Synthese (2008) 164:359–383
DOI 10.1007/s11229-008-9355-1

# Some remarks on the bearing of model theory on the theory of theories

**William Demopoulos**

Received: 12 March 2008 / Accepted: 2 June 2008 / Published online: 27 August 2008


**Abstract** The present paper offers some remarks on the significance of first order model theory for our understanding of theories, and more generally, for our understanding of the "structuralist" accounts of the nature of theoretical knowledge that we associate with Russell, Ramsey and Carnap. What is unique about the presentation is the prominence it assigns to Craig's Interpolation Lemma, some of its corollaries, and the manner of their demonstration. They form the underlying logical basis of the analysis.

**Keywords** Craig Interpolation Lemma · Structuralism · Ramsey sentence · Carnap sentence · Theoretical analyticity · Craig transcription

## 1 Introduction

The present paper offers some remarks on the significance of first order model theory for our understanding of theories, and more generally, for our understanding of certain accounts of the nature of theoretical knowledge. Some of the matters reported on here are treated in earlier publications of mine.[1] What is unique about the following presentation is the prominence it assigns to Craig's Interpolation Lemma, some of its corollaries, and the manner of their demonstration. They form the underlying logical basis of the analysis.

[1] The present paper is most closely related to Demopoulos and Friedman (1985) and my (2003a,b) and (2007).

W. Demopoulos (✉)
Philosophy Department, Talbot College, University of Western Ontario,
London, Ontario, Canada N6A 3K7
e-mail: wgdemo@uwo.ca

I will begin by reviewing my contribution to a paper I wrote with Michael Friedman. That paper revived an early criticism of Russell's structuralism that was presented in what was then a little-known essay by the eminent English topologist, M.H.A. Newman. Carnap's mature account of theoretical knowledge has a natural affinity with Russell's point of view; I will give an extended account of Carnap's reconstruction of the language of science, together with an important emendation of it by John Winnie. After describing Winnie's contributions, I will turn to a paper by Jane English which appeared shortly after Winnie's work, but in apparent ignorance of it. English sought to apply results of first order model theory to the problem of the underdetermination of theory by observation. I will show how English's analysis is based on a simple application of Craig's Lemma. I will then describe how a form of Newman's observation regarding Russell's structuralism applies both to Carnap's original reconstruction and to its emendation. (A variant of Carnap's reconstruction favored by the school of British structuralists is also discussed.) The basic idea for this application is illustrated in the standard proof of a corollary to Craig's Lemma; this corollary plays a central role in Winnie's analysis. I will conclude my discussion of Carnap by drawing a parallel between his analysis of theoretical knowledge and Poincaré's conventionalist account of geometry.

## 2 Russell's structuralism

There are three ideas belonging to Russell's logical investigations that had a profound effect on his views in epistemology: the theory of descriptions, the notion of a logical construction, and the concept of structure, or more precisely, of structural similarity. The theory of descriptions gave formal precision to Russell's concept empiricism, and the method of logical construction gave his phenomenalism its distinctive character. Both were therefore of great importance for the formulation of epistemological theses to which he was independently committed. The theory of descriptions retained its importance for Russell's epistemology throughout his career. But the epistemological interest of the method of logical construction was relatively short-lived; it achieved its fullest expression in the phenomenalist period spanned by *Our knowledge of the external world* and *The relation of sense data to physics*. Later applications of the method, such as the construction of space-time points, though important for achieving some sort of ontological economy, did not have the epistemological significance of its use in connection with phenomenalism. With the exception of the period represented by these two studies, Russell was always attracted to some form of Lockean realism, both in early work, like *Problems of philosophy*, and in his Tarner Lectures in the Philosophy of Science, published in 1927 as *The analysis of matter*. Work subsequent to *The analysis of matter* retains this Lockean perspective, but it is in this book that Russell's suggestions regarding the importance of structure for general epistemology are most fully developed. The concept of structure is absolutely essential to his argument—first announced in *Introduction to mathematical philosophy* (1919, pp. 61–62), and then elaborated at length in *The analysis of matter*—that, contrary to Berkeley and Kant, it is possible to maintain, within a strict empiricist theory of the scope of experience, that there is genuine knowledge of the material world and of things in themselves.

Of these three fundamentally logical ideas, the one that bears on the present study is the concept of structure. The doctrine to which it led is what I am calling *Russell's structuralism*. This doctrine was for a long time neglected; it was rediscovered by Grover Maxwell who championed its relevance to the debates of the third quarter of the last century involving scientific realism and the meaning of theoretical terms. Maxwell clearly perceived the connection between Russell's general epistemology and certain trends in the theory of theories, trends that will later occupy us at some length.[2] This section sets forth Russell's theory together with an important criticism of it.

Russell's structuralism would not have been possible without the foundational investigations that culminated in his logicism; and he would not have been convinced of its correctness had he not undertaken the extensive study of developments in physics that form the context of *The analysis of matter*. But although Russell's structuralism was informed by his reflections on mathematics and physics, the theory was not determined by these reflections but emerged in response to a traditional philosophical problem confronted by indirect or representational theories of perception: What is the character of our descriptive knowledge of the material world, if we lack acquaintance with that world? Since at least 1905, the idea that we lack such acquaintance was not an assumption Russell ever seriously doubted; nor, with the brief exception noted earlier, did he seriously doubt that there are material events and processes. And as early as *Problems of philosophy* Russell accepted Berkeley's criticisms of the Lockean view that our knowledge of material objects is divided between knowledge of their primary and secondary qualities. He agreed with Berkeley that there is no basis for supposing that we know the qualitative aspects of the constituents of the material world.[3] Unlike other less sophisticated exponents of similar views, Russell understood this to entail that we know neither the qualitative character of the properties of material events nor that of the relations among them. To have such qualitative knowledge would require acquaintance with the events of which they hold; but by hypothesis, this is precisely what we do not have.

Russell's earliest logical investigations into the concept of number led to the discovery that the notion of one-one correspondence is expressible with complete generality, being definable in pure second order logic. *As such, it is capable of being understood without acquaintance with any of the properties whose extensions it relates*. Similarity of properties or "equinumerosity" generalizes to similarity of relations—"structural similarity"—and this captures Russell's discovery of a kind of non-qualitative similarity that can be grasped without acquaintance with the relations whose fields it correlates. In particular we can have knowledge that a relation with which we are acquainted is structurally similar to one with which we are not acquainted; when we know of a relation that it is structurally similar to another relation, we are said to have *structural knowledge* of that other relation. Such knowledge is expressed by the assertion that there is a relation similar to one with which we are acquainted. Russell believed that

[2] Maxwell (1968) and (1970) remain among the most direct and readable papers on the subject.

[3] For the sake of simplicity, I am assuming a dualism of mental and material events. In a complete account of Russell's epistemology and metaphysics this assumption, and the conclusion just attributed to him in the text, would need to be qualified.

the discovery of structural knowledge yielded a solution to the problem that had eluded Locke, and that Kant had addressed with only limited success. Contrary to Locke, our knowledge of the material world is not qualitative. Contrary to Kant, it is possible to have significant knowledge of the material or "noumenal" world. The correct view is that of the relations of the material world our knowledge is always structural and never qualitative. Or as Russell also says, of the material world we can know its structure, but not its qualitative character. This, in essence, is Russell's structuralism.

There is a difficulty with this view that was elegantly stated by M.H.A. Newman in an article which appeared just one year after the publication of *The analysis of matter*. The observation on which Newman's criticism was based is a very simple one: Newman drew a contrast between the claim that a given relation has a certain structure—satisfies an appropriate condition—and the existential claim that there is a relation which satisfies this condition. He then observed that provided such an existential claim is consistent, it is true in *any* domain of sufficient cardinality. This however is not the case for claims regarding specific properties and relations: such claims of structural similarity are epistemologically significant precisely because they are assertions concerning the similarity of given relations, rather than bare existential claims to the effect that there is a relation which is similar to a given relation. In the context of Russell's Lockean realism, the difficulty this poses is that an assertion about the structure of relations holding among the parts of the material world with which we are not acquainted will be true in any sufficiently large model of those statements whose constituent terms refer only to events with which we are acquainted. Hence in a theory such as Russell's, modulo a cardinality claim, the *content* of our judgements regarding the material world coincides with the content of our reports on the objects of our acquaintance. Moreover, the cardinality assumption is represented as the only component of our knowledge of the material world that is not secured by logical considerations. But this is a wild distortion of what, pre-analytically, we take the epistemological significance of our judgements about the material world to be.

