On a Contrastive Criterion of Testability I: Defining 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 or a suggestion within Bayesianism. I argue that Sober's criterion entails that if one group of people is justified in believing a claim, every group is, and that it tacitly relies on an inconsistent interpretation of probabilistic inequalities. Furthermore, the criterion's restrictions on the use of auxiliary assumptions are in part redundant and in part unjustified. Most importantly, they are so weak that almost all theories can be contrastively tested. On the basis of these results, I suggest a modification of Sober's criterion that avoids these problems without abandoning Sober's core idea.

Contents

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

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 not been subjected to much scrutiny—all the more so because Sober's treatment of ID is widely discussed.[1] This article is meant to fill this gap.

To this end, I briefly discuss two criteria of empirical significance whose problems Sober's criterion is meant to avoid, namely falsifiability (§2.1) and a Bayesian definition of empirical significance (§2.2). To arrive at a precise and consistent definition of Sober's criterion, I discuss its implications, some of them unwanted (§3.1), and develop an interpretation of inequalities between possibly undefined probabilities that arguably follows from Sober's assumptions (§3.2). Then, I will argue that one of the criterion's restrictions on auxiliary assumptions is problematic (§4.1) and, more importantly, that the other one is not sufficient to avoid a trivialization of the criterion (§4.2). In light of these problems, I suggest two modified versions of Sober's criterion, a relative one that is provably non-trivial, and an absolute one to which the trivialization proof for Sober's criterion does not apply (§4.3).

2 Two criteria of empirical significance and their problems

Sober (2008, 154) conjectures that his criterion is "a step forward from the failed proposals of the logical positivists", but this is misleading because the logical positivists wanted to distinguish between meaningful and meaningless sentences (cf. Carnap 1963, §6.A). Sober (2010, 1), on the other hand, does not consider his definition of testability to provide a criterion of empirical significance precisely because "'[e]mpirical significance' suggests

[1] The only discussion that does not exclusively focus on the criterion's application to ID might be given by Justus (2010) in a book review.

that a sequence of terms has meaning iff it is empirically testable", a position to which he does not subscribe. Rather, Sober (2008, 149f) argues, meaningfulness is a semantic concept, while testability is epistemic. And furthermore:

> It seems clear that meaningfulness and testability are different. I suppose that the sentence "undetectable angels exist" is untestable, but the sentence is not meaningless gibberish. We know what it says, what logical relations it bears to other statements, and we can discuss whether it is knowable; none of this would be possible if the string of words literally made no sense.

Therefore Sober is rather improving on the demarcation criterion by Popper, who wanted to distinguish empirical from non-empirical statements, both of which can be meaningful (Popper 1935, §4; 1963, §II). I will follow Sober in the search for a demarcation criterion for empirical statements, but not in his choice of terminology. I think that 'empirical significance' differs from 'meaningfulness' enough to avoid confusion, and, as a technical term, is clearly meant as a placeholder for a concept that is yet to be explicated.

2.1 Falsifiability

In line with his search for a demarcation criterion, Sober (2007; 2008, §2.8) introduces his criterion as avoiding the problems of Popper's falsifiability criterion. Popper's criterion demands that, together with some initial conditions, "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 application of *modus tollens*, and assuming that the negation of an observation sentence is itself an observation sentence, Popper (1935, 86) arrives at the following definition:

> A theory is to be called 'empirical' or 'falsifiable' if it divides the class of all possible basic statements unambiguously into the following two non-empty subclasses. First, the class of all those basic statements with which it is inconsistent [. . .]; and secondly, the class of those basic statements which it does not contradict [. . .].

Since Popper allows the use of auxiliary assumptions for the derivation of observation sentences (cf. Lakatos 1974, 106f), possible basic statements ('observation sentences' in the following) are those observation sentences compatible with the auxiliary assumptions. This leads to

Definition 1. 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 $O \wedge A \vDash \neg H$.[2]

[2] I will follow Sober in treating observational claims and theories as single sentences, and the auxiliary

Here and in the following I will assume that for all theories H and auxiliary assumptions A it holds that $\Pr(H \wedge A) > 0$, and thus specifically that H and A are compatible ($H \wedge A \nvDash \bot$). Therefore, $H \wedge A$ must be compatible with some observation sentence, for otherwise it would entail negations of all observation sentences. Assuming with Popper that the negation of an observation sentence is itself an observation sentence, $H \wedge A$ would thus entail all observation sentences and their negations, that is, be inconsistent. Thus definition 1 captures Popper's formulation.

One problem with falsifiability, Sober (2008, 130, cf. 151) notes, is that no purely probabilistic statement is falsifiable: "Consider a simple example: the statement that a coin has a probability of .5 of landing heads each time it is tossed [...] is testable, but it does not satisfy Popper's criterion". To avoid this result, Popper (1935, ch. VIII) generalizes his criterion, considering a theory H falsified even if an event occurs that is possible but very improbable according to $H \wedge A$. But this generalization runs into problems as well, the most important of which is, for the sequel, the following: If a theory implies an observation sentence, then it allows a deductive inference via *modus ponens*, '$\{X, X \rightarrow Y\} \vDash Y$': The assumption of the theory X and the fact that it implies Y entail Y. Popper justifies his criterion of falsifiability by using the implication of an observation sentence in a *modus tollens*, '$\{\neg Y, X \rightarrow Y\} \vDash \neg X$'. The justification of the probabilistic generalization of his criterion would thus have to rely on a probabilistic version of *modus tollens*, in which a theory X is false or at the very least improbable if $\Pr(Y \mid X)$ is high and Y is false. But as Sober (2002, 69f) points out (cf. Sober 2008, §1.4):

> There is a "smooth [t]ransition" between probabilistic and deductive *modus ponens*; the minor premiss ("X") either ensures that Y is true, or makes Y very probable, depending on how the major premiss is formulated. In contrast, there is a radical discontinuity between probabilistic and deductive modus tollens. The *minor premiss* ("not-Y") guarantees that X is false in the one case, but has no implications whatever about the probability of X in the other.

