Carnap's Forgotten Criterion of Empirical Significance

JAMES JUSTUS
Florida State University
jjustus@fsu.edu

The waning popularity of logical empiricism and the supposed discovery of insurmountable technical difficulties led most philosophers to abandon the project to formulate a formal criterion of empirical significance. Such a criterion would delineate claims that observation can confirm or disconfirm from those it cannot. Although early criteria were clearly inadequate, criticisms made of later, more sophisticated criteria were often indefensible or easily answered. Most importantly, Carnap's last criterion was seriously misinterpreted and an amended version of it remains tenable.

1. Introduction: a philosophical revival

A criterion of empirical significance was the linchpin of philosophical progress for logical empiricists. Without it there seemed to be no principled way to distinguish metaphysics from subjects meriting serious philosophical attention, the sciences most importantly. Discoveries at the time—particularly Einstein's analysis of simultaneity (see Feigl 1969)—were also demonstrating that scientific theorizing requires careful attention to how hypotheses can be tested against observation. Given the unparalleled epistemic success of science, logical empiricists understandably thought philosophical theorizing should follow suit. Appeals to intuition or introspection were considered misguided and unreliable—as recent analyses suggest (e.g. Hintikka 1999, Margolis and Laurence 2003, Schwitzgebel 2008)—and a formal criterion was sought to fortify this judgement by clarifying the exact connection between observation and theory, whether in philosophy or science. It would thereby secure the legitimate epistemic status of theoretical science and some parts of philosophy, while revealing the empirical vacuity of others. The emerging tools of mathematical logic and the formal precision they afforded made finding such a criterion seem eminently attainable.

But the waning popularity of logical empiricism and apparent discovery of insurmountable technical problems led most philosophers to

Mind, Vol. 123 . 490 . April 2014
doi:10.1093/mind/fzu042 Advance Access publication 5 June 2014

abandon the search. In a defence of the similarly maligned concept of testability, Sober (1999, p. 47) captures (but rejects) the overwhelmingly negative sentiment: 'The concept of testability, like the analytic/synthetic distinction, is supposed to be a vestige of a bygone age, whose untenability we celebrate by speaking of the "demise of logical empiricism"'. A recent history of twentieth-century philosophy echoes this sentiment, and makes the sweeping claim that Church's (1949) *specific* (but devastating) criticism of Ayer's (1946) second criterion exposed insuperable difficulties with the *general* project to formulate an empirical significance criterion:

> For all intents and purposes, the collapse of Ayer's final formulation signaled the end of attempts to formulate the empiricist criterion of meaning in terms of either strong or weak verifiability. A few attempts were made to reformulate Ayer's criterion to save it from objections like the ones just considered [from Church and Israel Scheffler]. However, none proved successful. (Soames 2003a, p. 291)

However, scrutiny of the criticisms actually made of later significance criteria, some of which differ radically from Ayer's, provides little support for this view.

Although similarly redeeming analyses of amended versions of Ayer's (1946) criterion could be given (e.g. see Wright's adroit attempt in his 1986), this paper defends Carnap's final criterion of 1956, which has been much less discussed.[1] Section 2 highlights under-appreciated motivations for developing an empirical significance criterion, and explains how the 1956 criterion improved on Carnap's analysis in 'Testability and Meaning'. Section 3 describes the criterion and shows how its primary focus on *terms*, rather than *statements* as in Ayer's criterion, helps avoid difficulties the latter encountered. Section 4 argues Carnap's criterion was seriously misinterpreted and that, unlike Ayer's criterion and against prevailing opinion, criticisms targeting its formal details did not uncover fatal flaws. Section 5 concludes that the project to formulate such a criterion was far from the unmitigated failure it is often asserted to be.

2. Empirical significance

Although an anti-metaphysical agenda drove logical empiricists to develop an empirical significance criterion, there are other compelling

[1] Despite the generality of its asserted judgement, Soames's history does not consider Carnap's 1956 criterion.

motivations. Conceptualized more generally and less controversially, significance criteria simply attempt to indicate whether two domains are epistemically connected, for example, whether one statement class (e.g. observation statements) bears on the truth-values of another (e.g. hypotheses of theoretical science) (Wright 1986). Revealing the basis on which a certain type of content is possessed is the objective — *empirical content* for an empirical significance criterion — but other content types could be targeted (ethical, historical, legal, etc.) (Lewis 1988). This arguably constitutes the first step towards any general theory of empirical content, and content of other kinds. An empiricist theory of meaning, such as the view that non-analytic statements differ in meaning only insofar as they differ in empirical content, also seems to require such a criterion. These legitimate aspirations do not assume science can be demarcated from non-science, and the well-known problems confronting the latter do not redound to the former.

As canonical repositories of empirical content, a significance criterion should illuminate how scientific theories acquire this content through relations with observational data. In particular, Carnap thought the criterion would reveal how the empirical content of parts of theoretical science far removed from experience depends on the empirical content of parts more intimately tied to experience. These dependencies reveal the (partially) holistic nature of empirical content that previous significance criteria were rightly criticized for failing to recognize (Hempel 1950). Carnap understood this deficiency and designed the 1956 criterion to reflect this holism (see below).

Unlike Ayer (1946), whose criterion was intended to apply to natural languages directly, Carnap began by representing theories within a formal language he called a 'total language of science' (Carnap 1956, p. 40). It divides into a theoretical language L_T and an observation language L_O, with a corresponding partition in vocabulary. V_O is the class of non-logical terms of L_O and V_T is the class of *primitive* non-logical terms of L_T. Members of V_O designate observable properties and relations such as 'hard' and 'in physical contact with'. Variables of L_O range over similarly 'concrete, observable entities (e.g. observable events, things, or thing-moments)' (1956, p. 41), and Carnap required that the underlying logic of L_O permit at least one interpretation with a finite domain. L_O is also strictly extensional and constructive, that is, it contains no modalities and Carnap stipulated that every variable value must be designated by an individual term of L_O.

These restrictions reflected a narrow conception of observation, explicitly characterized later as 'directly perceived by the senses'

(Carnap 1966, p. 225). But Carnap recognized that different degrees of restrictivity about L_O were defensible. Towards the end of the 1956 analysis, the possibility of an observation language in a 'wider sense' was discussed (1956, Sect. 9). Such an observation language might include terms with explicit definitions of a 'special form, involving logical and causal modalities', although Carnap admitted 'the exact form of definitions of this kind is at present not yet sufficiently clarified, but still under discussion' (Carnap 1956, p. 64). Despite these enticing remarks, Carnap never pursued the task of developing an empirical significance criterion for such a robust observation language. The relationship between empirical significance and alternative observation languages—and modal languages more generally—remains largely unexplored.

