On a Contrastive Criterion of Testability II: The Material Inadequacy of Contrastive Testability

Sebastian Lutz*

Draft: 2011-04-10

Abstract

Elliott Sober has suggested his contrastive criterion of testability as an improvement over previous criteria of empirical significance like falsifiability and the standard Bayesian criterion of empirical significance. I argue that the criterion fails to meet four of the conditions of adequacy for a criterion of empirical significance that follow from Sober's position or are presumed in his arguments. I suggest to define empirical significance as empirical non-equivalence to a tautology, because this definition does meet the conditions of adequacy. Specifically, it is equivalent to the standard Bayesian criterion of empirical significance whenever all probabilities are defined and contains falsifiability as a special case. This latter feature is important because those conditions of adequacy that apply to criteria of deductive empirical significance single out falsifiability.

Contents

1 Introduction **2**

2 Contrastive testability **3**

3 Conditions of adequacy for a criterion of empirical significance **5**
3.1 Conditions of adequacy for a criterion of empirical significance 6
3.2 Falsifiability is the unique adequate criterion of deductive empirical significance . . 10

4 Contrastive testability and the conditions of adequacy **12**

5 Explicating probabilistic empirical significance **14**
5.1 Probabilistic empirical equivalence 15
5.2 Probabilistic observational assertions 16

*Theoretical Philosophy Unit, Utrecht University. sebastian.lutz@gmx.net. I thank James Justus, Jesse Mulder, and especially Thomas Müller and Elliott Sober for helpful comments and discussions.

6 Conclusion **20**

A Additional proofs **21**

1 Introduction

Over the last two decades, Elliott Sober (1990; 1999; 2007; 2008) has developed and defended a criterion of empirical significance, called ‘testability’, that is both promising and much needed. The promise stems from Sober’s defense of the criterion’s basic assumptions and the fact that it can deal with probabilistic theories and auxiliary assumptions. The need for a criterion of empirical significance stems from, for example, questions about the empirical significance of string theory (Smolin 2006, Woit 2006), scientific realism (Sober 1990), theism (Diamond and Litzenburg 1975; Martin 1990, §2), and the theory of intelligent design (ID), to which Sober (1999; 2007; 2008) applies his criterion. Given the possible applications of a criterion of empirical significance and the simultaneous wide-spread belief that the search for such a criterion has utterly failed (cf. Soames 2003, ch. 13), it is somewhat surprising that Sober’s criterion has neither been subjected to much scrutiny, nor led to further research into criteria of empirical significance; this is all the more surprising because Sober’s treatment of ID is widely discussed. Together with its companion piece (Lutz 2011), this article is meant to fill this gap.

Since Sober considers his definition of confirmation an explication (Sober 2008, 35) and his definition of testability is an outgrowth thereof, it is plausible that the latter is also meant as an explication.[1] His definition is thus an explicatum for the explicandum that could be circumscribed by pre-analytic terms like ‘having empirical content’, ‘making observational assertions’, ‘predicting experimental outcomes’, or ‘being an empirical theory’.[2] The latter term is particularly apt, because Sober (2007; 2008, §2.8) intends to improve upon Popper’s criterion of falsifiability, which was proposed as a demarcation criterion between empirical and non-empirical theories (Popper 1935, §4; 1963, §II).

One desideratum of an explication is that it must be possible to use the explicatum in place of the explicandum in the relevant contexts (Carnap 1950, §3; Hempel 1952, 663), which often leads to a variety of conditions of material adequacy that an explicatum has to fulfill. I will argue that Sober’s assumptions and the intended application of his criterion of empirical significance lead to six such conditions (§3.1), one of them the condition that a probabilistic criterion of empirical significance should contain falsifiability as a special case (§3.2). I then show that Sober’s definition of testability meets only two of the conditions of adequacy (§4).

[1] Sober (2010, 1) states as much in an unpublished note.

[2] I will distinguish between use and mention and between concepts and their names only when this improves clarity or readability.

In light of these shortcomings of contrastive testability, I suggest to define that a theory makes observational assertions if and only if it is not empirically equivalent to a tautology, and to call a theory empirically significant if and only if it makes observational assertions. This criterion of empirical significance relies only on concepts that Sober accepts and fulfills all six conditions of adequacy (§5).

2 Contrastive testability

According to Sober (2008, 24–30), whole theories typically cannot be assigned probabilities. With the help of auxiliary assumptions A, however, a theory H can often at least assign a probability $\Pr(O \mid H \wedge A)$ to an observation O, called the likelihood of H for O.[3] This consideration leads him to a criterion of empirical significance that only relies on likelihoods (Sober 2008, 152). With slight modifications that avoid some problems (Lutz 2011, §3.1), the criterion is given by

Definition 1. Theory H_1 can be *tested against* theory H_2 if and only if there are auxiliary assumptions A and an observation sentence O such that

(I) $\Pr(O \mid H_1 \wedge A)$ and $\Pr(O \mid H_2 \wedge A)$ are defined,

(II) $\Pr(O \mid H_1 \wedge A) \neq \Pr(O \mid H_2 \wedge A)$,

(III) A is justified, and

(IV) the justification of A

a) does not depend on H_1 or H_2 being justified and
b) does not depend on O or $\neg O$ being justified.

The conditions (III) and (IV) are meant to restrict the auxiliary assumptions to those that are "suitable" for testability (Sober 2008, 144). However, the restrictions are so weak that given the assumptions that Sober (2008, 145) makes in the justification of restriction (IVb), his criterion falls prey to a trivialization proof (Lutz 2011, §4.2).

To bracket the problems with the restrictions on the auxiliary assumptions, I suggest

[3] I will follow Sober in treating observational claims and theories as single sentences, and the auxiliary assumptions as a finite set thereof. Finite sets of sentences are identified with the conjunctions of their members. Mostly (and always for the auxiliary assumptions), the restriction to single sentences—in effect a restriction to finite axiomatizations—is only a matter of notational convenience; I will note whenever it is essential. I will furthermore always silently assume that $\Pr(C) \neq 0$ for any occurring conditional probabilities $\Pr(B \mid C)$, and that for any theory H and auxiliary assumptions A, $\Pr(H \wedge A) \neq 0$, so that specifically $A \nvDash \neg H$.

Definition 2. Theory H_1 can be *tested against* theory H_2 *relative to* auxiliary assumptions A if and only if there exists an observation sentence O such that $\Pr(O \mid H_1 \wedge A)$ and $\Pr(O \mid H_2 \wedge A)$ are defined and

$$\Pr(O \mid H_1 \wedge A) \neq \Pr(O \mid H_2 \wedge A) . \tag{1}$$

Note that H_1 can be tested against H_2 according to definition 1 if and only if H_1 can be tested against H_2 relative to auxiliary assumptions that fulfill restrictions (III) and (IV). The probability terms can be taken to have sets of reals from the unit interval as their values. Some of these values (e. g., sets other than intervals) may be unacceptable for probabilities, so that one or the other term in the inequality is undefined. Then, taking Sober's use of probabilities into account, there are two possible interpretations of the inequality (Lutz 2011, §3.2):

1. When all sets of reals are acceptable, the inequality is true if and only if the two sets differ. Otherwise, it is false.

2. When some sets of reals are unacceptable, the inequality is true if and only if its two sides are defined and different, or one side is defined and the other one is not. Otherwise, the inequality is false.

Sober assumes that some probabilities are undefined, that is, interpretation 2.

To arrive at a criterion of empirical significance not defined relative to the auxiliary assumptions, I suggest to restrict the sets of auxiliary assumptions to honest sets:

Definition 3. A is an *honest set* if and only if every $S \in A$ is a justified sentence, and A also contains every sentence on which the justification of S depends.