Therefore, while Popper is right to infer from the fact that a theory entails an observation sentence that the theory is falsifiable, he cannot infer the same from the fact that it assigns a high probability to an observation sentence (and thus a low probability to the sentence's negation). Thus the generalization of falsifiability to probabilistic theories has not been justified.

assumptions as a finite set thereof. Finite sets of sentences are identified with the conjunctions of their members. Mostly, 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. To allow sets of sentences and higher order logic in the definition of falsifiability, it 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.

2.2 Bayesian empirical significance

A more successful criterion of empirical significance for probabilistic theories has been suggested within Bayesianism, the position that non-deductive inferences should follow the rules of the probability calculus, and specifically that the confirmation of scientific theories should follow Bayes's theorem,

$$\Pr(H \mid O \wedge A) = \frac{\Pr(O \mid H \wedge A) \cdot \Pr(H \mid A)}{\Pr(O \mid H \wedge A) \cdot \Pr(O \mid \neg H \wedge A)} \ . \tag{1}$$

$\Pr(H \mid A)$ is the probability of the theory given only the auxiliary assumptions, that is, before the observation O is taken into account, and thus called the prior probability (of H). $\Pr(H \mid O \wedge A)$ is the probability of H after O is taken into account, and hence called the posterior probability.[3] $\Pr(O \mid H \wedge A)$ and $\Pr(O \mid \neg H \wedge A)$ are the likelihoods of H and $\neg H$, respectively (for O). The standard criterion of empirical significance suggested within Bayesianism is the following (cf. Sober 2008, 150):

Definition 2. *Observations are relevant for* theory H *relative to* auxiliary assumptions A if and only if there is an observation sentence O such that[4]

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

Sober (2008, 150, 24–30) argues that if H is, say, the theory of general relativity, it is well-nigh impossible to assign a probability to H, or assess the likelihood of $\neg H$, so that $\Pr(H \mid A)$, $\Pr(O \mid \neg H \wedge A)$, and $\Pr(H \mid O \wedge A)$ are often undefined. Bayes theorem (1) then becomes unusable, and definition 2 very questionable.

3 Sober's criterion of empirical significance

3.1 A precise formulation

Sober (2008, 152) suggests a criterion of empirical significance that avoids all of the problems discussed so far. Unlike falsifiability, his criterion does not render all probabilistic theories empirically non-significant and does not rely on a faulty probabilistic generalization of *modus tollens* for its justification; unlike the Bayesian criterion, it does not rely on the probabilities of whole theories or on likelihoods of the negations of theories. The criterion is the following:[5]

[3] Sober (2008, 8) calls $\Pr(H)$ the prior and $\Pr(H \mid O)$ the posterior probability and discusses Bayesianism without auxiliary assumptions. But he also argues that $\Pr(O \mid H)$, which would be used to determine $\Pr(H \mid O)$, is, unlike $\Pr(O \mid H \wedge A)$, almost never defined. Prior and posterior probabilities therefore have to be defined relative to auxiliary assumptions, lest Bayesianism be empty.

[4] I will always silently assume that for any occurring conditional probability $\Pr(B \mid C)$, $\Pr(C) \neq 0$.

[5] Here and in the following quotations, I will replace Sober's '&' by '$\wedge$'.

> Hypothesis H_1 can now be tested against hypothesis H_2 if and only if there exist true auxiliary assumptions A and an observation statement O such that (i) $\Pr(O|H_1 \wedge A) \neq \Pr(O|H_2 \wedge A)$, (ii) we now are justified in believing A, and (iii) the justification we now have for believing A does not depend on believing that H_1 is true or that H_2 is true and also does not depend on believing that O is true (or that it is false).

Sober (2008, 151) remarks that the "word 'now' marks the fact that whether a proposition has observational implications depends on the rest of what we are justified in believing, and that can change". However, the use of the indexical 'now' does not define testability relative to time in general, but relative to the specific time of the utterance. Thus 'H_1 can now be tested against H_2' is defined, but 'H_1 will be testable against H_2 within a decade' is not. 'Within a decade, the utterance "H_1 can now be tested against H_2" will be true', on the other hand, is defined. To avoid such cumbersome formulations, one can define contrastive testability as the three-place predicate 'H_1 can at time t be tested against H_2'. This achieves Sober's intention more explicitly.

The other indexical term, 'we', occurs only in the definiens, so that Sober's criterion is not a definition but rather a claim—and a false one at that. This is because the criterion violates the demand that in an explicit definition, any free variable of the definiens must also occur free in the definiendum, and thus Sober's criterion is creative (cf. Belnap 1993, 139): If for two theories H_1 and H_2, one referent of 'we' fulfills the definiens at t, the definiendum applies to H_1 and H_2 at t. But then the definiendum applies to H_1 and H_2 at t no matter the referent of 'we', and thus any referent fulfills the definiens for H_1 and H_2 at t. Hence according to Sober's criterion, it holds for any O, H_1, H_2, and A that fulfill (i): If one group is justified in believing A independently of H_1, H_2, O, and $\neg O$, then every group is.[6] To avoid this unintended implication, testability must be defined as both relative to time and relative to a group of people. It is thus as a four-place predicate.

As it stands, restriction (iii) on the auxiliary assumptions sounds like the demand that the justification of A must not depend on the fact that the truth of H_1, H_2, O or $\neg O$ is content of our beliefs. But very few statements are justified by the having of a belief, so that condition (iii) would be almost empty if this was meant. The restriction is therefore probably better expressed as the demand that the justifications for A must not depend on the fact that the belief in the truth of H_1, H_2, O or $\neg O$ *is justified*. For convenience, I will mostly drop the reference to beliefs in the following, and speak of justified sentences, rather than justified beliefs in the truth of propositions expressed by sentences.