Members of V_T, theoretical terms, designate unobservable properties and relations such as 'electron' and 'electromagnetic field.' According to Carnap (1956, p. 43), L_T 'includes all of mathematics that is needed in science and also all kinds of entities that customarily occur in any branch of empirical science'. Carnap noted that L_T could include modal operators but he thought a strictly extensional language provided sufficient expressive power for science, and thus was an appropriate language in which to formulate an empirical significance criterion.[2] Theories were characterized as finite sets of statements within L_T designated by their conjunction T. Carnap intended T to represent 'fundamental laws of science, and not other scientifically asserted sentences, e.g., those describing single facts' (1956, p. 51), but this was not explicitly required in his criterion.

A finite set of so-called correspondence rules designated by their conjunction C connects terms of V_T and V_O. Beyond requiring that each correspondence rule contain at least one V_T and V_O term, Carnap rightly did not restrict their form. His first major publication in English twenty years before, 'Testability and Meaning' (Carnap 1936–7), had been less flexible. There, Carnap showed how a particular

[2] This was an assumption of tractability, not philosophical principle or necessity:

> Other connectives, e.g., signs for logical modalities (e.g. logical necessity and strict implication) and for causal modalities (e.g. causal necessity and causal implication) may be admitted if desired; but their inclusion would require a considerably more complicated set of rules of logical deduction (as syntactical or semantical rules). (Carnap 1956, p. 42)

Unlike Quine's, Carnap's empiricism did not require extensionalism.

type of correspondence rule, a 'reduction postulate', could provide observational testing conditions for theoretical terms representing dispositions, such as Sx = 'x is soluble in water'. Two reduction postulates do so for Sx:

(RP1) $(\forall x)(O_1x \rightarrow (O_2x \rightarrow Sx))$

(RP2) $(\forall x)(O_3x \rightarrow (O_4x \rightarrow \neg Sx))$

where $O_1 - O_4$ designate observation terms. To illustrate, O_1x and O_2x might represent 'x is placed in water' and 'x dissolves in water,' respectively, while O_3 and O_4 might represent other observable conditions. If $O_1x \leftrightarrow O_3x$ and $O_2x \leftrightarrow \neg O_4x$, the pair reduces to a 'bilateral reduction statement':

(BRS) $(\forall x)(O_1x \rightarrow (O_2x \leftrightarrow Sx))$

Reduction postulates establish a very specific kind of connection between disposition and observation terms, but in 1936 Carnap believed they would reveal the observational basis of all terms of theoretical science (dispositional or otherwise). With sets of reduction postulates of greater scope and complexity (e.g. chains of nested reduction postulates; see Carnap 1936–7, Sect. 9) Carnap showed how relations between observation and theoretical terms could be clarified and expanded.

The analysis was influential, but also encountered serious difficulties. Two problems significantly shaped Carnap's 1956 criterion. First, Carnap (1936–7) considered *individual* bilateral reduction statements to be analytic because all their logical consequences containing only observation terms are logical truths.[3] But Hempel (1951, pp. 71–2) showed that *together* two (or more) bilateral reduction statements sometimes entail statements containing only observation terms that are not logically true. A bilateral reduction statement's analyticity and the nature of the 'bridge' it provides between theory and observation therefore depends upon what other bilateral reduction statements obtain. A bilateral reduction statement could be analytic in one theoretical context but synthetic in another composed of different reduction postulates. For Hempel, this suggested that empirical significance should be relativized to theoretical context, which Carnap's 1956 criterion does explicitly (see Sect. 3).

Carnap (1939) foresaw another deficiency which Hempel (1951) later publicized: reduction postulates could not capture the relationship

[3] This fails for reduction postulates. (RP1) and (RP2), for instance, entail $\neg \exists x(O_1x \,\&\, O_2x \,\&\, O_3x \,\&\, O_4x)$, which is not logically true.

between observation and many indispensable features of theoretical science. Reduction postulates constitute only one type of connection between observation and theory. How their simple conditional form could accommodate more complex and abstract scientific theorizing — metrics such as Euclidean distance used to describe unobservable intervals, principles of differential geometry and higher mathematics generally, various field concepts, and others — was entirely unclear from the 1936–7 analysis. Carnap (1956) recognized this shortcoming and adopted an appropriately permissive view about the form of correspondence rules connecting theoretical and observation terms. He thereby anticipated and circumvented Glymour's (1980, p. 62) criticism that focusing exclusively on reduction postulates presupposes an unjustifiably narrow observation-theory relationship.

Despite these improvements on 'Testability and Meaning', the framework still appears wanting from a contemporary perspective. Carnap's reliance on a notion of analyticity and an apparently sharp division between theory and observation seems especially problematic (see Quine 1951a and Sellars 1956). These legitimate concerns cannot be fully addressed here, but they have been elsewhere. Although it was once considered the death knell of logical empiricism, the attack on analyticity by Quine and others has lost much of its lustre (Creath 1991, Chalmers 2011). It is also widely supposed that logical empiricists presupposed a dubious foundationalist conception of observation and upheld a similarly questionable absolute dichotomy between observation and theory. Carnap had no such commitments (see Creath 1987). Rather, Carnap favoured the same context-dependent, pragmatic account of the relationship between theory and observation that grounds the ubiquitous testing of the former against the latter in scientific work (see Sober 2008, pp. 152–3). The canard that Carnap and logical empiricists in general held such an implausible view should be retired.

There is another worry about the framework: its representation of scientific theories as statements of first order logic seems simplistic and antiquated when gauged against the last half century of philosophical work on the function and form of models and theories in scientific practice. But criticizing Carnap's criterion on this basis misconstrues his objective, which was *explication* of the concept of empirical significance (Carnap 1956, p. 49). Rather than attempt to illuminate the often intuitive but nevertheless vague notion that observation is 'relevant' to a statement's truth value through informal methods,[4] explication in this case requires formulating a formal criterion that precisely demarcates the empirically significant from the non-significant (see Carnap 1950, Ch. 1). This, in turn, requires a high degree of precision about the potential structure of scientific theories and what they say about the observable world. Carnap recognized that the framework idealizes rather than straightforwardly describes most theoretical science, but the simplifying assumptions and quasi-axiomatic systematicity were intended to achieve the requisite precision to make the problem tractable. This tactic pervades philosophical and scientific work. Similar idealizations are frequently employed in scientific modelling to make mathematical representations of real-world systems tractable (Morgan and Morrison 1999). Carnap did not deny that more sophisticated representations of scientific theories and languages might necessitate modifications of the criterion he developed.[5] And even with its significant improvements over reduction postulates and Ayer's (1946) criterion (see Sect. 3), Carnap believed only that the 1956 criterion advanced, rather than completed, the general project to find an empirical significance criterion.