Arguably, an honest set is a set of sentences that, for all we know, could have been our set of background assumptions: It describes which set of justified beliefs we hold now (if A is simply the set of all our currently justified beliefs), or which set we could have held before we got to hold all our current justified beliefs. The subjunctive here excludes false starts, that is, beliefs that at one point were justified but later became unjustified. This is plausible, as it is of little interest whether a theory is empirically significant under the assumption that the world is different from the way it in fact is (as far as we know). In this sense, an honest set is one step in the accumulation of our currently justified beliefs.

Definition 3 allows

Definition 4. Theory H_1 can be *tested against* theory H_2 if and only if H_1 can be tested against H_2 relative to an honest set of auxiliary assumptions.

This criterion avoids the trivialization proof for definition 1. To distinguish clearly between concepts that are defined relative to auxiliary assumptions (as in definition 2) and those that are not (as in definition 4), I will refer to the latter sometimes as *absolute* concepts.

Sober calls his criterion simply 'testability', but the qualifier 'contrastive' distinguishes it clearly from the ordinary language term and emphasizes that, atypically, the empirical significance of one theory is defined relative to another. It may seem problematic to explicate a one-place predicate like 'makes observational assertions' by a two-place predicate like contrastive testability. However, many successful explications involve a change of the logical structure, as the explication of 'warm' by 'warmer than' and finally 'temperature' illustrates (Carnap 1950, §4). But unlike 'warmer than', contrastive testability is symmetric: The definiens is invariant up to logical equivalence under exchange of H_1 and H_2. Thus contrastive testability does not provide a means to decide which of two theories is what could be called 'more empirically significant'. And this may be a problem. What is more, in some passages Sober himself uses 'testability' like a one-place predicate. For instance, he claims that 'Undetectable angels exist' is untestable and that 'This coin has probability of .5 of landing heads' is testable, which is, strictly speaking, meaningless for a two-place predicate like contrastive testability. And both claims are important for Sober's line of argument, since he relies on the first to argue that testability is different from meaningfulness, and on the second to argue that falsifiability is not an adequate criterion of empirical significance.

Thus, even Sober seems to rely tacitly on a concept of empirical significance that is not captured by contrastive testability. I will argue in the following that contrastive testability as given by definitions 2 and 4, while formally correct, does not capture what Sober intends. The definitions are materially inadequate.

3 Conditions of adequacy for a criterion of empirical significance

In this section, I will argue that Sober's assumptions and his intended application of the criterion lead to six conditions of adequacy. Many of these relate empirical significance to concepts that rely on inferences and therefore have a deductive and a probabilistic formulation. This is because deductive inference (entailment: $B \vDash C$) clearly does not generalize probabilistic inference (e. g., $\Pr(B) = 1$ and $\Pr(C \mid B) = q$, thus $\Pr(C) = q$). But probabilistic inference also does not generalize deductive inference. For assume that the domain has infinite cardinality. Then it may be that $\Pr(C \mid B) = 1$, but there are cases in which B is true and C is false. This happens, for example, when the domain is the interval $[0, 2]$ with a uniform probability distribution, B is '$x \leq 1$', and C is '$x < 1$' (cf. Feller 1971, 33f).

This difference between the deductive and the probabilistic concept of inference generally leads to differences between the deductive and probabilistic formulations of the conditions of adequacy, which in turn may lead to one criterion of deductive empirical significance (in the following sometimes shortened to 'deductive criterion') and a separate criterion of probabilistic empirical significance ('probabilistic criterion'). A theory may then be called empirically significant (simpliciter) if and only if it is deductively or probabilistically empirically significant.

3.1 Conditions of adequacy for a criterion of empirical significance

(A) The criterion should not be trivial. A trivial definition, one that includes all or no objects of the domain, cannot be a good explicatum for a concept that is meant to include some, but not all objects of the domain. At the very least, a trivial explicatum is uninformative. Since Sober intends to distinguish between theories that are worthy to be pursued and theories that are not, his criterion must not be trivial, and he implicitly relies on this condition of adequacy when arguing for his own definition (Sober 2008, 145; cf. Lutz 2011, §4.2).

(B) The criterion should include all and only theories that make observational assertions. Popper (1935, 85) justifies his criterion of falsifiability with the assumption that all and only theories that make observational assertions are empirically significant (cf. Lutz 2011, §2.1). That assumption itself is justified by Ayer (1936, 97), who argues that "the purpose" of an empirical theory is "to enable us to anticipate the course of our sensations". If Ayer's argument is sound, empirical significance is a necessary and sufficient condition for making observational assertions.

Sober (2008, 130) states that "a testable statement makes predictions, either by deductively entailing that an observation will occur or by conferring a probability on an observational outcome."[4] Thus for Sober empirical significance is a sufficient condition for making observational assertions. Let this be condition (i). Sober also subscribes to the converse of condition (i) as can be seen from his claims that "[t]he problem with the hypothesis of intelligent design is [. . .] that it doesn't predict much of anything" (Sober 2008, §2.15) and that his "criticism of the design argument might be summarized by saying that the design hypothesis is untestable" (Sober 2008, 148).[5] Since Sober (2008, §2.12) infers the lack of empirical significance from the lack of observational assertions, his criticism of ID *relies* only on condition (i). However, Sober's criticism of Popper's falsifiability criterion does seem to rest on the converse of condition (i) for probabilistic assertions: 'This coin has probability of .5 of landing heads each time it is tossed' makes a probabilistic assertion, and its lack of falsifiability is a reason for Sober to reject Popper's criterion. This seems to assume that every theory that makes probabilistic assertions is empirically significant.

[4]Since Sober does not use 'prediction' to refer exclusively to claims about the future, I will treat it as synonymous with 'assertion'.

[5]When Sober (1999, 54) states that "hypotheses rarely make observational predictions on their own; they require supplementation by auxiliary assumptions if they are to be tested", he similarly seems to be treating testability and the making of observations simply as synonyms.

Sober (2008, 52, n. 29) further states two relations between deductive empirical significance and the making of deductive observational assertions:

> If a true observation sentence entails H [. . .] you can conclude without further ado that H is true; this is just *modus ponens*. And if H entails O and O turns out to be false, you can conclude that H is false [. . .]; this is just *modus tollens*.

These are two sufficient conditions for empirical significance, namely (ii) entailment *by* an observation sentence and (iii) entailment *of* an observation sentence. Condition (iii) is the converse of condition (i) for deductive assertions (cf. Sober 1999, 72, n. 14). Therefore, according to Sober all and only theories that make observational assertions are empirically significant.

Condition (ii), however, is incompatible with condition (i): For any sentence S and observation sentence O, $O \vDash O \vee S$, that is, $O \vee S$ is empirically significant according to condition (ii). But let S be such that it does not make observational assertions, that is, for any observation sentence O', $S \nvDash O'$, and S does not confer any probability on O'. Then, as a matter of logic, $O \vee S \nvDash O'$, so $O \vee S$ does not make deductive assertions. $O \vee S$ also does not confer a probability on any observation sentence, since the inferences one can draw from $O \vee S$ are weaker then those that one can draw from S, and S already does not allow to assign a probability to any observation sentence. Thus $O \vee S$ does not make any observational assertions and is therefore not empirically significant according to condition (i), which is incompatible with condition (ii). On pain of inconsistency, Sober therefore has to choose whether all theories entailed by observation sentences are empirically significant or whether all theories that are empirically significant make observational assertions. Given that his core argument against ID is that ID fails to make assertions, I take it that he would choose the latter.[6]

(C) The criterion should exclude all theories that are empirically equivalent to tautologies. What it means for a theory to make probabilistic assertions may not be completely clear, especially since some probabilities may be undefined. It will therefore be convenient to also have a plausible corollary of condition (B), starting from the observation that a tautology $\top$ makes no deductive assertions, since $B \wedge \top \vDash C$ only if $B \vDash C$, and makes no probabilistic assertions either, since adding a tautology to any set of sentences does not change the probabilities that can be assigned to the other sentences in the set. Therefore, tautologies should be excluded by any criterion of empirically significance.