A response to Newman that has recently gained currency argues that the scope of the quantified variable should be restricted to *natural* or *real* relations. Then it is by no means obvious that the condition $\ldots R \ldots$ is satisfied by *such* an $R$. However, there are two difficulties with this response. First, it makes a simple fact involving our judgements of epistemic significance depend on metaphysically difficult notions when it is evident that it does not. And secondly, it fails to capture the scope of Newman's observation, which is independent of a relation's "reality" or "naturalness": there is a distinction to be made between significant and non-significant claims regarding the structural similarity of relations even when one or both of the relations is wholly artificial.

The source of the difficulty that Newman uncovered is this: in Russell's theory the distinction between a claim concerning a given relation and one that is merely about some relation or other reduces to the distinction between a relation that is "given" to us by our acquaintance with it and one that is not given in this sense. On such a view, it is not possible for an assertion to be of a given relation unless we have acquaintance with that relation, and this means that assertions regarding relations that fall outside our acquaintance are necessarily general claims, claims of the form *there* *is an R, ... R ....*[4] What Russell's empiricism seeks to preserve is unexceptionable, namely the distinction between properties and relations that are given in experience and those that are not. But as Newman's discussion shows, a position according to which the legitimacy of claims about given relations can be made out only in terms of acquaintance or its absence is incapable of capturing the epistemological significance of common judgements about the structural similarity of relations. This observation must be acknowledged by any response to Newman on Russell's behalf.[5]

## 3 The concept of analyticity and the epistemic status of mathematical theories

Although there is a natural affinity between Russell's structuralism and Carnap's mature conception of theoretical knowledge, the considerations by which Russell and Carnap were led to their respective views are very different. Russell's theory resulted from the desire to establish a form of Lockean realism in theory of perception. But for Carnap realism was always a prime example of the kind of commitment that cannot be addressed by philosophical arguments. The issue at the center of Carnap's epistemological concerns is the controversy between rationalism and empiricism over the nature and scope of a priori knowledge. And his strategy for addressing this controversy consists in providing a reconstruction of the theories of pure mathematics and empirical science, thereby transforming an issue of traditional epistemology into a part of the theory of theories and the rational reconstruction of the language of science. The basic empirical science is physics, and given the centrality of mathematics to the formulation of its theories, the rational reconstruction of the language of physics necessarily requires an account of pure mathematics.

It has often been noted that Frege's and Russell's logicism together with Hilbert's axiomatic method were among the central influences on Carnap's thought about mathematics. Both of these movements derived a significant part of their philosophical interest from the perception that they might be exploited to address the rationalist claim that mathematics represents knowledge that is both synthetic and a priori. Logicism sought to address this claim by deriving arithmetic—in the broad sense, which includes analysis and the theory of the real numbers—from logic. But the attempt to account for our knowledge of arithmetic by representing it as an extension of logic foundered, in Frege's case, on the inconsistency of his system, and in the case of Russell, on the need to assume a posteriori principles. Logicism's legacy was therefore mixed: its central development—polyadic logic with multiple generality—proved an

[4] Newman's analysis strongly suggests that there is a close connection between Russell's structuralism and the idea of a Ramsey sentence. But neither Russell's book nor Newman's discussion of it make any mention of Ramsey's view of theories, let alone Ramsey sentences. *The analysis of matter* appeared in 1927 and Newman published his paper in 1928. The paper of Ramsey which introduces Ramsey sentences ("Theories") appeared posthumously several years later. When in 1931 Russell reviewed *The foundations of mathematics and other logical essays* he seems not to have noticed the essay.

[5] An adequate theory of knowledge must also accommodate some form of Russell's empiricist intuition regarding the role of the given in experience. Meeting this condition is a significant undertaking, one which I don't attempt here. Gupta's elegant treatise (2006) presents a promising framework for addressing this task while preserving common intuitions regarding theoretical knowledge.

indispensable tool for the logical analysis of mathematics. But logicism left unsolved the fundamental problem of responding to the rationalist claim that mathematics represents a counter-example to any theory of knowledge that seeks to minimize the importance of the synthetic a priori.

In his *Foundations of geometry* (1899/1971), Hilbert articulated a view of the geometrical axioms and their role in our understanding of the primitives of a mathematical theory that provided the key to Carnap's logicism. Hilbert's *Geometry* suggested an account of our knowledge of mathematics that is both independent of its relation to logic and free of the rationalist excesses of the synthetic a priori. The idea that emerged from Hilbert's work was that mathematical axioms *define* the primitives that occur in them. Hilbert's monograph demonstrated the fecundity of the idea that the language of geometry is freely interpretable subject only to the constraint that the interpretation must respect the logical category of its vocabulary. The subject matter of the axioms includes any system of objects which forms the basis of an interpretation under which the axioms come out true; the "truth" of the axioms thus consists in the existence of such an interpretation, and therefore amounts to nothing more than their *consistency*. The axioms of a mathematical theory are known a priori because they are free stipulations; hence the correct explanation of our knowledge of them is one that renounces any claim to their being synthetic truths about the world. On this interpretation of Hilbert, a theory of pure mathematics is not, in Quine's phrase, true by convention; it is not true at all: in the sense in which it is usually intended, 'true' is not properly applied to the theories of pure mathematics. And Hilbert's contention, that in mathematics truth and existence mean the same as consistency, must be understood in the context of a broader dialectical argument against the tradition represented perhaps most famously among his contemporaries by Frege: Call such theories true if you wish, but recognize that this can only mean that they are consistent. It was this view that Einstein had in mind[6] when, explicitly referring to Hilbert, he said of geometry that insofar as it is certain it does not refer to reality, and insofar as it refers to reality it is not certain. By this he clearly meant that the peculiar epistemic status of pure mathematics comes at the expense of its being true—at the expense of its "referring to reality." Carnap and the logical empiricists were deeply influenced by Hilbert's views and Einstein's celebrated endorsement of them.

To return to the significance of Hilbert's work for Carnap's elaboration of logicism, the account of the axioms of geometry that emerges from Hilbert's investigations is based on the observation that they are constitutive of their subject-matter. Carnap perceived that this idea might plausibly be extended to arithmetic. The contribution of logicism to the philosophy of pure mathematics could then be seen to lie not in its account of the axioms of arithmetic, but in its elucidation of mathematical reasoning as an elaboration of purely logical reasoning. The successful extension of Hilbert's view of geometrical axioms to a philosophy of mathematics that is a genuine alternative to one that holds mathematics to be synthetic a priori would then depend on whether the apriority attributed to the axioms could be shown to derive from their analyticity. This requires both an explication of analyticity and a demonstration that the explication

[6] In his often-quoted January 27, 1921 address to the Prussian Academy of Sciences, "Geometry and Experience."

supports the conclusion that the axioms are correctly characterized as analytic under the proposed explication.

For Carnap, the crucial notion that makes this a feasible strategy—the notion on which the successful explication of analyticity rests—is that of *factual content*. Carnap's solution to the problem of a priori knowledge is to argue that such knowledge lacks factual content and is, in this sense, analytic. *Analyticity thus becomes something that can attach to a mathematical theory independently of whether or not it is recoverable from the basic laws of logic*. More generally, Carnap's approach seeks to preserve the fundamentally empiricist intuition that reality is not accessible to reason alone, while admitting the presence of a genuinely a priori component to theoretical knowledge.[7] From this perspective, the essence of Carnap's conception of the analytic is that it is *non-factual*—not that it concerns matters of meaning or synonymy. That a component of our conceptual framework is a matter of meaning or definition would, of course, be worth noting because it is *one* way of showing that component to be non-factual in the relevant sense. But the weight of Carnap's conception of analyticity rests on the analysis of *factuality*—not on *meaning* or *synonymy*. This also explains the intimate connection that for Carnap obtains between the analytic and the conventional: conventions, insofar as they are reflective of our decisions rather than how things stand, are by their nature non-factual ingredients of our conceptual framework, and for this reason, belong to its analytic component.