Finally, the condition (i) on the likelihoods of H_1 and H_2 needs to be elucidated, given that Sober's critique of the Bayesian criterion of empirical significance assumes that some

[6] Keeping Sober's indexical formulation, it holds for any O, H_1, H_2, and A that fulfill (i): If for one group the claim 'We are now justified in believing A independently of H_1, H_2, O, and $\neg O$' is true, then the claim is true for every group.

likelihoods are undefined. *Prima facie*, one would expect that H_1 and H_2 cannot be tested against each other if and only if $\Pr(O \mid H_1 \wedge A) = \Pr(O \mid H_2 \wedge A)$ for all O and A that fulfill conditions (ii) and (iii). But this would mean that the lack of testability is transitive for any theories that are not used to justify their own or each other's auxiliary assumptions. And this is incompatible with Sober's remark that it is not clear that ID "can be tested against the Epicurean hypothesis that a mindless chance process gave vertebrates their eyes (or, for that matter, against the evolutionary hypothesis that the process of evolution by natural selection did the work)" (Sober 2008, 148). Assuming that the chance hypothesis can be tested against evolutionary theory (ET), if ID cannot be tested against either, the lack of testability is not transitive. The solution to this puzzle is that Sober (2010, 2f) interprets the inequality as true if and only if both likelihoods are defined and different. This interpretation, however, plays havoc with classical logic, for $p \neq q \boxminus \neg\, p = q$. Therefore, if the likelihoods p and q are defined and different, while the likelihood a is undefined, it follows from Sober's interpretation of the inequality that $p = a$ and $a = q$, while $p \neq q$. To avoid such inconsistencies, it is probably best to treat undefined likelihoods separately in the definition.

These considerations lead to

Definition 3 (Contrastive testability)**.** 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.

One could reformulate definition 3 to include a reference to times and groups of people, that is, define 'H_1 can be tested against H_2 *at time t by group g*' by relativizing 'justification' (and possibly 'dependence') to t and g. In similar cases, especially when the auxiliary assumptions are simply the background assumptions, these relativizations are typically suppressed because it is clear that the background assumptions and generally the set of justified sentences can change over time and from group to group. Thus I will do likewise.

Sober calls his criterion simply 'testability', but the qualifier 'contrastive' distinguishes it clearly from the ordinary language term and emphasizes that, atypically, the empirical significance of one theory is defined relative to another. It may seem problematic to explicate a one-place predicate like 'makes observational assertions' by a two-place predicate like contrastive testability. Frege (1918, 291), for example, objects to the explication of 'truth' as a correspondence relation on the grounds that the first is a one-place, the second a two-place predicate. However, many successful explications involve a change of the logical structure, as the explication of 'warm' by 'warmer than' and finally 'temperature' illustrates (Carnap 1950, §4). Hempel (1952, §10) even argues that the move from a classificatory to a comparative concept is often a sign of an investigation's maturity and, at another point, argues that empirical significance is comparative (Hempel 1965, 117).

3.2 Interpreting inequalities between probabilities

In the discussion of Sober's interpretation of the inequality in his criterion (§3.1), it has already become apparent that dealing with undefined probabilities is not an entirely straightforward matter. And even though definition 4 treats the case of undefined likelihoods explicitly, it still involves some lacunae. Specifically, if one of the likelihoods is not defined, it is not obvious how to treat the inequality (4), and not entirely obvious how to treat the whole definition. For the definition is logically a conjunction with the inequality as a conjunct, and it is unclear whether a conjunction with one undefined conjunct is undefined as well. It would thus be desirable if the inequality were never undefined, so that the usual rules of logic can apply. Luckily, Sober's assumptions arguably entail just such an interpretation of the inequality.

It is uncontentious that $\Pr(S\,|\,H)$ is defined when H assigns a real-valued probability to S. But as Sober himself states when arguing for the need for auxiliary assumptions, theories alone often do not assert anything, and thus do not assign a real-valued probability to any observation sentence. And even with auxiliary assumptions A, no theory will make assertions about everything.[7] Rather, $H \wedge A$ restricts the set of reals from the unit interval that can be assigned to *some* sentences S to a subset of the unit interval, possibly to one specific value $x \in [0,1]$, while the set of reals for some sentences will remain unrestricted. The conditional probability $\Pr(S\,|\,H \wedge A)$ can then either always be read as the set of reals that S can be assigned under the assumption of $H \wedge A$, from $[0,1]$ to proper subsets thereof down to the singleton set $\{x\}, x \in [0,1]$. Or $\Pr(S\,|\,H \wedge A)$ may be read as defined only when it is a set of some specific kind considered acceptable (e. g., an proper sub-interval of $[0,1]$ or a singleton set), and undefined in all other cases.

Depending on the treatment of formulas that contain undefined terms, the second reading of conditional probabilities leads to different interpretations of the inequality (4), given in table 1. Sober seems to assume the validity of classical logic, so that $\neg\varphi$ is false if and only if φ is true and tautologies are always true. This excludes some of the possible

[7]More precisely: For any even remotely plausible theory H, $H \wedge A$ can be observationally complete only for very restricted languages, where H is observationally complete if and only if for all observations O, $\Pr(O\,|\,H \wedge A) = x, x \in [0,1]$. It is thus always possible to expand the language by well-interpreted observation terms so that H fails to be observationally complete.

