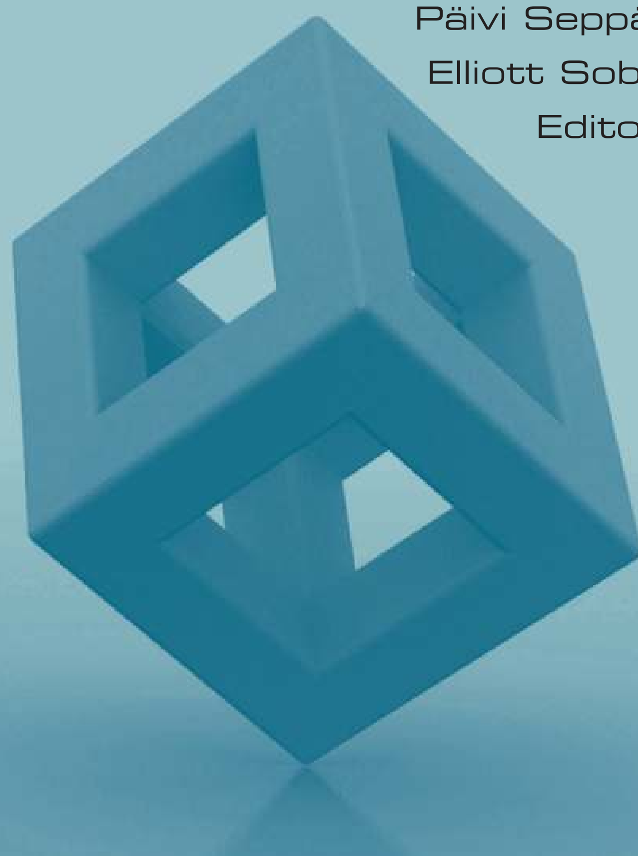
Logic, Methodology and Philosophy of Science

Proceedings of the Fifteenth International Congress

Hannes Leitgeb
Ilkka Niiniluoto
Päivi Seppälä
Elliott Sober
Editors



10 Logical empiricist reconstructions of theoretical knowledge

WILLIAM DEMOPOULOS *

Abstract. I argue that the logical empiricist account of theoretical knowledge exhibits a fundamental misconception about the character of the claims theories express. This is the idea that the application of truth to theoretical claims—whether they involve unobservable entities or the structure of space and time—rests on considerations that show them to possess an arbitrariness that does not attach to non-theoretical, empirical statements. This idea was completely explicit in the case of logical empiricism's conventionalist account of theoretical claims about the geometry of space and time, but it emerges as an unintended consequence of its partial interpretation account of theories about entities which transcend observation. The misconception stems from an incorrect assessment of the epistemic warrant theoretical claims enjoy and an incorrect assessment of the basis for our confidence in existence claims involving unobservable entities. The view I advocate allows that the value of a physical theory is often instrumental and independent of whether the theory is even approximately true. But this concession to instrumentalism is compatible with the idea that a theory's instrumental value can consist in facilitating the discovery of salient truths about reality, even a part of reality that is entirely hidden from observation. The argument to this conclusion rests on an analysis of the methodology of theory-mediated measurement and the role this methodology plays in securing fundamental existence claims of the kind we associate with Jean Perrin in the case of molecular reality and J. J. Thomson in connection

*The University of Western Ontario, London, Canada N6A 5B8

with the constitution of cathode rays. The larger study from which the present paper is drawn is an extended presentation of that argument, while the following discussion focuses on some of the main considerations that motivated its development.

Keywords: structuralist thesis, partial interpretation, logical empiricism, model-theoretic argument.

1 Introduction

The logical empiricists articulated what is arguably the first systematic development of a theory of theories. Their approach derived from the perspective afforded by modern logic and Hilbert's work on the axiomatic method. It was designed to address two key developments, the first concerned with abstract principles, the second with existence claims. In the case of abstract principles, Einstein's theories of relativity demonstrated the need to revise our conception of the epistemological status of the geometry of space and the chronometry of time; to understand the nature and epistemological significance of such a revision was pressing. As for existence claims, the early twentieth-century experimental success of atomism established the appeal to unobservable entities as a permanent feature of physical theory and raised the question of the nature of the support that attaches to claims which purport to be about entities that transcend observation.

The logical empiricist reconstruction of relativity was a corrective to the pretensions of the rationalist adherence to the synthetic a priori and the naiveté of an earlier empiricism; it was distinguished by its conventionalist interpretation of geometry and chronometry. The search for an account of theoretical claims which appeal to unobservable entities culminated in the partial interpretation view of theories, one of whose principal goals was to address the prima facie challenge to empiricism that such claims represent by clarifying their empirical status.

My focus is the partial interpretation reconstruction of theories and its elaboration by Carnap using the notion of the Ramsey-sentence of a theory. I will touch on work of David Lewis which is often seen as a contribution to this tradition, and I will articulate a central difficulty for the partial interpretation reconstruction that is suggested by one of Hilary Putnam's "model-theoretic arguments." The difficulty consists in the failure on the part of the partial interpretation and related reconstructions to provide an adequate account of our epistemic access to theoretical domains. To address this issue it is necessary to delve more deeply into the methodology by which such access has in fact been secured. My goal here is to motivate the importance of the problem of epistemic access. The larger study from which this paper is drawn is an extended discussion of this issue and the role theory-mediated measurement plays in its resolution.

