basil Blackwell, New York, NY. 18

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). 4, 8, 9, 10, 40, 52

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. 52

Przełęcki, M. (1958). W sprawie terminów nieostrych. *Studia Logica*, 8(1):313–317. 43

Przełęcki, M. (1969). *The Logic of Empirical Theories*. Monographs in Modern Logic Series. Routledge & Kegan Paul/Humanities Press, London/New York. 5, 8, 9, 20, 22

Przełęcki, M. (1974a). Empirical meaningfulness of quantitative statements. *Synthese*, 26:344–355. 5, 6, 7, 12, 19, 20, 21, 22, 29, 36

Przełęcki, M. (1974b). On model theoretic approach to empirical interpretation of scientific theories. *Synthese*, 26:401–406. 8

Psillos, S. (2000). Rudolf Carnap's 'Theoretical concepts in science'. *Studies in History and Philosophy of Science Part A*, 31:151–172. 8, 16

Putnam, H. (1962). What theories are not. In Nagel, E., Suppes, P., and Tarski, A., editors, *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress*. Stanford University Press, Stanford. 5

Raatikainen, P. (2010). Ramsification and inductive inference. *Synthese*. Print publication forthcoming. 17

Reichenbach, H. (1951). The verifiability theory of meaning. *Proceedings of the American Academy of Arts and Sciences*, 80(1):46–60. 5, 42

Roberts, F. S. and Franke, C. H. (1976). On the theory of uniqueness in measurement. *Journal of Mathematical Psychology*, 14:211–218. 23, 25

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. 8, 16

Rozeboom, W. W. (1970). The crisis in philosophical semantics. In Radner, M. and Winokur, S., editors, *Analyses of Theories and Methods of Physics and Psychology*, volume 4 of *Minnesota Studies in the Philosophy of Science*, pages 196–219. University of Minnesota Press, Minneapolis, MN. 5, 8

Ruja, H. (1961). The present status of the verifiability criterion. *Philosophy and Phenomenological Research*, 22(2):216–222. 7

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. 7, 18, 32

Salmon, W. C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press, Princeton. 37

Scheffler, I. (1968). Reflections on the ramsey method. *The Journal of Philosophy*, 65(10):269–274. Double issue: *On Scientific Hypotheses*. 17

Simpson, J. A. (1981). Foundations of metrology. *Journal of Research of the National Bureau of Standards*, 86(3):281–292. 6

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. 3

Sober, E. (1990). Contrastive empiricism. In Savage, C. W., editor, *Scientific Theories*, volume 14 of *Minnesota Studies in the Philosophy of Science*, pages 392–410. University of Minnesota Press, Minneapolis, MN. 3, 44, 45

Sober, E. (2008). *Evidence and Evolution: The Logic Behind the Science*. Cambridge University Press, Cambridge. 3, 7, 8

Suppe, F. (1972). What's wrong with the received view on the structure of scientific theories? *Philosophy of Science*, 39(1):1–19. 5

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. 21, 22, 23, 25

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. 23, 25

Tarski, A. (1944). The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research*, 4(3):341–376. 4, 33

Tuomela, R. (1973). *Theoretical Concepts*. Number 10 in Library of Exact Philosophy. Springer, Wien. 17

van Benthem, J. (1978). Ramsey eliminability. *Studia Logica*, 37:321–336. 13

van Fraassen, B. C. (1980). *The Scientific Image*. The Clarendon Library of Logic and Philosophy. Clarendon Press, Oxford. 5

Wassermann, G. D. (1978). Testability of the role of natural selection within theories of population genetics and evolution. *The British Journal for the Philosophy of Science*, 29:223–242. 3

Wójcicki, R. (1966). Semantical criteria of empirical significance. *Studia Logica*, 19:75–102. 43