It is worth emphasizing that these worries about the framework of Carnap's criterion are ancillary, at the periphery of the prevailing view that attempts to find such a criterion are doomed to fail. The core contention is that any proposed criterion would exhibit the kind of irredeemable logical deficiencies affecting Ayer's criterion. The concept of empirical significance itself, not the scaffolding in which it was presented, is invariably fingered as the failing. This may help explain why the undoing of Ayer's *specific* criterion is accorded such significance in *general* assessments of logical empiricists' search for such a criterion: the judgement was that the criterion is hopelessly flawed because the very concept of empirical significance is hopelessly flawed. The fact that criticisms targeting the particular formal details of Carnap's criterion (and that largely overlook the auxiliary framework accompanying it) are supposed to unequivocally and irrevocably expose its failings also reflects this sentiment (see Glymour 1980).

[4] After decisively criticizing the criterion found in the first edition of *Language, Truth, and Logic*, Isaiah Berlin (1939, p. 293) keenly warned that: 'Relevance is not a precise logical category, and fantastic metaphysical systems may choose to claim that observation data are "relevant" to their truth'.

[5] See the discussion of L_T and L_O above. Although most sciences do not exhibit this level of formal precision and systematicity, some do. Quantum mechanics, relativity theory, and perhaps formal learning theory are plausible candidates.

Before responding to those criticisms in section 4, the next section describes the 1956 criterion.

3. Carnap's 1956 criterion

Carnap presented the significance criterion in three 'definitions' (recall that T represents a theory and C a set of correspondence rules):

(D1) A theoretical term M is *empirically significant relative* to a class K with respect to $L_T, L_O, T,$ and $C =_{df}$
  (i) $K \subset V_T$
  (ii) $M \notin K$
  (iii) there are statements $S_M, S_K \in L_T$, and $S_O \in L_O$ such that:
    (a) S_M contains M as the only non-logical term
    (b) the non-logical terms in S_K belong to K
    (c) $(S_M \,\&\, S_K \,\&\, T \,\&\, C)$ is consistent
    (d) $(S_M \,\&\, S_K \,\&\, T \,\&\, C) \vDash S_O$
    (e) $\neg\,[(S_K \,\&\, T \,\&\, C) \vDash S_O$

(D2) M_n is empirically significant with respect to $L_T, L_O, T,$ and $C =_{df}$ there is a sequence of theoretical terms $\langle M_1, \ldots, M_n \rangle$ $(M_i \in V_T)$ such that every M_i is *significant relative* to $\{M_1, \ldots, M_{i-1}\}$ with respect to $L_T, L_O, T,$ and C

(D3) An expression A of L_T is an empirically significant statement of $L_T =_{df}$
  (i) A satisfies the rules of formation of L_T
  (ii) every non-logical term in A is significant as in (D2)

As with Ayer's criterion, these definitions explicate the idea that empirical significance requires making an observable predictive difference. For a term M to be empirically significant, there must be a statement S_M containing M as its only theoretical term (D1[a]) in which S_M is indispensable in the non-trivial derivation of at least one observation statement (D1[c–e]). The indispensability may be mediated: the derivation can depend on other theoretical terms K (D1[ii,iii]), those occurring in S_K(D1[b]). Given that Carnap and Ayer were attempting to capture the same idea, the broad similarity is unsurprising. It is the differences, however, that reveal the ingenious advances of Carnap's criterion.

Unlike Ayer's criterion, *term* significance is fundamental (D1, D2) and statement significance is derivative (D3); the significance of terms determines the significance of statements containing them. By making statement significance parasitic in this way, Carnap precludes criticisms that exploit the internal structure of non-atomic statements to establish disastrous results. For example, (D3[ii])'s requirement that statement significance requires *all* non-logical terms be significant prevents criticisms such as Church's (1949) use of the truth-functional position of S in $(\neg O_1 \,\&\, O_2) \vee (\neg S \,\&\, O_3)$ to show that Ayer's criterion judges almost any statement empirically significant.[6] The assertion that Carnap's criterion 'falls to a sophisticated version of Alonzo Church's counterexample (Church 1949) to Ayer's earlier attempt' (Anderson 2009, p. 17) misses this point.

Note that the criterion involves two relativizations. The 'with respect to' clauses of (D1) and (D2) relativize empirical significance to a language (L_T and L_O), and a theoretical context (T and C) as Hempel suggested. (D1) additionally relativizes the significance of the individual term M to a class of terms K; (D2) then describes how such dependencies between the significance of theoretical terms can occur. This is intended to reflect scientific methodology: theoretical terms and statements often bear on observations only in interdependent groups. Testing hypotheses in theoretical physics or evolutionary biology, for instance, usually requires methods of experimental design, data acquisition, and statistical analysis that are in turn informed by other theoretical concepts and results. To the extent dependencies between the significance of theoretical terms track these testing dependencies, (D2) captures one of the methodological insights categorized under the label the 'Duhem-Quine thesis'.

One implication of (D2) is critical and contrasts sharply with the Duhem-Quine thesis, at least in the strong form promulgated by Quine (1951a). To avoid problematic circularity, there must be some part(s) of theory that observation bears on directly, without the mediation of any other theoretical terms or statements. (D2) reflects this requirement by assuming at least one theoretical term is empirically significant through a direct connection with observation, that is, unfacilitated by other theoretical terms (when $\{M_1, \ldots, M_{i-1}\}$ is null).[7] Such directly significant terms would supply the first term in the

[6] Wright's (1986) notion of compact entailment deploys a similar strategy for circumventing such criticisms.