2 The partial interpretation account of theories

The partial interpretation account of theories holds that the vocabulary of a theory consists of an observational and a theoretical component, where the application of the observation-theory distinction is based on whether a vocabulary item is held to apply to entities in the intended domain of the theory which are observable—in which case the item belongs to the *observation vocabulary*—or unobservable, in which case it belongs to the *theoretical vocabulary*. In the classical formulation of Carnap (1956, pp. 41–42 and pp. 46–48), this distinction partitions vocabulary items into just these two classes. (The possibility of a third class of terms—*mixed vocabulary* items that are understood to apply to both observable and unobservable entities—will be addressed later.)

Given the foregoing distinction in vocabulary, the sentences of the language of a theory are divided into three classes: one consisting of sentences which are generated from just the observation vocabulary, another of sentences generated from just the theoretical vocabulary, and a third consisting of sentences generated from the combined observation and theoretical vocabularies. These are, respectively, the *observation sentences*, *theoretical sentences*, and *correspondence rules* of the language. A *theory* is the conjunction of a selection of theoretical sentences and correspondence rules. There are no special restrictions on the logic of a theory, and it may be either first-order or higher-order; the notion of a theory may also be generalized in various ways that are irrelevant to the conceptual issues on which I intend to focus.

Historically, the partial interpretation account of theoretical knowledge derives from the idea that theoretical terms are introduced by sentences which, taken by themselves, are indistinguishable in their epistemic status from the statements of a pure mathematical theory. Theoretical statements share with the statements of a mathematical theory the property that their interpretation is responsible only to the logical category of their constituent nonlogical constants. This view of theoretical statements is a consequence of the fact that the partial interpretation account is a continuation and extension to the theories of physics of the axiomatic tradition that Hilbert initiated in pure mathematics. Especially influential was Hilbert's contention that the primitives of a mathematical theory are whatever satisfies its axioms. This contention—that the postulates of a theory "implicitly define" its primitive notions—swept away the subjective associations that characterized an older tradition's understanding of a mathematical theory's primitives, even in the case of geometry, where they were thought to have a familiar "intuitive" content. The partial interpretation account sought to extend Hilbert's analysis of mathematical theories to physics by providing an account of the empirical content of the theoretical statements of physics that is based on the connections between theoretical terms and observation terms that are expressed by correspondence rules.

One can see in this brief sketch the two characteristic theses of the partial interpretation view: the first, its claim that only the observation vocabulary is completely understood; and the second, the correlative claim that the interpretation of the theoretical vocabulary is limited by constraints which depend only on the logical category of the theoretical terms and whatever restrictions the true observation sentences impose on the domain of unobservable entities over which the theoretical sentences and correspondence rules are evaluated. I will refer to this second claim as the *structuralist thesis*. We have yet to explain how the partial interpretation view conceives the relation between interpretations, true interpretations, and truth.

3 Carnap on Ramsey sentences and the explicit definition of theoretical terms

Carnap's mature reconstruction of the language of science[1] builds on and extends the partial interpretation view of theories. The central notion of this account is the *Ramsey sentence* of a theory: the sentence formed by replacing theoretical terms by (new) variables of the appropriate logical category, then closing the resulting formula by adding an existential quantifier for each of the new variables. It is a very short step from the two characteristic theses of the partial interpretation account of theories to the notion that a partially interpreted theory's Ramsey sentence captures its "factual content": the Ramsey sentence is observationally equivalent to the theory in the sense that any argument from the partially interpreted theory to a sentence of the observation language can be recovered using the Ramsey sentence instead.[2] Notice that the Ramsey sentence's use of variables in place of uninterpreted theoretical terms simply makes explicit the commitment of the partial interpretation account to the structuralist thesis. As Ramsey expressed it:

> So far . . . as *reasoning* is concerned, that the [transforms of the theoretical sentences and correspondence rules which constitute the matrix[3] of the Ramsey sentence of the theory] are not complete propositions makes no difference, provided we interpret all logical combinations as taking place within the scope of a [single existential] prefix. . . . For we can reason about the characters in a story just as well as if they were really identified,

[1]This is the reconstruction expounded in (Carnap, 1963) and extended in various ways to be described below in (Carnap, 1961). The publication date of (Carnap, 1963) is a poor guide to the work's date of composition since the publication of the volume in which it appeared was delayed for many years.

[2]The Ramsey sentence of a theory, like one of its Craig transcriptions, eliminates theoretical vocabulary, but unlike a Craig transcription, it retains the connections between observable properties and relations which are mediated by their association with theoretical properties and relations. This is why the difficulties which Hempel (1965, pp. 214–216) shows to be a necessary feature of the elimination of theoretical vocabulary for reconstructions based on the notion of a Craig transcription are not difficulties for Ramsey-sentence-based reconstructions. For further discussion see Demopoulos (2013, Chapter 7.2) and Putnam (2012).

[3]By the matrix of a Ramsey sentence I mean the formula which results when the "new" existential quantifiers are deleted.