Flew (1950, 258) goes so far to call every theory a tautology that does not make observational assertions, but this is clearly to strong: 'Borogroves are mimsy' is not a tautol-

[6]Note that the claim "There is an intelligent designer" is equivalent to "There is a human designer or there is a non-human designer" and thus analytically entailed by an observation sentence like "There are humans who design". Arguably, however, "There is a non-human designer" does not make an observational assertion, so that "There is an intelligent designer" does not either.

ogy, but on account of containing two undefined terms, does not make any observational assertions. Rather, any theory that makes the same observational assertions as a tautology should be taken as not empirical significant. Unlike condition of adequacy (B), which relies on some criterion for the making of observational assertions, condition (C) relies on a criterion for empirical equivalence, the making of *the same* observational assertions. Out of caution, one may treat the empirical non-equivalence to a tautology as a necessary, but not as a sufficient condition for empirical significance.

(D) The criterion should not rely on the probabilities of whole theories or likelihoods of the negations of whole theories. Sober (2008, 24–30) argues that for many theories H the probabilities $\Pr(H\,|\,A)$, $\Pr(H\,|\,O \wedge A)$, and $\Pr(O\,|\,\neg H \wedge A)$ are undefined (cf. Sober 1990, §III). A criterion that relies on these probabilities would therefore be unusable in many cases.

(E) The criterion should be equivalent to an adequate Bayesian criterion of empirical significance whenever all occurring probabilities are defined. Since Bayesianism relies on probabilities of whole theories and likelihoods of negations of whole theories, Sober rejects it as a general method of scientific inference. Instead, Sober (2008, 37) suggests *likelihoodism*, which relies only on the likelihoods of theories, but notes (cf. Sober 2008, 32):

> The likelihoodist is happy to assign probabilities to hypotheses when the assignment of values to priors and likelihoods can be justified by appeal to empirical information. Likelihoodism emerges as a statistical philosophy distinct from Bayesianism only when this is not possible.

Since there are criteria of empirical significance that have been developed within Bayesianism, this suggests that a probabilistic criterion of empirical significance should be equivalent to one of these Bayesian criteria whenever all probabilities are defined. This Bayesian criterion should, of course, fulfill all criteria of adequacy other than (D).

(F) The probabilistic criterion should contain as a special case an adequate criterion of deductive empirical significance that relies only on *modus ponens*. Sober (2002, 69f) sees a smooth transition between probabilistic and deductive *modus ponens*. More specifically, Sober (2008, 50) points out the following:

$$\text{(Update)}\quad \begin{array}{l} \Pr_{\text{then}}(H\,|\,O) \text{ is very high} \\ O \\ O \text{ is all the evidence we have gathered between then and now.} \\ \hline \Pr_{\text{now}}(H) \text{ is very high} \end{array}$$

> This is nothing other than the rule of updating by strict conditionalization. (Update) is a sensible rule, and it also has the property of being a generalization of deductive modus ponens.

As argued at the beginning of this section, (Update) is not, strictly speaking, a generalization of *modus ponens*. But at least when all and only sentences with probability 1 are certain, deductive and probabilistic inference coincide. This can be put more precisely as follows. Each structure $\mathfrak{M}$ of a language $\mathscr{L}$ of predicate logic assigns a truth value to each sentence in $\mathscr{L}$. If $\Pr_{\mathfrak{M}}$ is defined as the function that assigns 1 to all sentences true in $\mathfrak{M}$ and 0 to all sentence false in $\mathfrak{M}$, then $\Pr_{\mathfrak{M}}$ is a probability assignment (see appendix, claim 9). Call such probability assignments *truth value-like*. For truth value-like probability assignments, probabilistic inferences and deductive inferences coincide: The possible values of $\Pr(C \,|\, B)$ are restricted to 0 and 1, and $\Pr(C \,|\, B) = 1$ if and only if $B \vDash C$ (as always assuming that $\Pr(B) \neq 0$; see appendix, claim 10). Truth value-like probabilities may be assigned by fiat, but they also occur more or less naturally when there are no regularities whatsoever, so that no probabilities can be assigned to sentences that are not known to be true and thus have probability 1 or known to be false and thus have probability 0.[7]

In this sense, then, there can be a smooth transition between probabilistic and deductive inference. Given that all and only theories that make deductive or probabilistic assertions must be empirically significant by condition of adequacy (B), there must then also be a smooth transition between any criterion of probabilistic empirical significance and a criterion of deductive empirical significance that uses the implications of the theory only in a *modus ponens*. As I will say, the probabilistic criterion must contain as a special case a deductive criterion that relies only on *modus ponens*. Of course, the deductive criterion should fulfill all those conditions of adequacy that also have purely deductive formulations, that is, conditions (A), (B), and (C). To fulfill condition (B), it is enough for the deductive criterion to include all and only theories that make deductive assertions, because it is impossible that it could include theories that make only probabilistic assertions. Analogously, it is enough if the criterion excludes all theories that are deductively empirically equivalent to a tautology to meet condition (C).

Independently of any smooth transition in the case of *modus ponens*, it is clear that the criterion of empirical significance simpliciter should be a generalization of an adequate deductive criterion. Thus, when deductive and probabilistic inference coincide, the probabilistic criterion must not include theories that the deductive criterion excludes. For if it did, these theories would be included by the criterion of empirical significance simpliciter, and thus this criterion would not generalize the deductive criterion, but rather contradict it.

* * *

[7] This is arguably the case in Popper's approach to induction (cf. Salmon 1967, §II.3).

While I have tried to mention supporting arguments for these conditions of adequacy when possible, some of them remain controversial; especially condition (D) would be challenged by Bayesians. But these conditions all follow from Sober's basic assumptions or apply to Sober's criterion because of its intended application. Of course, it may be that these conditions of adequacy are incompatible, so that some have to be given up. This is the case for conditions (i) and (ii) discussed under condition of adequacy (B). But a criterion of empirical significance that is to be applied as Sober intends should fulfill as many of these conditions as possible.

3.2 Falsifiability is the unique adequate criterion of deductive empirical significance

Condition of adequacy (F) demands that the deductive criterion contained in the probabilistic criterion can be phrased in terms of *modus ponens* and fulfills all those conditions of adequacy that pertain to deductive criteria. In this section, I want to show that these conditions uniquely determine falsifiability.

Definition 5. A theory H is *falsifiable relative to* auxiliary assumptions A if and only if there is an observation sentence O with $A \nvDash \neg O$ such that $O \wedge A \vDash \neg H$.[8]

This definition can be combined with definition 3:

Definition 6. A theory H is *falsifiable* if and only if it is falsifiable relative to an honest set of auxiliary assumptions.