[7] Carnap (1956) scarcely discussed examples of direct significance. In the simplest case, a term's direct significance would be shown (in part) by it being the sole theoretical term in a correspondence rule. With respect to the theoretical term 'mass', 'one of the physical terms most closely related to observational terms' (1956, p. 67), Carnap (1956, p. 48) suggested that '[t]he rule may connect the theoretical term 'mass' with the observable term 'heavier than' as

sequence $\langle M_1, \dots, M_{i-1}\rangle$ and thereby undergird the significance of other theoretical terms according to (D2). Attempting to leverage this assumption into a systematic defect was the agenda of perhaps the most well-known criticism of Carnap's criterion due to Kaplan (1975).[8]

4. Formal challenges

Definition (D2) entails theories cannot have empirical significance without directly significant terms. Kaplan (1975) argued this fact establishes Carnap's criterion is too restrictive. The argument utilizes the concept of 'deoccamizing' $T\&C$. Deoccamization involves replacing all instances of some theoretical terms in $T\&C$ with disjunctions or conjunctions of new theoretical terms; an Occam-unfriendly addition of terms. To illustrate, one deoccamization of $(\forall x)(fx \rightarrow bx)$ on f and b yields $(\forall x)[(f_1x \vee f_2x) \rightarrow (b_1x \vee b_2x)]$ for new theoretical terms f_1, f_2, b_1, and b_2. Kaplan proved deoccamization preserves the deductive systematization of L_O by $T\&C$. For instance, if $T\&C$ entails $O_1 \rightarrow O_2$ where O_1 and O_2 are observation statements, any deoccamization of $T\&C$ also entails $O_1 \rightarrow O_2$. This seems to motivate Kaplan's (1975, p. 91) intuition that:

> Although we would not look with favor upon such a multiplication of entities beyond necessity, I think we would not say that a deoccamization of a theory and its connecting postulates can rob it of empirical content.

The intuition was not defended and the notion of 'empirical content' was left unspecified, but presumably the theoretical terms of $T\&C$ must be empirically significant for $T\&C$ to have empirical content. Kaplan's intuition therefore seems to be that new theoretical terms introduced by deoccamization of empirically significant terms should also be empirically significant.

---

follows: "If u is heavier than v, the mass of u′ (i.e, the mass of the coordinate region u′ corresponding to u) is greater than the mass of v′"', 'heavier' being an observation term.

[8] Directly significant terms may seem incompatible with Carnap's (1937) early endorsement of Duhem's original thesis. But besides Duhem restricting the original thesis to physics (Ariew 1984), the relevant portion of the endorsement was qualified: 'Further, it is, *in general*, impossible to test even a single hypothetical sentence' (Carnap 1937, p. 318; emphasis added). The claim it is *generally* impossible to test individual theoretical statements is consistent with some theoretical terms (though perhaps few) alone making observable predictions (as S_M and in conjunction with C), thereby establishing their direct significance. Directly significant terms are incompatible with Quine's (1951a) sweeping epistemological holism, but Carnap rejected this view and for good reasons others have done the same (e.g. Glymour 1980, Sober 2000).

To show Carnap's criterion was too restrictive, Kaplan argued as follows. If any members of V_T are empirically significant with respect to T, C, L_T, and L_O, at least one M_1 exists that is significant relative to null K ($K=\varnothing$). If $T\&C$ is deoccamized to $T'\&C'$ so that M_1x becomes $M_{11}x \vee M_{12}x$ (accompanied by additional quantifiers if necessary), each new disjunct is never found outside this disjunction in $T'\&C'$.[9] Since $M_{11}x$ and $M_{12}x$ are always connected by disjunction and never found alone, the logical relations that M_1 originally exhibited that satisfied (D1) relative to the null K cannot similarly satisfy (D1) for M_{11} or M_{12} individually. Thus, the sequence of theoretical terms $M_1, \dots, M_{i-1}$ required by (D2) would have no first member and no chain of implications establishing the empirical significance of successive theoretical terms *relative to* terms already shown significant would exist. Although deoccamization preserves deductive systematizations of L_O, Carnap's criterion judges every theoretical term of $T'\&C'$ empirically non-significant. It is thus too restrictive since deoccamization, Kaplan claimed, should not divest theoretical terms of empirical significance.

Although the formal inferences Kaplan made are unassailable, the intuition underlying his criticism is suspect. Consider 'equivalence-deoccamization' where $M_{i1}x \equiv M_{i2}x$ replaces each M_ix instead of conjunctions or disjunctions of new terms M_{i1} and M_{i2}. Like deoccamization, equivalence-deoccamization preserves the deductive systematization of L_O, and the initial member of $M_1, \dots, M_{i-1}$ required by (D2) cannot be found because (D1) is never satisfied for $K=\varnothing$. Kaplan's derivation therefore follows similarly for equivalence-deoccamization. But it is entirely unclear what ground preserving L_O's deductive systematization provides, following Kaplan, for the view that equivalence-deoccamization does not rob theoretical terms of empirical significance. For example, if a C-rule for a theoretical term S and observation terms D and W such as $\forall x[Wx \rightarrow (Dx \leftrightarrow Sx)]$ (e.g. 'if x is submerged in water, x dissolves iff x is soluble') were equivalence-deoccamized to $\forall x[Wx \rightarrow (Dx \leftrightarrow (S_1x \equiv S_2x))]$, the claim that the significance of S must ensure S_1 and S_2 are also significant seems entirely unmotivated. Equivalence-deoccamization is, after all, in some way treating S as the property of *equivalence* between S_1 and S_2, which differs decidedly from the properties S_1 and S_2 designate. One would expect this marked difference — and the differences ordinary deoccamization

[9] Although Kaplan (1975, p. 91) sometimes speaks of deoccamization as 'splitting' theoretical terms, note that the relevant splitting is at the sentential level.

imposes — to make a difference concerning significance. The fact that such a difference can affect significance is therefore unsurprising. Without a compelling reason to accept Kaplan's intuition, preservation of the deductive systematization of L_O seems to provide unreliable guidance about the empirical significance of theoretical terms.

Nor does there seem to be any other plausible basis for claiming deoccamization should preserve empirical significance. Although Glymour (1980, p. 44) endorsed Kaplan's criticism and thought the criterion was fatally flawed, his assessment of deoccamized theories actually supports the above response. Glymour (1980, p. 32) accurately described deoccamized theories as:

> just the sort of theories that theorists abjure; physicists say they have 'redundant quantities' or 'unobservable quantities' and regard them with suspicion and worse … without appropriate restrictions, the hypothetico-deductive view is committed to the legitimacy of deoccamized theories, and that commitment may not accord with either intuitive judgment or scientific practice.

According to Glymour, 'unrestricted hypothetico-deductivism' presumes that if observations confirm scientific theory T, they confirm all its deoccamizations. Glymour reasonably suggested that neither scientific practice nor intuition accord with this claim, which in turn suggests deoccamization should not preserve empirical significance. Glymour (1980) criticized the supposed hypothetico-deductivism underlying Carnap's criterion, but the criterion's evaluation of deoccamized theories as empirically non-significant accords with his assessment.