## 4 Factual content and the nature of empirical theories

Factual content is explicated in terms of the observational vocabulary of a theory, where, it is important to emphasize, the notions of *observational vocabulary item* and *observational sentence* are artifacts of the rational reconstruction of the language of the science under study. Putnam (1962) persuasively argued that the *un*reconstructed vocabulary items and the sentences formulated in terms of them are not easily classified as observational or theoretical in the required sense: they do not refer to *just* observable or unobservable events, a point that has been widely conceded within the logical empiricist and neo-logical empiricist traditions. But Putnam's observation, though correct, is largely irrelevant to the successful execution of a reconstructive program like Carnap's.

As just noted, the distinction that is required concerns *the vocabulary of the language of the reconstruction*. To make it out, it suffices that it should be possible to distinguish between observable and unobservable things and events, their properties and relations. So long as this distinction is admitted, it doesn't matter that the *un*reconstructed vocabulary in actual use is not a suitable instrument for drawing the intended dichotomy: Given a division of the domain of a possible model for the language of a theory into its observable and unobservable parts, we can introduce relation

[7] This empiricist intuition is also emphasized by Kant: "Thus all concepts and with them all principles, however a priori they may be, are nevertheless related to empirical intuitions, i.e., to *data* for possible experience. Without this they have no objective validity at all, but are rather a mere play, whether it be with representations of the imagination or of the understanding. One need only take as an example the concepts of mathematics" (Kant 1787, A239/B299).

symbols whose interpretation is restricted to the observable part of the domain.[8] The intended interpretation of the relation symbols of the theoretical vocabulary is the unobservable part of the domain of any model for the language of the theory. We call the observation vocabulary of the reconstruction the *O-vocabulary*, its theoretical vocabulary the *T-vocabulary*. An *O-sentence* is formed using only O-vocabulary items and a *T-sentence* is formed using only T-vocabulary items.

The artificiality of securing the observation-theory dichotomy in this way may be mitigated if we bear in mind that Carnap's project is a general epistemological one. Hempel captured the essential point when he wrote that Carnap sought to illuminate—indeed, set in relief—the classical empiricist project of

> showing that all our knowledge of the world derives from what is immediately given to us in the data of our direct experience. Stated in these general terms, the idea could be construed as a psycho-genetic claim concerning the development of man's conception of the world; but Carnap characteristically presented empiricism as a systematic-logical claim to the effect that all concepts suited to describe the world—and thus, all the concepts that could ever be required by empirical science, from physics to sociology and historiography—can be reduced, in a clearly specifiable sense, to concepts serving to describe the data of immediate experience or observation. . . . (Hempel 1975, pp. 1–2)

So understood, Carnap's project is importantly independent of how closely it addresses problems of the working scientist or how closely it adheres to ordinary intuitions about observational vocabulary. In this respect, rational reconstruction may be usefully compared to model theory and its fundamental preoccupation with the relation of mathematical structures to the formal languages with which they may be represented. Such investigations have borne importantly on questions within particular branches of mathematics that are of concern to working mathematicians, but their philosophical interest would not be diminished had they not done so. When Carnap's project is placed within a model theoretic framework, a salient feature of its approach to theory of theories is the division in vocabulary in terms of which it seeks to illuminate the role of the evidentiary basis for the interpretation and justification of theoretical claims.

It cannot be emphasized too strongly that the fundamental concept of Carnap's general epistemological project is that of *factual content*, and that its explication depends essentially on Carnap's proposals for reconstructing the language of physics. On Carnap's account of theories, theoretical terms make no contribution to the factual content of the theory to which they belong—a point that will be explained and justified in detail below. Rather, the theoretical terms acquire a content from the O-vocabulary of the language of the theory by the occurrence of mixed sentences, called "correspondence rules," which contain both T- and O-vocabulary items. A statement involving theoretical terms is properly regarded as a statement of an empirical theory only if such a theory contains correspondence rules that connect its constituent theoretical terms with observation terms.

8 For simplicity, we restrict ourselves to predicate and relation symbols.

Correspondence rules (C-rules) establish a correspondence between the holding of theoretical relations in the domain of unobservable events and the relations among observable events. They differ from the T-sentences by containing both O- and T-terms. In the absence of C-rules there is no non-arbitrary answer to the question whether a theoretical claim is true. This contention would be entirely commonplace if, instead of 'C-rules,' we wrote 'semantic interpretation,' taking the relativization to a specific interpretation to be the additional element that is required for the question of the truth of a theoretical claim to be well-posed. But when formulated in terms of C-rules the contention is a very different one, since it relativizes the truth of a T-sentence to an "epistemic interpretation" of the theoretical vocabulary in terms of the vocabulary of the evidentiary basis of the theory. Carnap and the logical empiricists rejected the idea that it is possible to address the problem of interpreting a theory as a theory about the material or actual world by giving an intended semantic interpretation.[9] From their perspective, the knowledge that the provision of such an interpretation requires is precisely what is expressed by the C-rules. The C-rules are therefore principles of epistemic interpretation, and without them, we would have no reason to suppose that we could even *understand* a semantic interpretation of the T-vocabulary in the domain of actual events. Here too the influence of Hilbert's *Geometry* was decisive: Hilbert had proposed that the axioms of a mathematical theory define the theory's primitives. It is possible to specify the "meanings" of the primitives of a *mathematical* theory by addressing only their logical category: such a specification *exhausts* what provision of a semantic interpretation of a mathematical theory achieves. But by contrast with a mathematical theory, a physical theory requires principles of *epistemic* interpretation. C-rules bridge the theoretical and observational vocabularies and secure the content of theoretical sentences in the evidential base; in doing so they elevate theoretical statements to the status of genuinely factual or synthetic claims.

The first phase of Carnap's rational reconstruction of the language of science is rooted in the division between T- and O-vocabulary and the corresponding division between T- and O-sentences it induces. The second phase of Carnap's reconstruction, to which we will soon turn, assumes that this regimentation of the vocabulary has been successfully effected and that the division between T- and O-predicates is exhaustive: there are no *mixed* primitive predicates—predicates that apply to both observable and unobservable events. The key distinction in primitive vocabulary is between observational and theoretical terms, and our understanding of the meaning of a primitive vocabulary item rests on the observability of its referent. A reconstruction such as Carnap's nevertheless allows for the formation of expressions that are about items that transcend our observation, allowing, for example, that a sentence built up out of exclusively O-vocabulary is about unobservable events. *This is not only entirely compatible with the view we are expounding but has long been regarded by its proponents as one of its principal strengths*. Only the sharp division in primitive O- and T-vocabulary is based on the observability or otherwise of the referents of the O- and T-terms. This point is often missed in discussions of the logical empiricist conception of theories, but it is characteristic of both it and its classical antecedents. What is vital to the epistemologi-

[9] This view is expressed with particular clarity in Hempel (1958/1965), p. 217.

cal point of the account is the sustainability of the dichotomy in primitive vocabulary on the basis of reference; *sentences* are O or T merely on the basis of the primitive vocabulary they contain.

If the exclusion of mixed primitive predicates were relaxed, a correlative question would naturally arise at a later stage, "Is our understanding of mixed predicates unproblematic in the way in which our understanding of O-predicates is held to be unproblematic, or do mixed predicates, like T-predicates, pose a special difficulty?" The reconstruction of theoretical knowledge we are now exploring excludes mixed primitive predicates from the language of the reconstruction or classifies them with T-predicates as requiring special consideration in any account of how they are understood. This is not an arbitrary choice, but is deeply imbedded in the epistemological basis and motivation of the reconstructive project. For the present, I will assume that there are no mixed primitive predicates in the language of the reconstruction. I will later discuss a contemporary proposal that weakens this restriction.

I should also mention that Carnap distinguishes between the case where the quantificational apparatus of the O-sentences is "nominalistic" from the case where it is allowed to include the full logical resources of the language, including variables for abstract (mathematical) objects, classes of them, classes of classes of them, etc. Since the separation into nominalistic and non-nominalistic languages plays no role in the following exposition, I will assume that the full complement of logical resources is available for the formation of the O-sentences.