Sober (1999, 48–57) defends many of the assumptions on which falsifiability depends against criticisms (see also Lutz 2010, §2). As for the conditions of adequacy:

To show that a criterion is not trivial and thus fulfills condition (A), it is enough to give a positive and a negative instance of the criterion. If S is the negation of an observation sentence, then S is falsifiable relative to $\varnothing$. If S is a first order sentence without identity and none of the terms in S occur in observation sentences, then S is not falsifiable relative to $\varnothing$, assuming that all observation sentences are of first order.[9] Thus definition 5 is not trivial. If S is the negation of an observation sentence, then S is also absolutely falsifiable according to definition 6, for it is falsifiable relative to $\varnothing$, which is an honest set according to definition 3. Thus there is a falsifiable sentence. I will not attempt to prove that there is a non-falsifiable sentence, because this would amount to finding a sentence that is not falsifiable relative to any honest set. The proof is immediate for tautologies, but impossible for contingent theories without more precise notions of justification and dependence.

[8] To allow sets of sentences and higher order logic, the definition must be phrased as "A theory H is falsifiable relative to assumptions A if and only if there is a set Ω of observation sentences with $\Omega \cup A \nvDash \bot$ such that $\Omega \cup H \cup A \vDash \bot$", where A and H are sets of sentences and $\bot$ is some contradiction.

[9] This follows from Craig's interpolation theorem (Hodges 1993, theorem 6.6.3).

That falsifiability includes all theories that make deductive assertions is already pointed out by Sober (1999, n. 14), who remarks that if "$H \wedge A$ deductively entails O, and A is known to be true, then, if we observe not-O, we can conclude that H is false." Note that Sober here assumes that the negation of an observation sentence is again an observation sentence. Under this assumption, condition (B) uniquely determines the criterion of relative falsifiability, if it is further assumed that a theory makes deductive assertions if and only if it entails observation sentences not entailed by the auxiliary assumptions alone:

Definition 7. A theory H *makes deductive observational assertions relative to* assumptions A if and only if there is an observation sentence O such that $H \wedge A \vDash O$ and $A \nvDash O$.[10]

Definition 7 is equivalent to Sober's definition of 'having observational implications' (Sober 2008, 151) except that the problem of suitable auxiliary assumptions is bracketed. Sober uses restrictions (III) and (IV) from definition 1 to determine the auxiliary assumptions, which, however, allows a trivialization proof (Lutz 2010, §9.2).

Definition 7 leads to

Claim 1. *If the negation of an observation sentence is again an observation sentence, then a theory H is falsifiable relative to A if and only if it makes deductive observational assertions relative to A.*

Proof. If O, A, and H are sentences, the proof is immediate. If O, A, and H are sets, the claim follows immediately from claim 11 (see appendix). □

The condition on observation sentences is not only implicitly assumed by Sober, but follows also from the most common restrictions on observation sentences (Lutz 2011, §4.2). Therefore relative falsifiability arguably meets condition of adequacy (B).

Claim 1 also establishes that absolute falsifiability meets condition (B) if and only if it is the case that a theory makes deductive assertions iff it makes deductive assertions relative to an honest set. But even if the definition of an honest set turns out to be wanting in some respect, there is no obvious reason to doubt that the auxiliary assumptions suitable for falsifiability are also suitable for the making of assertions. Rather, since background assumptions are usually considered to be independent from the concepts that rely on them, this is a fairly plausible conjecture. Under this conjecture, all equivalence results between relative concepts transfer to absolute concepts, and it will be silently assumed in

[10] To allow sets of sentences and higher order logic, the definition must be phrased as "A theory H makes deductive observational assertions relative to assumptions A if and only if there are a set Ω of observation sentences and an observation sentence O such that $\Omega \cup H \cup A \vDash O$ and $\Omega \cup A \nvDash O$". If a logic is compact, $\Omega \cup H \cup A \vDash O$ if and only if there is a finite set Ω' such that $\Omega' \cup H \cup A \vDash O$, which is equivalent to $H \cup A \vDash \bigwedge \Omega' \rightarrow \neg O$. Hence in first order logic this definition reduces to definition 7 if the set of observation sentences is closed under truth-functional composition.

the following. Claim 1 also establishes that both falsifiability and relative falsifiability can be phrased so that they rely only on *modus ponens*.

Condition of adequacy (C) is fulfilled because no theories that are deductively empirically equivalent to a tautology $\top$ are falsifiable:

Definition 8. Two theories H_1 and H_2 are *deductively empirically equivalent relative to* auxiliary assumptions A if and only if for all observation sentence O, $H_1 \wedge A \vDash O$ iff $H_2 \wedge A \vDash O$.[11]

Claim 2. *If the negation of an observation sentence is again an observation sentence, H is falsifiable relative to A if and only if H and $\top$ are not deductively empirically equivalent relative to A.*

Proof. H is not deductively empirically equivalent to $\top$ if and only if there is an observation sentence O such that either $H \wedge A \vDash O$ and $A \nvDash O$ or $H \wedge A \nvDash O$ and $A \vDash O$. Since the latter disjunct is logically impossible, this is equivalent to H making deductive observational assertions relative to A. Since the negation of an observation sentence is assumed to be observational, this is equivalent to H being falsifiable relative to A by claim 1.[12] □

Only falsifiability and equivalent criteria fulfill condition (B) in the deductive case, and so it is good news that falsifiability can be phrased in terms of *modus ponens* and fulfills all other conditions of adequacy that pertain to criteria of deductive empirical significance. To meet condition of adequacy (F), any criterion of relative probabilistic empirical significance must therefore contain relative falsifiability as formulated in definition 7 as a special case. Then the corresponding absolute criterion also contains absolute falsifiability as a special case.

4 Contrastive testability and the conditions of adequacy

Condition of adequacy (A) is that a criterion of empirical significance must not be trivial, and definition 1 does not meet this condition. Definition 4 does, however: Choose $A = \varnothing$, two non-observational, non-equivalent sentences S and S', and, for some observation sentence O, $H_1 \Dashv\vDash S \wedge \Pr(O) = p$ and $H_2 \Dashv\vDash S' \wedge \Pr(O) = q$ for some probabilities p and q. Then H_1 and H_2 are never equivalent, and H_1 can be contrastively tested against H_2 if and only if $p \neq q$, so that many theories can and many theories cannot be tested against each other relative to A. Since $\varnothing$ is an honest set, H_1 and H_2 can also be absolutely tested

[11] To allow sets of sentences and higher order logic, the definition must be phrased as "Two theories H_1 and H_2 are deductively empirically equivalent relative to assumptions A if and only if for every set Ω of observation sentences and every observation sentence O, $\Omega \cup H_1 \cup A \vDash O$ if and only if $\Omega \cup H_2 \cup A \vDash O$".

[12] The proof for sets of sentences is analogous, except for an additional existential quantification over sets of observation sentences.

against each other. Similarly to the case of absolute falsifiability, it is impossible to prove that there are non-equivalent theories that cannot be tested against each other relative to any honest set without more precise notions of justification and dependence.

Though non-trivial, contrastive testability fails to meet the two most important conditions of adequacy, conditions (B) and (C). That some theories that do not make probabilistic assertions and are probabilistically empirically equivalent to tautologies are contrastively testable can be inferred from an example that Sober (1999, n. 24) attributes, in a different context, to Greg Mougin:[13]

> Let H_1 = God created the eye, E = Jones is pregnant, A = Jones is sexually active, and H_2 = Jones used birth control. It is possible to test H_1 against H_2; given independently attested background assumptions A, E favors H_1 over H_2.

In the example, the observation sentence E is assigned one probability by the background assumptions alone (since H_1 is not about Jones at all), and another by the conjunction of the background assumptions and H_2. Now choose $H_1 \mathrel{\models\!\!\!\dashv} \top$. Then H_1 does not make any assertions and hence no observable ones, and it has trivially as much empirical content as a tautology. But H_1 can still be contrastively tested against H_2, both relative to A and absolutely, since the justification of A does not depend on E, H_1, or H_2.