Problems with the intuition underlying Kaplan's argument yield qualms but not a definitive response. Although Kaplan did not supply it, perhaps an intuitively compelling account of empirical content countering these considerations could be concocted. Fortunately, Creath (1976) has obviated the need for such a philosophical exploration. He showed how a generalization of Carnap's criterion to *sets* of terms (J below) could accommodate Kaplan's intuition:

(D1′) A theoretical term M is *significant relative* to a class K with respect to $L_T, L_O, T,$ and $C =_{df}$

- (i) $K \subset V_T$
- (ii) $M \notin K$
- (iii) there is a class J such that $J \subset V_T, M \in J, J \cap K = \varnothing$

(iv) there are statements $S_J, S_K \in L_T$, and $S_O \in L_O$ such that:
  (a) S_J contains members of J as the only non-logical terms
  (b) the non-logical terms in S_K belong to K
  (c) $(S_J \,\&\, S_K \,\&\, T \,\&\, C)$ is consistent
  (d) $(S_J \,\&\, S_K \,\&\, T \,\&\, C) \models S_O$
  (e) $\neg[(S_K \,\&\, T \,\&\, C) \models S_O]$
  (f) it is not the case that $(\exists J')(J' \subset J)$, and statements $S_{J'}, S_{K'} \in L_T$, and $S_{O'} \in L_O$ such that:
    (f1) $S_{J'}$ contains members of J' as the only non-logical terms
    (f2) the nonlogical terms in $S_{K'}$ belong to K
    (f3) $(S_{J'} \,\&\, S_{K'} \,\&\, T \,\&\, C)$ is consistent
    (f4) $(S_{J'} \,\&\, S_{K'} \,\&\, T \,\&\, C) \models S_{O'}$
    (f5) $\neg[(S_{K'} \,\&\, T \,\&\, C) \models S_{O'}]$

(D2′) M_n is significant with respect to L_T, L_O, T, and $C =_{df}$ there is a sequence of sets $\langle J_1, \ldots, J_n \rangle (M_n \in J_n$ and $J_i \subset V_T)$ such that every member of every set J_i is *significant relative* to $\bigcup_{p=1}^{p=(i-1)} J_p$ with respect to L_T, L_O, T, and C

Any term significant by Carnap's 1956 criterion is significant by (D1′) and (D2′). (D3) is as before. The main change is allowing sets of terms (e.g. deoccamized terms) to be empirically significant even if, taken individually, they are not on Carnap's original criterion. Condition (f) in particular ensures each theoretical term in J is required to establish the significance of M and thus (D1′[a–e]). Should Kaplan's intuition prove defensible, the generalized criterion would accommodate it while preserving the spirit of the original.

Kaplan (and Rozeboom, independently — see Rozeboom 1960) offered a second kind of criticism of Carnap's criterion: adding statements (even definitions) to $T \,\&\, C$ may change the empirical significance of theoretical terms, which is counterintuitive. Consider an example from Kaplan (1975) where $V_O = \{J_O, P_O, R_O\}$; L_O includes all statements of first-order logic with identity containing only non-logical constants from V_O; $V_T = \{B_T, F_T, G_T, H_T, M_T, N_T\}$; and L_T includes all statements of first-order logic with identity containing theoretical terms from V_T. Let T and C be:

(T) $[(\forall x)(H_T x \rightarrow F_T x)] \wedge [(\forall x)(H_T x \rightarrow (B_T x \vee \neg G_T x))] \wedge [(\forall x)(M_T x \equiv N_T x)]$

(C) $[(\forall x)(R_O x \rightarrow H_T x)] \wedge [(\forall x)(F_T x \rightarrow J_O x)] \wedge [(\forall x)(G_T x \rightarrow P_O x)]$

It is straightforward to show that G_T, F_T, and H_T are significant relative to $K=\varnothing$ (by (D1)) and therefore significant with respect to L_O, L_T, T, and C (by D2); B_T is significant relative to $K = \{G_T\}$; and, M_T and N_T are not empirically significant.

Now consider a conservative definitional extension T' in an extended vocabulary $V_{T'}$ and language $L_{T'}$ that adds two definitions to T:

(DF1) $(\forall x)(D_1 x \equiv (M_T x \,\&\, (\exists x) F_T x))$

(DF2) $(\forall x)(D_2 x \equiv (M_T x \to ((\exists x) G_T x))$

$S_M = \forall x D_1 x$, $S_K = \forall x(x = x)$, and $S_O = \exists x J_O x$ show D_1 is significant relative to $K=\varnothing$ (by (D1)) and therefore significant with respect to T', C, L_O, $L_{T'}$ (by D2). $S_M = \forall x D_2 x$, $S_K = \exists x D_1 x$, and $S_O = \exists x P_O x$ show D_2 is significant relative to $K = \{D_1\}$. $S_M = \forall x M_T x$, $S_K = \exists x D_2 x$, and $S_O = \exists x P_O x$ show M_T—previously not significant—is now significant relative to $K = \{D_2\}$ and therefore significant with respect to T', C,L_O, and $L_{T'}$. A similar definitional extension changes N_T from non-significant to significant. Definitional extensions could, therefore, bestow theoretical terms with empirical significance on Carnap's criterion. For Kaplan, this meant the criterion was unacceptably liberal because 'definitional extensions are ordinarily thought of as having no more empirical content than the original theory' (Kaplan 1975, p. 90).

Rozeboom (1960) made the converse criticism that the criterion was unacceptably restrictive. Although an explicit proof was not provided, Rozeboom correctly claimed that extending $T \,\&\, C$—not necessarily definitionally and thus not necessarily conservatively—could rob theoretical terms of empirical significance. If S_M, S_K, and S_O show M is empirically significant with respect to T, C, L_O, and L_T, then adding postulates (not necessarily definitions) to $T \,\&\, C$ may make S_M, S_K, and S_O fail (D1[e]) such that no other statements satisfy (D1). In effect, the augmented $T \,\&\, C$ appropriates any role M may have had in making an observable predictive difference. Moreover, if T' & C' is a maximally L_O-consistent extension of $T \,\&\, C$ in an extended language $L_{T'}$, no term of $L_{T'}$ can be significant. For any S_O in L_O, $T' \& C'$ entails either S_O or $\neg S_O$. Hence, for any S_O either: (D1[e]) fails because $T' \& C'$ already entails S_O, (D1[d]) fails because $S_M \,\&\, S_K \,\&\, T' \,\&\, C'$ is consistent and entails $\neg S_O$, or (D1[c]) fails because $S_M \,\&\, S_K \,\&\, T' \,\&\, C'$ is inconsistent. For Rozeboom, this indicated Carnap's criterion was too restrictive since extensions of $T \,\&\, C$ should not, intuitively, divest theoretical terms of empirical significance.