> provided we don't take part of what we say as about one story, part about another. (Ramsey, 1929, p. 232)

Carnap's mature reconstruction refines the doctrine of partial interpretation in two principal respects. As we have already noted, Carnap explicates the factual content of a partially interpreted theory in terms of its Ramsey sentence. But Carnap took things a step further by combining his account of the factual content of a theory with an explication of theoretical analyticity—analyticity relative to a theory—in terms of what has come to be known as the *Carnap sentence* of a theory: the conditional whose antecedent is the theory's Ramsey sentence and whose consequent is the partially interpreted theory. Before Carnap the distinction between the factual and analytic (and hence, non-factual) components of a theory followed the distinction between postulates and definitions. But since this distinction is inherently arbitrary, its utility for a dichotomy that is supposed to reveal our factual commitments may be doubted.

The Carnap sentence is justifiably regarded as analytic because it is a kind of "implicit definition" of the theoretical vocabulary, one that is provably non-factual in the sense that the only observation sentences it logically implies are logical truths. And as John Winnie (Winnie, 1970) later showed, the Carnap sentence, like a proper definition, satisfies a special noncreativity condition that is similar to the noncreativity condition that is customary for proper *explicit* definitions.[4]

Carnap advanced the Ramsey sentence not just as a clarification of the partial interpretation view of theories, but as a correct representation of how scientists understand their theoretical claims. They intend, Carnap held, an "indeterminate" claim, one that may have many interpretations under which it comes out true. As scientists understand them, theoretical claims are indeterminate as to the interpretation of their theoretical vocabulary and *any* representative class or relation which makes true the Ramsey sentence of the theory to which the claim belongs is as acceptable as any other. To narrow down the interpretation any further than is demanded by the truth of the Ramsey sentence would, for Carnap, violate the intentions of the scientist who constructed the theoretical system.

In (Carnap, 1961), which is one of his last papers on theoretical terms, Carnap converts the *implicit* definition of theoretical terms by the Carnap-sentence into a sequence of *explicit* definitions of them. But these explicit definitions do not eliminate—and were not intended by Carnap to eliminate—the indeterminateness of his earlier account. Indeed, Carnap formulates his explicit definitions in what he calls a "logically indeterminate" language. The language L_ε which he employs is a standard first- or higher-order language enriched with Hilbert's ε-operator and the extensional axioms which govern its use. There are two such axioms governing the use of Hilbert's ε-operator. Given a formula Fx in one free variable, the first axiom tells us that if there is something satisfying Fx, then there is an "ε-representative" of F, denoted '$\varepsilon_x(Fx)$', that is selected

[4]For an exposition of these matters, see (Demopoulos, 2007) and (Gupta, 2009).

by the choice function which interprets the ε-operator. The second axiom tells us that if the formulas Fx and Gx are extensionally equivalent, their ε-representatives are the same. That it should be possible to apply Hilbert's ε-operator to the Ramsey-sentence reconstruction of theories is a consequence of Carnap's observation that the Carnap sentence of a theory can be derived from a sentence that is in the same form as the first of the axioms for the ε-operator.

For Carnap the principal virtue of this proposal is that it incorporates the convenience of having the use of a theoretical vocabulary while retaining the characteristic indeterminateness of that vocabulary that is the hallmark of the partial interpretation view and of his mature reconstruction in terms of Ramsey and Carnap sentences. Thus he writes that the theoretical postulates and correspondence rules "are intended *by the scientist who constructs the system* to specify the meaning of [a theoretical term] to just this extent: if there is an entity satisfying the postulates, then [the term] is to be understood as denoting one such entity. Therefore the definition [of a theoretical term by means of Hilbert's ε-operator] gives to the indeterminate [theoretical term] just the intended meaning with just the intended degree of indeterminacy" (Carnap, 1961, p. 163, emphasis added).

4 A proposal of David Lewis and two theorems of John Winnie

David Lewis's (1970) is sometimes credited with having refined Carnap's and Ramsey's reconstructions and to have improved on Carnap's approach to the explicit definition of theoretical terms by showing how it might be possible to avoid multiple interpretations of the theoretical vocabulary under which the theory comes out true. Lewis maintained that allowing for what he calls "multiple realizations" concedes too much to instrumentalism. Lewis does not say why multiple realizablity is a concession to instrumentalism, but let us for the moment grant the point and consider how he makes the case that there are many theories which, if realizable, are *uniquely* realized. Lewis is clear that he must provide an independent defense of this contention, since the possibility of there being just one realization appears to be excluded by two theorems of Winnie (1967). Modulo the conceptually unimportant technical restriction that not all theoretical properties and not all theoretical relations are universal, Winnie shows that on the partial interpretation view of theories, if a theory has one realization, there is always another; and if a theory is realizable at all, it is arithmetically realizable.