By design, contrastive testability does not rely on prior probabilities or the likelihoods of the negation of theories and thus meets condition of adequacy (D). Contrastive testability fails to meet condition (E) simply because so far, no Bayesian criterion of empirical significance has been suggested that is equivalent to contrastive testability when all occurring probabilities are defined. Specifically, relative contrastive testability is not equivalent to the typical Bayesian criterion of empirical significance (cf. Sober 2008, 150) given by

Definition 9. *Observations are relevant for* theory H *relative to* auxiliary assumptions A if and only if there is an observation sentence O such that

$$\Pr(H \,|\, O \wedge A) \neq \Pr(H \,|\, A) \,. \tag{2}$$

That the two concepts are not equivalent is clear from their logical structures.

In principle, a probabilistic two-place predicate may contain a deductive one-place predicate as a special case. For example, if the probability assignments are truth value-like, '$\Pr(O \,|\, H_1) = .5 \vee \Pr(O \,|\, H_2) = 1$' is equivalent to '$H_2 \vDash O$' because the first argument, H_1, becomes irrelevant. In this sense, the two-place predicate of contrastive testability could therefore contain the one-place predicate of falsifiability as a special case. But since contrastive testability is symmetric, either both or neither of its two arguments are irrelevant for truth value-like assignments and thus it cannot meet condition (F).

[13] Unlike in the example by Salmon (1971, 29–88), it is this time not John Jones who is using birth control, but his wife.

H_1 can be tested against H_2 if and only if their defined likelihoods differ for at least one observation sentence. If only singleton sets of probabilities are acceptable (as Sober assumes for his criterion), this means that at least with respect to one observation, one of the two theories must be wrong. Arguably, then, contrastive testability explicates what it means for two theories to be probabilistically empirically incompatible for the special case that only singleton sets of likelihoods are acceptable. This is borne out by the comparison with

Definition 10. Theories H_1 and H_2 are deductively empirically incompatible relative to auxiliary assumptions A if and only if there there is an observation sentence O such that $H_1 \wedge A \vDash O$ and $H_2 \wedge A \vDash \neg O$.

Claim 3. *Let H_1 and H_2 be deductive theories, let all probability assignments be truth value-like, and let the set $\{0,1\}$ be unacceptable as a value of a likelihood. Then H_1 can be tested against H_2 relative to A if and only if H_1 and H_2 are deductively empirically incompatible relative to A.*

Proof. H_1 can be tested against H_2 if and only if there is an observation O such that the likelihood of one theory for O is 0, while the other one is 1. Without loss of generality, assume $\Pr(O \mid H_1 \wedge A) = 1$ and $\Pr(O \mid H_2 \wedge A) = 0$, that is, $\Pr(\neg O \mid H_2 \wedge A) = 1$. By claim 10 (see appendix), this holds if and only if $H_1 \wedge A \vDash O$ and $H_2 \wedge A \vDash \neg O$. □

Thus, if only singleton sets are acceptable as values of likelihoods, then contrastive testability contains as a special case a criterion for deductive theories that relies only on *modus ponens*. It is only the wrong one.

5 Explicating probabilistic empirical significance

Contrastive testability does not meet all criteria of adequacy, but that might just be because the criteria cannot all be met at once. I will argue that this is not so by suggesting a criterion of empirical significance that does meet all the conditions. First, however, I want to discuss an intuitively attractive but inadequate criterion.

One may think of defining that a theory is not empirically significant if and only if it cannot be tested against *any* theory. But this definition is inordinately inclusive. For assume that H_1 is not such that all assertions from suitable auxiliary assumptions become undefined, that is, $\Pr(O \mid A)$ and $\Pr(O \mid H_1 \wedge A)$ are defined (though possibly identical) for some O and some suitable A. Then H_1 is empirically significant if there is any H_2 such that $\Pr(O \mid H_2 \wedge A)$ is defined and different from $\Pr(O \mid A)$. For if $\Pr(O \mid A) = \Pr(O \mid H_1 \wedge A)$, H_1 can be tested against H_2, and if $\Pr(O \mid A) \neq \Pr(O \mid H_1 \wedge A)$, H_1 can be tested against any tautology. The premises of this argument are commonly true, for example, according to Sober, if H_1 is 'God created the eye', O is 'Jones is pregnant', and A is 'Jones is sexually active', for then H_2 can be 'Jones used birth control'. Choosing $H_1 \boxminus \top$, the argument shows that tautologies are empirically significant, which runs afoul of conditions of adequacy (B) and (C). It is also straightforward to show that conditions (E) and (F) are not met.

5.1 Probabilistic empirical equivalence

A more promising path to a criterion of probabilistic significance leads through the criterion of probabilistic empirically equivalence. To show that contrastive testability does not fulfill condition (C) it was sufficient to produce one contrastively testable theory that is probabilistically empirically equivalent to a tautology. To arrive at a criterion of empirical significance that provably fulfills (C), however, it is necessary to define probabilistic empirical equivalence.

Luckily, it is possible to explicate condition of adequacy (C) in line with Sober's position, for he states that "empirically equivalent theories have identical likelihoods" for any observation sentence O (Sober 1990, 399). Treating the case of undefined likelihoods explicitly, this leads directly to

Definition 11. Theories H_1 and H_2 are *probabilistically empirically equivalent relative to* auxiliary assumptions A if and only if for all observation sentences O,

(I) $\Pr(O \mid H_1 \wedge A)$ and $\Pr(O \mid H_2 \wedge A)$ are not defined or

(II) $\Pr(O \mid H_1 \wedge A)$ and $\Pr(O \mid H_2 \wedge A)$ are defined and $\Pr(O \mid H_1 \wedge A) = \Pr(O \mid H_2 \wedge A)$.

As defined, probabilistic empirical equivalence contains deductive empirical equivalence as a special case:

Claim 4. *Let H_1 and H_2 be deductive theories and let all probability assignments be truth value-like. Then H_1 and H_2 are probabilistically empirically equivalent relative to A if and only if H_1 and H_2 are deductively empirically equivalent relative to A.*

Proof. If all probability assignments are truth value-like, then interpretation 1 and interpretation 2 of inequality (1) are equivalent, independently of whether $\{0, 1\}$ is an acceptable set of probabilities. For if $\{0, 1\}$ is an acceptable set, the interpretations are trivially equivalent; if $\{0, 1\}$ is not acceptable, the inequality is false if and only if both likelihoods have the value $\{0\}$, $\{1\}$, or $\{0, 1\}$/undefined. Otherwise, the inequality is true.