There are at least three defences of Carnap's criterion. First, as Kaplan (1975, p. 93) himself noted, Carnap restricted the criterion to primitive, non-defined theoretical terms of V_T. It was therefore explicitly formulated to avoid criticisms based on leveraging definitional extensions. Of course, one could reasonably object that defined terms are and will continue to be essential in any 'language of science', and that an empirical significance criterion must account for them. Creath (1976) showed this could be accommodated with a simple modification: for defined terms to be empirically significant, terms within their *definiens* must be shown significant antecedently. But this fix is not fully satisfactory. It affirms a distinction between primitive and non-primitive terms now largely considered conventional or obsolete, and sharply distinguishes definitions from other theoretical postulates. Modern accounts of definition do not support this division, regarding definitions simply as postulates, albeit of a particular syntactic form (Feferman et al. 2000). A different response to the definitional extension criticism is therefore needed.

The second response rejects the claim that extending $T\&C$, definitionally or otherwise, should not change the empirical significance of theoretical terms. Consider Kaplan's criticism first. Although M_T and N_T were initially not empirically significant with respect to T, C, L_O, and L_T, DF1 and DF2 establish clear connections between them and observation terms. For example, DF2 and the third conjunct of C establish a connection between M_T and P_O *via* D_2 and relative to $K = \{D_1\}$. In particular, if at least one individual a is D_2—which S_K from Kaplan's criticism stipulates (see above)—the conditional on the right side of DF2 holds for that individual (i.e. $M_T a \rightarrow \exists x G_T x$). If, furthermore, all individuals are M_T—which S_M from Kaplan's criticism again stipulates—the right side of DF2 and the third conjunct of C (i.e. $\forall x[G_T x \rightarrow P_O x]$) together ensure $\exists x P_O x$ (S_O from Kaplan's criticism). Since the existential statement on the right side of DF1 (i.e. $\exists x F_T x$) and the second conjunct of C (i.e. $\forall x[F_T x \rightarrow J_O x]$) ensure D_1 is connected to J_O, and D_2 is significant relative to $K = \{D_1\}, M_T$ is significant with respect to T', C, L_O, and $L_{T'}$.

Rather than reveal a problem, the example shows directly how definitions can change whether a theoretical term *potentially* bears on observation statements. Recall, a *modal* claim is the target of assessment: *could* M_T make a predictive difference in the observable realm. That it could does not mean it does. Assessing the latter with respect to $T\&C$ requires plausibility restrictions on S_K and S_M: S_K must be true, highly probable, well-confirmed, or at least credible on other grounds, and S_M should also be reasonable. Whether M_T makes an actual observable predictive difference can only be gauged if S_K and S_M are restricted in this way. Letting S_M be $\forall x D_1 x$ or $\forall x D_2 x$ as in Kaplan's argument clearly fails this requirement. In contrast, the almost complete absence of strictures on the content and logical form of S_K and S_M in Carnap's criterion was intended to facilitate evaluation of the modal claim. Unconstrained specification of S_K and S_M — that is, specification irrespective of their truth value and other epistemic merits or demerits — permits evaluation of whether implausible, even absurd counterfactual S_K and S_M *could* establish a connection between M_T and observation. With the addition of DF1 and DF2, Kaplan's argument shows that the formerly non-significant M_T can play an essential role in deriving an observation statement, $\exists x P_O x$, albeit under far-fetched conditions (i.e. $\forall x D_1 x, \forall x D_2 x, \forall x M_T x$). Kaplan's (1975, p. 90) intuition that, 'definitional extensions are ordinarily thought of as having no more empirical content than the original theory' is only plausible if 'empirical content' is narrowly construed and S_K and S_M are similarly restricted to what is true, highly likely, or otherwise credible. The concept of empirical significance Carnap's criterion assesses, however, is a much broader modal notion that considers counterfactual scenarios.[10] Against it and a suitably broad conception of 'empirical content' Kaplan's own derivation shows why his intuition is untenable.[11]

It should now be evident why non-definitional extensions of $T\&C$ may change the significance of theoretical terms, contrary to Rozeboom's intuition. A simple example further illustrates the point. Consider:

$$(T) \quad [(\forall x)(Ax \vee Dx)]\&[(\forall x)(Bx \vee Ex)]\&[(\forall x)(Cx \vee Fx)]$$

$$(C) \quad [(\forall x)(Dx \rightarrow O_1 x)]\&[(\forall x)(Ex \rightarrow O_2 x)]\&[(\forall x)(Fx \rightarrow O_3 x)]$$

[10] Note that a modal concept (empirical significance) is being assessed within a non-modal logical framework. The ultimate tenability of Carnap's approach in 1956 to empirical significance therefore seems to depend on whether the modal reduces to the non-modal, or whether the latter can adequately represent the former (*cf.* Sider 2005). This important issue and potential difficulty is not, however, responsible for the target of this analysis: the widespread judgement that technical criticisms have shown Carnap's criterion of empirical significance is irredeemably flawed.

[11] Kaplan's claim about the empirical vacuity of definitions is questionable even on the narrow rendering of 'empirical content'. If definitions in scientific contexts are simply additional theoretical postulates of a particular syntactic form, they can be expected to sometimes affect the actual empirical content of what they definitionally extend.

where $A, B, C, D, E, F \in V_T$ and $O_1, O_2, O_3 \in V_O$. $S_M = \forall x(\neg Bx)$, $S_K = \forall x(x = x)$, and $S_O = \exists x O_2 x$ show B is significant relative to $K = \emptyset$. But if $\forall x(\neg Bx)$ is added to T, S_M, S_K, and S_O no longer show B is significant because (D1[e]) is violated ($T' \& C \vdash S_O$) and no other statements show B is significant. B becomes non-significant because the addition severs B's previous connection to O_2 through the second conjunct of T and C, rendering the theoretical term inert to observation. The example is admittedly simple, but it requires little insight to recognize that some scientific advances in which postulated theoretical entities or properties lose their empirical credentials—for example, abandoning concepts such as phlogiston, luminiferous ether, and absolute simultaneity[12]—seem to reflect the pattern Carnap's criterion captures, albeit in a much more complicated and interesting way.

