Criteria of Empirical Significance: A Success Story

Sebastian Lutz[*]

Draft: 2010–12–11

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; two criteria by Carnap and Sober are 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. Furthermore, the equivalence of some criteria allows for their individual justifications to be taken cumulatively and, together with the entailment relations between non-equivalent criteria, suggest criteria for general auxiliary assumptions, comparative criteria, and more liberal conceptions of observation.

Contents

[*]Theoretical Philosophy Unit, Utrecht University, The Netherlands. sebastian.lutz@gmx.net.

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 (1936, 38f) was quickly shown to be trivial; it was followed by a slew of amendments and new trivialization proofs succinctly summarized and extended by Pokriefka (1983), who cuts out the middleman and proves the triviality of his amendment himself (Pokriefka 1984). The latest contributions to this "puncture-and-patch industry" (Lewis 1988a, §XII) are two criteria by Wright (1986, 1989) and trivialization proofs by Lewis (1988a, §IV, n. 12), Wright (1989, §II), and Yi (2001).

This history has "done a lot to discredit the very idea of delineating a class of statements as empirical" (Lewis 1988a, §I), but does not show that *all* suggested criteria are trivial. Still, non-trivial criteria could be charged with arbitrariness, as Lewis (1988a, 127, footnote removed) does when he notes that the amendments of Ayer's criterion

> have led to ever-increasing complexity and ever-diminishing contact with any intuitive idea of what it means for a statement to be empirical. Even if some page-long descendant of Ayer's criterion [provably admitted] more than the observation-statements and less than all the statements, we would be none the wiser. We do not want just any class of statements that is intermediate between clearly too little and clearly too much. We want the right class.

This also holds for criteria that do not amend Ayer's criterion. Their multitude suggests that they are little more than arbitrary bipartitions of statements.

One goal of this article is to show that this charge of arbitrariness is unfounded, for many of the non-trivial criteria are equivalent or bear strong inferential relations to each other and to concepts from definition and measurement theory. Among the criteria are falsifiability and verifiability, but also more contemporary suggestions, the most recent being by Sober (2008). I will argue that there are essentially four major non-trivial criteria of empirical significance that stand in simple entailment relations to each other.

This is already a good reason to look more closely at the non-trivial 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. But there is also still a *need* for criteria of empirical significance. Sometimes a criterion is needed to state very clearly what is not generally under dispute, as in Sober's discussions of the empirical significance of claims about a designer of life whose intentions and abilities are unknown (Sober 1999, 2007, 2008). In other cases, a generally accepted endeavor is put under scrutiny, like string theory (Smolin 2006, Woit 2006), fish stock assessment theories (Corkett 2002), or natural selection (Wassermann 1978). The empirical significance of more philosophical positions like theism (Diamond and Litzenburg 1975) or realism and antirealism (Sober 1990) have also been investigated.

In the remainder of this article, I will discuss several criteria of empirical significance that turn out to be equivalent (§3.1) or nearly equivalent (§3.2) to falsifiability, and I will briefly describe verifiability (§4). Falsifiability and verifiability are more inclusive than the (universally panned) criterion demanding both (§5), which itself is more inclusive than the criterion of strong $\mathscr{O}$-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 $\mathscr{O}$-creativity thus make up the four major criteria of empirical significance. The entailment relations between them suggest the introduction of comparative concepts of empirical significance (§8.1). And since the different formulations of each major criterion have been arrived at by different considerations, their equivalence justifies a re-evaluation of their statuses (§8.2). The equivalences of the different formulations also allow to choose based on expedience which formulation to generalize. I suggest two generalizations, one of them weakening the assumptions about observations (§9.1). And modifying an otherwise trivial criterion by Elliott Sober leads to a generalization of falsifiability that takes general background assumptions into account. This generalization transfers directly to all other criteria (§9.2). In the end, these results will provide evidence that the search for criteria of empirical equivalence has been successful.

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.

Carnap explicates 'fruitful' as 'useful for the formulation of many universal statements', but this suggests that fruitfulness can be determined by counting sentences with universal quantifiers. The underlying idea has been put better 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". This suggests that the explicatum should connect in simple ways to concepts related to the explicandum, which, in the case of empirical significance, could be the notions of empirical content, confirmation, and measurement, for instance. To demand specific connections would lead to precise conditions of adequacy for the explicatum (cf. Tarski 1944, §4), but their discussion would take me too far. Explicata are correct or incorrect only *relative* to the conditions of adequacy, which are pragmatically chosen based on the intended application of the explicatum. Thus explications are not true or false claims, but more or less expedient suggestions (Popper 1935, 37f; 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 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 in fact suggest immediate generalizations beyond predicate logic. Second, not all major criticisms of the syntactic view are in fact justified (Lutz 2010a). 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 2010c). 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 observational terms $\mathscr{O}$ and theoretical terms $\mathscr{T}$, with sentences containing $\mathscr{V}$-terms called $\mathscr{V}$-sentences, sentences containing only $\mathscr{O}$-terms called observational or $\mathscr{O}$-sentences, and sentences containing only $\mathscr{T}$-terms called theoretical or $\mathscr{T}$-sentences. 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 to be ideal 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 an ideal 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) has suggested to achieve this result by simply taking all terms in the sciences as theoretical and introducing an artificial observational language. Friedman (1982, 276f) suggests a similar strategy to capture the notion of an empirical substructure with the help of a bipartition of the vocabulary. This is of specific interest because Van Fraassen (1980, §3.6) has conjectured that this is impossible. 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 observation 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, observation sentences do not have to be about observations in any sense of the word, but must only be unproblematic for the purposes at hand. According to Nielsen (1966, 15), for example, Flew's charge that theological statements are not falsifiable (Flew 1950) assumes that all and only "non-religious, straightforwardly empirical, factual statements" have primitive meaning. And Flew (1975, 274) claims that it is enough to assume that all and only statements about "anything which happens or which conceivably might happen in the ordinary world" have primitive meaning.[1]

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 $\Pi_{\mathscr{O}}$ for $\mathscr{O}$-terms (cf. Carnap 1952), and meaning postulates $\Pi_{\mathscr{T}}$ for $\mathscr{T}$-terms. The meaning postulates for $\mathscr{O}$-terms are $\mathscr{O}$-sentences, while those for $\mathscr{T}$-terms are $\mathscr{V}$-sentences, because they give the $\mathscr{T}$-terms' relations to each other and to $\mathscr{O}$-terms. Przełęcki (1974a, 345) argues that $\Pi_{\mathscr{T}}$ should be $\mathscr{O}$-conservative with respect to $\Pi_{\mathscr{O}}$, that is, $\Pi_{\mathscr{T}}$ should place no restrictions on $\mathscr{O}$-sentences or their interpretations beyond those given through $\Pi_{\mathscr{O}}$. 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

[1] These are only illustrations: I especially do not think that Flew's circumscription of the $\mathscr{O}$-sentences is precise enough, since theological statements might be considered to be about the "ordinary world".

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 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. 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'. I will further discuss the role of Π in §9.2. Π determines which $\mathscr{O}$-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 observational terms relate. To arrive at a formal definition, let $\mathfrak{A}|\mathscr{O}$ refer to the reduct of $\mathfrak{A}$ to $\mathscr{O}$, that is, the structure that results from eliminating the interpretations of all $\mathscr{T}$-terms from $\mathfrak{A}$. For an $\mathscr{O}$-structure $\mathfrak{A}_{\mathscr{O}}$, a structure $\mathfrak{B}$ with $\mathfrak{B}|\mathscr{O} = \mathfrak{A}_{\mathscr{O}}$ is called an expansion of $\mathfrak{A}_{\mathscr{O}}$ (Hodges 1993, 9). Any $\mathscr{O}$-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

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

Below, I will make a distinction between syntactic and semantic criteria of empirical significance based on whether the observations are described by sets of $\mathscr{O}$-sentences or by $\mathscr{O}$-structures. With $\mathscr{O}$-structures, observations can be described up to isomorphism, and with $\mathscr{O}$-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 ($\mathrm{Th}(\mathfrak{A}) \,\rightleftharpoons\, \mathrm{Th}(\mathfrak{B})$), that is, for all sentences φ, $\mathfrak{A} \vDash \varphi$ if and only if $\mathfrak{B} \vDash \varphi$. In first order logic, syntactic equivalence is called elementary equivalence (and is not equivalent to isomorphy).

2.2 On the explicandum

As I will not discuss different conditions of adequacy, I will not attempt a thorough elucidation of the explicandum of empirical significance. Rather, I will give a circumscription precise enough to counter some common criticisms.

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, 345f), 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.

The criteria are also not meant to determine the meaning of sentences as Ruja (1961), for instance, assumes in his critique. Rynin (1957, 51ff) 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, 149f). 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, which asserts the opposite (Rozeboom 1962, §II; Rozeboom 1970; Przełęcki 1969, §§5f; Przełęcki 1974b, 402f). If a sentence can be meaningful without being empirically significant, most of the criticisms by Hempel (1965b) are invalid (Hempel 1965c; Sober 2008, 149f). 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).

Gemes (1998, §1.4) argues that a criterion of empirical significance does not have to be a criterion of inductive confirmability as well. In the following, I will take the weaker stance that a criterion of deductive empirical significance may differ significantly from a criterion of inductive empirical significance, which in turn *may* differ from a criterion of inductive confirmability. And since I will discuss the criteria in the following as criteria of deductive empirical significance, not as criteria of inductive confirmability, 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 mainly rely on what Carnap sometimes calls the 'extended observation language', which contains all sentences that contain only logical and $\mathscr{O}$-terms (cf. Psillos

2000, 158f). The language thus also includes all quantified sentences, and thus "empirical laws" or "empirical generalizations" (Carnap 1966, 225–227). I will revisit restrictions on the observation sentences in §9.1.

3 Falsifiability

3.1 Syntactic criteria

As I will discuss in §9.2, Hempel's formulation of the falsifiability criterion deviates from Popper's original criterion in at least one crucial respect, but it can serve as a good starting point for my discussion. Hempel (1965b, 106) states his "requirement of complete falsifiability in principle" like this:

> A sentence has empirical meaning if and only if its negation is not analytic and follows logically from some finite logically consistent class of observation sentences.

Since I am here not interested in criteria of confirmability, I will drop Hempel's requirement that the set of observational 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 $\mathscr{T}$-sentences follow from a consistent set of $\mathscr{O}$-sentences, and therefore no $\mathscr{T}$-sentences have empirical meaning according to Hempel's definition. Finally, I will generalize the criterion for sentences to a criterion for sets of sentences because this allows the discussion of theories that cannot be finitely axiomatized and thus not be described in a single sentence. The generalization is straightforward: If α is a sentence and Γ a set thereof, then $\Gamma \vDash \neg\alpha$ if and only if $\Gamma \cup \{\alpha\} \vDash \bot$, 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.

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

Definition 4. A set Λ of $\mathscr{V}$-sentences is *syntactically falsifiable* if and only if it is falsified by a possible set of $\mathscr{O}$-sentences.

As noted, the qualifier 'syntactic' here does not refer to the use of syntactic deduction ('$\vdash$'), but to the syntactic description of observations (by 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 '$\mathscr{O}$-sentence' to be any sentence containing only $\mathscr{O}$-terms, this criterion, like all other syntactic criteria in the following, only presumes that the

$\mathscr{O}$-sentences form some distinguished set of sentences. The syntactic criteria thus 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 observational 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 $\mathscr{O}$-conservativeness, which is a necessary condition for explicit definitions (cf. Belnap 1993; Gupta 2009, §2.1).

Definition 5. A set Λ of $\mathscr{V}$-sentences is *syntactically $\mathscr{O}$-conservative with respect to* a set Δ of $\mathscr{V}$-sentences if and only if for any set Ω of $\mathscr{O}$-sentences and for any $\mathscr{O}$-sentence ω, $\Omega \cup \Lambda \cup \Delta \vDash \omega$ only if $\Omega \cup \Delta \vDash \omega$.

A set of $\mathscr{V}$-sentences is *$\mathscr{O}$-creative with respect to Δ* if and only if it is not $\mathscr{O}$-conservative with respect to Δ.

If a logic is compact, $\Omega \cup \Lambda \cup \Delta \vDash \omega$ if and only if there is a finite set Ω' such that $\Omega' \cup \Lambda \cup \Delta \vDash \omega$. This is equivalent to $\Lambda \cup \Delta \vDash \neg(\bigwedge \Omega' \to \omega)$, where $\bigwedge \Omega'$ is the conjunction of all elements of Ω'. Hence for first order logic, the use of the set Ω in definition 5 is superfluous if the set of observation sentences is closed under truth functional composition.

That the definition of a new term not in $\mathscr{O}$ must be $\mathscr{O}$-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 $\mathscr{O}$-conservative with respect to Π sanctions no inferences between $\mathscr{O}$-sentences that are not already sanctioned by Π. In the following, $\mathscr{O}$-conservativeness simpliciter is understood to be $\mathscr{O}$-conservativeness with respect to Π.

Popper's justification of falsifiability essentially starts from $\mathscr{O}$-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 an observation sentence is itself an observation sentence, he arrives at his definition of falsifiability:

Claim 1. *A set Λ of $\mathscr{V}$-sentences is syntactically falsifiable iff Λ is syntactically $\mathscr{O}$-creative with respect to Π.*

Proof. '$\Rightarrow$': If $\Omega \cup \Lambda \cup \Pi \vDash \bot$, then $\Omega \cup \Lambda \cup \Pi \vDash \omega$ for any observation sentence ω. Since $\Omega \cup \Pi \nvDash \bot$, there is some ω such that $\Omega \cup \Pi \nvDash \omega$.

'$\Leftarrow$': For ω and Ω with $\Omega \cup \Lambda \cup \Pi \vDash \omega$ and $\Omega \cup \Pi \nvDash \omega$, $\Omega \cup \{\neg\omega\} \cup \Pi \nvDash \bot$ and $\Omega \cup \{\neg\omega\} \cup \Lambda \cup \Pi \vDash \bot$. □

The relation between falsifiability and $\mathcal{O}$-creativity provides a justification for Reichenbach's (and Nielsen's and Flew's) claim that the $\mathcal{O}$-sentences only need to be unproblematic: The theory of definition and the concept of $\mathcal{O}$-creativity are independent of the meaning of the $\mathcal{O}$-terms.

Sticking with the standard interpretation of $\mathcal{O}$-sentences, 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" (suitable subsidiary hypotheses for falsifiability being analytic and observational). It is one of the cruel jokes of philosophical terminology that in this quote, Hempel (1965b, 106) describes Ayer's two criteria of verifiability. Accordingly, the justification that Ayer provides for his criteria complements Popper's justification. Ayer (1936, 97ff) argues that the function of an empirical hypothesis is to predict experiences, and thus arrives at his first criterion of empirical significance, namely that "the mark of a genuine factual proposition [is] that some 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, 38f).

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 $\mathcal{O}$-sentences, which makes the criterion equivalent to $\mathcal{O}$-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'[2] and the criterion is meant as a necessary and sufficient condition, this can be paraphrased as

Definition 6. A $\mathcal{V}$-sentence α is *directly verifiable* if and only if α is an $\mathcal{O}$-sentence or there is some set Ω of $\mathcal{O}$-sentences and an $\mathcal{O}$-sentence ω such that $\Omega \cup \{\alpha\} \cup \Pi \vDash \omega$ and $\Omega \cup \Pi \nvDash \omega$.

Without any assumptions about the set of observation sentences, this follows immediately:

Claim 2. *A $\mathcal{V}$-sentence α is directly verifiable iff α is an $\mathcal{O}$-sentence or is syntactically $\mathcal{O}$-creative with respect to Π.*

[2] This is what Ayer seems to do, since he calls translations from one language into another 'logically equivalent' (Ayer 1946, 6f). 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.

The condition that α may be an $\mathscr{O}$-sentence is not redundant because α may be analytic and therefore not $\mathscr{O}$-creative with respect to Π.