Therefore, it suffices to prove the claim for interpretation 1. H_1 is probabilistically empirically equivalent to H_2 relative to A if and only if for all observation sentence O, $H_1 \wedge A$ restricts the probability to the same set of values as $H_2 \wedge A$. Since for any H, $\Pr(S \mid H \wedge A) = 0$ if and only if $\Pr(\neg S \mid H \wedge A) = 1$, this is the case if and only if for every sentence, $H_1 \wedge A$ restricts the probability to 1 iff $H_2 \wedge A$ does. By claim 10 (see appendix), this holds if and only if $H_1 \wedge A$ and $H_2 \wedge A$ entail the same observation sentences, that is, if H_1 and H_2 are deductively empirically equivalent relative to A. $\square$

Another reason to consider definition 11 a good explication of probabilistic empirical equivalence is that, if all occurring probabilities are defined, it bears the same relation to the Bayesian criterion of empirical significance given in definition 9 as the criterion of deductive empirical equivalence bears to falsifiability: If two theories are deductively empirically equivalent, then either both or neither are deductively empirically significant (see appendix, claim 12). Analogously, the following holds:

Claim 5. *If all occurring probabilities are defined and H_1 is probabilistically empirically equivalent to H_2 relative to auxiliary assumptions A, then, relative to A, observations are relevant for H_1 if and only if observations are relevant for H_2.*

Proof. If O is any observation sentence for which $\Pr(O \,|\, H_1 \wedge A) = \Pr(O \,|\, H_2 \wedge A)$, then $\Pr(O \,|\, H_1 \wedge A) = \Pr(O \,|\, A)$ if and only if $\Pr(O \,|\, H_2 \wedge A) = \Pr(O \,|\, A)$. Therefore $\Pr(H_1 \,|\, A) = \Pr(H_2 \,|\, O \wedge A)$ if and only if $\Pr(H_2 \,|\, A) = \Pr(H_2 \,|\, O \wedge A)$ (see appendix, claim 13). Thus observations are relevant for both H_1 and H_2 or for neither. □

5.2 Probabilistic observational assertions

It follows from claims 1 and 2 that a theory makes deductive observational assertions if and only if it is not deductively empirically equivalent to a tautology. This suggests that conditions of adequacy (B) and (C) are in fact equivalent, so that a theory makes probabilistic observational assertions relative to auxiliary assumptions A if and only if it is not probabilistically empirically equivalent to a tautology relative to A according to definition 2. This leads to

Definition 12. Theory H *makes probabilistic observational assertions relative to* auxiliary assumptions A if and only if there exists an observation sentence O such that

(I) $\Pr(O \,|\, H \wedge A)$ is defined if and only if $\Pr(O \,|\, A)$ is not defined or

(II) $\Pr(O \,|\, H \wedge A)$ and $\Pr(O \,|\, A)$ are defined and $\Pr(O \,|\, H \wedge A) \neq \Pr(O \,|\, A)$.

Claim 6. *H makes probabilistic assertions relative to A according to definition 12 if and only if H is not probabilistically empirically equivalent to a tautology relative to A according to definition 11.*

Proof. Let $\ulcorner Ox \urcorner$ stand for $\ulcorner x$ is an observation sentence$\urcorner$, $\ulcorner Dxy \urcorner$ stand for $\ulcorner \Pr(x \,|\, y \wedge A)$ is defined$\urcorner$, and $\ulcorner Exyz \urcorner$ for $\ulcorner \Pr(x \,|\, y \wedge A) = \Pr(x \,|\, z \wedge A) \urcorner$. Then it is straightforward to prove that

$$\exists x \{Ox \wedge [(Dxy \longleftrightarrow \neg Dxz) \vee (Dxy \wedge Dxz \wedge \neg Exyz)]\} \\ \mathrel{\models\!\!\!\!\dashv} \neg \forall x \{Ox \rightarrow [(\neg Dxy \wedge \neg Dxz) \vee (Dxy \wedge Dxz \wedge Exyz)]\} \quad (3)$$

Since $\Pr(O \,|\, \top \wedge A) = \Pr(O \,|\, A)$, the claim follows. □

With the help of definition 3 of an honest set, one can give

Definition 13. Theory H *makes probabilistic observational assertions* if and only if H makes probabilistic observational assertions relative to an honest set A of auxiliary assumptions.

It may be considered problematic that a theory H makes assertions if assuming H makes it impossible to assign a probability to an observation sentence that otherwise would be assigned a probability by the auxiliary assumptions. To render such theories empirically non-significant, the biconditional in condition (I) of definition 12 could be made into a conjunction. However, it is very plausible that for a theory H that makes no predictions, $\Pr(O \mid H \wedge A)$ is defined whenever $\Pr(O \mid A)$ is defined. For if $\Pr(O \mid A)$ is defined, then, given A, one must expect a specific regularity of occurrences of O. But if $\Pr(O \mid H \wedge A)$ is undefined, one must expect a breakdown of this regularity, and this expectation is plausibly a prediction. An example would be the prediction that under specific circumstances, some law fails that was assumed to hold universally. It certainly is pragmatically relevant when some observation sentence can, contrary to the auxiliary assumptions, not be assigned a probability.

Definitions 12 and 13 rely only on concepts that Sober uses himself, and should therefore be conceptually unproblematic for him. He does not discuss the domains of applicability of the concepts, but with one exception, the domains can just be assumed to be the same for definitions 12 and 13 as they are for contrastive testability. The exception is the term $\Pr(O \mid A)$. Sober could argue that the concept of a likelihood cannot be applied to tautologies because the auxiliary assumptions themselves assign probabilities to no or too few observation sentences. Sober (2008, 29f) in fact shortly discusses $\Pr(O)$, but not in connection with auxiliary assumptions. The discussion therefore does clearly not apply to definition 12, and it does not apply to definition 13 because H is testable if there are *some*, not necessarily tautological, suitable auxiliary assumptions A such that $\Pr(O \mid H \wedge A)$ differs from $\Pr(O \mid A)$. Sober's original definition 1 and definition 3 of an honest set put no restrictions on individual elements of A except that they be justified (in Sober's definition, independently of a specific observation sentence and two specific theories). Therefore whole theories can be included in the auxiliary assumptions. Since Sober introduces auxiliary assumptions to allow for actual scientific practice, and assertions made by scientific theories in fact often rely on other scientific theories as auxiliary assumptions, such an inclusion obeys letter and spirit of Sober's criterion. Since scientific theories H are supposed to make observational assertions, $\Pr(O \mid H \wedge A)$ will often be defined. And the inclusion of H into the auxiliary assumptions is then just the notational change to $\Pr(O \mid A^*)$ with $A^* \mathrel{\vDash\!\!\!\Dashv} H \wedge A$.

It now follows from condition of adequacy (B) that all and only theories that fulfill definition 12 (13) are probabilistically empirically significant (relative to auxiliary assumptions A). Without any basic conceptual problems, a theory can then be defined to be empirically significant (relative to A) if and only if it it makes observational assertions (relative to A), that is, if and only if it makes probabilistic observational assertions (relative to A) or it makes deductive observational assertions (relative to A).

As argued in §3.2, falsifiability fulfills all appropriate conditions of adequacy. I now want to show that this new definition of probabilistic empirical significance does, too. That it is non-trivial and thus meets condition of adequacy (A) is easily shown since sentences without any terms that occur in observation sentences are not testable relative to $\varnothing$. And in the example with Jones's pregnancy, the theory that Jones uses birth control (H_2) has a different likelihood in conjunction with the auxiliary assumption A that Jones is sexually active than A alone, and therefore H_2 is testable relative to A. This example also shows that there are positive instances of absolute testability. As in the case of absolute falsifiability, it is impossible to prove that there is a non-tautological theory that makes no probabilistic predictions relative to any honest set without more precise notions of justification and dependence.

That analogues of restrictions (III) and (IV) of definition 1 cannot be substituted for the restriction to honest sets can be shown as follows:[14] Let the observation sentence O and the auxiliary assumptions A be such that $\Pr(O\,|\,A)$ is defined. For instance, A might express that 1 out of 10 vases of some kind breaks when dropped from a specific height, and O express that the vase does not break on some specific drop. Let furthermore S be such that it entails O. For instance, S might express that that specific vase does not break when dropped a hundred times from a that height. Then S is justified independently of O when the vase has been dropped 99 times without breaking, so that O, S, and A fulfill conditions (III) and (IV) for any theory that is not related to vases. For a theory that makes assertions about vases, it should not be difficult to find analogous sentences O', S', and A' about something else. Since S and A are justified independently of O, $\neg O$, and H, so is $A^* \mathrel{\rightleftharpoons} (\neg H \vee S) \wedge A$. Since $\Pr(O\,|\,A^*) \neq \Pr(O\,|\,H \wedge A^*)$, H would make predictions if restrictions (III) and (IV) of definition 1 were substituted for the restriction to honest sets.