A third response appeals to Carnap's broader views. Carnap (1956) insisted that terms are significant *within particular languages* and *for particular T and C.* Although his criterion yields cogent judgements for the transformations of $T \& C$ Kaplan and Rozeboom considered, Carnap was not proposing a criterion applicable across all changes in theory and/or language. Rather, the criterion explicated the concept of empirical significance *in particular theoretical contexts and languages*, not significance *simpliciter*. The same relativization characterizes Carnap's explication of analyticity:

> [with respect to] a transition from a language L_n to a new language L_{n+1} ... My concept of analyticity as an explicandum has nothing to do with such a transition. It refers in each case to just one language; 'analytic in L_n' and 'analytic in L_{n+1}' are two different concepts. (1963, p. 921)

Carnap did not refrain from attempting to formulate a criterion applicable across all changes in theory and/or language out of excessive caution or restraint. Rather, it reflected a deep commitment to a particular philosophical methodology: the internal-external framework described in 'Empiricism, Semantics, and Ontology' (Carnap 1947a).

The methodology depends on the concept of a framework and a sharp distinction between questions external to a framework and questions internal to it. A framework is a complete 'way of speaking': a comprehensive set of linguistic and derivational rules that determines what statements are expressible, what inferences are permissible, and what standards are appropriate for assessing claims

[12] Reichenbach (1958) argued Einstein's analysis of simultaneity rendered the concept of absolute simultaneity empirically non-significant and raised doubts about the significance of other concepts.

involving truth, justification, evidence, and other semantic and epistemological concepts. The 'total language for science' in which the 1956 criterion is formulated is such a framework, one Carnap believed was sufficient for science.[13] Questions internal to this framework (T, C, L_T, and L_O in this context) are logico-mathematical or empirical and are usually answered by the framework's properties themselves (for the former) or by analyzing observational data with the statistical methods the framework permits (for the latter) (Carnap 1947a, pp. 208–9).

External questions, however, concern pragmatic decisions about adopting different frameworks. Their answers are therefore not true or false, but instead depend entirely on the purposes for which the framework is employed and its efficiency achieving them. The changes to T and C that Kaplan and Rozeboom consider constitute framework changes. To demand that the empirical significance of theoretical terms remain invariant across such changes would strike Carnap as confused, mistaking a pragmatic external issue for an internal one. Whether a term is empirically significant is determined by the theoretical context and language in which it occurs — that is, the framework. The 1956 criterion captures this dependency. Convictions about empirical significance somehow floating free from a framework — for example, Kaplan and Rozeboom's intuitions about how empirical significance should change across frameworks (as T & C changes) — rely on philosophical commitments Carnap's (1947a) methodology counsels against and attempts to supplant.

Of course, if Carnap's preferred methodology is misguided, all bets are off. And there are formidable obstacles to overcome, such as Quine's influential critique (Quine 1951b) concerning what Carnap could have meant by 'framework' (see also Eklund 2009). Such an ambitious task cannot be undertaken here, but it is worth noting that, as it has concerning analyticity, Quine's final word on Carnap on philosophical methodology has lost much of its finality (Price 2009).

[13] The languages L_1 and L_2 from the *Logical Syntax of Language* are frameworks of different expressive power for mathematics (Carnap 1937). See Goldfarb and Ricketts (1992) for a compelling defence of Carnap's methodology in philosophy of mathematics.

5. Conclusion

Although optimistic in 1950,[14] a year later Hempel lamented:

> All the criteria considered so far attempt to explicate the concept of empirical significance by specifying certain logical connections which must obtain between a significant sentence and suitable observation sentences. It now seems that this type of approach offers little hope for the attainment of precise criteria of meaningfulness. (1951, p. 64)

Hempel's pessimism pervades contemporary assessments of the search for a criterion of empirical significance. Glymour (1980) judges it a failure and cites Rozeboom 1960 and Kaplan 1975 as definitive proof that Carnap's criterion fails. Without analysis of Carnap's (1956) criterion (or Wright's 1986 significant improvement on Ayer 1946), Soames (2003a, p. 297) characterizes the search as a 'failed attempt'. Without intending hyperbole, Kaplan's criticism has been claimed to '[mark] the philosophical end to logical positivism as the dominant force in philosophy in the twentieth century' (Anderson 2009, p. 17). Eclipsed by the tendentious analytic–synthetic debate in the 1950s and 1960s, later more sophisticated criteria received little scrutiny, and the sentiment that the entire project was an instructive but categorical failure ossified.

The analysis presented here suggests a more sanguine evaluation. Appropriately amended, Carnap's criterion exhibits no formal flaw akin to those exposed in Ayer's and remains defensible. Nonetheless, the project may ultimately prove untenable. The formal precision from which significance criteria derive normative force also exposes them to technically ingenious critique. But Rozeboom (1960) and Kaplan (1975) are not the Church (1949) of Carnap's criterion. At least for now, Carnap's enduring optimism (see his 1963) concerning the search for a formal criterion of empirical significance remains sound.[15]

[14] '[I]t is to be hoped that before long some of the open problems encountered in the analysis of cognitive significance will be clarified and that then our last version of the empiricist meaning criterion will be replaced by another, more adequate one' (Hempel 1950, p. 62).

[15] Thanks to Rick Creath, Cory Juhl, Hannes Leitgeb, Sebastian Lutz, Mark Sainsbury, Sahotra Sarkar, Elliott Sober, and Thomas Uebel for helpful comments over this paper's leisurely gestation. Audiences at Sydney University, Australian National University, and the 2010 US Philosophy of Science Association meeting in Montréal provided similarly valuable feedback. In part, this work was generously supported by the Sydney Centre for the Foundations of Science.

References

Achinstein, Peter and Stephen Barker (eds) 1969: *The Legacy of Logical Positivism*. Baltimore, MD: Johns Hopkins Press.

Almog, Joseph and Paolo Leonardi (eds) 2009: *The Philosophy of David Kaplan*. New York: Oxford University Press.

Anderson, C. Anthony 2009: 'David Kaplan: Formal Aspects of His Work'. In Almog and Leonardi 2009, pp. 11–24.

Ariew, Roger 1984: 'The Duhem Thesis'. *The British Journal for the Philosophy of Science*, 35, pp. 313–25.

Ayer, A. J. 1946: *Language, Truth, and Logic*. 2nd edition. New York: Dover Publications Inc.