There are, as I have said, at least two phases to Carnap's reconstruction to consider. The first phase introduces a distinction between T- and O-vocabulary which extends to a distinction in theoretical and observational sentences in the intended way. It must be stressed that the formulation of the language of physics this assumes is already heavily reconstruction-dependent. The conjunction of the correspondence rules and theoretical postulates comprising a theory is then given by

$$\boldsymbol{TC}(O_1, \ldots, O_m; T_1, \ldots, T_n),$$

where $O_1, \ldots, O_m$ and $T_1, \ldots, T_n$ are the O- and T-predicates introduced at the first phase of the reconstruction. Notice, we are supposing that the correspondence rules and theoretical postulates are finite in number. As observed in (Craig and Vaught 1958), this is an assumption that can always be met, though doing so may incur the cost of increasing the strength of the underlying logic of the theory.[10]

The *second*, and distinctively Carnapian phase of the reconstruction consists, in the first instance, in replacing a theory $\boldsymbol{TC}$ with its *Ramsey sentence* $\mathbf{R}(\boldsymbol{TC})$,

$$\exists X_1, \ldots, \exists X_n \boldsymbol{TC}(O_1, \ldots, O_m; X_1, \ldots, X_n),$$

i.e., in replacing the theoretical predicates $T_1, \ldots, T_n$ of $\boldsymbol{TC}$ with variables $X_1, \ldots, X_n$ of the appropriate type and arity and existentially generalizing over the

[10] Many of the central metalogical results relevant to the discussion to follow are contained in Craig and Vaught (1958). For a summary of the relevant definitions and theorems, see van Bentham (1978) especially pp. 327–329, from Definition 3.9 through the discussion following Theorem 3.14, which can be read independently of the rest of van Bentham's review article.

new variables. Such sentences were first discussed in Ramsey (1929), but the general idea of expressing what we would today call "satisfiability in a model" by a higher-order existentially quantified sentence was a common practice in the logical tradition of the 1930s.[11] From a model-theoretic perspective, the innovation of the Ramsey sentence consists in the fact that it uses a higher-order sentence to express satisfiability in a model relative to a fixed interpretation of a part of the language, namely, the O-vocabulary.

The Ramsey sentence of a theory is important for Carnap because it and the theory imply the same O-sentences, since if there is an assignment to the variables $X_1, \ldots,$ $X_n$ which satisfies the matrix

$$\boldsymbol{TC}(O_1, \ldots, O_m; X_1, \ldots, X_n)$$

of the Ramsey sentence, then this assignment can form the basis of an interpretation of the theory's theoretical terms $T_1, \ldots, T_n$ under which it is true. Hence, if an O-sentence is not a consequence of **R**(***TC***), neither is it a consequence of ***TC***. The fact that ***TC*** and **R**(***TC***) imply the same O-sentences motivates the proposal that the Ramsey sentence of a theory represents its factual content.

Carnap's account of the conventional or analytic component of ***TC*** requires the notion of the *Carnap sentence* **C**(***TC***) of a theory, namely, the conditional whose consequent is ***TC*** and whose antecedent is ***TC***'s Ramsey sentence.[12] The conjunction **R**(***TC***) $\wedge$ **C**(***TC***) is obviously logically equivalent to ***TC***. Carnap argues that the Carnap sentence is analytic on the ground that all of its O-consequences—all the sentences in the O-vocabulary it logically implies—are logically true (L-true). The Carnap sentence is in this sense *observationally uninformative*, a fact that can be readily verified.[13] There is therefore a straightforward sense in which the Carnap sentence can be said to have no factual content and can, therefore, properly be regarded as analytic.

The reconstruction that emerges from these considerations thus divides ***TC*** into two components, **R**(***TC***) and **C**(***TC***), the first expressive of ***TC***'s factual content, the second merely a stipulation controlling the use of its theoretical vocabulary $T_1, \ldots, T_n$ and expressive of the theory's analytic component, where a sentence is *analytic*—or more precisely, *analytic in* ***TC***—if it is a consequence of just **C**(***TC***).

There are three desiderata Carnap imposes on a reconstruction that seeks to incorporate such a division into analytic and factual components:

($i$) The conjunction of the factual and analytic components of ***TC*** is logically equivalent to ***TC***.

[11] This is true even of Tarski's seminal papers. See Hodges (2004, p. 97) .

[12] An early presentation of the Carnap sentence is given in Carnap's 1959 Santa Barbara lecture to the Pacific Division of the American Philosophical Association (recently published with an historical introduction as Psillos (2000). The idea is developed in Carnap's *Replies*; see also Carnap (1966), the text based on Carnap's UCLA lectures, edited by Martin Gardner).

[13] If **C**(***TC***) implies X, then ¬**R**(***TC***) implies X and ***TC*** implies X. But since X is an O-sentence, **R**(***TC***) also implies X. Hence, **R**(***TC***) ∨¬**R**(***TC***) implies X, and X is L-true. Notice that, since for any ***T***, **R**(***T***) is O-equivalent to ***T***, the O-uninformativeness of **C**(***TC***) is equivalent to the claim that the Ramsey sentence of the Carnap sentence of ***TC*** is L-true.

$(ii)$ The factual component is O-equivalent to $\boldsymbol{TC}$.
$(iii)$ The analytic component is observationally uninformative.

These desiderata are naturally fulfilled by Carnap's proposed rational reconstruction in terms of the Ramsey and Carnap sentences of a phase one reconstructed theory—of a theory whose formulation respects the division into T- and O-sentences.

## 5 Winnie's emendation of Carnap's proposed reconstruction

There is a difficulty with Carnap's reconstruction which was noted by John Winnie in his (1970). To understand Winnie's observation, recall that $\boldsymbol{TC}$ is a conjunction of T-postulates and C-rules. Now suppose we take one of the T-postulates, $\boldsymbol{T}_i$ say, and propose the new and equivalent reconstruction, $\mathbf{R}(\boldsymbol{TC}) \wedge [\mathbf{C}(\boldsymbol{TC}) \wedge \boldsymbol{T}_i]$, which takes $[\mathbf{C}(\boldsymbol{TC}) \wedge \boldsymbol{T}_i]$ as its analytic component. For this proposal to be allowed, $[\mathbf{C}(\boldsymbol{TC}) \wedge \boldsymbol{T}_i]$ must be shown to satisfy Carnap's third desideratum. That, in many interesting cases it does, is the content of the following proposition of Winnie's:

*Suppose* $\boldsymbol{TC}$ *is satisfiable and that* $\boldsymbol{T}_i$ *is a T-sentence logically implied by* $\boldsymbol{TC}$. *Then* $\boldsymbol{C}(\boldsymbol{TC}) \wedge \boldsymbol{T}_i$ *is O-uninformative*.

Winnie's proof[14] depends on the following corollary to Craig's Lemma:

*Let X and Y be sentences of a first order language without equality such that (i) X and Y share no nonlogical vocabulary, (ii) X logically implies Y, and (iii) X is satisfiable. Then Y is logically true*.

I will return to this corollary and its proof. For now, let us verify its role in the proof of Winnie's proposition.

We must show that if $\mathbf{C}(\boldsymbol{TC}) \wedge \boldsymbol{T}_i$ implies an observational X, then X is L-true. Since, by hypothesis, $\mathbf{C}(\boldsymbol{TC})$ implies $\boldsymbol{T}_i \rightarrow$ X, $\neg\mathbf{R}(\boldsymbol{TC})$ implies $\boldsymbol{T}_i \rightarrow$ X. Now $\boldsymbol{TC}$ implies $\mathbf{C}(\boldsymbol{TC})$, and by hypothesis $\boldsymbol{TC}$ implies $\boldsymbol{T}_i$; hence, since $\boldsymbol{TC}$ implies X and X is an O-sentence, $\mathbf{R}(\boldsymbol{TC})$ *also* implies X; and therefore $\mathbf{R}(\boldsymbol{TC})$ implies $\boldsymbol{T}_i \rightarrow$ X. Thus $\boldsymbol{T}_i$ implies X. Now $\boldsymbol{T}_i$ and X share no nonlogical vocabulary and $\boldsymbol{T}_i$ is satisfiable. It therefore follows by the corollary that X is L-true and that $\mathbf{C}(\boldsymbol{TC}) \wedge \boldsymbol{T}_i$ is O-uninformative.