Lewis's response to Winnie rests on two features of his conception of the language in which theories are formulated. First of all, Lewis follows the partial interpretation account by dividing the vocabulary of a theory into two parts, which he calls the "O-vocabulary" and the "T-vocabulary" of the theory. However, Lewis's "O-T distinction" is not the distinction between observation and theoretical vocabulary of the partial interpretation account. Lewis's distinction concerns *old* vocabulary, vocabulary which is understood prior to the formulation of the vocabulary-introducing theory; and the contrast Lewis's distinction draws between old and T- or *new* vocabulary has nothing to do with observation or observability. In principle, Lewis's O-T distinction could be completely orthogonal to the observational-theoretical distinction of Carnap and the doctrine of partial interpretation. A second, related, difference involves Lewis's notion of an "O-mixed term." This is a notion that does real work for Lewis, but before explaining it, some further background regarding the partial interpretation view and its relation to Lewis's O-T distinction is necessary.

In his exposition of the partial interpretation reconstruction, Winnie includes in addition to the observation and theoretical predicates a separate category of *mixed predicates*, predicates that apply to both observable and unobservable entities. Lewis has the notion of a *mixed term*, and of special importance are those he calls "O-mixed" terms. These are terms which, like Winnie's mixed predicates, can apply to both observable and unobservable entities. But their characterization—unlike Winnie's characterization of observation and theoretical predicates—has nothing to do with observability. And while Winnie's mixed predicates are distinguished from observation predicates, Lewis's O-mixed terms count as O-terms, and as such, are assumed to be fully understood whether they apply to observable or unobservable entities; therefore the interpretation of O-terms—whether they are unmixed and refer only to observable entities, or are mixed and refer also to unobservable entities—must be preserved as we pass from one realization of a theory to another.

The situation is altogether different for Winnie and for the standard view. As we will soon see in greater detail, when Winnie establishes the existence of alternative realizations by means of a permutation map, it is only the entities in the observable part of the domain that cannot be permuted, and it is only the interpretation of the observation predicates—predicates which apply only to observable entities—that must be the same in any realization of a partially interpreted theory. No such requirement applies to the interpretation of theoretical predicates; nor does it apply to mixed predicates.

Since Winnie's permutation map is defined as the identity on observable entities, it is trivially true that observable relations are isomorphic to their images under his mapping. But although it is trivially true that Winnie's map is an isomorphism between observable relations, it is not part of Winnie's argument that *every* one-one map from the observable part of the domain onto itself can be extended to an isomorphism on the properties and relations which interpret the observation predicates of the theory. However the situation is different for one-one maps from the unobservable part of the domain onto itself and the relations which interpret the theory's theoretical predicates.

To establish that a partially interpreted theory has many models if it has one model, Winnie uses a construction based on a mapping from and onto the domain of a model of the theory that permutes at least one pair of unobservable entities so that it changes the image, under the mapping, of the interpretation of at least one theoretical predicate. By the structuralist thesis—and this is the observation Winnie's proof rests on—the theoretical predicates and relations can always be understood so that the relations which interpret them are *defined* by their images under an arbitrary one–one mapping from and onto the unobservable part of the domain. As for the properties and relations which interpret mixed predicates, their images are unchanged only by the action of the mapping from and onto the observable part of the domain. Such a construction defines the theoretical and mixed properties and relations of a new structure as the images of the properties and relations of a structure we know to be a model of the theory. Since the new properties and relations are *by construction* isomorphic to the properties and relations of a model of the theory, the structure which they define must also be a model of the theory.

To see why Winnie's argument does not affect Lewis's claims about uniqueness of realization, let us suppose that electrons and protons are unobservable entities, that electrons are (strictly) smaller than protons, and that 'smaller than' is a term of Lewis's O-mixed vocabulary; any realization must therefore preserve the interpretation of 'smaller than.' It follows that the construction of a realization which, like Winnie's, interchanges an electron and a proton, is ruled out. In fact, the only case in which Lewis's and Winnie's notions of realization coincide is the case in which Lewis's O-terms and T-terms have (respectively) only observable and only unobservable entities in their extensions, and where therefore there are no mixed terms. In such a case, even on Lewis's account, a theory must have more than one realization if it has any realization at all. We might put this by saying that Lewis's approach to theories and the definition of theoretical terms is at best only "accidentally" affected by Winnie's results.

To sum up our discussion of Winnie and Lewis, Winnie's mixed predicates form a special category distinct from observation predicates, and only the observable entities in the interpretation of mixed and observation predicates are unaffected by his permutation map. But for Lewis mixed terms can be classified as O-mixed terms, and in order to avoid Winnie's permutation argument, it suffices that there should be a suitably rich collection of O-mixed terms. In fact it may suffice that there should be *one* term that stands for a relation which is such that all the various types of theoretical entity are comparable in terms of this relation. In our example of electrons and protons, it is sufficient that the particles of one kind are (strictly) smaller along some dimension than those of the other—assuming, of course, that *smaller than* along this dimension is picked out by an O-mixed term. It is therefore not at all implausible that for Lewis a realizable theory can always be extended by the introduction of O-mixed terms and appropriate postulates involving them to a theory that is *uniquely* realizable.

