1.4 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 'Replies' 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.[24] In doing so he might be understood 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 realizability 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 given in his 1976 APA presidential address. 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

[24] 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 [1956a Section VII] purely logico-mathematical entities, e.g. natural numbers, classes of such, classes of classes, etc." Carnap (1963, p. 963). This is also the understanding of the theoretical terms belonging to a partially interpreted theory that, in his survey of possible positions regarding existential hypotheses, Feigl (1950) calls "(Va) Formalistic Phenomenalism or Syntactical Positivism." In this paper Feigl sought to stress the continuity between traditional approaches to the problem of our knowledge of the external world and various views of theories and the nature of theoretical knowledge. As its title suggests, the paper's focus is the status of the existence claims that are central to modern physical theories of the constitution of matter.

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, the conditions under which a partially interpreted theory can be shown to be satisfiable suffice to show that the theory is *true*.

Let M be a model of the observational consequences of our theory such that the domain of M is the "standard domain" of observable entities. To obtain a model of the observation sentences, theoretical sentences, and correspondence rules, we can exploit a folklore result reported by van Benthem that assures us that there is an "abstract" model N which is an extension of M to a model of the whole theory.[25] Given N and a single nonlogical assumption, it is possible to define a model M* of the theory which, like N, is also an extension of M, but has as its domain the standard domain of observable and unobservable (i.e. theoretical) entities. On the "purely contingent" assumption that the domain of M* is of the same cardinality as the domain of N, there is a one-one onto mapping which is the identity on the common observable part of the domain of M* and the domain of N, and is arbitrary from the "unobservable" part of N onto the unobservable part of M*. Let the observable relations of M* be the same as those of N, and define theoretical relations for M* as the images, under this one-one correspondence, of the theoretical properties and relations which interpret the theoretical predicates of the theory in N. This defines a model M* which extends M by the addition of theoretical relations which are defined over the standard domain of unobservable entities, and which are *by definition* isomorphic to the theoretical relations of N; since M* extends M if N does, and since N is a model of the theory, so also therefore is M*. But since this construction of the theoretical relations of M* meets all of the conditions that the partial

---

[25] "Abstract" in the sense that its proof does not exclude the possibility that part of the domain of the model is arithmetical. See van Benthem (1978) where this result occurs 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 L_t.) Note that the lemma assumes that the vocabulary of the theory consists only of theoretical and observational terms. We will soon consider the effect of adding mixed terms.

interpretation account is capable of imposing on such relations, the fact that M* models the theory suffices to show that the theory is not merely true in M*, but *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 are most naturally understood to support conceptually distinct difficulties for the view. My use of 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. My claim is rather that 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 standard 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 the fact that 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 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 realizability 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 can 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 these entities are observable or unobservable. This is likely why, 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.[26] 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.

It might be thought 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 could be further exploited to characterize true *empirical* theories as those that are not merely satisfiable in some 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

[26] Lewis's response is developed in his (1983) and (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 realizability 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."

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 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.

1.5 Ramsey on Russell's analysis of matter and the partial interpretation of theories

As Michael Friedman and I noted, Putnam's argument is reminiscent of an objection raised by M. H. A. Newman against the causal theory of perception Russell advanced in