Values of likelihoods		$\Pr(O \mid H_1 \wedge A) \neq \Pr(O \mid H_2 \wedge A)$								
$\Pr(O \mid H_1 \wedge A)$	$\Pr(O \mid H_2 \wedge A)$	1	2	3	4	5	6	7	8	9
x	y	$T\ldots$								
x	x	$F\ldots$								
x	U	T	T	T	F	F	F	U	U	U
U	U	T	F	U	T	F	U	T	F	U

Table 1: The nine possible interpretations of the inequality in definition 6 depending on the values of the probabilities, where x, y with $x \neq y$ are acceptable values for likelihoods, 'T' stands for 'true', 'F' stands for 'false', and 'U' stands for 'undefined'.

interpretations: Considering again the case where p and q are defined and different, while a is undefined, it is clear that interpretations 4–6 (5 being Sober's) are inconsistent because they lead to $p = a$, $a = q$, and $p \neq q$. When a is undefined, interpretations 1, 3, 7, and 9 do not render $a \neq a$ false, and thus they are also excluded. Interpretation 8 is excluded if one demands that classical logic be truth-preserving and at least one disjunct of a true disjunction be true. For then, if p and q are defined and identical, while a and b are undefined, $p = q$ and $a = b$ are true, and entail $a = p \vee b \neq q$, which according to interpretation 8 has two undefined disjuncts. The remaining interpretation 2 can be seen as following from the introduction of the special value 'undefined' for probability-terms.

Under these assumptions, there are thus two possible interpretations of the inequality:

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

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

It is clear that the inequality is true more often for interpretation 1 than for interpretation 2, since in interpretation 1 it is true whenever the set on one side differs from the set on the other side, but also when there is a difference between two sets that are unacceptable under the second reading of the likelihoods. It is also clear that Sober does not subscribe to interpretation 1, since in that case, there are no undefined likelihoods. In fact, he developed his concept of contrastive testability under the assumption that only singleton sets are acceptable (Sober 2010, 3).

With these two readings of probability and the corresponding interpretations of the inequality, definition 3 is now indeed defined in all cases because the inequality is always either true or false.

4 The restrictions on the auxiliary assumptions

Definition 3 is formally correct, but is nonetheless problematic. Specifically, I will argue that there are serious problems with the restrictions (IVa) and (IVb) that Sober places on the auxiliary assumptions.

Sober (1999, 54) introduces auxiliary assumptions into the definition because "hypotheses rarely make observational predictions on their own; they require supplementation by auxiliary assumptions if they are to be tested" (cf. Sober 2007, 5f; Sober 2008, 144).[8] But this "raises the question of which auxiliary assumptions we should use to render a theory testable. What makes an auxiliary assumption 'suitable'?" (Sober 2008, 144). Restrictions (IVa) and (IVb) are answers to these questions.

4.1 Dependence on the theories

Sober (2008, 145) at one point simply states that the need for restriction (IVa) is obvious, and elaborates in another passage that without it, his criterion would beg the question (Sober 2007, 6). But this, at least, is not obvious. Arguments must not in general allow their conclusion among their premises (that is, beg the question) because otherwise every claim could be shown to be true, and the concept of an argument would be trivial. But even without restriction (IVa), it is not possible to simply assume that H_1 can be tested against H_2 when the criterion is applied. In fact, I want to show that (IVa) is often ineffective or redundant, and in general lacks a justification.

There are a number of cases in which restriction (IVa) is ineffective. Obviously, two theories that are not contrastively testable *without* (IVa) cannot be contrastively testable *with* (IVa). For if there are no auxiliary assumptions A that fulfill (IVb), there are also no auxiliary assumptions that fulfill (IVb) and (IVa). And the converse usually also holds: Two theories that are contrastively testable without (IVa) cannot fail to be so with (IVa). It is typical, for example, that (i) theories H_1 and H_2 are incompatible, and (ii) the auxiliary assumptions used to determine the likelihoods of H_1 at most depend on H_1, not on the competing theory H_2 (and vice versa). Now, for restriction (IVa) to do any work, there has to be an observation O such that there are no auxiliary assumptions that fulfill conditions (I)–(IV), but there are justified auxiliary assumptions A that fulfill restriction (IVb), $\Pr(O \,|\, H_1 \wedge A) \neq \Pr(O \,|\, H_1 \wedge A)$, and both likelihoods are defined. By assumption (ii), the auxiliary assumptions are of the form $A \vDash\!\dashv A_1 \wedge A_2$, where A_1 does not depend on H_2 and A_2 does not depend on H_1. Thus $H_1 \to A_1$ and $H_2 \to A_2$ are justified without assuming H_1 or H_2. Therefore, $A' := (H_1 \to A_1) \wedge (H_2 \to A_2)$ fulfils restriction (IVa), and since by assumption (i), $H_1 \wedge A' \vDash\!\dashv H_1 \wedge A$ and $H_2 \wedge A' \vDash\!\dashv H_2 \wedge A$,

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

$\Pr(O \mid H_1 \wedge A') \neq \Pr(O \mid H_2 \wedge A')$ so that H_1 can be tested against H_2.[9]

That restriction (IVa) is often ineffective is unproblematic however, because it is also redundant in the important cases. To see this, note that restriction (III) demands that A be justified, and that a sentence whose justification depends on another sentence B be justified only if B is justified. (Giving up this relation between 'justified' and 'depend' would render Sober's restriction (IV) altogether empty.) Thus an auxiliary assumption whose justification depends on H_1 fulfills the definiens of definition 3 only if H_1 itself is justified, and analogously for H_2. Typically, however, the question of empirical significance does not even come up for theories that are already justified. Indeed, Sober assumes that a theory is confirmed only if it has been tested, and this is possible only if it is contrastively testable. Assuming that only confirmed theories are justified, restriction (IVa) therefore goes beyond restriction (III) *only* when the question of empirical significance has already been answered.[10]

This general argument is not countered by the example that Sober (2008, 145) adduces to show the need for restriction (IVa). In it, he envisions Jones being tried for a murder, with a size 12 shoe print, cigar ash, and .45 Colt shells found at the crime scene. When considering whether Smith may be the culprit instead, Sober notes, one must not simply conclude that the evidence favors Smith over Jones on the basis of the *assumption* that Smith is a Colt-owning smoker with size 12 feet, while Jones is not.