Lewis's deployment of the Ramsey sentence and a modified form of the Carnap sentence shares a strong formal affinity with Carnap's reconstructions: Lewis's adoption of the Ramsey sentence suggests a commitment to the idea that a model of our understanding of new or T-terms—including the case in which the new terms are associated with the introduction of a new class of entities—is adequate if it captures the inferential connections of the statements containing the new terms with the sentences in the O-vocabulary. But although Lewis's old vocabulary may well include "observation" terms, the connections between T-terms and O-terms which Lewis's mixed statements express is not explicitly proposed as a connection with observation vocabulary, but merely as a connection with vocabulary that is understood. And for Lewis, by contrast with the partial interpretation or Ramsey-sentence reconstructions of the logical empiricists, so far as our understanding of *old* vocabulary is concerned, it might be based entirely on our grasp of the inferential connections into which its items enter—independently of whether or not these include inferential connections with observation sentences.[5] It is clear therefore that despite their formal affinity, the absence of an epistemological motivation underlying Lewis's reconstruction of theories makes his account very different from both the partial interpretation reconstruction and from Carnap's various refinements of it.

Carnap and the advocates of partial interpretation take it as a desideratum of an adequate reconstruction of theoretical knowledge that it should address the empirical basis of theoretical claims. On the partial interpretation reconstruction, this problem is addressed by the provision of an explanation of our understanding of theoretical claims in terms of the connection correspondence rules establish between theoretical and observation vocabulary. Carnap's Ramsey-sentence reconstruction dissolves the problem of how we come to understand the meanings of terms which apply to unobservable entities by *eliminating* theoretical terms in favor of variables. But this dissolution of the problem is merely an emendation—not a rejection—of the partial interpretation view, an emendation that preserves the structuralist thesis. Indeed, Carnap's transition to the Ramsey sentence reconstruction rests on the recognition that the partial interpretation view subscribes to this thesis. For if, in addition to whatever restrictions the true observation sentences impose on the domain of a model of a partially interpreted theory, the constraints on the interpretation of theoretical vocabulary appeal only to the logical categoryof the theoretical terms, then there can be no objection to their replacement by variables, and the problem of accounting for how theoretical terms are understood then simply disappears. As for the empirical basis for theoretical *claims*—as opposed to our understanding of theoretical *vocabulary*—and the explanation of their difference from the claims of pure mathematics, modulo the elimination of theoretical vocabulary, these issues are addressed by Carnap's Ramsey-sentence reconstruction much as they are by the partial interpretation reconstruction. Instead of appealing to the association of theoretical claims with observation sentences by the mediation of correspondence rules, the empirical basis of such claims is accounted for by the association of the Ramsey-sentence transforms of theoretical claims with observation sentences that is effected by the Ramsey-sentence transforms of correspondence rules.

The centrality of the problem of the empirical basis of theoretical claims to the logical empiricists' reconstruction of theories serves not only to distinguish their ap-

[5]See (Gupta, 2013, especially section 11) for a critical perspective on "inferential-role" models of meaning.

proach from Lewis's. It also shows why it is so misleading to characterize their account—whether partial interpretation, Ramsey-sentence, or one of Carnap's later reconstructions—to have incorporated a "syntactic view" of theories. For the logical empiricists, theories are indeed linguistic objects. But to call the logical empiricist account "syntactic" misses the fact that it is first and foremost a reconstruction of theoretical knowledge that purports to illuminate how observation bears on the empirical character and evidential support of theoretical claims. Evidently, neither goal can be successfully addressed without going beyond syntax—as indeed the principal logical empiricist reconstructions do when they assume that the observation language is interpreted.

This contrasts with the situation in mathematical logic where a theory is defined as a set of sentences in a language about which we assume *only* that its syntax and underlying logic are completely explicit. The logical tradition is also motivated by an epistemological problem which is no less fundamental to its notion of a theory than the epistemological motivation underlying logical empiricism is to its conception of a theory. The logical tradition sought to show that a *finitary* notion of proof suffices for the reconstruction of all mathematical reasoning—even reasoning within theories whose intended interpretation is over an infinite domain of arbitrarily large cardinality. But here the restriction to syntax in the characterization of a theory is entirely natural, since it is essential to the successful positive solution of the problem which motivates the logical tradition that a theory should be represented as a purely syntactic object. Although the logical empiricist approach to theories was profoundly influenced by the logical tradition, its goals were different: it sought to build a platform for the representation of the theoretical claims of physics that would be capable of illuminating their content and the basis on which they are understood and evaluated. In particular, it sought to show how observation must be a central component of an adequate empiricist solution to this problem. The nature of the questions the logical empiricist approach sought to address demanded—and were recognized by its proponents as demanding—a notion of theory that includes more than the purely syntactic conception of the logical tradition.

5 Hilary Putnam's model-theoretic argument

We have seen how, by contrast with Lewis, Carnap was prepared to accept multiple realizability as a point *in favor* of partial interpretation and Ramsey-sentence reconstructions. In his "Reply to Hempel" in the Schilpp volume devoted to his work Carnap even went so far as to endorse an arithmetical interpretation of the Ramsey sentence of a theory as the *correct* understanding of what, on his reconstruction, the sentence asserts. As Carnap put it,

> I agree with Hempel that the Ramsey sentence does indeed refer to theoretical entities . . . . However, it should be noted that these entities are not unobservable physical objects like atoms, electrons, etc., but rather (at

> least in the form of the theoretical language which I have chosen in [(Carnap, 1956, Section VII)] purely logico-mathematical entities, e.g. natural numbers, classes of such, classes of classes, etc. (Carnap, 1963, p. 963).