In his second definition, Ayer (1946, 13) proposes

> to say that a statement is indirectly verifiable if it satisfies the following conditions: first, that in conjunction with certain other premises [Γ] it entails one or more directly verifiable statements [β] which are not deducible from these other premises alone; and secondly, that these other premises do not include any statement that is not either analytic, or directly verifiable, or capable of being independently established as indirectly verifiable.

Since analytic entailment already allows the inclusion of Π in the premises of an inference, Π can be dropped from the auxiliary assumptions Γ. In the special case that Γ is a set of $\mathscr{O}$-sentences and β an $\mathscr{O}$-sentence as well, indirect verifiability reduces to direct verifiability (cf. Pokriefka 1983),[3] so that Γ can contain $\mathscr{O}$-sentences instead of directly verifiable sentences. Ayer's criterion can then be stated as

Definition 7. A $\mathscr{V}$-sentence α is *indirectly verifiable* if and only if there is a set Γ of indirectly verifiable or $\mathscr{O}$-sentences and a sentence β that is directly verifiable such that $\{\alpha\} \cup \Gamma \cup \Pi \vDash \beta$ and $\Gamma \cup \Pi \nvDash \beta$.

Church (1949) shows that for any sentence, as long as there are three logically independent observational sentences, the sentence or its negation is indirectly verifiable, a trivialization that is possible even if β is required to be an $\mathscr{O}$-sentence. This trivialization can be avoided by restricting both Γ and β to $\mathscr{O}$-sentences, but this more exclusive version of indirect verifiability then again just amounts to $\mathscr{O}$-creativity and falsifiability.

In connection with his first criterion, Ayer (1936, 38) argues that a "hypothesis cannot be conclusively confuted any more than it can be conclusively verified", but that a sentence is verifiable "if it is possible for experience to render it probable" (Ayer 1936, 37). Ayer (1936, 99) then argues that "if an observation to which a given proposition is relevant conforms to our expectations, the truth of that proposition is confirmed. [Then] one can say that its probability has been increased." 'Probability' is here not used in its mathematical sense, but as a measure of our "confidence" in a proposition (Ayer 1936, 100). Thus 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]This holds even without the assumption needed for Church's trivialization proof given below, simply by restricting Γ and β; hence I consider it an innocent observation that does not just transform one trivial criterion into another one.

3.2 Semantic criteria

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

Definition 8. A set Λ of $\mathscr{V}$-sentences is *semantically $\mathscr{O}$-conservative with respect to* a set Δ of $\mathscr{V}$-sentences if and only if for each $\mathscr{O}$-structure $\mathfrak{A}_{\mathscr{O}}$ for which there is a $\mathscr{V}$-structure $\mathfrak{B} \vDash \Delta$ with $\mathfrak{B}|\mathscr{O} = \mathfrak{A}_{\mathscr{O}}$, there is also a $\mathscr{V}$-structure $\mathfrak{C} \vDash \Delta \cup \Lambda$ with $\mathfrak{C}|\mathscr{O} = \mathfrak{A}_{\mathscr{O}}$.

Definition 8 is slightly more general than that given, for example, by Przełęcki (1974a, 345), so that it allows for any $\mathscr{V}$-sentence in Δ. A description of the generalization is given in appendix A. Note that, like the other semantic definitions up to §8.1, 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 observational information:

Claim 3. *A set Λ of $\mathscr{V}$-sentences is syntactically $\mathscr{O}$-conservative with respect to Δ iff for each $\mathscr{O}$-structure $\mathfrak{A}_{\mathscr{O}}$ for which there is a $\mathscr{V}$-structure $\mathfrak{B} \vDash \Delta$ with $\mathfrak{B}|\mathscr{O} \equiv \mathfrak{A}_{\mathscr{O}}$, there is a $\mathscr{V}$-structure $\mathfrak{C} \vDash \Delta \cup \Lambda$ with $\mathfrak{C}|\mathscr{O} \equiv \mathfrak{A}_{\mathscr{O}}$.*

Proof. '⇒': Assume $\mathfrak{A}_{\mathscr{O}}$ is syntactically equivalent to a structure that can be expanded to a model $\mathfrak{B}$ of Δ. Then choose $\Omega \cup \{\neg\omega\}$ equivalent to $\mathrm{Th}(\mathfrak{A}_{\mathscr{O}})$. I follows that $\mathfrak{B} \vDash \Omega \cup \{\neg\omega\} \cup \Delta$ and thus $\Omega \cup \Delta \nvDash \omega$. By syntactic $\mathscr{O}$-conservativeness, $\Omega \cup \Lambda \cup \Delta \nvDash \omega$, so there is a $\mathfrak{C} \vDash \Omega \cup \{\neg\omega\} \cup \Lambda \cup \Delta \mathrel{\models\!\!\!\dashv} \mathrm{Th}(\mathfrak{A}_{\mathscr{O}}) \cup \Lambda \cup \Delta$. Thus there is a $\mathfrak{C} \vDash \Lambda \cup \Delta$ such that $\mathfrak{C}|\mathscr{O} \equiv \mathfrak{A}_{\mathscr{O}}$.

'⇐': Let $\Omega \cup \Delta \nvDash \omega$. Choose $\mathfrak{A} \vDash \Omega \cup \Delta \cup \{\neg\omega\}$; by assumption, there is a $\mathfrak{C} \vDash \Lambda \cup \Lambda$ with $\mathfrak{C}|\mathscr{O} \equiv \mathfrak{A}|\mathscr{O}$ and thus $\mathfrak{C} \vDash \Omega \cup \Lambda \cup \Delta \cup \{\neg\omega\}$, so that $\Omega \cup \Lambda \cup \Delta \nvDash \omega$. □

This suggests

Claim 4. *A set Λ of $\mathscr{V}$-sentences is semantically $\mathscr{O}$-conservative with respect to Δ only if Λ is syntactically $\mathscr{O}$-conservative with respect to Δ. The converse does not hold in first order logic.*

Proof. '⇒': From claim 3 because $\mathfrak{A}|\mathscr{O} = \mathfrak{B}|\mathscr{O}$ only if $\mathfrak{A}|\mathscr{O} \equiv \mathfrak{B}|\mathscr{O}$.

'⇍': See appendix B. □

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

A short overview of mainly philosophical treatments of the relation is given in appendix B. Because of the difference between syntactic and semantic $\mathscr{O}$-conservativeness, it may not always be possible to bipartition the set of analytic sentences Π such that $\Pi_{\mathscr{T}}$ is semantically $\mathscr{O}$-conservative with respect to $\Pi_{\mathscr{O}}$: If Π is only syntactically conservative with respect to $\Pi_{\mathscr{O}}$, there are some $\mathscr{O}$-models of $\Pi_{\mathscr{O}}$ that cannot be expanded to models of Π, and there is no $\mathscr{O}$-sentence that excludes all and only those structures when added to $\Pi_{\mathscr{O}}$.

The analogy between syntactic and semantic $\mathscr{O}$-conservativeness suggests a semantic criterion of falsifiability analogous to syntactic falsifiability.

Definition 9. An *$\mathcal{O}$-structure $\mathfrak{A}_{\mathcal{O}}$ falsifies* a set Λ of $\mathcal{V}$-sentences if and only if for all $\mathfrak{C} \vDash \Pi$ with $\mathfrak{C}|\mathcal{O} = \mathfrak{A}_{\mathcal{O}}$, $\mathfrak{C} \nvDash \Lambda$.

In other words, a structure $\mathfrak{A}_{\mathcal{O}}$ falsifies Λ if and only if Λ is false in every possible structure that is an expansion of $\mathfrak{A}_{\mathcal{O}}$.

Definition 10. A set Λ of $\mathcal{V}$-sentences is *semantically falsifiable* if and only if it is falsified by a possible $\mathcal{O}$-structure.

Now the following holds:

Claim 5. *A set Λ of $\mathcal{V}$-sentences is semantically falsifiable iff Λ is semantically $\mathcal{O}$-creative with respect to Π.*

Proof. Λ is semantically $\mathcal{O}$-creative with respect to Π iff there is an $\mathfrak{A}_{\mathcal{O}}$ that has an expansion $\mathfrak{B} \vDash \Pi$ (which is always the case since Π is consistent) and every expansion $\mathfrak{C} \vDash \Pi$ of $\mathfrak{A}_{\mathcal{O}}$ is such that $\mathfrak{C} \nvDash \Lambda$. This holds iff Λ is semantically falsifiable. □

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

David Lewis argues that one of his explications of 'partial aboutness' is closely connected to syntactic falsifiability. To see that it is even more closely connected to semantic falsifiability, consider first Lewis's explication of 'aboutness' as supervenience. According to Lewis (1988b, 136), a "statement is entirely about some subject matter iff its truth value supervenes on that subject matter. Two possible worlds which are exactly alike so far as that subject matter is concerned must both make the statement true, or else both make it false". Assuming that possible worlds are all and only those worlds in which all analytic sentences are true, and assuming that all statements can be expressed by sets of $\mathcal{V}$-sentences, there is a one-to-one mapping from possible worlds to $\mathcal{V}$-structures (cf. Kemeny 1963, §IV). Lewis does not explicate what it means for possible worlds to be "exactly alike" with respect to observation (except that 'being exactly alike' is an equivalence relation), so I suggest the following: Two possible worlds are exactly alike if and only if the reducts of their corresponding structures to $\mathcal{O}$ are identical. This leads to

Definition 11. A set Λ of $\mathcal{V}$-sentences is *about observation* if and only if for any $\mathcal{V}$-structures $\mathfrak{A}, \mathfrak{B} \vDash \Pi$ with $\mathfrak{A}|\mathcal{O} = \mathfrak{B}|\mathcal{O}$ it holds that $\mathfrak{A} \vDash \Lambda$ iff $\mathfrak{B} \vDash \Lambda$.

To distinguish aboutness more clearly from partial aboutness, I will also sometimes speak of sentences being *entirely* about observation when they are about observation.

Lewis (1988b, §VII, footnote removed) suggests to weaken definition 11 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 observation'. 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—in this case, the observations. 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 E of possible worlds that it excludes. In the model theoretic paraphrase, the content of a set Λ of sentences is thus given by $E_\Lambda := \{\mathfrak{A} \mid \mathfrak{A} \vDash \Pi \text{ and } \mathfrak{A} \nvDash \Lambda\}$. The content of Λ is about observation iff Λ itself is about observation, which is the case iff for any two $\mathfrak{B}, \mathfrak{C} \vDash \Pi$ with $\mathfrak{B}|\mathcal{O} = \mathfrak{C}|\mathcal{O}$, $\mathfrak{B} \in E_\Lambda$ iff $\mathfrak{C} \in E_\Lambda$. The parts of the content of Λ are then defined as the subsets of E_Λ, which leads to

Definition 12. *Part of the content* of a set Λ of $\mathcal{V}$-sentences is *about observation* if and only if there is a non-empty set of structures $F \subseteq E_\Lambda := \{\mathfrak{A} \mid \mathfrak{A} \vDash \Pi \text{ and } \mathfrak{A} \nvDash \Lambda\}$ such that for any two $\mathfrak{B}, \mathfrak{C} \vDash \Pi$ with $\mathfrak{B}|\mathcal{O} = \mathfrak{C}|\mathcal{O}$, $\mathfrak{B} \in F$ iff $\mathfrak{C} \in F$.

Lewis does not demand F to be non-empty, but without this restriction, part of the content of every sentence is about observation. 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 observation iff the sentence is syntactically falsifiable.[4] This relation can be made more precise with

Claim 6. *Part of the content of a set Λ of $\mathcal{V}$-sentences is about observation iff Λ is semantically falsifiable.*

[4]To be more precise, since Lewis does not demand F to be non-empty, he can show that part of a statement's content is about observation iff the statement is incompatible with some statement entirely about observation. But according to definition 11 and Lewis (1988b, 141) himself, contradictions are entirely about observation, and since contradictions are incompatible with every statement, this shows that his definition is trivial. Demanding F to be non-empty excludes contradictions.

Proof. '⇒': Assume part $F \subseteq E_\Lambda$ of Λ's content is about observation. Define $\mathfrak{A}_\mathscr{O} := \mathfrak{A}|\mathscr{O}$ for some $\mathfrak{A} \in F$. Since $\mathfrak{A} \in F$ and according to definition 12 either all $\mathfrak{B}$ with $\mathfrak{B}|\mathscr{O} = \mathfrak{A}_\mathscr{O}$ are in F or none is, all such $\mathfrak{B}$ are in F. Since all such $\mathfrak{B}$ are also in E_Λ, $\mathfrak{B} \nvDash \Lambda$, and the possible structure $\mathfrak{A}_\mathscr{O}$ falsifies Λ.

'⇐': Assume Λ is semantically falsified by $\mathfrak{A}_\mathscr{O}$. Define $F := \{\mathfrak{B} \mid \mathfrak{B} \vDash \Pi \text{ and } \mathfrak{B}|\mathscr{O} = \mathfrak{A}_\mathscr{O}\}$. Since $\varnothing \neq F \subseteq E_\Lambda$, part of Λ's content is about observation. □

Because of claims 1, 5, and 6, the relation between syntactic falsifiability and Lewis's definition for sets of sentences part of whose content is about observation is the same as that between syntactic and semantic $\mathscr{O}$-creativity, which is given in claim 3.

A sentence whose content is partly about observation could also be said to have some observational content, and indeed this is essentially how Carnap (1928, 327f) described a criterion of meaningfulness at the time of the Vienna circle (see page 29). Decades later, he argued that, absent sentences already established as analytic, the observational content of a sentence α is given by its Ramsey sentence $\mathrm{R}(\alpha)$, the existential quantification over all $\mathscr{T}$-terms in α (Psillos 2000). $\mathrm{R}(\alpha)$ plausibly describes α's observational content because it entails the same $\mathscr{O}$-sentences as α itself (Rozeboom 1962, 291ff). Now, just as a criterion of the meaning of a set of sentences is not a criterion of empirical significance, neither is a description of the observational content of such a set. Something weaker is needed, namely a criterion to determine when the observational content is non-empty. Since anything that is already entailed by the analytic sentences is not an empirical claim, this suggests

Definition 13. If Π can be finitely axiomatized, let $\tilde{\Pi}$ be this axiomatization. Then a $\mathscr{V}$-sentence α has *$\mathscr{O}$-content* if and only if $\tilde{\Pi} \nvDash \mathrm{R}(\alpha \wedge \bigwedge \tilde{\Pi})$.

Under this definition, Carnap's later notion of $\mathscr{O}$-content squares well with the notion of falsifiability:

Claim 7. *If Π can be finitely axiomatized, then a $\mathscr{V}$-sentence α has $\mathscr{O}$-content iff α is semantically $\mathscr{O}$-creative with respect to Π.*

Proof. A sentence α is Ramseyfied by substituting every $\mathscr{T}$-term $T_i, 1 \leq i \leq n$ in α by a variable X_i and existentially quantifying over each X_i, leading to $\exists X_1 \ldots X_n \alpha[T_1/X_1, \ldots, T_n/X_n]$. Define $g : \{T_i\}_{1 \leq i \leq n} \to \{X_i\}_{1 \leq i \leq n}, T_i \mapsto X_i$.