First note that in this example the question is which theory can be inferred from the evidence, not which observations are asserted by the theory; that is, the example revolves around a question of confirmation, not empirical significance. More important in the following is that the belief about Smith's shoe size would be excluded from the auxiliary assumptions even without restriction (IVa). This is because, first, the belief that Smith is the murderer (H_1) is itself not justified, and thus cannot justify anything. Second, Smith's murdering someone does not allow to conclude anything about her shoe size. This conclusion also requires the belief that there was a size 12 shoe print at the crime scene (O). In other words, the justification of the auxiliary assumption cannot depend on H_1, for then it would be excluded from A by restriction (III), and it is excluded by restriction (IVb) anyway, because its justification depends on O.

Restriction (IVa) is included in definition 3 for more serious reasons than fictitious murder trials with careless jurors. It is meant to address an argument in defense of the contrastive testability of ID against evolutionary theory (ET) that Sober (1999, 65, note and line break removed) describes as follows (cf. Sober 2008, 143–146):

> [A]dvocates of the design argument should not be confident that they know

[9]This inference assumes that H_1, H_2, A_1, and A_2 can be finitely axiomatized. The assumption can be somewhat alleviated by splitting A into $A^* \cup \{A_1\} \cup \{A_2\}$, where A^* does not depend on either H_1 or H_2 and can be infinite. Then A' can be defined as $A^* \cup \{H_1 \rightarrow A_1\} \cup \{H_2 \rightarrow A_2\}$.

[10]Of course, one may want to justify or confirm an already justified or confirmed theory *further*, but this is then not a question of contrastive testability any more.

> what characteristics God would have wanted to give to organisms on earth if he had created them. Creationists may be tempted to respond to this challenge simply by inspecting the life we see around us and saying that God wanted to create *that*. After all, if life is the result of God's blueprint, can't we infer what the blueprint said by seeing what the resulting edifice looks like? [But you] can't just *assume* that God created organisms, and you also can't *assume* that if God created organisms he would have made them with such-and-such characteristics.

Analogously to the murder trial, the justification of the auxiliary assumption about God's intentions in the creationists' argument depends both on the assumption that God exists and the observational assumption that life is as we see it around us, like "that".[11] Therefore it is excluded from A by restriction (III) because it is not justified until the belief in God is justified. And if a description of life as we see it around us is given as the observation sentence O for which the likelihoods of ID and ET differ, then the auxiliary assumption is also excluded by restriction (IVb) for dependence on O.

4.2 Dependence on observation sentences

Sober (2008, 145) justifies restriction (IVb) as follows (cf. Sober 2007, 6):

> If O is true, so is the disjunction "either H_1 is false or O is true". If you use this disjunction as your auxiliary assumption A_1, then it turns out that the conjunction $H_1 \land A_1$ entails O. This allows H_1 to make a prediction about O even when H_1 has nothing at all to do with O. The same ploy can be used to obtain auxiliary assumptions A_2 so that the conjunction $H_2 \land A_2$ also entails O. Using propositions A_1 and A_2 as auxiliary assumptions leads to the conclusion that the two hypotheses H_1 and H_2 both have likelihoods of unity.

As it stands, this argument proves nothing about the relevance of restriction (IVb) for the definition of contrastive testability, since it only shows that for one specific auxiliary assumption, $A \boxminus A_1 \land A_2$, both theories' likelihoods are 1. But to show that H_1 cannot be tested against H_2, their likelihoods have to be identical for *all* auxiliary assumptions that fulfill restrictions (III) and (IVa). (Furthermore, if the goal was to arrive at the *same* likelihood for both theories, $A \boxminus O$ would achieve the same result.)

But the ingenuity of the choice of A_1 is exactly that, if H_1 and H_2 are completely unrelated to O, the likelihood of $H_1 \land A_1$ is 1, while the likelihood of $H_2 \land A_1$ is not. Reconceptualized in this way, Sober's case for restriction (IVb) is a typical trivialization proof, since it shows that without it, any two theories can be tested against each other.

[11] In disanalogy to the murder trial, the question in this case is indeed which observations the theory asserts, not which theory the observations confirm.

The argument has four tacit assumptions, however. First, a sentence S (here: $\neg H_1 \vee O$) logically entailed by a justified sentence J (here: O) is also justified, since otherwise S might be excluded by restriction (III). Second, S depends for its justification only on J, for otherwise, S might be excluded by restriction (IVa) or, implausibly, for its dependence on $\neg O$ by restriction (IVb). Third, $\Pr\big(O \,|\, H_2 \wedge (\neg H_1 \vee O)\big) \neq 1$. Finally, $\Pr\big(O \,|\, H_2 \wedge (\neg H_1 \vee O)\big)$ must be defined, for otherwise condition (I) is not fulfilled.

The fourth tacit assumption is probably false, since neither H_1 nor H_2 are related to O. By considering additional auxiliary assumptions A, however, one can arrive at a modification of the proof that has plausible premises. Let A be such that it is unrelated to H_1 and H_2, and $\Pr(O \,|\, A)$ is defined. Then a plausible fourth tacit assumption is that for any S unrelated to H_1 and H_2, $\Pr(O \,|\, H_2 \wedge (\neg H_1 \vee S) \wedge A)$ is therefore defined as well: Conjoining H_2 with A does not render the conditional probability of O undefined, since H_2 is not related to O. Conjoining $\neg H_1 \vee S$ with $H_2 \wedge A$ arguably does not render the conditional probability of O undefined, either, because the inferences one can draw from $H_1 \vee S$ are weaker then those that one can draw from H_1, and H_1 is already unrelated to O. Choosing $S = O$ and incorporating A in all premises now allows to give a corrected version of Sober's proof, which can also be recovered from the proof of claim 1 below.