Carnap might well be understood therefore to have also anticipated and embraced the content of the second of Winnie's two theorems as an acceptable consequence of his view of the factual content of theories.

In light of these considerations, let us for the moment set to one side the issues connected with arithmetical interpretations and multiple realizablity and turn our attention to a closer examination of the question, 'How do theories, whose characteristic feature is that their theoretical claims transcend observation, acquire their empirical status?' This brings us to Hilary Putnam's model-theoretic argument, by which I mean his first such argument, the one developed in (Putnam, 1977). I should emphasize that my primary interest is the actual argument, rather than Putnam's uses of it; these are all more various than the application I will isolate. As I understand its significance, the argument shows that the answer to our question given by the doctrine of partial interpretation and its close descendants is incompatible with the thesis that when theories which transcend observation are true, they express salient truths about unobservable entities. The fault with all these views stems from their failure to satisfactorily address the basis for our epistemic access to theoretical domains.

The model-theoretic argument consists of a simple technical argument and an observation. The technical argument establishes that any model of a theory's observational consequences can be extended to a model of the theory's theoretical sentences and correspondence rules, where the domain of this extension is the standard domain of observable and unobservable entities. The argument which establishes this conclusion also supports an observation, namely, that on the partial interpretation reconstruction of theories, as well as Carnap's various refinements of it, the conditions under which a partially interpreted theory can be shown to be *satisfiable* suffice to show that the theory is *true*.

The simple technical argument exploits a construction like Winnie's. The conclusion of the argument follows from a well-known folklore result that assures us that any model of a theory's observational consequences can be extended to a model of the theory.[6] This is true, in particular, when the model of the theory's observational consequences is defined over the standard domain of observable entities. But so far as

[6]The result is reported in (van Benthem, 1978) as

Lemma 3.2 For any L_o–structure M, if M is a model of the L_o–consequences of T, then there exists an $L_o \cup L_t$ -structure N such that N is a model of T and the reduction of N to L_o is an elementary extension of M.

Here L_o and L_t, are first-order languages with equality all of whose vocabulary is, respectively, observational and theoretical. ($L_o \cup L_t$ is the language generated by the observational and theoretical vocabulary of L_o and Lt.) The lemma assumes that the vocabulary of the theory consists only of theoretical and observational terms.

the folklore result is concerned, the “theoretical” or “unobservable” part of the extension might well be “abstract”—even number-theoretic. Nevertheless, if we are given such a “partially abstract” model, we can use a construction based on a one-one onto mapping which is the identity on the observable part of the domain of this model and is arbitrary from its theoretical part onto the theoretical domain of non-abstract unobservable entities. To do so we require the additional and contingent assumption that the cardinalities of the theoretical domains of the two models are the same. Relative to this assumption, Putnam’s argument proceeds by defining the theoretical properties and relations over the standard domain of observable and unobservable entities as the images of the properties and relations of the abstract model under a mapping that is the identity map on the observable part of the domain of the partially abstract model, and is arbitrary from its theoretical part onto the theoretical part of the domain of non-abstract theoretical entities. But since the defined properties and relations meet all the conditions that the partial interpretation account is capable of imposing on theoretical properties and relations, the argument shows not merely that the theory is satisfiable, but that, on the partial interpretation account, it is *true*.

The arguments of Winnie and Putnam both exploit the same technical idea in their respective definitions of the theoretical relations which interpret the theoretical predicates of a partially interpreted theory. But their arguments support what are conceptually very different difficulties for the view. Putnam’s argument is not directed at the existence of *multiple* realizations; nor does it concern the existence of an *arithmetical* model of a partially interpreted theory. Rather, Putnam’s argument takes us from a cardinality assumption, and the existence of what might well be an arithmetical model of the kind explored by Winnie, to the conclusion that, on the partial interpretation view, the fact that a theory is *satisfiable* over the intended domain of observable and theoretical entities suffices to show that it is *true*.

The significance of the model-theoretic argument has been the subject of an extensive discussion. But whatever the resolution of the many controversies the argument has generated involving “metaphysical realism”, “intended” reference, or the “indeterminacy” of reference, it seems clear that the notion that the truth of what is asserted about unobservable entities might depend only on their *number* runs counter to one of the simplest and least contentious convictions of “realism” and, indeed, of common sense. This is the conviction that if a theory is true, this is because its theoretical claims have captured a salient aspect of the reality they seek to describe, an aspect that goes beyond any mere question of cardinality.

The partial interpretation account of theories claims to reconstruct the empirical status of a theory’s theoretical statements using only the theory’s logico-mathematical framework and the apparatus of correspondence rules. But when we are restricted to just these resources, the truth of theoretical claims reduces to their satisfiability in any sufficiently large model of the true observation sentences. And this shows that the reconstruction has failed to correctly represent the nature of the epistemological status of a theory’s theoretical claims. It has failed because the epistemic basis for such an assertion of satisfiability is entirely different from what is required by an assertion of truth. The idea that the claim that a theory is true should depend only on a cardinality constraint and a logical argument fails to adequately separate the epistemic basis for the truth of the theoretical assertions of an empirical theory from the epistemic basis for the mere satisfiability of the "abstract" assertions of a purely mathematical theory over a given domain.