This observation regarding $\boldsymbol{T}_i$ and $\mathbf{C}(\boldsymbol{TC})$ can be iterated through all the T-postulates of $\boldsymbol{TC}$. There is, therefore, nothing to exclude the acceptability of a reconstruction which, like $\mathbf{R}(\boldsymbol{TC}) \wedge [\mathbf{C}(\boldsymbol{TC}) \wedge \boldsymbol{T}_1 \wedge \ldots \wedge \boldsymbol{T}_k]$, takes the conjunction of *all* the T-postulates $\boldsymbol{T}_1 \wedge \ldots \wedge \boldsymbol{T}_k$ of $\boldsymbol{TC}$ to be part of the analytic component. This consequence was wholly unanticipated by Carnap and it forms the justification for my earlier claim (Sect. 4) that, on a reconstruction like Carnap's, theoretical terms make no contribution to the factual content of the theory to which they belong.

To address this difficulty, Winnie (1970, p. 150) proposes revising Carnap's reconstruction by adding a fourth desideratum. This additional desideratum can be motivated by the classical theory of definition according to which, if $\boldsymbol{T}$ is a set of axioms in the

[14] Cp. section V of the appendix to Winnie (1970) as well as the discussion on pp. 149–150, where the page numbers refer to its reprinting in Hintikka (1975). Notice also my departure from Winnie's terminology.

language $\boldsymbol{L}$ and $T$ is a term not in $\boldsymbol{L}$, a definition $\mathrm{X}(T)$ of $T$ (in $\boldsymbol{T}$) is *noncreative* if, whenever a sentence of $\boldsymbol{L}$ is a logical consequence of $\boldsymbol{T} \wedge \mathrm{X}(T)$, it is a consequence of $\boldsymbol{T}$ alone. Winnie imposes a similar condition on the analytic component of $\boldsymbol{TC}$:

(*iv*) The analytic component of $\boldsymbol{TC}$ is observationally noncreative in $\boldsymbol{TC}$,

where a sentence X *is observationally noncreative in* $\boldsymbol{TC}$ just in case $\boldsymbol{TC}$ logically implies X, and for any Y such that $\boldsymbol{TC}$ logically implies Y, every O-consequence of $\mathrm{X} \wedge \mathrm{Y}$ is an O-consequence of Y. Observational noncreativity is clearly a special case of the non-creativity requirement for definitions. Given Carnap's conception of the factual content of a theory, observational noncreativity is evidently *the* noncreativity condition that truths which are analytic in $\boldsymbol{TC}$ should satisfy.

It can be shown[15] that Winnie's desideratum *(iv)* is satisfied by Carnap's original proposal to represent the analytic component of $\boldsymbol{TC}$ by $\mathbf{C}(\boldsymbol{TC})$, but not by any of the problematic extensions $[\mathbf{C}(\boldsymbol{TC}) \wedge \boldsymbol{T}_i]$. Indeed, the consequence class of the Carnap sentence characterizes *exactly* the sentences that are O-noncreative in $\boldsymbol{TC}$: any sentence not implied by the Carnap sentence will, when added to $\mathbf{C}(\boldsymbol{TC})$, be O-creative in $\boldsymbol{TC}$. This is arguably the principal interest of the Carnap sentence for the project of rational reconstruction. The consequence class of the Carnap sentence is the largest subclass of the class of O-uninformative sentences of $\boldsymbol{TC}$ that are O-noncreative in $\boldsymbol{TC}$. The condition of O-noncreativity rules out adding to the analytic component any sentence not implied by $\mathbf{C}(\boldsymbol{TC})$, and this is sufficient to keep some T-sentences from forming part of the analytic component of the theory.

There are many positive features of Carnap's reconstruction when it is revised in the manner just reviewed. We may note three.

The holism of the approach means that the division of factual and analytic components is independent of the formalization. Since the consequence class of the Carnap sentences of two distinct, but logically equivalent, formulations of a theory are the same, so also are their characterizations of analytic and factual sentences. As Winnie (1970, p. 149) has observed, the relativity that attaches to the notion *postulate of* $\boldsymbol{T}$ does not attach to the notion *analytic in* $\boldsymbol{T}$. Moreover, as Hempel (1963, p. 703) has emphasized, the ambiguous status of C-rules makes the application of the analytic/synthetic distinction to the sentences of a phase one reconstruction highly problematic. Indeed, it calls into question the very possibility of a non-arbitrary dichotomy of sentences of this phase into analytic and factual statements. These difficulties are completely avoided when the distinction is applied at the second reconstructive phase—when the analytic sentences are represented as the consequence class of the Carnap sentence of the theory.

Factuality is explicated in terms of observability, since, among other things, a factual sentence is explained as one having nontautological *observational* consequences. But although Carnap uses observability to explicate factuality, he does not propose to reduce the referent of a theoretical term to its observable effects, nor does he equate the meaning of a theoretical sentence with its observational consequences. In these

15 See Winnie (1970) or the appendix to my (2007).

respects, his proposal is not a standard verificationist one. But it is also not a weak confirmationist theory of meaning of the sort considered in (Quine 1951, Sect. 5), since according to such a theory, the meaning of a T-sentence is given by the class of its O-consequences. Carnap's account of the meaning of T-terms by means of the Carnap sentence is a variant of the doctrine of implicit definition by postulates.[16] But it differs importantly from that doctrine by its retention of the classical distinction between postulates and definitions, and by its provision of an implicit definition of the T-terms that is compatible with the requirement that a definition not be factually creative. For, if we grant Carnap's explication of factuality, then the Carnap sentence is not only without factual content, it does not *create* factual content. It is this feature that makes it plausible to propose that the Carnap sentence is properly regarded as a kind of definition, since non-creativity is the key feature that definitions have traditionally been charged with exhibiting in order for them to be regarded as analytic. By contrast, an implicit definition of T-terms by postulates will typically be creative in this sense.

The problem of explicating factual content is shown not to require a solution to the vexed problem of the meaning of the theoretical vocabulary, the problem of how terms which purport to refer to unobservable entities can be understood. Carnap's proposal addresses this problem by replacing all the theoretical predicates with variables, and then implicitly defining them with the Carnap sentence; hence, aside from the matter of their logical category, their elimination means that there is no need to appeal to theoretical terms—nor therefore to their meaning—in order to provide for the factual content of a theory. This is perhaps the point at which Hilbert's influence is most evident.

## 6 Empirical adequacy and truth

Borrowing a terminology Carnap employs in another context (Carnap 1950), let us distinguish two problems of underdetermination, one "external," the other "internal." We understand the *internal* problem to be the price of empirical inquiry under the constraint of finite resources. The data accessible to us can always be improved, and the parameters employed in their description can always be replaced with new parameters. Underdetermination at one stage in this process of refinement might be resolved at a later stage by the expansion of the available observational information, but some underdetermination is always present because of the incompleteness of our data.

By contrast, the *external* problem of underdetermination concerns the ideal case where *all* observational data are in. In her (1973), English observed that within a Ramsey sentence reconstruction of theories such as Carnap's the notion that there could be an external problem of underdetermination rests on a mistaken supposition. Although English fails to locate the correct metalogical justification for the conclusions she draws—in particular, she fails to locate the role of Craig's Interpolation Lemma in the argument—her observation is correct: For Carnap, given complete observational

[16] This way of expressing the point was suggested to me by a remark of Anil Gupta's in his entry, 'Definitions,' for the *Stanford encyclopedia of philosophy*.

knowledge, two theories can be incompatible only if there is a sentence expressible in their common observational vocabulary which is implied by one theory while its negation is implied by the other.

To see this, recall first that Craig's Lemma tells us that if $X_i$ $(i = 1, 2)$ are first order sentences in $\boldsymbol{L}_i$ such that $\boldsymbol{L} = \boldsymbol{L}_1 \cap \boldsymbol{L}_2$ and $X_1 \wedge X_2$ has no model, then there is a sentence X of $\boldsymbol{L}$ such that $X_1$ implies X and $X_2$ implies $\neg$X. Now consider two first order theories $\boldsymbol{T}_1, \boldsymbol{T}_2$ with disjoint T-vocabularies but coincident O-vocabularies, such that $\boldsymbol{T}_1 \cup \boldsymbol{T}_2$ is inconsistent. By the Compactness Theorem, if $\boldsymbol{T}_1 \cup \boldsymbol{T}_2$ is inconsistent, there are finite subsets $\Sigma_i \subseteq \boldsymbol{T}_i$ $(i = 1, 2)$ such that $\Sigma_1 \cup \Sigma_2$ is inconsistent. Let $X_i$ be the conjunction of the sentences in $\Sigma_i$. Then $X_1$ implies $\neg X_2$. Hence, by Craig's Lemma there is a sentence X such that $X_1$ implies X and X implies $\neg X_2$, and X is an O-sentence formulated in $\boldsymbol{L}$, the common observational vocabulary of $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$. Hence, $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ cannot both be compatible with all observations: if X holds, then $\boldsymbol{T}_2$ is false, and if it does not hold, $\boldsymbol{T}_1$ is false. Hence, if $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ are inconsistent, they cannot be compatible with the same "data," if by 'data' we mean what is expressible in the common O-vocabulary of these two first order theories.