'⇐': Assume that α has no $\mathscr{O}$-content. Since $\mathrm{R}(\alpha \wedge \bigwedge \tilde{\Pi})$ is an $\mathscr{O}$-sentence, $\tilde{\Pi} \vDash \mathrm{R}(\alpha \wedge \bigwedge \tilde{\Pi})$ if and only if $\mathrm{R}(\tilde{\Pi}) \vDash \mathrm{R}(\alpha \wedge \bigwedge \tilde{\Pi})$. Then for any $\mathfrak{A}_\mathscr{O}$, $\mathfrak{A}_\mathscr{O} \vDash \mathrm{R}(\bigwedge \tilde{\Pi})$ only if $\mathfrak{A}_\mathscr{O} \vDash \mathrm{R}(\alpha \wedge \bigwedge \tilde{\Pi})$. Thus for any $\mathfrak{A}_\mathscr{O}$, if there is a satisfaction function v mapping each variable $X_i, 1 \leq i \leq n$ to an extension of the same type over $|\mathfrak{A}_\mathscr{O}|$ such that $\mathfrak{A}_\mathscr{O}, v \vDash \bigwedge \tilde{\Pi}[T_1/X_1, \ldots, T_n/X_n]$, there is a satisfaction function v' such that $\mathfrak{A}_\mathscr{O}, v' \vDash (\alpha \wedge \bigwedge \tilde{\Pi})[T_1/X_1, \ldots, T_n/X_n]$. Now assume that $\mathfrak{A}_\mathscr{O}$ can be expanded to a model $\mathfrak{B} \vDash \tilde{\Pi}$. $\mathfrak{B} = \langle |\mathfrak{B}|, f \rangle$ with domain $|\mathfrak{B}|$ and a function f that maps every $\mathscr{V}$-term to an extension of the same type in $|\mathfrak{B}|$. Since $|\mathfrak{B}| = \left|\mathfrak{A}_\mathscr{O}\right|$, any extension v

of $f|\{T_1,\dots,T_n\}\circ g^{-1}$ to all variables of the language is a satisfaction function such that $\mathfrak{A}_{\mathcal{O}}, v \vDash \bigwedge \tilde{\Pi}[T_1/X_1,\dots,T_n/X_n]$. By assumption, there is then a satisfaction function v' such that $\mathfrak{A}_{\mathcal{O}}, v' \vDash (\alpha \wedge \bigwedge \tilde{\Pi})[T_1/X_1,\dots,T_n/X_n]$. Then any extension f of $v'|\{X_1,\dots,X_n\}\circ g$ to all $\mathcal{T}$-terms can be used to expand $\mathfrak{A}_{\mathcal{O}}$ to a model of $\bigwedge \tilde{\Pi} \wedge \alpha$, and therefore α is semantically $\mathcal{O}$-conservative with respect to $\tilde{\Pi}$.

'⇒': Similar. □

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 an analogous way to his formulation of falsifiability leads to

Definition 14. A *set Ω of $\mathcal{V}$-sentences verifies* a set Λ of $\mathcal{V}$-sentences if and only if $\Omega \cup \Pi \vDash \Lambda$.

Definition 15. A set Λ of $\mathcal{V}$-sentences is *syntactically verifiable* if and only if there is a possible set Ω of $\mathcal{O}$-sentences that verifies Λ.

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 8. *A $\mathcal{V}$-sentence α is syntactically verifiable iff $\neg\alpha$ 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 16. An *$\mathcal{O}$-structure $\mathfrak{A}_{\mathcal{O}}$ verifies* a set Λ of $\mathcal{V}$-sentences if and only if for all $\mathfrak{C} \vDash \Pi$ with $\mathfrak{C}|\mathcal{O} = \mathfrak{A}_{\mathcal{O}}$, $\mathfrak{C} \vDash \Lambda$.

Definition 17. A set Λ of $\mathcal{V}$-sentences is *semantically verifiable* if and only if there is a possible $\mathcal{O}$-structure that verifies Λ.

And again, the following can easily be shown to hold:

Claim 9. *A $\mathcal{V}$-sentence α is semantically verifiable iff $\neg\alpha$ is semantically falsifiable.*

The relations between syntactic and semantic falsifiability described in claims 3 and 4 therefore transfer to the verifiability of sentences. Furthermore, claims 9 and 7 entail the following:

Claim 10. *If Π can be finitely axiomatized, let $\tilde{\Pi}$ be this axiomatization. Then a $\mathcal{V}$-sentence α is semantically verifiable iff $\tilde{\Pi} \nvDash \mathrm{R}(\neg\alpha \wedge \bigwedge \tilde{\Pi})$.*

Even if the observational content of α is considered to be the set of possible $\mathscr{O}$-sentences or $\mathscr{O}$-structures that *verify* α, however, $\mathrm{R}(\neg\alpha \wedge \tilde{\Pi})$ is not the observational content of α. Rather, it can be shown similarly to the proof of claim 7 that the possible $\mathscr{O}$-structures that verify α are the models of $\neg\mathrm{R}(\neg\alpha \wedge \tilde{\Pi})$. And this sentence is also analytically entailed by the same $\mathscr{O}$-sentences as α, since $\omega \wedge \bigwedge \tilde{\Pi} \vDash \alpha$ if and only if $\neg\alpha \wedge \bigwedge \tilde{\Pi} \vDash \neg\omega$, which holds if and only if $\mathrm{R}\big(\neg\alpha \wedge \bigwedge \tilde{\Pi}\big) \vDash \neg\omega$, and thus if and only if $\omega \vDash \neg\mathrm{R}\big(\neg\alpha \wedge \bigwedge \tilde{\Pi}\big)$. The connection to claim 10 is rather that α is verifiable if and only if its observational content is not empty, which it is if Π entails that $\neg\mathrm{R}(\neg\alpha \wedge \tilde{\Pi})$ has no models, that is, $\tilde{\Pi} \vDash \mathrm{R}(\neg\alpha \wedge \tilde{\Pi})$.[5]

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

Claim 11. *A set Λ of $\mathscr{V}$-sentences is syntactically verifiable iff there is a possible $\mathscr{O}$-structure $\mathfrak{A}_{\mathscr{O}}$ such that Λ is verified by each possible $\mathscr{O}$-structure syntactically equivalent to $\mathfrak{A}_{\mathscr{O}}$.*

Proof. '⇒': Assume that the possible set of $\mathscr{O}$-sentences Ω verifies Λ. Then for every $\mathfrak{B} \vDash \Omega \cup \Pi$, $\mathfrak{B} \vDash \Lambda$. Since Ω is possible, there is some such $\mathfrak{B}$. Choose $\mathfrak{A}_{\mathscr{O}} = \mathfrak{B}|\mathscr{O}$. Then every $\mathfrak{C}$ with $\mathfrak{C}|\mathscr{O} \equiv \mathfrak{A}_{\mathscr{O}}$ is such that $\mathfrak{C} \vDash \Omega$. Since for every possible $\mathscr{O}$-structure syntactically equivalent to $\mathfrak{A}_{\mathscr{O}}$, there is such a $\mathfrak{C}$, every possible $\mathscr{O}$-structure syntactically equivalent to $\mathfrak{A}_{\mathscr{O}}$ verifies Λ.

'⇐': Assume that every possible $\mathscr{O}$-structure syntactically equivalent to $\mathfrak{A}_{\mathscr{O}}$ verifies Λ. Choose $\Omega \mathrel{\vDash\!\!\!\dashv} \mathrm{Th}(\mathfrak{A}_{\mathscr{O}})$. Since $\mathfrak{A}_{\mathscr{O}}$ is possible, $\Omega \cup \Pi$ has a model, and thus Ω is possible. By assumption, $\mathfrak{B} \vDash \Omega \cup \Pi$ only if $\mathfrak{B} \vDash \Lambda$, and thus Ω verifies Λ. □

As in the case of falsifiability, semantic verifiability is like syntactic verifiability, except that the observational 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.

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 observational sentences, but rejects it as too strong. Rynin (1957, 51) also rejects such a finite version of this criterion. Furthermore, Sober (2008, 149f) has argued that the negations of empirically significant sentences do not have to be themselves empirically significant—unlike, arguably, the negations of meaningful sentences. In that case, Hempel's motivation cannot be used for criteria of empirical significance.

[5] If $\mathrm{R}(\alpha \wedge \tilde{\Pi})$ is akin to the "canonical commitment of the content that" α as described by Peacocke (1986, 47), then $\neg\mathrm{R}(\neg\alpha \wedge \tilde{\Pi})$ is akin to the "canonical ground for the content that" α.

A real, though small, advantage of the criterion is that a sentence that is both verifiable and falsifiable is automatically analytically contingent, and therefore the criterion can be formulated without demanding analytic contingency explicitly. Probably the main reason for using this criterion is that it is a sufficient condition for empirical significance for both proponents of falsifiability and proponents of verifiability (see Kitts (1977) for an example of this kind of argument). This dialectical advantage and the slight convenience in formulation cannot, however, outweigh the criterion's lack of other justifications.

6 Strong $\mathscr{O}$-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. But, first, Hempel and Rynin explicitly only reject criteria that rely on finite sets of observational 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:[6]

Definition 18. An *$\mathscr{O}$-structure* $\mathfrak{A}_{\mathscr{O}}$ *determines* a set Λ of $\mathscr{V}$-sentences if and only if for all $\mathscr{V}$-structures $\mathfrak{B}, \mathfrak{C} \vDash \Pi$ with $\mathfrak{B}|\mathscr{O} = \mathfrak{C}|\mathscr{O} = \mathfrak{A}_{\mathscr{O}}$ it holds that $\mathfrak{B} \vDash \Lambda$ iff $\mathfrak{C} \vDash \Lambda$.

Definition 19. A set Λ of $\mathscr{V}$-sentences is *strongly semantically $\mathscr{O}$-determined* if and only if it is determined by every possible $\mathscr{O}$-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 $\mathscr{O}$-determined and analytically contingent.

The truth value of a strongly semantically $\mathscr{O}$-determined set Λ of sentences is fixed by any interpretation of the observational terms in any domain, because Λ is either true in all possible models that expand such an $\mathscr{O}$-structure, or it is false in all such models. Hence

Claim 12. *A set Λ of $\mathscr{V}$-sentences is strongly semantically $\mathscr{O}$-determined if and only if every possible $\mathscr{O}$-structure either falsifies or verifies Λ.*

As Przełęcki (1974a, 346f) already notes, this definition is very exclusive. If, for example, the theoretical term T_1 is conditionally defined by $\{\forall x[O_1 x \rightarrow (O_2 x \longleftrightarrow T_1 x)]\} =: \Pi$, and '$O_1$', '$O_2$', and '$o$' are observational terms, then the possible structure $\mathfrak{A}_{\mathscr{O}} = \langle\{1,2\},\{\langle O_1,\{1\}\rangle,\langle O_2,\{1\}\rangle,\langle o,2\rangle\}\rangle$ does not determine $T_1(o)$. Therefore, $T_1(o)$ is not strongly semantically $\mathscr{O}$-determined. This is unsurprising, because, in Lewis's terminology, the definition includes only sentences that are entirely about observation:

[6] 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.

Claim 13. *A set Λ of $\mathscr{V}$-sentences is strongly semantically $\mathscr{O}$-determined iff Λ is about observation.*

Proof. '$\Rightarrow$': Assume $\mathfrak{B}, \mathfrak{C} \vDash \Pi$, $\mathfrak{B}|\mathscr{O} = \mathfrak{C}|\mathscr{O}$. Then $\mathfrak{B}|\mathscr{O}$ is a possible $\mathscr{O}$-structure, and by assumption, $\mathfrak{B} \vDash \Lambda$ iff $\mathfrak{C} \vDash \Lambda$.

'$\Leftarrow$': Assume $\mathfrak{A}_{\mathscr{O}}$ is a possible $\mathscr{O}$-structure. For any two possible $\mathscr{V}$-structures $\mathfrak{B}$, $\mathfrak{C}$ with $\mathfrak{B}|\mathscr{O} = \mathfrak{A}_{\mathscr{O}}$ and $\mathfrak{C}|\mathscr{O} = \mathfrak{A}_{\mathscr{O}}$, $\mathfrak{B}|\mathscr{O} = \mathfrak{C}|\mathscr{O}$ and thus, by assumption, $\mathfrak{B} \vDash \Lambda$ iff $\mathfrak{C} \vDash \Lambda$. □

That the criterion is relevant despite being exclusive is shown by a justification very attuned to the needs of the measuring scientist and questions of symmetry. Suppes (1959, 131) begins his justification with the idea that

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

The numerical quantities are functions, and transformations that lead only from one adequate function to another are appropriate (Suppes 1959, 132). To be adequate, a function has to fulfill the conditions of adequacy for the measurement it represents. Suppes (1959, 135) states the conditions for functions m representing mass measurement as

$$\begin{aligned}\Pi_{\text{mass}} := \{&\forall x \forall y\,(x \precsim y \leftrightarrow mx \leq my),\\ &\forall x \forall y\,(m(x * y) = mx + my)\}\,,\end{aligned} \tag{1}$$

where '$\precsim$' stands for 'is at most as heavy as', '$*$' stands for physical combination, and x and y are silently understood to range over physical objects. Suppes (1959, 135) notes that "the functional composition of any similarity transformation φ with the function m yields a function $\varphi \circ m$ which also satisfies" Π_{mass}, where a similarity transformation in Suppes's sense is also called a positive linear transformation. Therefore, Suppes (1959, 138) suggests that

> a formula S [...] is empirically meaningful [...] if and only if S is satisfied in a model $\mathfrak{M}$ [...] when and only when it is satisfied in every model [...] related to $\mathfrak{M}$ by a similarity transformation.

To connect Suppes's criterion to Lewis's and thereby to Przełęcki's, note first that $\precsim$ and $*$ play the role of observational terms with some set of axioms $\Pi_{\mathscr{O}}$ (Suppes 1959, 135, n. 7), and m is the sole theoretical term. Now let $\mathfrak{B}[m/\varphi \circ m^{\mathfrak{B}}]$ be the structure that $\mathfrak{B}$ becomes when m is interpreted by $\varphi \circ m^{\mathfrak{B}}$ instead of $m^{\mathfrak{B}}$. Suppes's criterion of adequacy can then be paraphrased like this:[7]

[7]Przełęcki's paraphrase is slightly different, for one because he aims to prove its equivalence with definition 19, not definition 11, but also because his definition of $\mathscr{O}$-conservativeness is slightly less general (see appendix A).

Definition 20 (Empirically meaningful statements about mass)**.** Assume the standard interpretation for arithmetical terms. Then a $\mathscr{V}$-sentence α is *empirically meaningful* if and only if for any $\mathfrak{B} \vDash \Pi_{\mathscr{O}} \cup \Pi_{\text{mass}}$ and any $\mathfrak{C}$, if $\mathfrak{C} = \mathfrak{B}[m/\varphi \circ m^{\mathfrak{B}}]$ and φ is a positive linear transformation, then $\mathfrak{B} \vDash \alpha$ if and only if $\mathfrak{C} \vDash \alpha$.

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 $\mathscr{O}$-determinacy as well.

Claim 14. *Assume $\mathscr{T} = \{m\}$, $\Pi_{\mathscr{T}} = \Pi_{\text{mass}}$ and the standard interpretation for arithmetical terms. Then a $\mathscr{V}$-sentence α is empirically meaningful according to definition 20 iff α is about observation.*

Proof. Przełęcki (1974a, 349). □

In claim 14, the interpretations of $+$ and $\leq$ are assumed to be fixed. Przełęcki (1974a, 347f) assumes that this is ensured by a semantic restriction on the possible structures. One could also ensure the standard interpretation with the usual axioms in second order logic.