Sober does not show why the reference to $\neg O$ in restriction (IVb) is necessary, but his trivialization proof can be repeated for $\neg O$. Since $\Pr(O \,|\, H_i \wedge A) = 1 - \Pr(\neg O \,|\, H_i \wedge A), i \in \{1,2\}$, A can be justified with O, while the likelihoods of H_1 and H_2 would differ for $\neg O$. This proof assumes that the negation of an observation sentence is itself an observation sentence, which is fairly uncontroversial. It is not only tacitly assumed by Popper and Sober (e. g. 1999, n. 14), but also fulfilled by the most common restrictions on observation sentences: It holds if all and only sentences with a specific non-logical vocabulary are observational (cf. Psillos 2000, 158f), if all and only molecular sentences with a specific vocabulary are observational (cf. Carnap 1937, §23), and if all and only sentences are observational whose quantifiers are relativized to observable objects (cf. Carnap 1956, §II; Friedman 1982, 276f). A sentence could also be considered observational if and only if it is about observations, and according to Lewis (1988, 140f), if a sentence is about observations, so is its negation. All these restrictions even entail that the set of observation sentences is closed under truth-functional composition.

While restriction (IVb) is necessary to avoid trivialization of definition 3, it is not sufficient. Specifically, any two theories can be tested against each other as long as one of them can be finitely axiomatized and thus phrased as one (possibly inordinately long) sentence:[12]

Claim 1. *Let H_1 and H_2 be theories and O, S, and A be any sentences such that*

1. *O is an observation sentence,*

[12] Since the proof relies on the use of the negation of one of the theories, this restriction is essential.

2. *$S \vDash O$,*

3. *S and A are justified independently of O, $\neg O$, H_1, and H_2,*

4. *S and A are unrelated to H_1 and H_2,*

5. *$\mathrm{Pr}(O\,|\,A)$ is defined, and*

6. *$\mathrm{Pr}(O\,|\,H_2 \wedge \neg H_1 \wedge A) \neq 1$.*

Given Sober's tacit assumptions, H_1 and H_2 can then be tested against each other.

Proof. Choose O, S, and A such that conditions 1–6 hold. Since S and A are justified, so is $A^* \,\,\text{⫤⊨}\,\, (\neg H_1 \vee S) \wedge A$ by Sober's first tacit assumption. It follows from Sober's second tacit assumption that, since the justifications of S and A do not depend on O, $\neg O$, H_1, or H_2, neither does the justification of A^*. Therefore A^* fulfills restrictions (III) and (IV) of definition 3.

Now, from $\mathrm{Pr}(O\,|\,H_2 \wedge \neg H_1 \wedge A) \neq 1$ and $S \vDash O$ it follows that $\mathrm{Pr}\big(O\,|\,H_2 \wedge (\neg H_1 \vee S) \wedge A\big) \neq 1$:

$$\begin{aligned}
&\mathrm{Pr}\big(O\,|\,H_2 \wedge (\neg H_1 \vee S) \wedge A\big) \\
&\qquad = \frac{\mathrm{Pr}(O \wedge H_2 \wedge \neg H_1 \wedge A) + \mathrm{Pr}(O \wedge H_2 \wedge S \wedge A) - \mathrm{Pr}(O \wedge H_2 \wedge \neg H_1 \wedge S)}{\mathrm{Pr}(H_2 \wedge \neg H_1 \wedge A) + \mathrm{Pr}(H_2 \wedge S \wedge A) - \mathrm{Pr}(H_2 \wedge \neg H_1 \wedge S \wedge A)} \\
&\qquad = \frac{\mathrm{Pr}(O \wedge H_2 \wedge \neg H_1 \wedge A) + c}{\mathrm{Pr}(H_2 \wedge \neg H_1 \wedge A) + c}\,. \qquad (3)
\end{aligned}$$

The last term equals 1 if and only if $\mathrm{Pr}\big(O\,|\,H_2 \wedge \neg H_1 \wedge A\big) = 1$. Since $\mathrm{Pr}(O\,|\,A)$ is defined, $\mathrm{Pr}(O\,|\,H_2 \wedge (\neg H_1 \vee S) \wedge A) \,\text{⫤⊨}\, \mathrm{Pr}(O\,|\,H_2 \wedge A^*)$ is defined by the fourth tacit assumption.

Since furthermore $\mathrm{Pr}(O\,|\,H_1 \wedge A^*) = 1$, it holds that $\mathrm{Pr}(O\,|\,H_1 \wedge A^*) \neq \mathrm{Pr}(O\,|\,H_2 \wedge A^*)$, where both probabilities are defined. H_1 and H_2 can therefore be tested against each other. □

Note that for $\mathrm{Pr}(H_1 \wedge H_2 \wedge A) = 0$, condition 6 simplifies to '$\mathrm{Pr}(O\,|\,H_2 \wedge A) \neq 1$'. The corrected version of the trivialization proof that Sober uses to justify restriction (IVb) can be recovered by dropping S in condition 3 and choosing $S = O$. Then condition 2 is trivially true, 3 amounts to Sober's restriction (IV), 4 and 5 are the antecedents of the fourth tacit assumption, and 6 is equivalent to the third tacit assumption.