But now Carnap's Ramsey sentence reconstruction allows this fact about first order theories to be exploited for *any* two first order theories whose languages distinguish separate T- and O-vocabularies and whose O-vocabularies coincide. This follows because the essential idea behind the use of the Ramsey sentence in reconstructing a theory is that only the logical category of the theoretical terms is relevant to their theoretical role. Thus so long as the logical category of the original vocabulary items is preserved, someone who grants that the Ramsey sentence of a theory captures its factual content can have no objection to a uniform replacement of the T-vocabularies of $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ with new nonlogical constants. Such a replacement makes the T-vocabularies of the theories disjoint and thereby ensures that $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ satisfy the conditions that are required for the application of Craig's Lemma. The representation of the factual content of a theory by its Ramsey sentence therefore implies that the notion that two theories might be compatible with the same data and yet conflict with one another must be given up, so that a conflict in some T-postulate must always be reflected in an O-sentence. Given Carnap's conception of factual content, *complete* observational knowledge *excludes* underdetermination. For Carnap there is no external problem of underdetermination because empirical adequacy with respect to all possible observational evidence is the same as truth.

## 7 The factuality of theoretical sentences

Carnap's adherence to the idea that the factual content of a theory is captured by its observational content is an abiding feature of his view. But his understanding of this thesis contains a subtlety which is briefly noted in the recently published transcription of his 1959 Santa Barbara lecture:

> One of the most important characteristics of the T-terms, and therefore of all sentences containing T-terms—at least if they occur not in a vacuous way—is that their interpretation is not a complete one, because we cannot specify in an

> explicit way by just using observational terms what we mean by the 'electromagnetic field.' We can say: if there is a distribution of the electromagnetic field in such and such a way, then we will see a light-blue, and if so and so, then we will see or feel or hear this and that. But we cannot give a sufficient and necessary condition entirely in the observational language for there being an electromagnetic field, having such and such a distribution. Because, in addition to observational consequences, the content is too rich; it contains much more than we can exhaust as an observational consequence. (Psillos 2000, p. 159)

Carnap does not further elaborate on the nature of the richness in content of theoretical concepts to which this last sentence appeals. But it is important to recognize that he is *not* saying that the richness of theoretical concepts is at variance with the idea that the Ramsey sentence of a theory captures its factual content. The Santa Barbara lecture, like its later elaboration in his *Replies* (1963), explicitly assumes that the observation language is extended—unrestricted in its logical and mathematical resources—and advances the thesis that under this assumption, the Ramsey sentence fully captures a theory's factual content. I believe this passage is an allusion to a point elaborated in detail by Hempel (1958/1965, Section 9) in his discussion of a well-known observation of Craig's regarding the axiomatizability of recursively enumerable sets of sentences: Given a theory $\boldsymbol{T}$ whose vocabulary is divided between T- and O-terms, and whose O-consequences are recursively enumerable, there exists a recursive set of sentences in just the O-vocabulary of $\boldsymbol{T}$ from which all of its O-consequences follow. Call such an axiomatization a *Craig transcription* of $\boldsymbol{T}$. Hempel shows that while a Craig transcription has the same O-consequences as $\boldsymbol{T}$, it does not preserve those connections between O-properties (and relations) that are expressed by the T-sentences and C-rules of $\boldsymbol{T}$. Such sentences relate O-properties to one another without establishing deductive connections between O-sentences, and the relations between properties they assert can be exploited by inductive arguments based on $\boldsymbol{T}$. By contrast with $\boldsymbol{T}$, *and by contrast also with the Ramsey sentence of* $\boldsymbol{T}$, a theory's Craig transcriptions are incapable of supporting such inductive arguments. This contrast, though emphasized by Hempel in his discussion of theories and their Craig transcriptions is not addressed in his discussion of Ramsey sentences and Craig transcriptions. Indeed, one could easily be led to conclude from Hempel's discussion that the difficulty with Craig transcription is attributable to its elimination of theoretical *terms*. But the problem is not that it eliminates T-terms, but that, unlike the Ramsey sentence, a Craig transcription fails to provide analogues of the sentences which contain them. Thus although the Ramsey sentence *also* replaces $\boldsymbol{T}$ by a theory whose vocabulary is wholly observational, the Ramsey sentence preserves the connections between O-properties which are expressed by $\boldsymbol{T}$ and which are lost in the transition to any of its Craig transcriptions.[17] When

[17] In 'The theoretician's dilemma' Hempel writes:

> So far, we have examined the eliminability of theoretical concepts and assumptions only in the context of deductive systematization: we considered an interpreted theory . . . exclusively as a vehicle [for] establishing deductive transitions among observational sentences. However, such theories may also afford means of inductive systematization . . .; an analysis of this function will yield a further argument against the elimination of theoretical expressions by means of Craig's method. (Hempel 1958/1965, p. 214)

Carnap says that the content of a theoretical construct like the electromagnetic field is too rich to be exhausted by its observational consequences, he is alluding to the fact that its content cannot be captured by a reconstruction which merely preserves the O-consequences of the theory to which it belongs. But this is precisely where the Ramsey sentence represents an advance over a reconstruction by Craig transcription.[18]

The nature of the special existence assumptions the Ramsey sentence incorporates is given a particular interpretation by Carnap in his response to Hempel, where he says of the Ramsey sentence that while it

> does indeed refer to theoretical entities by the use of abstract variables[, ...] it should be noted that these entities are ... purely logico-mathematical entities, e.g. natural numbers, classes of such, classes of classes, etc. Nevertheless, **R**(***TC***) is obviously a factual sentence. It says that the observable events in the world are such that there are numbers, classes of such, etc. which are correlated with the events in a prescribed way and which have among themselves certain relations; and this assertion is clearly a factual statement about the world (*Replies*, p. 963).

This reply expresses a view of theoretical properties and relations that appears to reject the idea that the higher order variables range over a domain built upon unobservable things and events. But in his (1966, p. 255), Carnap offers a clarification:

> [... physicists may, if they so choose,] evade the question about [the existence of electrons] by stating that there are certain observable events, in bubble chambers and so on, that can be described by certain mathematical functions, within the framework of a certain theoretical system. ... [T]o the extent that [the theoretical system] has been confirmed by tests, it is justifiable to say that there are instances of certain kinds of events that, in the theory, are called "electrons."

Such remarks suggest that the conception of the factuality of the theoretical postulates that underlies Carnap's reconstruction—even with Winnie's emendation—may not

Footnote 17 continued
If we understand "theoretical expressions" to include the T-*sentences* of the theory, then Hempel has likely seen the contrast between a Craig transcription and a Ramsey sentence. But if in this passage Hempel merely means to call attention to the elimination, by Craig transcription, of theoretical *terms*, he has likely missed the essential difference. As is well-known, Hempel's contribution to the Schilpp volume was an earlier version of 'The theoretician's dilemma,' and in this earlier presentation he seems to be discussing only the elimination of theoretical terms:

> But I think there is yet another reason why science cannot dispense with theoretical terms in this fashion. Briefly, it is this: The application of scientific theories in the predication [sic] and explanation of empirical findings involves not only deductive inference, i.e. the exploitation of whatever deductive connections the theory establishes among statements representing potential empirical data, but it also requires procedures of an inductive character, and some of these would become impossible if theoretical terms were avoided. (Hempel 1963, pp. 699–700)

The passage would be decisive except for the phrase "in this fashion" which concludes the first sentence. Nevertheless, it remains the case that Hempel never explicitly emphasizes the contrast between Craig and Ramsey noted in the text.