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 $\mathfrak{E} = \langle |\mathfrak{E}|, \{\langle R_1, R_1^{\mathfrak{E}}\rangle, \ldots, \langle R_p, R_p^{\mathfrak{E}}\rangle\}, \{\langle \circ_1, \circ_1^{\mathfrak{E}}\rangle, \ldots, \langle \circ_q, \circ_q^{\mathfrak{E}}\rangle\}\rangle$, sometimes called 'empirical', to another relational system $\mathfrak{F} = \langle |\mathfrak{F}|, \{\langle Q_1, Q_1^{\mathfrak{F}}\rangle, \ldots, \langle Q_p, Q_p^{\mathfrak{F}}\rangle\}, \{\langle *_1, *_1^{\mathfrak{F}}\rangle, \ldots, \langle *_q, *_q^{\mathfrak{F}}\rangle\}\rangle$, sometimes called 'formal'. A homomorphism (an element of $\hom(\mathfrak{E}, \mathfrak{F})$) is a function $\mu : |\mathfrak{E}| \to |\mathfrak{F}|$ such that for all $a_1^{\mathfrak{E}}, \ldots, a_{k_i}^{\mathfrak{E}}, a^{\mathfrak{E}}, b^{\mathfrak{E}} \in |\mathfrak{E}|$ with $i = 1, \ldots, p$ and for all $j = 1, \ldots, q$ it holds that

$$\begin{gathered} R_i^{\mathfrak{E}}(a_1^{\mathfrak{E}}, \ldots, a_{k_i}^{\mathfrak{E}}) \text{ iff } Q_i^{\mathfrak{F}}(\mu(a_1^{\mathfrak{E}}), \ldots, \mu(a_{k_i}^{\mathfrak{E}}))\ , \\ \mu(a^{\mathfrak{E}} \circ_j^{\mathfrak{E}} b^{\mathfrak{E}}) = \mu(a^{\mathfrak{E}}) *_j^{\mathfrak{F}} \mu(b^{\mathfrak{E}})\ . \end{gathered} \tag{2}$$

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, §5.1). 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 $\mathfrak{A}|\mathscr{E}_E$ is the substructure of $\mathfrak{A}|\mathscr{E}$ whose domain is $E^{\mathfrak{A}}=|\mathfrak{E}|$.[8] The relativization theorem then says that for every formula φ of $\mathscr{E}$ ($\mathscr{F}$) and its relativization $\varphi^{(E)}$ ($\varphi^{(F)}$) 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). In a relativization $\varphi^{(P)}$, all quantifications in the formula φ are restricted by the predicate P. For any set Λ of formulas, define $\Lambda^{(P)}$ as the set of the relativization of the elements of Λ.

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

$$\begin{aligned}\Pi_{\text{adeq}}:=&\{\forall a(Ea\to Ffa)\}\cup\\&\bigcup_{i=1}^{p}\left\{\forall a_1\dots\forall a_{k_i}\left[Ea_1\wedge\dots\wedge Ea_{k_i}\to\left(R_ia_1\dots a_{k_i}\leftrightarrow Q_ifa_1\dots fa_{k_i}\right)\right]\right\}\cup\\&\bigcup_{j=1}^{q}\left\{\forall a\forall b\left(Ea\wedge Eb\to f(a\circ_j b)=fa*_j fb\right)\right\}\ .\end{aligned}\tag{3}$$

[8]A substructure of a structure $\mathfrak{A}$ is a structure $\mathfrak{B}$ for the same terms and with $|\mathfrak{B}|\subseteq|\mathfrak{A}|$ that agrees with $\mathfrak{A}$ on all interpretations of the terms over $|\mathfrak{B}|$ (Hodges 1993, 6f). $\mathfrak{A}$ is called an extension of $\mathfrak{B}$.

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, let f furthermore map any elements of $|\mathfrak{F}|$ to $|\mathfrak{E}| \cup |\mathfrak{F}|$. All in all, $\mathfrak{A}$ is determined by a set Π_{scale} that entails Π_{adeq}, the restrictions on $\mathfrak{F}$ and possible empirical structures, and 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 reduction to $\mathscr{E} \cup \mathscr{F} \cup \{f\} =: \mathscr{E}\mathscr{F}f$ and $\langle \lambda x . Ex \vee Fx \rangle =: EF$ (Hodges 1993, 203), so that $\mathfrak{A}|\mathscr{E}\mathscr{F}f_{EF}$ exists. Because of the relativization theorem, $\mathfrak{A}|\mathscr{E}\mathscr{F}f_{EF} \vDash \Lambda$ if and only if $\mathfrak{A} \vDash \Lambda^{(EF)}$ for any $\mathscr{E}\mathscr{F}f$-sentence Λ. Defining $\Lambda^{\mathfrak{A}}$ to be the set theoretic conditions on the extensions of the terms in Λ that have to hold for Λ to be true in $\mathfrak{A}$, it therefore holds for any set Λ of $\mathscr{E}\mathscr{F}f$-sentences that $\mathfrak{A} \vDash \Lambda^{(EF)}$ if and only if $\Lambda^{\mathfrak{A}|\mathscr{E}\mathscr{F}f_{EF}}$ is true for the scale $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$, where $\mathfrak{A}|\mathscr{E}_E = \mathfrak{E}$ and $\mathfrak{A}|\mathscr{F}_F = \mathfrak{F}$.

The definition of admissible transformation argued for by Roberts and Franke (1976) can now be paraphrased as such:

Definition 21. 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 22. A set Λ 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 $\mathfrak{A} \vDash \Lambda$ iff $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \vDash \Lambda$.

This is indeed a generalization of the original definition by Suppes and Zinnes (1963):

Claim 15. *A set* $\Lambda^{(EF)}$ *of* $\mathscr{E}\mathscr{F}f$*-sentences is strongly invariant if and only if* $\Lambda^{\mathfrak{A}|\mathscr{E}\mathscr{F}f_{EF}}$ *is meaningful for any scale* $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$ *according to Suppes and Zinnes (1963).*

Proof. '⇒': Assume that $\Lambda^{(EF)}$ is strongly invariant. For scale $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$, construct $\mathfrak{A}$ so that $\mathfrak{A} \vDash \Pi_{\text{scale}}$ and $\mathfrak{A}|\mathscr{E}_E = \mathfrak{E}$ and $\mathfrak{A}|\mathscr{F}_F = \mathfrak{F}$. Now assume that $\Lambda^{\mathfrak{A}|\mathscr{E}\mathscr{F}f_{EF}}$ is true for $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$ and ψ is an admissible transformation for $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$. Then $\mathfrak{A} \vDash \Lambda^{(EF)}$ and by assumption, for any admissible φ relative to $\mathfrak{A}$, $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \vDash \Lambda^{(EF)}$, and thus $\Lambda^{\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]|\mathscr{E}\mathscr{F}f_{EF}}$ is true for the scale $\langle \mathfrak{E}, \mathfrak{F}, \mu \rangle$. Now, ψ is admissible relative to $\mathfrak{E}$, $\mathfrak{F}$, and μ only if $\psi\left(f^{\mathfrak{A}}||\mathfrak{E}|\right)$ fulfills equations 3. And then some extension φ of ψ with $\varphi(f^{\mathfrak{A}})||\mathfrak{E}| = \psi\left(f^{\mathfrak{A}}||\mathfrak{E}|\right)$ is admissible relative to $\mathfrak{A}$, for which $\Lambda^{(\mathfrak{A}|\mathscr{E}\mathscr{F}f_{EF})[f/\psi(\mu)]} = \Lambda^{\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]|\mathscr{E}\mathscr{F}f_{EF}}$. Therefore, $\Lambda^{(\mathfrak{A}|\mathscr{E}\mathscr{F}f_{EF})[f/\psi(\mu)]}$ is true for any admissible transformation ψ relative to $\mathfrak{E}$, $\mathfrak{F}$, and μ. By an analogous proof, $\Lambda^{\mathfrak{A}|\mathscr{E}\mathscr{F}f_{EF}}$ is false for $\langle\mathfrak{E},\mathfrak{F},\mu\rangle$ only if $\Lambda^{(\mathfrak{A}|\mathscr{E}\mathscr{F}f_{EF})[f/\psi(\mu)]}$ is false for any admissible transformation ψ.

'$\Leftarrow$': Assume that $\Lambda^{\mathfrak{A}|\mathscr{E}\mathscr{F}f_{EF}}$ is meaningful. Now assume that $\mathfrak{A} \vDash \Pi_{\text{scale}}$, $\mathfrak{A} \vDash \Lambda^{(EF)}$ and φ is an admissible transformation relative to $\mathfrak{A}$. Then $\Lambda^{\mathfrak{A}|\mathscr{E}\mathscr{F}f_{EF}}$ is true for scale $\langle\mathfrak{E},\mathfrak{F},\mu\rangle$ with $\mathfrak{E} = \mathfrak{A}|\mathscr{E}_E$ and $\mathfrak{F} = \mathfrak{A}|\mathscr{F}_F$, and by assumption, so is $\Lambda^{\mathfrak{A}|\mathscr{E}\mathscr{F}f_{EF}[f/\psi(\mu)]}$ for any admissible transformation ψ relative to $\mathfrak{E}$, $\mathfrak{F}$, and μ. Now, φ is admissible relative to $\mathfrak{A}$ only if $\varphi(f^{\mathfrak{A}})|\,|\mathfrak{E}|$ fulfills equations 3 and thus is admissible relative to $\mathfrak{E}$, $\mathfrak{F}$, and μ. Thus $\Lambda^{\mathfrak{A}|\mathscr{E}\mathscr{F}f_{EF}[f/\varphi(f^{\mathfrak{A}})]} = \Lambda^{\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]|\mathscr{E}\mathscr{F}f_{EF}}$ is true in $\langle\mathfrak{E},\mathfrak{F},\mu\rangle$, and hence $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \vDash \Lambda^{(EF)}$ for any admissible transformation φ relative to $\mathfrak{A}$. By an analogous proof, $\mathfrak{A} \nvDash \Lambda^{(EF)}$ only if $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \nvDash \Lambda^{(EF)}$ for any admissible transformation φ. □

Admissible transformations are defined relative to the observations and the analytic sentences Π. Strong invariance avoids the dependence on specific observations by a universal quantification over all possible observations, and is therefore defined relative to the set of admissible transformations and thus relative to the analytic sentences. Strong invariance is thus a symmetry relative to the analytic sentences.

Now Przełęcki's result can be generalized:

Claim 16. *Assume $\mathscr{V} = \mathscr{A}$, $\mathscr{O} = \mathscr{E}$ and $\Pi = \Pi_{\text{scale}}$. Then a set Λ of $\mathscr{E}\mathscr{F}f$-sentences is strongly invariant iff Λ is about observation.*

Proof. '$\Leftarrow$': 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 $\mathfrak{B}|\mathscr{O} = \mathfrak{A}|\mathscr{O}$ such that $\mathfrak{B}|\mathscr{O} \cup \{f\} = \mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]|\mathscr{O} \cup \{f\}$ $(*)$, which can be shown as follows: By definition 21, $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] \vDash \Pi_{\text{scale}}$, and since $f \notin \mathscr{O}$, $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]|\mathscr{O} = \mathfrak{A}|\mathscr{O}$. 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 $\mathfrak{B}|\mathscr{O} = \mathfrak{A}|\mathscr{O}$, $\mathfrak{A} \vDash \Lambda$ iff $\mathfrak{B} \vDash \Lambda$. 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 $\mathfrak{D}|\mathscr{O} = \mathfrak{C}|\mathscr{O}$ and $\mathfrak{D} = \mathfrak{C}[f/\varphi(f^{\mathfrak{C}})]$. Therefore, by assumption, $\mathfrak{C}[f/\varphi(f^{\mathfrak{C}})] \vDash \Lambda$ iff $\mathfrak{C} \vDash \Lambda$.

'$\Rightarrow$': First, note that for any $\mathfrak{A},\mathfrak{B} \vDash \Pi_{\text{scale}}$ with $\mathfrak{B}|\mathscr{O} = \mathfrak{A}|\mathscr{O}$, there is some transformation φ admissible relative to $\mathfrak{A}$ such that $\mathfrak{B}|\mathscr{O} \cup \{f\}$ is isomorphic to $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]|\mathscr{O} \cup \{f\}$ $(**)$, which can be shown as follows: Since $\mathfrak{A}|\mathscr{F}$ is isomorphic to $\mathfrak{B}|\mathscr{F}$, assume without loss of generality that $\mathfrak{A}|\mathscr{F} = \mathfrak{B}|\mathscr{F}$. Now choose $\varphi(\mu) = f^{\mathfrak{B}}$ for any function μ. Then φ is admissible relative to $\mathfrak{A}$ because $\mathfrak{A}[f/\varphi(f^{\mathfrak{A}})] = \mathfrak{B} \vDash \Pi_{\text{adeq}}$.

Now assume that $\mathfrak{B}|\mathscr{O} = \mathfrak{A}|\mathscr{O}$ and $\mathfrak{B} \vDash \Pi_{\text{scale}}$. By $(**)$, there is some admissible φ such that $\mathfrak{B}|\mathscr{E}\mathscr{F}f = \mathfrak{A}[f/\varphi(f^{\mathfrak{A}})]|\mathscr{E}\mathscr{F}f$. Therefore, if Λ is strongly invariant, $\mathfrak{B} \vDash \Lambda$ iff $\mathfrak{A} \vDash \Lambda$. □

Like strong invariance, strong $\mathscr{O}$-determinacy is thus a symmetry relative to the analytic sentences Π.

To arrive at a syntactic version of strong $\mathcal{O}$-determinacy, it is helpful to look at the line of reasoning that led to definition 11. There, a set of sentences is taken to be about observation if and only if its truth value is identical in any two worlds that are exactly alike so far as observation 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 observational structures, which is borne out by claim 3 for falsifiability and claim 11 for verifiability. To arrive at an analogous relation for strong $\mathcal{O}$-determinacy, I thus suggest

Definition 23. A *set Γ of $\mathcal{V}$-sentences determines* a set Λ of $\mathcal{V}$-sentences if and only if $\Gamma \cup \Pi \vDash \Lambda$ or $\Gamma \cup \Lambda \cup \Pi \vDash \bot$.

Definition 24. A set Ω of $\mathcal{O}$-sentences is *maximal* if and only if for every $\mathcal{O}$-sentence ω, $\Omega \cup \Pi \vDash \omega$ or $\Omega \cup \Pi \vDash \neg\omega$.

Then one can formulate

Definition 25. A set Ω of $\mathcal{V}$-sentences is *strongly syntactically $\mathcal{O}$-determined* if and only if it is determined by every possible and maximal set of $\mathcal{O}$-sentences.

As in the case of falsifiability and verifiability, the difference between syntactic and semantic strong $\mathcal{O}$-determinacy is that between isomorphy and syntactical equivalence:

Claim 17. *A set Λ of $\mathcal{V}$-sentences is strongly syntactically $\mathcal{O}$-determined iff for any $\mathcal{V}$-structures $\mathfrak{A}, \mathfrak{B} \vDash \Pi$ with $\mathfrak{A}|\mathcal{O} \equiv \mathfrak{B}|\mathcal{O}$, it holds that $\mathfrak{A} \vDash \Lambda$ iff $\mathfrak{B} \vDash \Lambda$.*

Proof. '⇒': Let $\mathfrak{A}, \mathfrak{B} \vDash \Pi$ and $\mathfrak{A}|\mathcal{O} \equiv \mathfrak{B}|\mathcal{O}$. Then $\mathfrak{A}, \mathfrak{B} \vDash \mathrm{Th}(\mathfrak{B}|\mathcal{O}) =: \Omega$. Ω is maximal and possible, thus, by assumption, $\Omega \cup \Pi \vDash \Lambda$ or $\Omega \cup \Lambda \cup \Pi \vDash \bot$. Thus $\mathfrak{A} \vDash \Lambda$ if and only if $\mathfrak{B} \vDash \Lambda$.