Conditions 1–3 are impossible to fulfill if a justification can proceed only deductively from observation sentences, because then the justification of a sentence depends on every observation sentence it entails. However, since Sober's criterion is meant to be applicable to inductive theories, it is plausible that auxiliary assumptions can also be inductively justified. In that case, it is easy to find sentences O, S, and A that fulfill all the requirements. For instance, let A express that 1 out of 10 vases of some kind breaks when dropped from a specific height. Let furthermore S express that a specific vase of that kind does not break when dropped a hundred times from a that height, and O express that the vase does not break on the hundredth drop. Then S is justified independently of O when the vase is dropped 99 times without breaking, so that O, S, and A fulfill conditions 1–6 for any two theories that are not related to vases. Since according to Sober (1999, 54), "hypotheses rarely make observational predictions on their own", that includes almost all theories. But even for two theories that make assertions about vases, it should not be difficult to find other observations that neither they nor their negations assert with probability 1, but that can be asserted by enumerative induction.

Considering the somewhat problematic status of the fourth tacit assumption, the result of this section can be seen as a dilemma: Either the tacit assumptions (especially the fourth) hold or they do not hold. If they do not hold, Sober's restriction (IVb) has not been justified. If they do, the restriction is to weak to avoid trivialization.

4.3 New definitions

Sober's restriction (IVa) is unjustified where it is not redundant or ineffective, and restrictions (III) and (IV) together are to weak to avoid trivialization. Clearly, the search for general restrictions on the auxiliary assumptions poses a host of subtle problems. To bracket these problems, I suggest

Definition 4. 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

$$\Pr(O \mid H_1 \wedge A) \neq \Pr(O \mid H_2 \wedge A) . \tag{4}$$

This definition is not trivial: Choose $A = \varnothing$, two non-observational, non-equivalent sentences S and S', and, for some observation sentence O, $H_1 \mathrel{\text{⫤⊨}} S \wedge \Pr(O) = p$ and $H_2 \mathrel{\text{⫤⊨}} S' \wedge \Pr(O) = q$ 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 contingent theories can and many contingent theories cannot be tested against each other relative to A. Definition 4 makes it necessary, however, to decide on a case by case basis which auxiliary assumptions are suitable. This may be a good preliminary strategy, because often the suitable auxiliary assumptions are reasonably clear. For instance, often the suitable auxiliary assumptions are those that could feature as background assumptions.

Eventually, of course, it would be helpful to have a general criterion for suitable auxiliary assumptions and define absolute contrastive testability as contrastive testability relative to suitable auxiliary assumptions. To this end, I suggest the following. The proof of claim 1 is a modification of a trivialization proof for Sober's criterion of 'having observable implications' (Lutz 2010, §9.2) and leads to a similar diagnosis. Sober's proof and that of claim 1 rely on the possibility to include a sentence ($\neg H_1 \vee O$ or $\neg H_1 \vee S$) in A that is justified by another one (O or S) that is itself not included in A. Both trivialization proofs can therefore be blocked by explicating 'suitable auxiliary assumptions' as 'honest set of auxiliary assumptions':

Definition 5. 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.

Note that this definition only uses concepts that already occur in Sober's definition 3 of contrastive testability and that A can be a proper subset of the set of all justified sentences.

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 5 allows to modify Sober's criterion of testability as follows:

Definition 6. 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.

To distinguish definition 6 from definition 4 more clearly, I will sometimes also speak of *absolute* contrastive testability.

The proof that relative testability is not trivial also shows that not any two theories are absolutely testable against each other, for $\varnothing$ is an honest set according to definition 5. I will not attempt to prove that there is a non-falsifiable sentence, because this would amount to finding two theories that are not contrastively testable relative to any honest set. The proof is immediate for equivalent theories, but impossible for other pairs of theories without more precise notions of justification and dependence.[13]

Definition 6 is at least as exclusive as Sober's definition 3, however. The restriction of the auxiliary assumptions in definition 6 to honest sets entails restriction (III) of definition 3. And while the restriction to honest sets does not entail restriction (IVb), it precludes all trivializations precluded by that restriction: Two theories H_1 and H_2 fail to be contrastively testable because of restriction (IVb) only if for any S whose inclusion in A would lead to differing likelihoods for some O, the justification of S depends on O or $\neg O$. In that case, (IVb) ensures that H_1 and H_2 are not contrastively testable. The restriction of A to honest sets leads to the same result, because if the justification of S depends on O (or

[13]Equivalent theories are excluded from the conclusion of claim 1 because of premise 6 and the assumption that for any conditional probability $\Pr(B \mid C)$, $\Pr(C) \neq 0$ (see n. 4).

$\neg O$) and A is honest, then $O \in A$ (or $\neg O \in A$). Thus $\mathrm{P}(O \mid H_1 \wedge A) = 1 = \mathrm{P}(O \mid H_2 \wedge A)$ (or $\mathrm{P}(O \mid H_1 \wedge A) = 0 = \mathrm{P}(O \mid H_2 \wedge A)$). As an example, take the sentence $(\neg H_1 \vee O) \wedge A$ of Sober's trivialization proof. Restriction (IVb) excludes $(\neg H_1 \vee O) \wedge A$ from the auxiliary assumptions A^*, so that the likelihoods of H_1 and H_2 for O cannot differ because of $(\neg H_1 \vee O) \wedge A$. The restriction to honest sets, on the other hand, leads to the inclusion of O in A^*, so that the likelihoods do not differ, either. Unlike restriction (IVb), the restriction to honest sets also leads to identical likelihoods if $(\neg H_1 \vee S) \wedge A \in A^*$ is justified by a sentence $S \vDash O$, thereby precluding the proof of claim 1. Specifically, premise 6 will be false because $A^* \vDash O$.

Since it is not clear in which case restriction (IVa) is meant to preclude trivialization, or in general, which problem it is meant to solve, I cannot show that definition 5 can fulfill the role of restriction (IVa). Given the restriction's questionable role and justification, this should not be considered a drawback of definition 6. If there is a justification for restriction (IVa), however, one can modify definition 6 by defining contrastive testability as contrastive testability relative to an honest set that does not include H_1 or H_2. This restriction entails restriction (IVa).