The conclusion we have just reached should perhaps have been anticipated, given the origin of the partial interpretation view in Hilbert's conception of the foundations of geometry. In his correspondence with Frege, Hilbert defended the idea that satisfiability in a sufficiently large domain is a suitable surrogate for the "truth" of a mathematical theory. But whatever its plausibility for theories of pure mathematics, the methodological demands we impose on the theoretical claims of physics cannot be captured by so weak a requirement, not at least if we wish to preserve the methodological difference between physics and pure mathematics. An advocate of partial interpretation might respond to this objection by recalling that a physical theory will qualify as true not if it is merely satisfiable, but only if it is satisfiable in a model which is an extension of a domain that forms the basis for a model of the true observation statements. By contrast, the domains which bear witness to the "truth" of a mathematical theory need not have any connection with such a model. For an advocate of partial interpretation, the theories of physics are true because they are *empirically adequate* in the sense that they have observational consequences, all of which are true; but a theory of pure mathematics is not necessarily associated with *any* observation language and is not required to be empirically adequate.

However this response misses the point of the model-theoretic argument as we have presented it: provided the domain over which a partially interpreted theory involving unobservable entities is interpreted includes the domain of the model of the true observation sentences, it is a consequence of the partial interpretation view that the method of argument by which we are able to establish the "truth" of a purely mathematical claim over a given domain also suffices to establish the truth of a theoretical claim.

The model-theoretic argument puts us in a position to see why—pace Lewis—the multiple realizablity which afflicts partial interpretation and Ramsey-sentence reconstructions is largely tangential to the question of realism. For suppose we are given a realization of the sort Putnam's argument shows is possible. Given such a model, we have seen how, following Lewis (1970), we might rule out alternative realizations by supplementing the partially interpreted theory with a judicious selection of O-mixed terms and appropriate assumptions involving them. Notice however that this is compatible with the possibility that a theory is true only because it has a realization that models its observational consequences and is the right size. So in light of Putnam's argument, uniqueness of realization is insufficient to ensure the widely held conviction that if our physical theories are true, this is because they succeed in isolating salient truths about the entities with which they deal, independently of whether they are observable or unobservable. In any case, when, long after his 1970 paper on theoretical terms, Lewis came to address Putnam's argument, he did not appeal to O-mixed terms to resolve the problem he took the argument to pose but based his reply on a distinction among possible realizations. The core assumption of Lewis's response is that a theory is *true* only if it is true relative to a realization whose properties and relations are *natural*. Since there is nothing in Putnam's construction of his interpretation of theoretical predicates which requires that they should be natural properties and relations, Lewis argued that the construction fails to show that the theory's theoretical claims are, in the relevant sense, true.[7]

It has been suggested that we should adapt Lewis's reply to the model-theoretic argument and supplement the partial interpretation account by restricting the class of admissible realizations to those that involve natural relations, thereby distinguishing, in the way Lewis proposes, true theories from theories that are merely satisfiable in a domain that extends a model of the true observation statements. Lewis's distinction might be further exploited to characterize true *empirical* theories as those that are not merely satisfiable in some such realization or other, but are true because they are true in a realization whose relations are natural. But we should be cautious about accepting Lewis's suggestion as an adequate response to the model-theoretic argument or as a guide for emending the partial interpretation view.

To begin with, Lewis's reply to Putnam leaves unresolved the problem of how we are able to make significant claims about relations that are *not* "natural". Even if we have no interest in theorizing about such relations, an adequate response to the model-theoretic argument should nevertheless explain how it is *possible* to do so without the assertion of the truth of such a theory collapsing into an assertion of its satisfiability over a domain—even a domain that extends the model of the observational consequences of the "theory" of such a natural relation. Indeed, as Fraser MacBride has remarked, on the assumption that we achieve knowledge of natural relations only with the progress of science—and perhaps only after many distractions involving nonnatural relations—anyone following Lewis's suggestion *must* have an interest in how we manage to make significant, but as it happens, misguided claims about nonnatural relations.

But secondly, and more importantly, addressing Putnam's argument by appealing to Lewis's proposal obscures the difficulty the model-theoretic argument raises for partial interpretation and Ramsey-sentence reconstructions. The problem with these approaches is not their failure to designate certain properties and relations as natural, but the fact that they are too weak to explain the difference between the epistemological status of theoretical and purely mathematical claims. But then the partial interpretation framework for addressing how theories are warranted must *also* fail to capture the

[7]Lewis's response is developed in (Lewis, 1983) and (Lewis, 1984). See also (Merrill, 1980), which Lewis cites as having influenced his reply. (Psillos, 1999, pp. 67–68) has argued that Lewis's requirement of uniqueness is an anticipation of his later appeal to natural relations in his reply to Putnam. But it is clear from our earlier discussion that for Lewis all the heavy lifting to establish unique realizablity is done by assumptions which are independent of the requirement that the properties and relations which interpret the O- and T-vocabularies are natural, or that they "carve nature at its joints."

methodology by which claims about unobservables are established, and this shows that a different approach to these two issues is required. An emendation of the view based on Lewis's reply to Putnam only succeeds in recording the fact that we *do* distinguish mathematical claims from the theoretical claims of physics, but it has nothing to contribute to our understanding of the methodology by which we make this distinction. Nor does it contribute to our understanding of how we successfully gain epistemic access to theoretical domains in order to warrant our claims about them.