'⇐': Assume Ω is possible and maximal. Then for any $\mathfrak{A}, \mathfrak{B} \vDash \Omega \cup \Pi$, $\mathfrak{A}|\mathcal{O} \equiv \mathfrak{B}|\mathcal{O}$. Therefore, by assumption, $\mathfrak{A} \vDash \Lambda$ iff $\mathfrak{B} \vDash \Lambda$ and thus either all $\mathfrak{A} \vDash \Omega \cup \Pi$ are models of Λ or none is. Thus $\Omega \cup \Pi \vDash \Lambda$ or $\Omega \cup \Lambda \cup \Pi \vDash \bot$. □

This entails

Claim 18. *If a set Λ of $\mathcal{V}$-sentences is strongly syntactically $\mathcal{O}$-determined, then Λ is strongly semantically $\mathcal{O}$-determined.*

Proof. From claims 13 and 17 because $\mathfrak{B}|\mathcal{O} = \mathfrak{B}|\mathcal{O}$ only if $\mathfrak{A}|\mathcal{O} \equiv \mathfrak{B}|\mathcal{O}$. □

The relation of strong $\mathcal{O}$-determinacy to falsifiability and verifiability is given by

Claim 19. *Let Λ be a set of strongly syntactically (semantically) $\mathscr{O}$-determined $\mathscr{V}$-sentences. Then Λ is syntactically (semantically) falsifiable/verifiable iff Λ is not analytically true/false.*[9]

Proof. '$\Rightarrow$': Immediate.

'$\Leftarrow$': Assume $\Pi \not\models \Lambda$. Then for some $\mathfrak{A}$, $\mathfrak{A} \models \Pi$ and $\mathfrak{A} \not\models \Lambda$. If Λ is syntactically $\mathscr{O}$-determined, then $\mathrm{Th}(\mathfrak{A}|\mathscr{O}) \cup \Lambda \cup \Pi \models \bot$ because $\mathrm{Th}(\mathfrak{A}|\mathscr{O}) \cup \Pi \not\models \Lambda$. Thus $\mathrm{Th}(\mathfrak{A}|\mathscr{O})$ falsifies Λ. If Λ is semantically $\mathscr{O}$-determined, then for all $\mathfrak{B} \models \Pi$ with $\mathfrak{B}|\mathscr{O} = \mathfrak{A}|\mathscr{O}$, $\mathfrak{B} \not\models \Lambda$. Thus $\mathfrak{A}|\mathscr{O}$ falsifies Λ.

The proofs for verifiability are analogous. □

7 Weak $\mathscr{O}$-determinacy

Since Przełęcki considers strong semantic $\mathscr{O}$-determinacy too exclusive, he suggests a straightforward weakening of definition 19:

Definition 26. A set Λ of $\mathscr{V}$-sentence is *weakly semantically $\mathscr{O}$-determined* if and only if it is determined by a possible $\mathscr{O}$-structure.

The motivation for the criterion is clear: The truth value of a strongly semantically $\mathscr{O}$-determined sentence is fixed for any $\mathscr{O}$-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 observation. The idea to take a sentence to be partially about observation if it partially supervenes on observation 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 $\mathscr{O}$-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 α partially

[9] 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 iff it is not analytically true, and Λ is verifiable iff 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 19 therefore has four conjuncts.

supervenes on any M. To exclude such classes, Lewis demands that X contain a majority of the worlds in which α is true and a majority of the worlds in which α 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 iff $P(X|\alpha) > \frac{1}{2}$ and $P(X|\neg\alpha) > \frac{1}{2}$. Under some additional assumptions, the notion of partial supervenience that results is equivalent to the standard probabilistic criterion that α is empirically significant iff $P(\alpha|\omega) \neq P(\alpha)$ for some observational sentence ω.

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 α worlds and $\neg\alpha$ worlds. It is much more in keeping with the goal of explicating *empirical* significance to place only observational restrictions on X. The minimal requirement is thus that X be closed under observational equivalence, such that for any world that is in X, every $\mathscr{O}$-equivalent world is also in X. This condition already avoids trivialization, and furthermore does not lead to complications when statements are taken to be expressed by *sets* of sentences, which are not easily negated. To partially supervene on observation, Λ thus has to be assigned the same truth value by all members of a set X closed under observational equivalence.

Definition 27. A set Λ of $\mathscr{V}$-sentences *partly supervenes on observation* if and only if there is some non-empty set X of possible $\mathscr{V}$-structures such that for any $\mathfrak{A} \in X$, all $\mathfrak{B} \vDash \Pi$ with $\mathfrak{B}|\mathscr{O} = \mathfrak{A}|\mathscr{O}$ are in X, and for any $\mathfrak{A}, \mathfrak{B} \in X$ with $\mathfrak{A}|\mathscr{O} = \mathfrak{B}|\mathscr{O}$, $\mathfrak{B} \vDash \Lambda$ iff $\mathfrak{A} \vDash \Lambda$.

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

Claim 20. *A set Λ of $\mathscr{V}$-sentences is weakly semantically $\mathscr{O}$-determined iff Λ partly supervenes on observation.*

Proof. If $\mathfrak{A}_{\mathscr{O}}$ is possible and determines Λ, choose X as the set of possible expansions of $\mathfrak{A}_{\mathscr{O}}$. If Λ partly supervenes on observation, then any $\mathfrak{A}|\mathscr{O}$ with $\mathfrak{A} \in X$ is possible and determines Λ. □

Przełęcki (1974a, 347) points out that under his assumption that $\Pi_{\mathscr{T}}$ is $\mathscr{O}$-conservative with respect to $\Pi_{\mathscr{O}}$, definition 26 has a very conspicuous formulation: A $\mathscr{V}$-sentence α is weakly semantically $\mathscr{O}$-determined iff $\{\alpha\} \cup \Pi_{\mathscr{T}}$ or $\{\neg\alpha\} \cup \Pi_{\mathscr{T}}$ is semantically $\mathscr{O}$-creative with respect to $\Pi_{\mathscr{O}}$. Because sets of sentences are not easily negated, this formulation is neither as general nor as conspicuous as

Claim 21. *A set Λ of $\mathscr{V}$-sentences is weakly semantically $\mathscr{O}$-determined iff Λ is semantically verifiable or semantically falsifiable.*

Proof. Assume Λ is semantically verifiable or semantically falsifiable. This holds iff there is a possible $\mathfrak{A}_{\mathscr{O}}$ such that Λ is true in all structures $\mathfrak{B} \vDash \Pi$ with $\mathfrak{B}|\mathscr{O} = \mathfrak{A}_{\mathscr{O}}$ or false in all of them, that is, Λ has the same truth value in all of these structures. This is equivalent to Λ being weakly semantically $\mathscr{O}$-determined. □

With claims 5 and 8, this means that a sentence α is weakly semantically $\mathscr{O}$-determined iff α or $\neg\alpha$ is semantically $\mathscr{O}$-creative with respect to Π. This is Przełęcki's claim, reformulated using the generalized definition of $\mathscr{O}$-conservativeness (see appendix A).

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

Definition 28. A set Λ of $\mathscr{V}$-sentences is *weakly syntactically $\mathscr{O}$-determined* if and only if it is determined by some possible and maximal set of $\mathscr{O}$-sentences.

This definition relates to that of weak semantic $\mathscr{O}$-determinacy in the usual way:

Claim 22. *A set Λ of $\mathscr{V}$-sentences is weakly syntactically $\mathscr{O}$-determined iff there is some $\mathscr{O}$-structure $\mathfrak{A}_{\mathscr{O}}$ such that for all structures $\mathfrak{B}, \mathfrak{C} \vDash \Pi$ with $\mathfrak{B}|\mathscr{O} \equiv \mathfrak{C}|\mathscr{O} \equiv \mathfrak{A}_{\mathscr{O}}$, it holds that $\mathfrak{B} \vDash \Lambda$ iff $\mathfrak{C} \vDash \Lambda$.*

Proof. '⇒': Choose $\Omega := \mathrm{Th}(\mathfrak{A}_{\mathscr{O}})$ and proceed as in the proof of claim 17.

'⇐': Choose some $\mathfrak{A}_{\mathscr{O}} \vDash \Omega$ and proceed as in the proof of claim 17 □

And analogously to the semantic case, the following holds:

Claim 23. *A set Λ of $\mathscr{V}$-sentences is weakly syntactically $\mathscr{O}$-determined iff Λ is syntactically verifiable or syntactically falsifiable.*

Proof. '⇒': Immediate.

'⇐': If Ω verifies or falsifies Λ, Ω is possible. Thus $\Omega \cup \Pi$ can be extended to a possible and maximal set of $\mathscr{O}$-sentences. □

As the disjunction of falsifiability and verifiability, weak syntactic $\mathscr{O}$-determinacy has occurred often in the history of philosophy, albeit repeatedly sailing under false colors. The illicit reflagging often occurred with the help of the prediction criterion of confirmation. 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 tedious,[10] 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 $C_1, C_2, \ldots, C_n$ 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.

For the trivialization proof, assume that C allows the direct incomplete reducibility of at least one sentence γ. For any sentence α, if γ is directly incompletely reducible to Ω, so is $\gamma \wedge \alpha$, which can therefore be in C_1. Then α can be completely reduced to $C_1 := \{\gamma \wedge \alpha\}$ because $\{\gamma \wedge \alpha\} \vDash \alpha$ and $\{\gamma \wedge \alpha\}$ is a finite subset of itself. Thus the confirmation of α is directly reducible to C_1, whose confirmation is directly reducible to C, and therefore the confirmation of α is reducible to C.

Now, the confirmation of S is reducible to a class of $\mathscr{O}$-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 [$\mathscr{O}$] and the negations of these sentences" (Carnap 1936, 435f); 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).[11] Thus if some confirmation class Ω allows the direct incomplete reducibility of at least one sentence γ, the confirmation of α is reducible to $\mathscr{O}$. In that case α 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 α, 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 $\mathscr{O}$ as the concept of confirmability, and this is just the disjunction of falsifiability and verifiability restricted to a special class of $\mathscr{O}$-sentences.

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.

[10] That this holds for most definitions in the article may explain why, as far as I know, no concept introduced in "Testability and Meaning" besides that of reduction sentences has been used since.

[11] I will discuss the relevance of contravalidity in §9.2.

Since Ayer (1936, 37f) 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", thereby implicitly assuming the prediction criterion of confirmation.

In an early work, Carnap (1928, 327f) avoids the prediction criterion by leaving the concept of confirmation undefined. He writes:

> If a statement *p* expresses the content of an experience E, and if the statement *q* is either the same as *p* or can be derived from *p* and prior experiences, either through deductive or inductive arguments, then we say that *q* is "supported by" the experience E. [. . .] A statement *p* is said to have "factual content", if experiences which would support *p* or the contradictory of *p* are at least conceivable, and if their characteristics can be indicated.

Carnap's examples indicate that quantified $\mathscr{O}$-sentences describe conceivable experiences, so that in my terminology, Carnap considers a sentence to have factual content if and only if it is verifiable, falsifiable, confirmable or disconfirmable. Questions of confirmation are outside the scope of this article, so that, as far as this article is concerned, Carnap suggests to consider a sentence empirically significant if and only if it is weakly $\mathscr{O}$-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 observational sentences, Rynin's criterion is therefore equivalent to weak syntactic $\mathscr{O}$-determinacy.

Let me conclude this section with a puzzling observation that suggests that Hempel was not overly diligent in his dismissal of the search for a criterion of empirical significance.

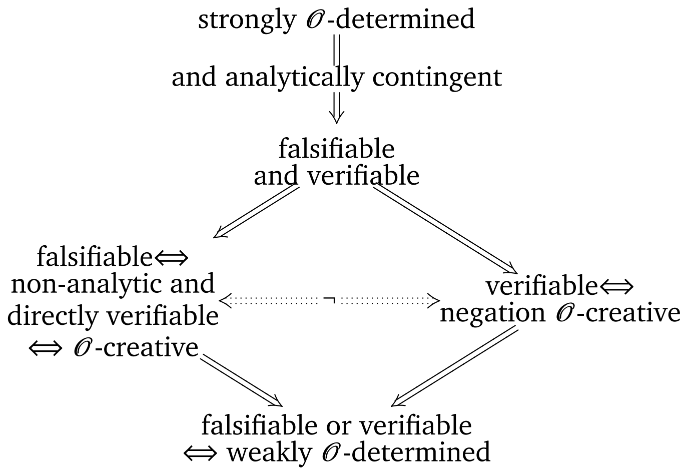

Figure 1: Relations between the syntactic definitions. The equivalence holds for direct verifiability and the negation of sets of sentences whenever the concepts are defined. A strongly $\mathscr{O}$-determined set of sentences is also weakly $\mathscr{O}$-determined even if not analytically contingent.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

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, which is symmetric under negation and more inclusive than both falsifiability and verifiability.

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 $\mathscr{O}$-determinacy, falsifiability, and verifiability as the four major criteria of empirical significance. With this overview, it is now easy to consider the implications of the entailment relations and equivalences.

8.1 Six comparative concepts of empirical significance

The entailment relations between the criteria show that there can be stronger and weaker criteria of empirical significance, and suggest that there may be criteria of comparative empirical significance. Hempel (1965b, 117) similarly states that "cognitive significance in a system is a matter of degree". He sees this as a reason to dispose of the concept altogether, and "instead of dichotomizing this array [of systems] into significant and non-significant

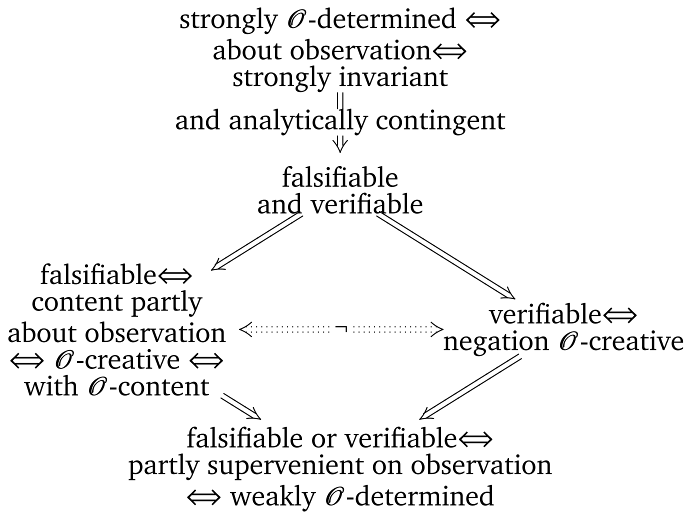

Figure 2: Relations between the semantic definitions. The equivalence holds for strong invariance, empirical content, and the negation of sets of sentences whenever the concepts are defined. A strongly $\mathcal{O}$-determined set of sentences is also weakly $\mathcal{O}$-determined even if not analytically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

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).

As the split of strong and weak $\mathcal{O}$-determinacy into falsifiability and verifiability shows, a comparative explicatum for empirical significance will probably have to be partial, in that not all criteria can be compared with respect to their inclusiveness without further assumptions. Therefore I suggest

Definition 29. A set Λ of $\mathcal{V}$-sentences is *at least as syntactically (semantically) falsifiable/ verifiable/$\mathcal{O}$-determined* as a set Γ of $\mathcal{V}$-sentences if and only if every possible set of

$\mathcal{O}$-sentences (possible $\mathcal{O}$-structure) that falsifies/verifies/$\mathcal{O}$-determines Γ also falsifies/verifies/$\mathcal{O}$-determines Λ.

The partial order of the subset relation transfers to 'being at least as falsifiable/verifiable/$\mathcal{O}$-determined', in both its syntactic and its semantic guise, for each set Π of analytic sentences. 'At least as syntactically falsifiable' is called 'falsifiability of at least as high a degree' by Popper (1935, §33), who also notes that this order is partial (Popper 1935, §34).

There is a second reason why Hempel should not have dismissed the search for criteria of empirical significance so easily. For each set Π, each relation in definition 8.1 has natural greatest and, more importantly, least elements.

Claim 24. *A set Λ of $\mathcal{V}$-sentences is analytically false/analytically true/analytically false or analytically true if and only if Λ is at least as syntactically (semantically) falsifiable/syntactically (semantically) verifiable/syntactically $\mathcal{O}$-determined as any other set of $\mathcal{V}$-sentences.*

Proof. '⇒': Immediate.

'⇐': If Λ is not analytically false, it is not syntactically (semantically) at least as falsifiable as $\bot$. Analogously for verifiability and $\top$.