That the notion of an honest set plausibly explicates the notion of possible background assumptions provides a justification of definition 6 that is independent of the trivialization proofs: If relative contrastive testability is a good explication of empirical significance, then absolute contrastive testability explicates what it means for a theory to be empirically significant in our current epistemic situation (as determined by our background assumptions) or a possible situation on our way to our current epistemic situations.

The independent justification of definition is not only relevant because the proof of claim 1 rests, like Sober's justification of restriction (IVb), on somewhat contentious assumptions. More importantly, it justifies the hope that the definition is right, while the amendments in light of the trivialization proofs at best allow to claim that the modified definitions are not obviously wrong. Of course, definition 6 might still allow the proof that any two non-equivalent theories can be absolutely tested against each other. In response to such a proof, one can fall back on definition 4 until a better explication of 'suitable' is found than definition 5. In general, any results that determine which auxiliary assumptions are suitable, or which assumptions are possible background assumptions, can be used directly as a substitute for definition 5 (cf. Lutz 2010, §9.2).

5 Conclusion

I have argued that Sober's definition of contrastive testability is settled with a host of problems: It entails that if one group of people is justified to hold some belief, all groups of people are justified to hold that belief, and it rests on an inconsistent interpretation of inequalities between likelihoods. Furthermore, one of its restrictions on the auxiliary assumptions is almost always redundant, often ineffective, and has in general not been justified. Most importantly, the restrictions on the auxiliary assumptions are too weak, leaving almost any to theories contrastively testable.

Considering these problems, the criteria for empirical significance that I suggest deviate surprisingly little from Sober's original definition. The definition of contrastive testability relative to a set of auxiliary assumptions simply avoids the question which auxiliary assumptions are suitable. The definition of contrastive testability simpliciter relies on the notion of honest sets, and it avoids the problems discussed in connection with Sober's original definition. Both criteria retain the essential feature of Sober's criterion, that empirical significance is contrastive.

As argued in the beginning, it is not obviously a shortcoming that the criteria explicate a concept usually considered to be classificatory as a two-place relation. It is plausible that a comparative explication of empirical significance is helpful (Popper 1935, §33; Lutz 2010, §8.1). But unlike 'warmer than' and 'more significant 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, his claims that 'Undetectable angels exist' is untestable and that 'This coin has probability of .5 of landing heads' is testable are, strictly speaking, meaningless according to his own definition. 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.

With the suggested modifications, contrastive testability is consistently defined and, arguably, non-trivial. But whether it is adequate as a criterion of empirical significance is still an open question.

References

Belnap, N. (1993). On rigorous definitions. *Philosophical Studies*, 72:115–146. 6

Carnap, R. (1937). Testability and meaning—continued. *Philosophy of Science*, 4(1):2–35. 13

Carnap, R. (1950). *Logical Foundations of Probability*. University of Chicago Press, Chicago. References are to the 2nd edition (Carnap 1962). 8

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. 13

Carnap, R. (1962). *Logical Foundations of Probability*. University of Chicago Press, Chicago, 2nd edition. 19

Carnap, R. (1963). Replies and systematic expositions. In Schilpp (1963), pages 859–1016. 2

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

Frege, G. (1918). Der Gedanke: eine logische Untersuchung. *Beiträge zur Philosophie des deutschen Idealismus*, 1:58–77. References are to the translation (Frege 1956). 8

Frege, G. (1956). The thought: A logical inquiry. *Mind*, 65(259):289–311. 19

Friedman, M. (1982). Review. *The Journal of Philosophy*, 79(5):274–283. Review of (Van Fraassen 1980). 13

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. 8

Hempel, C. G. (1965). Empiricist criteria of cognitive significance: Problems and changes. In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, pages 81–119. The Free Press, New York. 8

Justus, J. (2010). Evidentiary inference in evolutionary biology. *Biology and Philosophy*, forthcoming. Review of (Sober 2008). 2

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. 3

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

Lewis, D. (1998). *Papers in Philosophical Logic*. Cambridge Studies in Philosophy. Cambridge University Press, Cambridge. 20

Lutz, S. (2010). Criteria of empirical significance: A success story. Preprint: `http://philsci-archive.pitt.edu/8423/`. 15, 17, 18

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

Popper, K. (1963). The demarcation between science and metaphysics. In Schilpp (1963), pages 183–226. 3

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). 3, 4, 18, 20

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. 20

Psillos, S. (2000). Rudolf Carnap's 'Theoretical concepts in science'. *Studies in History and Philosophy of Science Part A*, 31:151–172. 13

Schilpp, P. A., editor (1963). *The Philosophy of Rudolf Carnap*, volume 11 of *The Library of Living Philosophers*. Open Court Publishing Company, Chicago and LaSalle, IL. 19, 20

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

Sober, E. (1999). Testability. *Proceedings and Addresses of the American Philosophical Association*, 73(2):47–76. 2, 10, 11, 13, 15

Sober, E. (2002). Intelligent design and probability reasoning. *International Journal for Philosophy of Religion*, 52(2):65–80. 4

Sober, E. (2007). What is wrong with intelligent design? *The Quarterly Review of Biology*, 82(1):3–8. 2, 3, 10, 12

Sober, E. (2008). *Evidence and Evolution: The Logic Behind the Science*. Cambridge University Press, Cambridge. 2, 3, 4, 5, 6, 7, 10, 11, 12, 19

Sober, E. (2010). Comments on Sebastian Lutz's "On Sober's criterion of contrastive testability". Unpublished typescript. 2, 7, 9

van Fraassen, B. C. (1980). *The Scientific Image.* The Clarendon Library of Logic and Philosophy. Clarendon Press, Oxford. 19

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.