Luckily, definition 13 is stricter than it would be with analogues of restrictions (III) and (IV). The restriction of the auxiliary assumptions in definition 4 to honest sets entails restriction (III). And while the restriction to honest sets does not entail restriction (IVb), it precludes all trivializations precluded by that restriction: A theory H fails to make probabilistic assertions because of restriction (IVb) only if for any S whose inclusion in A would lead to $\Pr(O\,|\,H \wedge A) \neq \Pr(O\,|\,A)$ for some O, the justification of S depends on O or $\neg O$. In that case, (IVb) ensures that H makes no probabilistic assertion. The restriction of A to honest sets leads to the same result, because if the justification of S depends on O (or $\neg O$) and A is honest, then $O \in A$ (or $\neg O \in A$). Thus $\mathrm{P}(O\,|\,H \wedge A) = 1 = \mathrm{P}(O\,|\,A)$ (or $\mathrm{P}(O\,|\,H \wedge A) = 0 = \mathrm{P}(O\,|\,A)$). Unlike restriction (IVb), the restriction to honest sets also leads to identical likelihoods if $(\neg H \vee S) \in A$ is justified by a sentence $S \vDash O$, thereby precluding the trivialization proof in the previous paragraph. The restriction to honest sets

[14] The analogues of the restrictions feature only one theory H rather than two theories H_1 and H_2.

also precludes any trivialization that restriction (IVa) could preclude, for if an element of A depends on H, $H \in A$, so that $\Pr(O \mid H \wedge A) = \Pr(O \mid A)$ for all O.

Definitions 12 and 13 trivially fulfill condition of adequacy (B). Because of claim 6, they meet condition (C) as well. Since condition (C) only states that empirical equivalence to a tautology is a sufficient condition for empirically non-significance, (C) is also met if the biconditional of condition (I) in definition 12 is substituted by a conjunction, so that more theories are empirically non-significant.

By design, definitions 12 and 13 fulfill condition of adequacy (D). Condition of adequacy (E) is met because of

Claim 7. *If all occurring probabilities are defined, then H makes probabilistic observational assertions if and only if observations are relevant for H.*

Proof. For all observations O, $\Pr(O \mid H \wedge A) \neq \Pr(O \mid A)$ if and only if $\Pr(H \mid O \wedge A) \neq \Pr(H \mid A)$ (see appendix, claim 13). Therefore there is an O such that $\Pr(O \mid H \wedge A) \neq \Pr(O \mid A)$ if and only if there is an O such that $\Pr(H \mid O \wedge A) \neq \Pr(H \mid A)$. □

Definition 12 fulfills condition of adequacy (F), because it generalizes falsifiability, or rather definition 7:

Claim 8. *Let H be a deductive theory and let all probability assignments be truth value-like. Then H makes probabilistic observational assertions relative to A if and only if H makes deductive observational assertions relative to A.*

Proof. Since interpretation 1 and interpretation 2 of the inequality in condition (II) are equivalent, it suffices to prove the claim for interpretation 1. H does not make probabilistic assertions relative to A if and only if for every observation sentence O, $H \wedge A$ restricts the probability to the same set of values as A. This is the case if and only if for every observation sentence, $H \wedge A$ restricts the probability to 1 iff A does. By claim 10 (see appendix), this holds if and only if $H \wedge A$ and A entail the same observation sentences, that is, if and only if H makes no deductive observational assertions relative to A. □

Note that the proof also holds if the biconditional of condition (I) in definition 12 is substituted by a conjunction, because then H makes no observational assertions if and only if for all O, $H \wedge A$ restricts the probabilities to the same set of values as A, or A restricts the probabilities more than $H \wedge A$. But the latter is impossible since A restricts the probabilities of an observation sentence to $\{0\}$ or $\{1\}$ only if $H \wedge A$ does. If the suitable auxiliary assumptions for falsifiability are given by honest sets, condition (F) is also fulfilled by definition 13.

Therefore, definition 12 and arguably definition 13 fulfill all conditions of adequacy that Sober wants a criterion of empirical significance to meet. Additionally, they also make it possible to evaluate one theory, ID for example, independent of another one like ET.

Finally, the claims in which Sober uses testability as a one-place predicate are not only meaningful, but also correct: 'Undetectable angels exist' arguably makes no observational assertions relative to any honest set of sentences, and 'This coin has a probability of .5 of landing heads each time it is tossed' makes observational assertions, for it assigns a probability of .5 to an observation sentence relative to $\varnothing$.

6 Conclusion

Contrastive testability fails as a criterion of probabilistic empirical significance because it fails to meet four criteria of adequacy that follow from Sober's position and the intended application of contrastive testability: It includes some theories that make no observational assertions, and some theories that are empirically equivalent to tautologies. It is not equivalent to a Bayesian criterion of empirical significance when all probabilities are defined, and it does not contain falsifiability as a special case. This last property is important because a criterion of probabilistic empirical significance should contain *some* adequate criterion of deductive empirical significance as a special case, and falsifiability is the only adequate criterion.

Given that contrastive testability is not an adequate criterion of empirical significance, I have suggested to consider a theory empirically significant if and only if it does not make observational assertions. This definition fulfills all six conditions of adequacy, and in particular contains both falsifiability and the Bayesian criterion of empirical significance as special cases. The criterion could be called a synthesis, as it is acceptable for falsificationists, Bayesianists, and likelihoodists alike. The hope is that it will also lead to agreement on the empirical significance of string theory, scientific realism, theism, and ID.

A Additional proofs

Claim 9. *For every language $\mathscr{L}$ and every $\mathfrak{M}$,* $\mathrm{Pr}_{\mathfrak{M}} : \mathfrak{P}\mathscr{L} \to \{0,1\}, \mathrm{Pr}_{\mathfrak{M}}(\Sigma) = 1 \Leftrightarrow \mathfrak{M} \vDash \Sigma$ *is a probability assignment.*

Proof. Show that for all $\Sigma, \Xi \in \mathscr{L}$ and any $\mathfrak{M}$ it holds:

1. $\mathrm{Pr}_{\mathfrak{M}}(\Sigma) \geq 0$,
2. $\mathrm{Pr}_{\mathfrak{M}}(\{\top\}) = 1$, and
3. if Σ and Ξ are finite and $\Sigma \cup \Xi \vDash \bot$, $\mathrm{Pr}_{\mathfrak{M}}(\{\bigwedge \Sigma \vee \bigwedge \Xi\}) = \mathrm{Pr}_{\mathfrak{M}}(\Sigma) + \mathrm{Pr}_{\mathfrak{M}}(\Xi)$.

1 and 2 are immediate. 3 holds because for $\Sigma \cup \Xi \vDash \bot$, $\mathfrak{M} \nvDash \Sigma$ or $\mathfrak{M} \nvDash \Xi$, so that $\mathrm{Pr}_{\mathfrak{M}}(\bigwedge \Sigma \vee \bigwedge \Xi) = 1$ if and only if either $\mathfrak{M} \vDash \Sigma$ or $\mathfrak{M} \vDash \Xi$ but not both, which holds if and only if $\mathrm{Pr}_{\mathfrak{M}}(\Sigma) = 1$ or $\mathrm{Pr}_{\mathfrak{M}}(\Xi) = 1$ but not both, that is, $\mathrm{Pr}_{\mathfrak{M}}(\Sigma) + \mathrm{Pr}_{\mathfrak{M}}(\Xi) = 1$ □

Claim 10. *For any sets $\Sigma, \Xi \subseteq \mathscr{L}$ of sentences, $\Sigma \vDash \Xi$ if and only if for all $\mathfrak{M}$ it holds: If* $\mathrm{Pr}_{\mathfrak{M}}(\Sigma) \neq 0$ *then* $\mathrm{Pr}_{\mathfrak{M}}(\Xi \mid \Sigma) = 1$.