[18] The discussion of this paragraph is indebted to Michael Friedman and Thomas Uebel. A question raised in correspondence with them led me to investigate the precise relationship between ramsification and Craig transcription, and the significance of the different ways in which they eliminate T-vocabulary.

be sufficiently robust. The difficulty is similar to the one pointed out by Newman in connection with Russell's structuralism. For Carnap, theoretical sentences, though factual, are *almost* logical truths, and hence, are almost analytic. It is often suggested that the effect of ramsification is to call into question the status of realism. Since neither realism nor its alternatives are theses Carnap seeks to defend or dispute, this would be a difficulty at most for his metaphysical neutrality.[19] But a more basic difficulty with ramsification—and the difficulty I wish to emphasize here—is whether, having identified the factual content of a theory with its Ramsey sentence, one can claim to have captured pre-analytic intuitions about the nature of our knowledge of the truth of theoretical statements. This is arguably the fundamental methodological issue that a reconstruction of our theoretical knowledge must successfully address, and it bears on Carnap's claim to have captured the notion of factuality that attaches to theoretical assertions.

We can begin to see how the "quasi-analyticity" of theoretical claims arises by reviewing the proof of the corollary to Craig's Lemma which figured importantly in the analysis of Carnap's notion of observational uninformativeness, namely

> *Let X and Y be sentences of a first order language without equality such that (i) X and Y share no nonlogical vocabulary, (ii) X logically implies Y, and (iii) X is satisfiable. Then Y is logically true*.

It will be recalled that the corollary implies that on a reconstruction like Carnap's, a purely theoretical statement—one with no O-vocabulary items—has only L-true observational consequences. Of greater interest, in the present context, is how its standard proof[20] exploits the disjointness of the T- and O-vocabularies to effect a certain model construction. Arguing toward a contradiction, suppose that there is a model **M** in which Y fails so that ¬Y holds in **M**. Since X is satisfiable, it too has a model **N**, which we may assume is of the same cardinality as **M**. Let $f$ be a one-one onto map from the domain $N$ of **N** to the domain $M$ of **M**, and for each $n$-ary relation symbol $R$ occurring in X define an $n$-ary relation $R^M$ on $M$ by the condition, $\langle a_1, \ldots, a_n\rangle$ is in $R^M$ iff $\langle f^{-1}a_1, \ldots, f^{-1}a_n\rangle$ is in $R^N$. Since X holds in **N**, this expands **M** to a model **M*** for the vocabulary of X in which X holds when its relation symbols are interpreted by the relations $R^M$. Since X and Y have disjoint vocabularies, ¬Y is true in **M*** iff ¬Y is true in **M**. Thus X ∧¬Y holds in M*, contrary to the hypothesis that X implies Y.

Notice that the language ***L*** for which the corollary holds is restricted: it is a language without equality. If ***L*** contained equality, ¬Y might hold only in a model of finite cardinality; but if, for example, X holds only in infinite models, the argument will

---

[19] The importance of Carnap's metaphysical neutrality is stressed by Thomas Uebel in his "Carnap, explication and the problems of Ramseyfication", (currently in preparation), and by Michael Friedman, in his contribution to the present number of this journal. I agree with Uebel and Friedman regarding Carnap's neutralism on metaphysical issues like realism; however, I do not believe that his neutrality in matters of metaphysics extended to neutrality concerning what counts as a factual claim. Nor do I believe that Carnap's "principle of tolerance" regarding systems of logic requires neutrality regarding the domain of the factual; what seems rather to be the case is that tolerance in logical matters is justified by the non-factuality of the analytic.

[20] Our proof-sketch is based on Robinson (1974, 5.1.8).

break down. Restricting the language prevents this, since satisfiability then implies satisfiability in a countably infinite model. This restricts the generality of the corollary, but it does not restrict its philosophical interest. The only effect of the restriction on ***L*** that we require is

> (*) *The reconstruction applies to sentences which are true in countably infinite models if they are true at all.*

Taking ***L*** to be without equality is a simple way of ensuring (*), but it is not strictly necessary: we can simply impose the requirement (*) directly and proceed to avail ourselves of equality and the expressive resources it brings.

Now suppose there is a model **M** in which the O-sentences hold. Then provided only that the cardinality of **M** is not unduly restricted by the O-sentences, we can, just as in the proof of the corollary to Craig's Lemma, expand **M** to a model **M*** in which the T-sentences are also true. The sense of "almost analytic"—and even, "almost L-true"—that applies to the T-sentences is this:

> *Modulo a logical assumption of satisfiability and an empirical assumption about cardinality, it follows that if the O-sentences are true, the T-sentences are also true.*

The simple proof of the technical part of this claim follows exactly our argument for the corollary. But the main interest of the claim is conceptual, rather than technical, and it derives from the fact that the relations which, when assigned to the T-predicates occurring in a T-sentence ***T***, make ***T*** true-in-**M*** also make ***T*** *true*. Such an inference is permitted by Carnap's reconstruction because the content of ***T*** is reduced to the purely existential assertion that *there are* relations corresponding to the relational expressions of ***T*** which make it true. So long as ***T*** is satisfiable, this will hold as a matter of logic in *any* model of the O-sentences, provided only that the model is large enough.[21]

For Hilbert truth and consistency coincide for theories of pure mathematics, but there is no indication that Hilbert maintained this view of applied mathematics; such theories were for him constrained in a way in which theories of pure mathematics are not. Here it is interesting to consider Poincaré's view of geometry as a theory of the structure of space—which is of course a piece of *applied* mathematics—since Poincaré's view regarding it has a great deal in common with Hilbert's conception of theories of *pure* mathematics. Poincaré argued that since Bolyai-Lobachevsky geometry has a true interpretation in any model of Euclidian geometry, the choice between one or the other geometry as the correct theory of space can be shown to be a matter of convention. Poincaré's full argument leaves unaltered the interpretation of the physical vocabulary but invokes modifications of physical assumptions regarding the presence of forces; however it is clear that the relative interpretability of one geometry in the other is the argument's key premise, and it is clear that it was its discovery

[21] This observation does not depend on Carnap's alleged commitment to what has been called "the syntactic view of theories." This is argued at length in section 5 of my (2003a), where the observation is shown to be equally true of views (such as constructive empiricism) that are based on the so-called "semantic view of theories."

that led Poincaré to the conclusion that the choice between geometries is not a factual question.[22]

Now it is of great importance to the viability of Carnap's drawing of the analytic-synthetic distinction that *some*, at least, of the T-sentences not be part of the analytic or non-factual component of his reconstruction.[23] We have seen how this can be secured by Winnie's observation that T-sentences are typically O-creative. And as we have also seen, the Carnap sentence is naturally and untendentiously regarded as expressing a convention, since it has no non-tautological observational consequences and satisfies Winnie's non-creativity condition. But there is a much closer connection between Poincaré's conventionalism regarding geometry and Carnap's general analysis of theories than Carnap's use of the terms 'factual' and 'non-factual' or 'conventional' would otherwise suggest, and it is only the slightest thread that separates Carnap from a conventionalism about the theoretical that parallels Poincaré's view of geometry. Carnap's position on the truth of theoretical sentences is that provided the theoretical sentences are consistent, they are true, since they are interpretable as true in an expansion of any model of the O-consequences of the theory to which they belong, subject only to the provision that the domain of the expansion is large enough. To extend this argument to *another* set of T-sentences it is necessary to ensure that there is no overlap between its T-vocabulary and the T-vocabulary of the original theory. But given Carnap's acceptance of the Ramsey sentence reconstruction, this must always be possible, since if there is no objection to replacing T-terms with variables, there can be no objection to replacing them with new constants. The argument which establishes the truth of the T-sentences of the original theory in this model, can then be extended to this and many other sets of T-sentences as well. The O-sentences form a fixed and stable set of truths, but the situation is altogether different with T-sentences. Many such sets are available and in Carnap's framework there appears to be no non-conventional basis for choosing among them.

## 8 Modifying the Ramsey sentence

By contrast with Carnap, the partition of a theory's primitive non-logical vocabulary favored by the school of British structuralists[24] distinguishes between O-terms, T-terms, and terms which refer to properties and relations that hold of both observable and unobservable entities—the so-called *mixed* terms. It is clear that if one includes the mixed predicates within the choice of terms to ramsify, the result is the same as if

[22] These remarks are not intended to do justice to Poincaré's complex view. For two recent contrasting studies of Poincaré on geometry see Ben-Menahem (2006) and DiSalle (2006).