If Λ is neither analytically false nor analytically true, there are a structure $\mathfrak{A} \vDash \Pi \cup \Lambda$ and a structure $\mathfrak{B} \vDash \Pi$ with $\mathfrak{B} \nvDash \Lambda$. Choose $\Gamma := \Omega := \mathrm{Th}(\mathfrak{A}|\mathcal{O}) \cap \mathrm{Th}(\mathfrak{B}|\mathcal{O})$. Then Ω determines Γ but not Λ. □

This shows that the comparative notions connect fruitfully to analyticity.

Strong semantic $\mathcal{O}$-determinacy connects very straightforwardly to 'semantically more determinate than', because all and only sets of sentences semantically determined by every $\mathcal{O}$-structure are at least as semantically $\mathcal{O}$-determined as any other:

Claim 25. *A set Λ of $\mathcal{V}$-sentences is strongly semantically $\mathcal{O}$-determined if and only if Λ is at least as semantically $\mathcal{O}$-determined as any other set of $\mathcal{V}$-sentences.*

Falsifiability, verifiability, and weak $\mathcal{O}$-determinacy are immediately connected to their comparative counterparts:

Claim 26. *A set Λ of $\mathcal{V}$-sentences is not syntactically (semantically) falsifiable/verifiable/weakly $\mathcal{O}$-determined if and only if Λ is at most as syntactically (semantically) falsifiable/verifiable/$\mathcal{O}$-determined as any other $\mathcal{V}$-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.

8.2 The justifications of the criteria

Many of the circumscriptions of the explicandum and methodological presumptions discussed in the preliminaries provide arguments for the feasibility of a criterion of empirical significance. The relations between the criteria suggest that the criteria are already to a certain extent adequate.

For one, the equivalences of many of the criteria to falsifiability, verifiability, or $\mathcal{O}$-determinacy help to counter the charge of arbitrariness that Lewis (1988b, 127) and, presumably, Hempel (1965b, §4) have put forth. They suggest that the explicated notions are robust under a change of formalism from predicate logic to model theory to set theory, and a change of formulation within each formalism. This provides an argument analogous to (though less spectacular than) that for the successful explication of 'computability', in which the equivalence of different definitions is cited as evidence for their adequacy (Barker-Plummer 2009, Copeland 2008).

The equivalences also show that especially Lewis's charge, that many criteria of empirical significance have strayed too far from the intuitive explicandum, does not apply to the criteria discussed here: Strong semantic $\mathcal{O}$-determinacy is equivalent to aboutness, semantic falsifiability is equivalent to partial aboutness of content, and weak semantic $\mathcal{O}$-determinacy is arguable equivalent to partial supervenience, all of which are meant to capture the ordinary language notion. And although verifiability is not equivalent to any of Lewis's criteria, it occurs with falsifiability in the disjunction that makes up weak $\mathcal{O}$-determinacy (claim 23). It is thus at least the link connecting two different notions of partial aboutness.

The close connection of the criteria to ordinary language may prompt another criticism: that the criteria are of little use in the sciences, the way the ordinary language notion of 'fish', which includes the likes of whales and dolphins, is of little use in biology (Carnap 1950, §3). If the sciences are taken to include mathematics, then the equivalence of falsifiability to $\mathcal{O}$-conservativeness already provide a rebuttal, for the notion of definition is essential in mathematics, and $\mathcal{O}$-conservativeness is essential for the notion of definition (Belnap 1993). At least for sentences, claim 9 shows the relevance of verifiability by its connection to falsifiability through the negation of sentences, and claim 23 shows the relevance of weak $\mathcal{O}$-determinacy by its connection to both criteria through their disjunction. Lest one argue that it is only falsifiability that is really needed, I provide the following appeal to authority: Church (1949) only proofs that every sentence *or its negation* is empirically significant according to Ayer's criterion of indirect verifiability, and this result was considered important in the philosophical community. If such a proof were possible for falsifiability, it would, because of claims 9 and 23, show that every sentence is weakly $\mathscr{O}$-determinate. Thus $\mathscr{O}$-determinacy has been considered closely enough related to falsifiability that the triviality of the one suggests the triviality of the other.

Without relying on the importance of mathematics, one can argue that definitions are similarly important in the natural sciences. Additionally, claim 14 and claims 15 and 16 show that at least strong $\mathscr{O}$-determinacy is important for the concepts of measurement because it generalizes strong invariance. Its close relation to weak $\mathscr{O}$-determinacy suggests that the latter criterion is important within measurement theory as well, in effect stating that a numerical statement is weakly $\mathscr{O}$-determinate if and only if its truth value is for some observations invariant under the admissible transformations (cf. Przełęcki 1974a, 350). Similarly, one may demand that the statement be false or be true for all admissible transformations, thus arriving at a special case of falsifiability or verifiability in terms of admissible transformations and thus measurements.

Conversely, the equivalence to a special case of strong $\mathscr{O}$-determinacy protects strong invariance against the charge that it is ad-hoc. There is always the possibility to be mislead by the special conditions of a context, in this case the features of measurements, but the equivalence shows that strong invariance is a special case of a much more generally motivated criterion.

There is also the more general charge of irrelevance. For even if a criterion is close to some intuitive notion and a generalization of some concept defined for scientific application, the intuitive notion and the scientific concept may be applicable in their domain but irrelevant. Such a charge becomes more difficult to sustain the more concepts rely on or connect to the criterion. The relations shown in figures 1 and 2 suggest what would be irrelevant as well if the criteria discussed here were irrelevant, and it is doubtful that the notion of meaningfulness in measurement, the notion of empirical content as explicated by the Ramsey sentence, and the notions of aboutness and partial aboutness are all of no use.

Finally, there is what Gemes (1998, §1.1) has called “the problem of past failures”. The failure of Ayer's first criterion and subsequent amendments, the argument goes, is a good basis to inductively conclude that future criteria will be failures, too. I have already noted that not all criteria have been shown to be trivial, and the equivalences between the criteria discussed here make it easy to show that none of them is trivial: Assume $\mathscr{O} = \{O, o\}$, $\mathscr{T} = \{P, Q, b\}$, and $\Pi = \{\forall x[Ox \leftrightarrow \neg Px], a = o\}$. Then Pa is empirically significant according to all the criteria, and Qb is not. It is also notable that none of the criteria are amendments of Ayer's criteria (direct verifiability being only the first half of Ayer's second criterion). The induction on past failures therefore does not obviously apply.

The equivalences also provide a positive justification rather than a defense, because now the arguments in favor of each individual formulation turn out to be arguments for the same criterion. Thus Przełęcki's general argument and Suppes's measure theory-specific argument already lead to strong $\mathcal{O}$-determinacy, and Lewis's argument for aboutness adds additional support. Lewis's analysis of the term 'partly' also provides a justification for the differences between the major criteria: The underlying explicandum is ambiguous because 'partly about observation' is ambiguous. Falsifiability is accordingly supported by one specific disambiguation of 'partly about observation', 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 $\mathcal{O}$-determinacy also support the same criterion. Verifiability, while historically not often defended by itself, again receives some justification through its role in weak $\mathcal{O}$-determinacy.

Furthermore, the relations between the criteria suggest that the criteria fulfill Carnap's desiderata for explications (§2.1). They are certainly more precise than the phrase 'empirically meaningful', and some of the formulations are fairly simple—at least, they are not "page-long", as Lewis (1988a, 127) feared. In fact, the equivalences allow the application of different formulations according to expedience. The equivalences to Lewis's ordinary language notions suggest that the criteria are similar to their explicandum "in such a way that, in most cases in which the explicandum has been used, the explicatum can be used".[12] I have also suggested to use conditions of adequacy instead of Carnap's demand for fruitfulness, or, lacking conditions of adequacy, use as a proxy Hempel's demand that the desideratum allow developing a comprehensive and sound theoretical system. The relations between the comparative and classificatory notions of empirical significance (claims 24 and 26) and the relations listed above to counter the charge of irrelevance are steps in that direction. The following section provides more evidence that the criteria allow the development of a comprehensive theoretical system, for it shows how they can be generalized with comparative ease.

9 Generalizations

The presumptions of the semantic criteria of empirical significance are comparably strong: They assume predicate logic, a bipartition of the vocabulary and a set Π containing only analytic sentences. As already noted, the syntactic criteria of empirical significance can be defined with any distinguished set of observational sentences, and the equivalence proofs

[12] I actually think that this demand is problematic because it suggests that there is something sacred about the current usage of a term, as Laudan (1986, 120) has argued: More than half of current uses must be captured. Rather, I would suggest that the current use of the term only provide some conditions of adequacy. Here, I rely on Lewis's ordinary language intuitions 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.

at most rely on the set being closed under truthfunctional composition. At least falsifiability is furthermore already generalized to all systems of logic in which conservativeness is defined. In the following, I will suggest two further generalizations by weakening the restrictions on the observations and the restrictions on Π.

9.1 General observational sentences and structures

Many classical syntactic criteria allow only finite sets of molecular or atomic $\mathscr{O}$-sentences to falsify or verify a $\mathscr{V}$-sentence (Popper 1935, Hempel 1965b, Rynin 1957). Since there may be infinitely many non-equivalent observation sentences, this restriction cannot be captured by restricting the set of observation sentences. To accommodate this restriction and others, I suggest to consider a set of sentences one of observation sentences if and only if it is in $\mathbf{\Omega}$, where $\mathbf{\Omega}$ can be determined as needed. Any element of $\mathbf{\Omega}$ will be called 'set of $\mathbf{\Omega}$-sentences'. This allows the following

Definition 30. A set Λ of $\mathscr{V}$-sentences is *syntactically falsifiable/verifiable/weakly determined in* $\mathbf{\Omega}$ if and only if there is a possible set of $\mathbf{\Omega}$-sentences that falsifies/verifies/determines Λ.

If $\mathbf{\Omega}$ contains all sets of $\mathscr{O}$-sentences, definition 30 is equivalent to the conjunction of definitions 4, 15, and 25.

To achieve a similar versatile definition for semantic criteria of empirical significance, I suggest to use a set $\mathbf{\Omega}$ of sets of structures. Elements of $\mathbf{\Omega}$ will be called 'sets of $\mathbf{\Omega}$-structures'. Whether $\mathbf{\Omega}$ contains sets of structures or sets of sentences will be clear from context. For sets of $\mathbf{\Omega}$-structures, the subset-relation is the analogue of the entailment relation for sets of sentences. In analogy to the compatibility of a set of sentences with Π, a set of $\mathscr{V}$-structures is possible if and only if its intersection with the models of Π is not empty. In other words, a set of $\mathscr{V}$-structures is possible if and only if one of its elements is possible according to definition 2. To define this notion also for structures of proper subsets of $\mathscr{V}$, I suggest

Definition 31. The possible subset of a set of structures Γ is the set of possible structures in Γ. Γ is possible if and only if its possible subset is not empty.

This leads, in analogy to definitions 9, 16, and 18 to

Definition 32. A *set of structures determines* (*falsifies/verifies*) a set Λ of $\mathscr{V}$-sentences in $\mathbf{\Omega}$ if and only if all elements $\mathfrak{A}$ and $\mathfrak{B}$ ($\mathfrak{A}$) of its possible subset are such that $\mathfrak{A} \vDash \Lambda$ iff $\mathfrak{B} \vDash \Lambda$ ($\mathfrak{A} \vDash \Lambda$ / $\mathfrak{A} \nvDash \Lambda$).

Finally, this suggests in analogy to definitions 10, 17, and 26

Definition 33. A set Λ of $\mathscr{V}$-sentences is *semantically falsifiable/verifiable/weakly determined in* $\mathbf{\Omega}$ if and only if there is a possible set of $\mathbf{\Omega}$-structures that falsifies/verifies/determines Λ.

If $\mathbf{\Omega}$ contains all sets of $\mathscr{O}$-structures, definition 33 is equivalent to the conjunction of definitions 10, 17, and 26.

It is also possible to define a generalization of strong $\mathscr{O}$-determinacy:

Definition 34. A set Ω of $\mathbf{\Omega}$-sentences ($\mathbf{\Omega}$-structures) is maximal if and only if there is no possible set Γ of $\mathbf{\Omega}$-sentences ($\mathbf{\Omega}$-structures) such that $\Gamma \cup \Pi \vDash \Omega$ and $\Omega \cup \Pi \nvDash \Gamma$ (the possible subset of Γ is a proper subset of the possible subset of Ω).

If $\mathbf{\Omega}$ contains all sets of $\mathscr{O}$-sentences, this definition is equivalent to definition 24.

As a generalization of definitions 19 and 25, I thus suggest

Definition 35. A set Λ of $\mathscr{V}$-sentences is *strongly semantically (syntactically) determined in* $\mathbf{\Omega}$ if and only if it is determined by every possible maximal set of $\mathbf{\Omega}$-sentences ($\mathbf{\Omega}$-structures).

If $\mathbf{\Omega}$ contains all sets of $\mathscr{O}$-sentences, this definition is equivalent to definition 25. If $\mathbf{\Omega}$ contains all singleton sets of $\mathscr{O}$-structures, this definition is equivalent to definition 19.

Based on these concepts, one can construct generalized notions of comparative determinacy, falsifiability, and verifiability:

Definition 36. A set Λ of $\mathscr{V}$-sentences is *syntactically (semantically) at least as falsifiable/ verifiable/determined in* $\mathbf{\Omega}$ as a set Γ of $\mathscr{V}$-sentences if and only if every possible set of $\mathbf{\Omega}$-sentences ($\mathbf{\Omega}$-structures) that falsifies/verifies/determines Γ in $\mathbf{\Omega}$ also falsifies/verifies/ determines Λ in $\mathbf{\Omega}$.

Since $\mathbf{\Omega}$ is not further determined, it is only possible to proof a simple weak analogy to claim 24:

Claim 27. *If a set Λ of $\mathscr{V}$-sentences is analytically false/analytically true/analytically false or analytically true, then Λ is at least as syntactically (semantically) falsifiable/verifiable/ determined in $\mathbf{\Omega}$ as any other set of $\mathscr{V}$-sentences.*

In complete analogy to claim 26, the following holds:

Claim 28. *A set Λ of $\mathscr{V}$-sentences is not syntactically (semantically) falsifiable/verifiable/ weakly determined in $\mathbf{\Omega}$ if and only if Λ is at most as syntactically (semantically) falsifiable/ verifiable/determined in $\mathbf{\Omega}$ as any other set of $\mathscr{V}$-sentences.*

Proof. The claim holds for all criteria because only the empty set is a subset of every set. □

Claims 27 and 28 show that the generalized comparative concepts, in partial analogy to the more specific ones, relate fruitfully to their classificatory counterparts. An exception is the notion of strong semantic determinacy in $\mathbf{\Omega}$. Unlike strong semantic $\mathscr{O}$-determinacy, but like strong syntactic $\mathscr{O}$-determinacy, it does not identify the greatest elements of its comparative counterpart. Thus the generalizations of strong syntactic and semantic $\mathscr{O}$-determinacy bring the concepts closer together.

9.2 General background assumptions

With reference to Duhem (1914), Sober (2007, 5) notes that scientific theories, "on their own, do not make testable predictions. One needs to add 'auxiliary propositions' to the theories one wishes to test". Typically, these auxiliary sentences are not taken to be just analytic sentences, and therefore the above definitions would still exclude almost all scientific theories even though the criteria are non-trivial. Gemes (1998, §1.2) calls this "the challenge from holism" and points out that the logical empiricists were acutely aware of it.[13]

One simple way to meet the challenge from holism is to consider the empirical significance of the union of the theory and the auxiliary sentences. But this approach does not always allow to infer the empirical (non-)significance of a theory from the (non-)significance of the theory and its auxiliary sentences. For a subset of a falsifiable or $\mathcal{O}$-determined set may itself not be falsifiable or $\mathcal{O}$-determined, respectively, and a subset of a non-verifiable or non-$\mathcal{O}$-determined set may itself be verifiable or $\mathcal{O}$-determined, respectively. This is a problem because a criterion of empirical significance is meant to determine whether a specific theory is significant, while this approach determines only the empirical significance of huge sets of theories, possibly including most of science and everyday knowledge.