Proof.

$$\Sigma \vDash \Xi \Leftrightarrow \forall \mathfrak{M}\big[\mathfrak{M} \vDash \Sigma \Rightarrow \mathfrak{M} \vDash \Xi\big] \tag{4}$$

$$\Leftrightarrow \forall \mathfrak{M}\big[\mathrm{Pr}_{\mathfrak{M}}(\Sigma) = 1 \Rightarrow \mathrm{Pr}_{\mathfrak{M}}(\Xi) = 1\big] \tag{5}$$

$$\Leftrightarrow \forall \mathfrak{M}\left[\mathrm{Pr}_{\mathfrak{M}}(\Sigma) \neq 0 \Rightarrow \mathrm{Pr}_{\mathfrak{M}}(\Xi \mid \Sigma) = \frac{\mathrm{Pr}_{\mathfrak{M}}(\Xi)}{\mathrm{Pr}_{\mathfrak{M}}(\Xi \cup \Sigma)} = 1\right] \tag{6}$$

□

Claim 11. *If $H \cup A \nvDash \bot$ and the negation of an observational sentence is observational, the following holds: There is a set Ω of observation sentences such that $\Omega \cup A \nvDash \bot$ and $\Omega \cup H \cup A \vDash \bot$ if and only if there are a set Ω of observation sentences and an observation sentence O such that $\Omega \cup H \cup A \vDash O$ and $\Omega \cup A \nvDash O$.*

Proof. '$\Rightarrow$': If $\Omega \cup H \cup A \vDash \bot$, then $\Omega \cup H \cup A \vDash O$ for any observation sentence O. Since $\Omega \cup A \nvDash \bot$, there is some O such that $\Omega \cup A \nvDash O$.

'$\Rightarrow$': For O and Ω with $\Omega \cup H \cup A \vDash O$ and $\Omega \cup A \nvDash O$, $\Omega \cup \{\neg O\} \cup A \nvDash \bot$ and $\Omega \cup \{\neg O\} \cup H \cup A \vDash \bot$. □

Claim 12. *If the negation of an observation sentence is again an observation sentence, and H_1 is deductively empirically equivalent to H_2 relative to A, then, relative to A, H_1 is falsifiable if and only if H_2 is falsifiable.*

Proof. Assume that for all observation sentences O and sets of observation sentence Ω, $H_1 \cup \Omega \cup A \vDash O$ if and only if $H_2 \cup \Omega \cup A \vDash O$. Then, for all Ω and O, $H_1 \cup \Omega \cup A \vDash O$ and $\Omega \cup A \nvDash O$ if and only if $H_1 \cup \Omega \cup A \vDash O$ and $\Omega \cup A \nvDash O$. Thus there are Ω and O such that $H_1 \cup \Omega \cup A \vDash O$ and $\Omega \cup A \nvDash O$ if and only if there are Ω and O such that $H_1 \cup \Omega \cup A \vDash O$ and $\Omega \cup A \nvDash O$. By claim 1, this means that H_1 is falsifiable relative to A if and only if H_2 is falsifiable relative to A. □

Claim 13. *If* $\Pr(H \mid A)$ *is defined, then* $\Pr(H \mid O \wedge A) = \Pr(H \mid A)$ *if and only if* $\Pr(O \mid H \wedge A) = \Pr(O \mid A)$.

Proof. The claim follows immediately from

$$\frac{\Pr(H \mid O \wedge A)}{\Pr(H \mid A)} = \frac{\Pr(O \mid H \wedge A)}{\Pr(O \mid A)} \,. \tag{7}$$

□

References

Ayer, A. J. (1936). *Language, Truth and Logic*. Victor Gollanz, London, 1st edition. Page numbers refer to the second edition (Ayer 1946). 6

Ayer, A. J. (1946). *Language, Truth and Logic*. Victor Gollanz, London, 2nd edition. 22

Carnap, R. (1950). *Logical Foundations of Probability*. University of Chicago Press, Chicago. References are to the 2nd edition (Carnap 1962). 2, 5

Carnap, R. (1962). *Logical Foundations of Probability*. University of Chicago Press, Chicago, 2nd edition. 22

Diamond, M. L. and Litzenburg, T. V., editors (1975). *The Logic of God: Theology and Verification*. Bobbs-Merill, Indianapolis, IN. 2, 22

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, volume 2. John Wiley & Sons, Inc., New York, N.Y., 2nd edition. 5

Flew, A. (1950). Theology and falsification. *University*, 1:1–8. References are to the reprint (Diamond and Litzenburg 1975, 257–259). 7

Hempel, C. G. (1952). *Fundamentals of Concept Formation in Empirical Sciences*, volume II,7 of *Foundations of the Unity of Science. Toward an International Encyclopedia of Unified Science*. University of Chicago Press, Chicago and London. References are to the two-volume edition. 2

Hodges, W. (1993). *Model Theory*, volume 42 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge. 10

Lutz, S. (2010). Criteria of empirical significance: A success story. Preprint: `http://philsci-archive.pitt.edu/8423/`. 10, 11

Lutz, S. (2011). On a contrastive criterion of testability I: Defining contrastive testability. Forthcoming. 2, 3, 4, 6, 11

Martin, M. (1990). *Atheism: A Philosophical Justification*. Temple University Press, Philadelphia. 2

Popper, K. (1963). The demarcation between science and metaphysics. In Schilpp, P. A., editor, *The Philosophy of Rudolf Carnap*, volume 11 of *The Library of Living Philosophers*, pages 183–226. Open Court Publishing Company, Chicago and LaSalle, IL. 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). 2, 6, 23

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. 23

Salmon, W. C. (1967). *The Foundations of Scientific Inference*. University of Pittsburgh Press, Pittsburgh, PA. 9

Salmon, W. C. (1971). *Statistical Explanation and Statistical Relevance*. University of Pittsburgh Press, Pittsburgh, PA. With contributions by Richard C. Jeffrey and James G. Greeno. 13

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. 2

Soames, S. (2003). *The Dawn of Analysis*, volume 1 of *Philosophical Analysis in the Twentieth Century*. Princeton University Press, Princeton, NJ. 2

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. 2, 8, 15

Sober, E. (1999). Testability. *Proceedings and Addresses of the American Philosophical Association*, 73(2):47–76. 2, 6, 7, 10, 11, 13

Sober, E. (2002). Intelligent design and probability reasoning. *International Journal for Philosophy of Religion*, 52(2):65–80. 8

Sober, E. (2007). What is wrong with intelligent design? *The Quarterly Review of Biology*, 82(1):3–8. 2

Sober, E. (2008). *Evidence and Evolution: The Logic Behind the Science*. Cambridge University Press, Cambridge. 2, 3, 6, 8, 11, 13, 17

Sober, E. (2010). Comments on Sebastian Lutz's "On Sober's criterion of contrastive testability". Unpublished typescript. 2

Woit, P. (2006). *Not Even Wrong: The Failure of String Theory and the Search for Unity in Physical Law*. Basic Books, New York. 2

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