6 Conclusion

There are two questions that an adequate account of theoretical knowledge must address: (i) 'What is the methodology by which we gain epistemic access to theoretical domains?' and (ii) 'How are the theoretical claims of physics about such domains distinguished from the claims of pure mathematics?' In so far as partial interpretation and Ramsey-sentence reconstructions are unable to adequately address the second of these two questions, it is evident that they have failed to marshal the resources necessary for addressing the first. The burden of the larger study from which the present discussion is drawn is to explain a key physical methodology that makes theoretical domains epistemically accessible. This methodology clarifies how the properties and relations—more generally, the parameters—which enter into theoretical claims are empirically well-founded by robust theory-mediated measurements. Theoretical domains are epistemically accessible because the parameters which qualify their constituents are empirically well-founded. This account of epistemic accessibility is capable of supporting existential hypotheses involving entities which transcend observation without, however, necessarily also supporting—or even requiring—the truth of the theories which employ such hypotheses. And it is also capable of explaining the basis on which we distinguish physical claims about theoretical domains from the claims of pure mathematics. The role of robust theory-mediated measurements has been largely missed by the models of empirical adequacy and predictive success that have dominated partial interpretation and other traditions in the philosophy of science; and it has also been missed by some of logical empiricism's severest critics.

Acknowledments. Support from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged. I wish to thank Thomas Uebel for presenting this paper to the CLMPS on my behalf.

Bibliography

Carnap, R. (1956). The methodological character of theoretical concepts. In H. Feigl, & M. Scriven (Eds.), *The foundations of science and the concepts of psychology and psychoanalysis, Minnesota Studies in the Philosophy of Science Vol 1*. (pp. 38–76). Minneapolis: University of Minnesota Press.

Carnap, R. (1961). On the use of Hilbert's ε-operator in scientific theories. In Y. Bar-Hillel (Ed.), *Essays in the foundations of mathematics dedicated to A. A. Fraenkel* (pp. 156–164). Jerusalem: The Magnus Press of Hebrew University.

Carnap, R. (1963). Replies and systematic expositions. In P. A. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 859–1013). La Salle IL: Open Court.

Demopoulos, W. (2007). Carnap on the rational reconstruction of scientific theories. In M. Friedman, & R. Creath (Eds.), *The Cambridge companion to Carnap* (pp. 248–272). Cambridge: Cambridge University Press.

Demopoulos, W. (2013). *Logicism and its philosophical legacy*. Cambridge: Cambridge University Press.

Gupta, A. (2009). Definition. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. (Spring 2009 Edition). url: plato.stanford.edu.

Gupta, A. (2013). The relationship of experience to thought. *The Monist*, *96*, 252–294.

Hempel, C. G. (1965). The theoretician's dilemma: A study in the logic of theory construction. In *Aspects of scientific explanation* (pp. 173–226). New York: The Free Press.

Lewis, D. (1970). How to define theoretical terms. *Journal of Philosophy*, *67*, 427–445.

Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, *61*, 343–377.

Lewis, D. (1984). Putnam's paradox. *Australasian Journal of Philosophy*, *62*, 221–236.

Merrill, G. (1980). The model-theoretic argument against realism. *Philosophy of Science*, *47*, 69–81.

Psillos, S. (1999). *Scientific realism: How science tracks truth*. London: Routledge.

Putnam, H. (1977). Realism and reason. *Proceedings and addresses of the American Philosophical Association*, *50*, 483–498.

Putnam, H. (2012). A theorem of Craig's about Ramsey sentences. In M. DeCaro, & D. Macarthur (Eds.), *Philosophy in an age of science* (pp. 277–279). Cambridge MA: Harvard University Press.

Ramsey, F. P. (1929). Theories. In R. Braithwaite (Ed.), *The foundations of mathematics and other logical essays by Frank Plumpton Ramsey* (pp. 212–236). (Published 1931). Paterson NJ: Littlefield and Adams.

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

Winnie, J. (1967). The implicit definition of theoretical terms. *British Journal for the Philosophy of Science*, *18*, 223–229.

Winnie, J. (1970). Theoretical analyticity. In R. Cohen, & M. Wartofsky (Eds.), *Boston Studies in the Philosophy of Science, Vol. VIII* (pp. 289–305). Dordrecht and Boston: Reidel.

Author biography. William Demopoulos is professor emeritus from The University of Western Ontario, where he taught from 1972 to 2012. He has held numerous visiting positions, among them visiting professorships at Harvard University and the University of California, Irvine. He was recently the recipient of a Killam Fellowship. Background to his contribution to these *Proceedings* can be found in his recently published collection of essays, *Logicism and its Philosophical Legacy*, Cambridge University Press. He is currently at work on a monograph entitled *On Theories*.