Ayer's definition of indirect verifiability is an attempt at meeting this challenge from holism in a different way, by defining empirical significance relative to a set Π containing not only analytic sentences, but also other empirically significant sentences. This kind of recursive definition is suggestive given its success in the theory of definitions, for if a term P is definable in $\mathcal{O}$-terms, then any term definable in $\mathcal{O}$-terms and P is also definable in $\mathcal{O}$-terms alone, and if a set Λ of sentences is translatable into a set of $\mathcal{O}$-sentences, then any sentence translatable into a set of sentences containing only $\mathcal{O}$-terms and Λ is also translatable into a set of $\mathcal{O}$-sentences. For strong syntactic $\mathcal{O}$-determinacy, such a recursion works as well, because the truth value of an $\mathcal{O}$-determined set Λ of sentences is a function of the truth values of observational sentences, and thus, if the truth value of any set of sentences is a function of the truth values of Λ and observational sentences, it is also a function of the truth values of observational sentences alone. But for any of the weaker criteria of empirical significance, this recursion breaks down.

The problem of past failures clearly does not allow the conclusion that all criteria will fail. But the puncture-and-patch industry as summarized in §1 consists exclusively of recursive criteria, most recursive criteria suggested so far are trivial, and no recursive criterion has been shown to be non-trivial. The correct inference to draw from the problem of past failures may thus be that there is no adequate *recursive* criterion of empirical significance.

Another reason to question the search for a recursive criterion is that recursive criteria

[13]Gemes (1998, §2) attempts to solve this problem for Ayer-type criteria with his concept of natural axiomatization.

do not seem to address the challenge from holism. In a recursive criterion, the auxiliary sentences can contain any empirically significant sentence, even those that are known to be false. But the challenge from holism consists in the need for other *true* (or at least justified) sentences to evaluate a theory.

Surprisingly enough, there have been few justifications of the assumption that a criterion of empirical significance should be recursive. In defense of his second criterion of empirical significance, Ayer (1946, 12) states that only by taking auxiliary sentences into account can "hypotheticals" be rendered empirically significant. But Ayer then simply chooses empirically significant auxiliary sentences without even considering justified auxiliary sentences. As far as I know, this is the only justification that has been provided for recursive criteria.

Sober (2008, 151, §2.14) suggests a criterion that uses justified auxiliary sentences. He points to the triviality of Ayer's first criterion and, with "some fear [of] stumbling into the same old quagmire", suggests instead:

> Proposition P now has observational implications if and only if there exist true auxiliary assumptions A, and an observation statement O, such that (i) $P\&A$ entails O, but A by itself does not entail O, (ii) we now are justified in believing A, and (iii) the justification we now have for believing A does not depend on believing that P is true (or that it is false), and also does not depend on believing that O is true (or that it is false).

Sober argues for the requirement that A be justified independently of O as follows. Let P be any sentence, and O a justified true observational sentence. Then $\neg P \lor O$ is true and justified, and if the choice $A := \{\neg P \lor O\}$ were allowed by the definition, P would have observational implications. In short, the criterion is trivial without the restriction.[14]

Sober's trivialization argument is incomplete because he assumes, but does not prove, that $\neg P \lor O \nvDash O$ for some O. Lewis (1988a) notes that if $\neg P \lor O \vDash O$ for all O, then P is a logical truth. Hence the restriction on A is justified under the assumption that not only logical truths should lack observational implications. The argument is also no strict proof as long as 'justification' and 'dependence of a justification' are not defined. Whether it is valid depends on whether, for example, O justifies $\neg P \lor O$ independently of $\neg O$, P, and $\neg P$ for any P.

More problematic is that the criterion itself is arguably trivial: For, assume that P is any sentence and Q a true non-observational sentence that entails O and is justified independently of O, $\neg O$, P, and $\neg P$. Then, in analogy to Sober's trivialization argument, $A := \{\neg P \lor Q\}$ is allowed by his definition, and thus, unless it is logically true, P has observational implications. Unlike Sober's trivialization argument, this argument assumes that

[14] Sober (2008, 144f) provides a trivialization argument for the case that A depends on P only for his explication of 'P is testable against Q', so its relevance for the definition at hand is not immediately clear. Furthermore, there are reasons to assume that the argument fails (Lutz 2010b).

the justification for a sentence does not always have to depend on *all* of the observational sentences it entails, but this is a rather weak assumption about justification.

Sober's and my trivialization arguments rely on the same trick: Some sentence (O or Q) is used to infer another ($\neg P \vee O$ or $\neg P \vee Q$) by irrelevant disjunction, but only the inferred sentence is included in A. One way of avoiding this specific trivialization is to disallow irrelevant disjunctions, for example by restricting justificatory inferences to relevant deductions as developed by Schurz (1991). I will suggest another way that is more suitable to the current case, starting from an observation that Schurz (1991, §2.2, parenthetical remark in the original) makes in his justification of relevant deduction. He notes that if one knows that A and tells someone else that $A \vee B$, one misleads the hearer because

> in practical speech situations the hearer assumes that if the speaker tells him a disjunction, say $A \vee B$, then the speaker's knowledge K_s about A and B is indeed *incomplete*, i. e. both $\neg A$ and $\neg B$ and thus also A and B are possible in K_s [because, given $K_s \vdash_L A \vee B$, $K_s \nvdash_L A/B$ implies $K_s \nvdash_L \neg B/\neg A$, respectively]. [. . .] The irrelevant conclusion together with this implicit assumption causes in the hearer an expectation which is not only irrelevant but *wrong*—

namely that it is possible that B. '$\vdash_L$' here refers to deductive inference in predicate logic.

My suggestion is to avoid this wrong expectation by disallowing the speaker to tell the hearer that $A \vee B$ but not A, if he knows that A. If the speaker asserts auxiliary sentences Π, this leads to

Definition 37. Π is an *honest set* if and only if every $\varphi \in \Pi$ is a justified sentence, and Π also contains every sentence on which the justification of φ depends.

Somewhat less precisely, I will also speak of "honest auxiliary sentences" rather than "honest sets of auxiliary sentences". One might paraphrase this requirement on Π more intuitively as the demand that Π contain *all* auxiliary sentences that are accepted, either actually or in some counterfactual situation that is of interest. Note that an honest set of auxiliary sentences can be finite, because analytic sentences do not need justifications, and some sentences may be justified by entities that are not sentences (e. g. observations). Coherentists may argue that the sentences in a finite set can justify each other without such primitively justified sentences.

To avoid trivialization of Sober's criterion, I therefore suggest the following modification:

Definition 38. A set Λ of $\mathscr{V}$-sentences *has observational implications* if and only if Λ is $\mathscr{O}$-creative with respect to an honest set containing all analytic sentences.

This is syntactic falsifiability, with Π taken to be an honest set of justified or analytic sentences rather than just the set of analytic sentences. All other definitions discussed above, including the generalizations in §9.1, can be reinterpreted analogously. It is clear that in the intuitive paraphrase of definition 37, the counterfactual case is important because otherwise, an accepted theory could have no observational implications at all since anything it entails would (I assume) be accepted and therefore in Π.

Sober's definition and definition 38 differ beyond their restrictions on the auxiliary sentences. Unlike Sober's criterion, $\mathcal{O}$-creativity is defined for set of sentences and includes sets that only allow new inferences from infinite sets of observational sentences (in the case of higher order logic). I do not think that Sober's criterion was designed with this distinction in mind, however, so this modification is only of a technical nature. Sober also does not assume that observational sentences are determined by their vocabulary. But as noted, this restriction can be accommodated in artificial languages, is not essential for syntactic criteria, and is completely avoided in the generalizations in §9.1. Finally, the definition does not include a reference to the current point in time, because time dependence follows from the definition's dependence on justified auxiliary sentences. Since the set of justified auxiliary sentences changes over time, so does the set of sentences that are $\mathcal{O}$-creative with respect to them.

Ignoring these rather formal differences, definition 38 is at least as exclusive as Sober's definition: Condition (ii) is entailed by the demand that the auxiliary sentences form an honest set. Condition (iii) for one excludes sentences P that are $\mathcal{O}$-creative relative to auxiliary sentences whose justification depends on P. In definition 38, those sentences lack observational implications because they are already contained in the auxiliary sentences, which therefore already entail everything that their conjunction with P entails. Condition (iii) furthermore excludes sentences P that only entail observational sentences O on which the justification of the auxiliary sentences depends. In definition 38, these observational sentences are also contained in the auxiliary sentences, so that they are already entailed by the auxiliary sentences alone. There is also hope that definition 38 is strictly more exclusive than Sober's definition, and thus not trivial, for the use of honest auxiliary sentences blocks the trivialization arguments given above. This is because in the arguments, $\neg P \vee O$ and $\neg P \vee Q$ are justified by O and Q, respectively. If O or, respectively, Q are included in Π, Π alone entails O.

I have not shown that reinterpreting Π as an honest set of auxiliary sentences leads to non-trivial criteria of significance, and the triviality of Sober's criterion might suggest that the new criteria will suffer the same fate as Ayer's. However, there are important differences between Ayer's criterion and the new ones. First, the criteria's core ideas—falsifiability, verifiability, and $\mathcal{O}$-determinacy relative to a set Π—are not trivial. Therefore, in the case of a trivialization proof, one can always fall back on falsifiability, verifiability, or $\mathcal{O}$-determinacy, and find a new restriction on the auxiliary sentences Π. (Without a criterion of justification, empirical significance can be defined in a precise way only *relative* to a set Π, and it has to be decided on a case-by-case basis whether Π is acceptable.) Second, and in keeping with my criticism of recursive definitions of empirical significance,

Π is not assumed to be determined by the criterion itself. This blocks trivialization proofs that rely on recursive definitions. From these two differences follows a third: Problems can only occur with the definition of 'honest set', and amendments of the criteria can accordingly be restricted to this definition. In this sense, falsifiability, verifiability, and $\mathcal{O}$-determinacy are already good criteria of empirical significance. What is missing is a good criterion of justification for sets of sentences.

Historically, the interpretations of Π have been varied. Przełęcki's definitions explicitly assume Π to contain all and only analytic sentences. Rynin seems to define his criterion with analytic sentences in mind, and, as argued above, Ayer's definition of direct verifiability relies on analytic inferences, too. Lewis's definitions rely on the concept of possible worlds, and whether these are determined by analytic sentences is up for discussion. Suppes is silent on the matter. Carnap (1936, 443), on the other hand, considers observation sentences possible only if they are compatible with the laws of physics, that is, only if they are not contravalid. And in another passage, Carnap (1935, 11) writes:

> A proposition *P* which is not directly verifiable can only be verified by direct verification of propositions deduced from *P* together with other already verified propositions.

This is almost Ayer's definition of indirect verifiability, except for one crucial difference: The "other propositions" are not only required to be verifiable, but actually verified. Unlike Ayer, Carnap does not define 'verifiability' recursively, but rather relative to a set of justified propositions.[15]

Popper (1935, §3, emphasis changed) is very explicit about the role of justified sentences in his conception of falsification:[16]

> [T]here is the testing of the theory by way of empirical applications of the conclusions which can be derived from it.
>
> [. . .] With the help of other statements, *previously accepted*, certain singular statements—which we may call 'predictions'—are deduced from the theory [. . .]. Next we seek a decision as regards these (and other) derived statements by comparing them with the results of practical applications and experiments. [I]f the decision is negative, or in other words, if the conclusions have been falsified, then their falsification also falsifies the theory from which they were logically deduced.

[15]Given the prediction criterion of confirmation, however, Carnap's definition is still trivial: For any P and any two $\mathcal{O}$-sentences $\omega \nvdash \omega'$, where ω is true, $\{(P \rightarrow \omega') \land \omega\} \vdash \omega$ and is thus verified by ω. Therefore, P is indirectly verifiable.

[16]Lakatos (1974, 106f) also stresses that for Popper, falsifiability is relative to background assumptions, and gives further references to Popper's remarks on the topic.

With the inclusion of previously accepted sentences in Π, the viability of Popper's criterion depends on the conception of acceptance, as do the viability of definition 38 and Sober's criterion. Indeed, apart from the specific restriction (iii) on auxiliary sentences, Sober's criterion is essentially Popper's falsifiability criterion. But requirement (iii), if taken to be the only restriction on the set of accepted sentences, trivializes Sober's criterion. Therefore it seems fair to say that Sober's criterion, insofar as it is successful, is anticipated by Popper.

In a review, Nott (1959) concludes about Popper's dissolution of the problem of induction:

> One cannot help feeling that if [*The Logic of Scientific Discovery*] had been translated as soon as it was originally published[,] philosophy in this country might have been saved some detours. Professor Popper's thesis has that quality of greatness that, once seen, it appears simple and almost obvious.

This is the right conclusion, but the wrong thesis. It is doubtful that Popper's dissolution of the problem of induction is indeed simple and obvious (cf. Salmon 1967, §II.3). But the preceding analysis suggests that if Popper's falsifiability criterion, rather than Ayer's, had been the basis of further research into the problem of demarcation, philosophy might have been saved the "sorry history of unintuitive and ineffective patches" (Lewis 1988b, §I) that discredited the very idea of empirical significance.

10 Conclusion

I am now in a position to defend the title of this article. The belief that the search for a criterion of empirical significance has been a failure is usually based on the problem of past failures. I have given an alternative view on this search, mostly based on criteria that have not been shown to be trivial. I have already argued that the criteria successfully explicate their explicandum (§8.2), specifically with respect to their fruitful connection to each other, comparative criteria of empirical significance, Ramsey sentences, and concepts from measurement and definition theory. And because of the inferential relations between the criteria, it is easy to see when an analysis or justification of one criterion transfers to another.

The inferential relations also allow for a more informed search for generalizations of the criteria. Two such generalizations, to general observational sentences and structures and to general auxiliary sentences, have been developed in this article. My hope is that the relations will also provide guidance in the search for criteria that can be applied to probabilistic theories.

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

Acknowledgments

Parts of this article were developed during stays at the Department of Philosophy, University of Western Ontario, Canada, and the Tilburg Center for Logic and Philosophy of Science, Tilburg University, The Netherlands. Parts were presented at the *Research Seminar in Epistemology and Philosophy of Science* at the Tilburg Center for Logic and Philosophy of Science, the *Lunch Talk of the Theoretical Philosophy Unit* at Utrecht University, The Netherlands, and the *PhDs in Logic II*-conference in Tilburg. Many thanks to the participants and to Reinhard Muskens, Thomas Müller, Harrie de Swart, Stefan Wintein, and Alana Yu for helpful discussions.

A A generalization of Przełęcki's definition of $\mathscr{O}$-conservativeness

Definition 8 of semantic $\mathscr{O}$-conservativeness is slightly more general than Przełęcki's, who suggests

Definition 39 (Przełęcki)**.** A set Λ of $\mathscr{V}$-sentences is *semantically $\mathscr{O}$-conservative with respect to* Δ if and only if for each $\mathscr{O}$-structure $\mathfrak{A}_{\mathscr{O}} \vDash \Delta$ there is a $\mathscr{V}$-structure $\mathfrak{C} \vDash \Delta \cup \Lambda$ with $\mathfrak{C}|\mathscr{O} = \mathfrak{A}_{\mathscr{O}}$.