Bell, David and Wilhelm Vossenkuhl (eds) 1992: *Science and Subjectivity*. Berlin: Akademie Verlag.

Carnap, Rudolf 1936–7: 'Testability and Meaning'. *Philosophy of Science*, 3, pp. 419–71, 4, pp. 1-40.

——1937: *The Logical Syntax of Language*. London: Routledge Press.

——1939: *Foundations of Logic and Mathematics*. Chicago: University of Chicago Press.

——1947a: 'Empiricism, Semantics, and Ontology'. In Carnap 1947b, pp. 205–21.

——1947b: *Meaning and Necessity*. Chicago: University of Chicago Press.

——1950: *Logical Foundations of Probability*. Chicago: University of Chicago Press.

——1956: 'The Methodological Character of Theoretical Concepts'. In Feigl and Scriven 1956, pp. 38–76.

——1963: 'Intellectual Autobiography'. In Schlipp 1963, pp. 3–86.

——1966: *Introduction to the Philosophy of Science*. New York: Dover Publications.

Chalmers, David 2011: 'Revisability and Conceptual Change in "Two Dogmas of Empiricism"'. *Journal of Philosophy*, 108, pp. 387–415.

Chalmers, David, David Manley, and Ryan Wasserman (eds) 2009: *Metametaphysics: New Essays on the Foundations of Ontology*. New York: Oxford University Press.

Church, Alonzo 1949: 'Review of the Second Edition of *Language, Truth and Logic*'. *Journal of Symbolic Logic*, 14, pp. 52–3.

Creath, Richard 1976: 'Kaplan on Carnap on Significance'. *Philosophical Studies*, 30, pp. 393–400.

——1987: 'Some Remarks on "Protocol Sentences"'. *Noûs*, 21, pp. 471–75.

——1991: 'Every Dogma Has Its Day'. *Erkenntnis*, 35, pp. 347–89.
Eklund, Matti 2009: 'Carnap and Ontological Pluralism'. In Chalmers, Manley, and Wasserman 2009, pp. 130–56.
Feferman, Solomon, Harvey M. Friedman, Penelope Maddy, and John R. Steel 2000: 'Does Mathematics Need New Axioms?' *Bulletin of Symbolic Logic*, 6, pp. 401–46.
Feigl, Herbert and Michael Scriven (eds) 1956: *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*. Minneapolis: University of Minnesota Press.
Feigl, Herber 1969: 'The Origin and Spirit of Logical Positivism'. In Achinstein and Barker 1969, pp. 3–24.
Glymour, Clark 1980: *Theory and Evidence*. Princeton: Princeton University Press.
Goldfarb, Warren and Thomas Ricketts 1992: 'Carnap and the Philosophy of Mathematics'. In Bell and Vossenkuhl 1992, pp. 61–78.
Hempel, Carl 1950: 'Problems and Changes in the Empiricist Criterion of Meaning'. *Revue Internationale de Philosophie*, 11, pp. 41–63.
——1951: 'The Concept of Cognitive Significance: A Reconsideration'. *Proceedings of the American Academy of Arts and Sciences*, 80, pp. 61–77.
Hintikka, Jaakko (ed.) 1975: *Rudolf Carnap, Logical Empiricist*. Dordrecht: D. Reidel Publishing Company.
Hintikka, Jaakko (ed.) 1999, "The Emperor's New Intuitions", 'The Emperor's New Intuitions'. *Journal of Philosophy*, 96, pp. 127–47.
Kaplan, David 1975: 'Significance and Analyticity'. In J. Hintikka 1975, pp. 87–94.
Lewis, David 1988: 'Statements Partly about Observation'. *Philosophical Papers*, 17, pp. 1–31.
Loux, Michael and Dean Zimmerman (eds) 2005: *Oxford Handbook of Metaphysics*. New York: Oxford University Press.
Macdonald, Graham and Crispin Wright (eds) 1986: *Fact, Science, and Morality*. Oxford: Basil Blackwell.
Margolis, Eric and Stephen Laurence 2003: 'Should We Trust Our Intuitions? Deflationary Accounts of the Analytic Data'. *Proceedings of the Aristotelian*, Society, 103, pp. 299–323.
Morgan, Mary S. and Margaret Morrison 1999: *Models as Mediators*. Cambridge: Cambridge University Press.
Oberdan, Thomas 1990: 'Positivism and the Pragmatic Theory of Observation'. *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1, pp. 25–37.

Price, Huw 2009: 'Metaphysics After Carnap: The Ghost Who Walks?' In Chalmers, Manley, and Wasserman 2009, pp. 320–46.

Quine, W. V. 1951a: 'Two Dogmas of Empiricism'. *Philosophical Review*, 60, pp. 20–43.

——1951b: 'On Carnap's Views on Ontology'. *Philosophical Studies*, 2, pp. 65–72.

Reichenbach, Hans 1958: *The Philosophy of Space and Time*. New York: Dover.

Rozeboom, William 1960: 'A Note on Carnap's Meaning Criterion'. *Philosophical Studies*, 11, pp. 33–8.

Schlipp, Paul (ed.) 1963: *The Philosophy of Rudolf Carnap*. Peru: Open Court Press.

Schwitzgebel, Eric 2008: 'The Unreliability of Naive Introspection'. *Philosophical Review*, 117, pp. 245–73.

Scriven, Michael, Paul Feyerabend, and Grover Maxwell (eds) 1956: *Minnesota Studies in the Philosophy of Science*. Vol. 1. Minneapolis: University of Minnesota Press.

Sellars, Wilfrid 1956: 'Empiricism and the Philosophy of Mind'. In Scriven, Feyerabend, and Maxwell 1956, pp. 253–329.

Sider, Theodore 2005: 'Reductive Theories of Modality'. In Loux and Zimmerman 2005, pp. 180–208.

Soames, Scott 2003a: 'The Rise and Fall of the Empiricist Criterion of Meaning'. In Soames 2003b, pp. 271–99.

——2003b: *Philosophical Analysis in the Twentieth Century*. Princeton: Princeton University Press.

Sober, Elliott 1999: 'Testability'. *Proceedings of the American Philosophical Association*, 73, pp. 47–76.

——2000: 'Quine's Two Dogmas'. *Proceedings of the Aristotelian Society*, 74, pp. 237–99.

——2008: *Evidence and Evolution: The Logic behind the Science*. Cambridge: Cambridge University Press.

Wright, Crispin 1986: 'Scientific Realism, Observation and the Verification Principle'. In Macdonald and Wright 1986, pp. 247–74.