Semi-formally, this can be paraphrased as

$$\forall \mathfrak{A}_{\mathscr{O}} [\mathfrak{A}_{\mathscr{O}} \vDash \Delta \Rightarrow \exists \mathfrak{C}(\mathfrak{C} \vDash \Delta \cup \Lambda \wedge \mathfrak{C}|\mathscr{O} = \mathfrak{A}_{\mathscr{O}})] \ , \tag{4}$$

which is obviously restricted to observational sentences in Δ due to '$\mathfrak{A}_{\mathscr{O}} \vDash \Delta$' in the antecedent of the implication. Equivalently,

$$\forall \mathfrak{A}_{\mathscr{O}} [\exists \mathfrak{B}(\mathfrak{B}|\mathscr{O} \vDash \Delta \wedge \mathfrak{B}|\mathscr{O} = \mathfrak{A}_{\mathscr{O}}) \Rightarrow \exists \mathfrak{C}(\mathfrak{C} \vDash \Delta \cup \Lambda \wedge \mathfrak{C}|\mathscr{O} = \mathfrak{A}_{\mathscr{O}})] \ , \tag{5}$$

but here the restriction of $\mathfrak{B}$ in '$\mathfrak{B}|\mathscr{O} \vDash \Delta$' is gratuitous, and dropping it leads to definition 8.

B The relation of syntactic and semantic $\mathscr{O}$-conservativeness

That syntactic $\mathscr{O}$-conservativeness does not entail semantic $\mathscr{O}$-conservativeness in first order logic is known in model theory as the difference between elementary and pseudo-elementary classes (Hodges 1993, §5.2). Elementary classes are simply those whose members are models of a set Γ of first order sentences ($\{\mathfrak{A} \mid \mathfrak{A} \vDash \Gamma\}$), while pseudo-elementary classes are those whose members are reducts, to a fixed subvocabulary, of the members of an elementary class ($\{\mathfrak{A}|\mathscr{O} \mid \mathfrak{A} \vDash \Gamma\}$). In light of claim 3, it is easy to see that a set Λ is syntactically $\mathscr{O}$-conservative if and only if $\{\mathfrak{A}_{\mathscr{O}} \mid \mathfrak{A}_{\mathscr{O}} \vDash \Pi|\mathscr{O}\} = \{\mathfrak{A}_{\mathscr{O}} \mid \mathfrak{A}_{\mathscr{O}} \vDash \Lambda \cup \Pi|\mathscr{O}\}$, where $\Pi|\mathscr{O}$ is the set of $\mathscr{O}$-sentences entailed by Π, but semantically $\mathscr{O}$-conservative if and only if $\{\mathfrak{A}|\mathcal{O} \mid \mathfrak{A} \vDash \Pi\} = \{\mathfrak{A}|\mathcal{O} \mid \mathfrak{A} \vDash \Lambda \cup \Pi\}$. Mal′tsev introduced pseudo-elementary classes in 1941 (see Hodges 1993, 207, 260).

Przełęcki (1969, 52f) was possibly the first to mention this distinction in connection with philosophy of science, specifically with respect to questions of concept formation and the theory of definition. He does mention a proof, personally communicated by C. C. Chang, that semantic $\mathcal{O}$-creativity does not entail syntactic $\mathcal{O}$-creativity.

Przełęcki and Wójcicki (1971, §1) reproduce a proof by Łoś (1955):

Proof of claim 4. Let

$$\mathfrak{A}_{\mathcal{O}} := \langle \mathbb{N}, +, \cdot, 0, 1 \rangle \tag{6}$$

be the standard model of arithmetic and define $\Delta := \mathrm{Th}(\mathfrak{A}_{\mathcal{O}})$. Δ is therefore complete. Define $\mathcal{T} := \{P\}$ for the predicate symbol P and

$$\alpha := P0 \wedge \forall x[Px \rightarrow Px + 1] \wedge \exists x \neg Px\ . \tag{7}$$

$\{\alpha\} \cup \Delta$ is consistent, but there are models of Δ, for example $\mathfrak{A}_{\mathcal{O}}$, that cannot be expanded to a model of $\{\alpha\} \cup \Delta$. □

The difference between syntactic and semantic $\mathcal{O}$-creativity was first popularized in the philosophy of science by Sneed (1971). Subsequent discussions were often phrased in terms of the Ramsey-eliminability of theoretical terms (cf. Rynasiewicz 1983, §1; van Benthem 1978).

With a reference to Przełęcki (1969), van Benthem (1982, §2.1.3) gives the following proof with a more intuitive natural language interpretation:

Proof of claim 4. Let Δ be a complete axiomatization of the theory of 0 and successor:

$$\begin{aligned}\Delta := \{&\neg\exists x\, sx = 0, \forall x \forall y(sx = sy \rightarrow x = y), \forall x(x \neq 0 \rightarrow \exists y\, x = sy)\} \cup \\ &\{\neg\exists x_0 \exists x_1 \exists x_2 \ldots \exists x_{n-1} \exists x_n (sx_0 = x_1 \wedge sx_1 = x_2 \wedge \cdots \wedge sx_{n-1} = x_n \wedge sx_n = x_0)\}_{n \in \mathbb{N}}\ .\end{aligned} \tag{8}$$

Δ can be understood as the theory that time proceeds without loops by a one-to-one successor function s starting from 0. Define $\mathcal{T} := \{\prec, E\}$ with a 'before' relation, an 'early' predicate, and a set of sentences stating the following: 0 is early, the successor of an early time is also early, each time is earlier than some late (not early) time, and any time later than a late time is itself late:

$$\alpha := E0 \wedge \forall x(Ex \rightarrow Esx) \wedge \forall x \exists y(x \prec y \wedge \neg Ey) \wedge \forall x \forall y(x \prec y \wedge \neg Ex \rightarrow \neg Ey)\ . \tag{9}$$

Any finite subset of $\{\alpha\} \cup \Delta$ has a model, and by compactness, $\{\alpha\} \cup \Delta$ itself has a model. However, $\langle \mathbb{N}, S, 0 \rangle$ is a model of Δ that cannot be expanded to a model of $\{\alpha\} \cup \Delta$. □

The difference between semantic and syntactic falsifiability was recently again discussed in connection with Ramsey sentences by Ketland (2004, 297f) and Demopoulos (2010, §2).

References

Ayer, A. J. (1936). *Language, Truth and Logic*. Victor Gollanz, London, 1st edition. Page numbers refer to the second edition (Ayer 1946). 2, 10, 11, 28, 29

Ayer, A. J. (1946). *Language, Truth and Logic*. Victor Gollanz, London, 2nd edition. 10, 11, 39, 46

Barker-Plummer, D. (2009). 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 2009 edition. 33

Belnap, N. (1993). On rigorous definitions. *Philosophical Studies*, 72:115–146. 9, 33

Carnap, R. (1928). *Scheinprobleme in der Philosophie: Das Fremdpsychische und der Realismusstreit*. Weltkreis-Verlag, Berlin-Schlachtensee. References are to the translation (Carnap 1967a). 15, 29

Carnap, R. (1935). *Philosophy and Logical Syntax*. Number 70 in Psyche Miniatures, General Series. Kegan Paul, Trench, Trubner & Co., London. References are to the corrected reprint (Carnap 1967b). 42

Carnap, R. (1936). Testability and meaning. *Philosophy of Science*, 3(4):420–468. 27, 28, 42

Carnap, R. (1937). Testability and meaning—continued. *Philosophy of Science*, 4(1):2–35. 27

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). 4, 31, 33

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. 46

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. 7

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. 8

Carnap, R. (1967a). *The Logical Structure of the World. Pseudoproblems of Philosophy.* University of California Press, Berkeley and Los Angeles. Translation by Rolf A. George. 46

Carnap, R. (1967b). *The Logical Syntax of Language*. Routledge & Kegan Paul Ltd, London. Reprinted with corrections. Translation by Amethe Smeaton (Countess von Zeppelin). 46

Church, A. (1949). Review of Ayer's *Language, Truth, and Logic*, 2nd edition. *The Journal of Symbolic Logic*, 14(1):52–53. 11, 34

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

Demopoulos, W. (2010). Three views of theoretical knowledge. *The British Journal for the Philosophy of Science*. 45

Diamond, M. L. and Litzenburg, T. V., editors (1975). *The Logic of God: Theology and Verification*. Bobbs-Merill, Indianapolis, IN. 3, 47

Duhem, P. (1914). *La théorie physique, son objet et sa structure*. Marcel Rivière, Paris, 2nd edition. References are to the translation (Duhem 1954). 38

Duhem, P. (1954). *The Aim and Structure of Physical Theory*. Princeton University Press, Princeton, NJ. 47

Flew, A. (1950). Theology and falsification. *University*, 1:1–8. References are to the reprint (Diamond and Litzenburg 1975, 257–259). 5

Flew, A. (1975). "Theology and falsification" in retrospect. In Diamond and Litzenburg (1975), pages 269–283. 5

Friedman, M. (1982). Review. *The Journal of Philosophy*, 79(5):274–283. Review of (Van Fraassen 1980). 5

Gemes, K. (1998). Logical content and empirical significance. In Weingartner, P., Schurz, G., and Dorn, G., editors, *Proceedings of the 20th International Wittgenstein Symposium*, Vienna. Hölder-Pichler-Tempski. 7, 11, 28, 34, 38

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, Stanford, spring 2009 edition. 9

Hempel, C. G. (1950). Problems and changes in the empiricist criterion of meaning. *Revue Internationale de Philosophie*, 11:41–63. 29

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*. The University of Chicago Press, Chicago and London. References are to the two volume edition. 4, 31

Hempel, C. G. (1965a). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. The Free Press, New York. 48

Hempel, C. G. (1965b). Empiricist criteria of cognitive significance: Problems and changes. In Hempel (1965a), pages 81–119. 7, 8, 10, 16, 30, 33, 36

Hempel, C. G. (1965c). Postscript (1964) on cognitive significance. In Hempel (1965a), pages 120–122. 7, 17, 30

Hempel, C. G. (1965d). Studies in the logic of confirmation. In Hempel (1965a), pages 3–51. 11

Hempel, C. G. and Oppenheim, P. (1936). *Der Typusbegriff im Lichte der neuen Logik. Wissenschaftstheoretische Untersuchungen zur Konstitutionsforschung und Psychologie*. Sijthoff, Leiden. 31

Hodges, W. (1993). *Model Theory*, volume 42 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge. 6, 21, 22, 44, 45

Kemeny, J. G. (1963). Analyticity versus fuzziness. *Synthese*, 15(1):57–80. 6, 13

Ketland, J. (2004). Empirical adequacy and ramsification. *British Journal for the Philosophy of Science*, 55:287–300. 45

Kitts, D. B. (1977). Karl popper, verifiability, and systematic zoology. *Systematic Zoology*, 26(2):185–194. 18

Kuipers, T. A. F., editor (2007). *General Philosophy of Science—Focal Issues*, volume 1 of *Handbook of the Philosophy of Science*. Elsevier, Amsterdam. 4

Lakatos, I. (1974). Falsification and the methodology of scientific research programmes. In Lakatos, I. and Musgrave, A., editors, *Criticism and the Growth of Knowledge*, volume 4 of *Proceedings of the International Colloquium in the Philosophy of Science*, pages 91–196. Cambridge University Press, Cambridge. 42

Laudan, L. (1986). Some problems facing intuitionist meta-methodologies. *Synthese*, 67(1):115–129. 35

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, 10, 35, 39

Lewis, D. (1988b). Statements partly about observation. *Philosophical Papers*, 17:1–31. References are to the reprint (Lewis 1998, 125–155). 10, 13, 14, 25, 33, 43

Lewis, D. (1998). *Papers in Philosophical Logic*. Cambridge Studies in Philosophy. Cambridge University Press, Cambridge. 49

Łoś, J. (1955). On the extending of models I. *Fundamenta Mathematicae*, 42. 45

Lutz, S. (2010a). On a straw man in the philosophy of science: A defense of the received view. Presented at the *British Society for the Philosophy of Science Annual Meeting*, (Dublin, Ireland, July 9th, 2010). `http://philsci-archive.pitt.edu/archive/00005497/`. 4

Lutz, S. (2010b). On Sober's criterion of contrastive testability. Forthcoming. 39

Lutz, S. (2010c). What's right with a syntactic approach to theories and models? Presented at *EPSA09: 2nd Conference of the European Philosophy of Science Association (Amsterdam, The Netherlands, October 22nd, 2009)*. `http://philsci-archive.pitt.edu/archive/00005264/`. 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

Nielsen, K. (1966). On fixing the reference range of 'God'. *Religious Studies*, 2(1):13–36. 5

Nott, K. (1959). Review of Karl Popper: *The Logic of Scientific Discovery*. *Times Literary Supplement*. Published on the 27th of February. Republished online 2009-01-08. Retrieved 2009-11-16, 18:26h. 43

Peacocke, C. (1986). *Thoughts: an Essay on Content*, volume 4 of *Aristotelian Society*. Basil Blackwell, New York, NY. 17

Pokriefka, M. L. (1983). Ayer's definition of empirical significance revisited. *Analysis*, 43(4):166–170. 2, 11

Pokriefka, M. L. (1984). More on empirical significance. *Analysis*, 44(2):92–93. 2

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, 7, 9, 32, 36, 42, 50

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. 50

Przełęcki, M. (1969). *The Logic of Empirical Theories*. Monographs in Modern Logic Series. Routledge & Kegan Paul/Humanities Press, London/New York. 5, 7, 18, 45

Przełęcki, M. (1974a). Empirical meaningfulness of quantitative statements. *Synthese*, 26:344–355. 5, 7, 12, 18, 20, 26, 34

Przełęcki, M. (1974b). On model theoretic approach to empirical interpretation of scientific theories. *Synthese*, 26:401–406. 7

Przełęcki, M. and Wójcicki, R. (1971). Inessential parts of extensions of first-order theories. *Studia Logica*, 28(1):83–95. 45

Psillos, S. (2000). Rudolf Carnap's 'Theoretical concepts in science'. *Studies in History and Philosophy of Science Part A*, 31:151–172. 7, 15

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

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

Roberts, F. S. and Franke, C. H. (1976). On the theory of uniqueness in measurement. *Journal of Mathematical Psychology*, 14:211–218. 20, 22

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. 7, 15

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, 7

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

Rynasiewicz, R. A. (1983). Falsifiability and the semantic eliminability of theoretical languages. *The British Journal for the Philosophy of Science*, 34(3):225–241. 45

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, 17, 29, 36

Salmon, W. C. (1967). *The Foundations of Scientific Inference*. University of Pittsburgh Press, Pittsburgh, PA. 43

Schurz, G. (1991). Relevant deduction. *Erkenntnis*, 35:391–437. 40

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

Sneed, J. D. (1971). *The Logical Structure of Mathematical Physics*. D. Reidel Publishing Co., Dordrecht, The Netherlands. 45

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

Sober, E. (1999). Testability. *Proceedings and Addresses of the American Philosophical Association*, 73(2):47–76. 3

Sober, E. (2007). What is wrong with intelligent design? *The Quarterly Review of Biology*, 82(1):3–8. 3, 38

Sober, E. (2008). *Evidence and Evolution: The Logic Behind the Science*. Cambridge University Press, Cambridge. 3, 7, 17, 39

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. 19, 21, 22

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. 20, 21, 22

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

van Benthem, J. (1978). Ramsey eliminability. *Studia Logica*, 37:321–336. 45

van Benthem, J. (1982). The logical study of science. *Synthese*, 51:431–472. 45

van Fraassen, B. C. (1980). *The Scientific Image.* The Clarendon Library of Logic and Philosophy. Clarendon Press, Oxford. 5, 47

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

Woit, P. (2006). *Not Even Wrong: The Failure of String Theory and the Search for Unity in Physical Law*. Basic Books, New York. 3

Wright, C. (1986). Scientific realism, observation and the verification principle. In Macdonald, G. and Wright, C., editors, *Fact, Science and Morality: Essays on A. J. Ayer's Language, Truth and Logic*, pages 247–274. Basil Blackwell, Oxford. 2

Wright, C. (1989). The verification principle: Another puncture—another patch. *Mind*, 98(392):611–622. 2

Yi, B.-U. (2001). Compact entailment and Wright's verification principle. *Mind*, 110(437):414–421. 2

Typeset in Bitstream Charter with pdfLaTeX, mathdesign, and microtype.