[23] The point is expressed with particular directness in his Santa Barbara lecture: "... we want to make a distinction between logical truth and factual truth. I believe that such a distinction is very important for the methodology of science. I believe that the distinction between pure mathematics on the one side and physics, which contains mathematics but in applied form, on the other side, can only be understood if we have a clear explication of the distinction that in traditional philosophy is known under the terms analytic and synthetic, or necessary truth and contingent truth, or however you may express it" (Psillos 2000, p. 160).

[24] My discussion of British structuralism is based on the paper of Ketland (2004) and the response to it by Cruse in his (2005a, b).

one neglected to distinguish a special category of mixed predicates. But if one leaves the mixed predicates constant, the Ramsey sentence that results is *not* one which holds in any sufficiently large model of just the O-sentences for the simple reason that what the original theory claimed regarding *mixed* properties and relations must also hold; and there may be models of the O-sentences that are not models of the sentences that contain mixed-vocabulary. By not ramsifying on mixed predicates, the truth of sentences containing mixed and O-predicates becomes as much a condition for the truth of the Ramsey sentence as it was for the truth of the original (*un*ramsified) theory. On this view, the content of a theory is not given by its O-consequences but by its consequences in both its mixed and O-vocabulary. Hence, when *factual content* is explicated by the modified Ramsey sentence it is no longer true that the factual content of the theory is expressed by its observational content. Content now extends to assertions involving terms that refer to unobservable parts of the domain of any interpretation of the language of the theory—although not to assertions whose terms refer *just* to unobservable parts of the domain.

Does this modification make a difference to the conception of the truth of a theory's theoretical claims? It was part of Newman's original observation against Russell that assertions concerning the structure of given relations often express significant truths. Newman argued that this contrasts sharply with what happens when such assertions are replaced by purely existential claims. And indeed, it remains the case that the modified Ramsey sentence reconstruction represents the theoretical claims of the original theory as true, provided only that such claims are satisfiable in a sufficiently large model of the theory's mixed and O-consequences; hence for a reconstruction based on the modified Ramsey sentence, truth and satisfiability effectively coincide for theoretical claims. But this is essentially Newman's observation, and so far as I can see, it is as compelling when urged against the modified Ramsey sentence reconstruction as it was against Russell.

The debate surrounding the bearing of Newman's paper on the modified Ramsey sentence proposal of British structuralism has recently been addressed by Jeffrey Ketland (2004). At the risk of appearing ungenerous to an author who has taken such care to improve on my paper with Friedman, I have three difficulties with Ketland's analysis. Let me begin by quoting from Ketland's concluding section;

> [It has been claimed] that 'if our theory is consistent, and if all its purely observational consequences are true, then the truth of the Ramsey sentence follows as a theorem of set theory or second order logic' [W. Demopoulos and M. Friedman, Philos. Sci. 52 (1985), no. 4, 621–639 (p. 635)]. This corresponds in our terminology to the claim that if $\boldsymbol{\theta}$ is satisfiable and weakly empirically adequate, then $\mathbf{R}(\boldsymbol{\theta})$ is true. And this is not quite the case, because weak empirical adequacy is properly weaker than empirical adequacy. However, the [Demopoulos–Friedman] claim is almost correct. For we have shown that the truth of the Ramsey sentence is equivalent to a sort of combination of empirical adequacy and a Newman-esque domain cardinality constraint. Indeed, the 'structural content' of a theory $\boldsymbol{\theta}$, at least if it is identified with what $\mathbf{R}(\boldsymbol{\theta})$ 'adds' to the claim that $\boldsymbol{\theta}$ is empirically adequate, is just this Newman-esque cardinality constraint. This leaves the structural realist in a sticky position, given that the epistemological intention was to

> provide an interesting third way between anti-realism and realism: the position collapses to something very close to antirealism. As Friedman and Demopoulos put it: ramsification 'trivializes physics: it threatens to turn the empirical claims of science into mere mathematical truths'(Ketland 2004, p. 299).

The qualification Ketland cites as necessary but missed in Demopoulos–Friedman is mentioned in numerous places, including the passage (from p. 635) quoted by Ketland. When Ketland quotes this passage earlier (on p. 294), he quotes it in full, and includes the qualification that the domain must have the right cardinality. When he quotes it in his conclusion, he leaves out the qualification. Admittedly we go on to say (in the full text from p. 635) that if the chosen domain does not have the right cardinality, consistency will take us to one that does; but in the context of all the other explicit acknowledgements of the necessity of a cardinality constraint, this should obviously be understood as meaning that from consistency and the existence of an *abstract* model of the right size we can find a *concrete* one of the right size. So it's at least misleading to suggest that the paper misses the importance of such a qualification and that only with Ketland's analysis has it been corrected.

Regarding the Carnapian framework that is implicit in these discussions, Ketland rejects mixed relations, as well as theoretical and observational relations, for their "oddity." The claim (on p. 289, footnote 3) that there are problems with the concept of a language whose intended semantics is "carved up" according to the epistemic status of what its terms refer to may well be correct, but its problems are not those that Ketland raises. In particular it is not enough to observe that it is counter-intuitive when laid against the fact that "many scientifically significant relations and quantities . . . such as mass, length, duration, [and] location . . . will 'decompose' into . . . strangely different relations depending upon the observational status of their relata (*ibid*, footnote 5)." This misses the point of Carnap's project, which is after all a reconstructive one, and so, not committed to the preservation of every common intuition regarding the use of 'observable' and 'unobservable.' Nor, for that matter, is it in any way precluded that it might be reasonable to advance explications of the terms in Ketland's list that differ from their ordinary scientific counterparts by not being continuous with observation terms. The goal here is to explain how theories look from the perspective of assumptions which have an independent philosophical interest; in the present case from the perspective of a strict empiricism. One may object to these assumptions, but it is no argument to dismiss the framework for its oddity, and to then ignore its fundamental point. Very few philosophical conceptions of any interest would pass the test of not presenting some oddity to ordinary intuition.

Ketland makes his task against British structuralism too easy by ramsifying on the mixed predicates. Ramsifying away the mixed terms is evidently contrary to the intentions of this school. It is indeed a failing of the approach that it has given no satisfactory explanation of why it is entitled to leave the mixed terms constant, but it is clear that it defines itself in opposition to Russell and Carnap by its refusal to identify the factual content of a theory with its observational content. Despite its painstakingly meticulous formality, Ketland's analysis is very highly restricted in its scope: it deals only with the claim that structural realism is justified in calling itself a species of realism, and then only on the highly questionable assumption that mixed terms can be ramsified away. By focusing on realism, Ketland's objection misses the kernel of Newman's original observation: structuralism is incapable of accurately representing the truth of theoretical claims because it takes their truth to collapse into satisfiability in a sufficiently large domain. This is hardly what we take the truth of theoretical claims to consist in, since we characteristically—and rightly—distinguish them from those of *pure* mathematics. A reconstruction which fails to acknowledge this, is not merely odd, it misses what is arguably one of the chief desiderata of an adequate philosophy of the exact sciences.

## 9 A personal note

I met William Craig when I was a graduate student at the University of Minnesota. I've not seen him since, and I have no reason to suppose that he remembers our meeting, which occurred in the spring or autumn of 1965 when the Minnesota Center for Philosophy of Science held a conference on the nature of theories. Craig was among the conference's invited participants, and the list of invited speakers included C.G. Hempel, Norwood Russell Hanson, and Paul Feyerabend. Herbert Feigl and Grover Maxwell—both of them truly noble souls—were, respectively, Director and Associate Director of the Center. I mention my meeting Craig because the circumstances carry a story about him that is important to my intellectual life and I hope of sufficiently general interest to be worth telling in the context of a collection of essays commemorating his life and work.

Not atypically, my fellow graduate students and I were principally observers of the conference. Feyerabend's attendance, in particular, was noted by us and his forceful presence had an effect. Craig must have observed the phenomenon of Feyerabend more than once before. They were colleagues at Berkeley, and Feyerabend's career was then very much in its ascendancy. I was experiencing the Feyerabend-phenomenon as a very young student, and I assumed that the experience of others was much as it was for me and my fellow students. Although it would never have occurred to me then, I can now see how Feyerabend might well have exasperated his peers.

It came as a very pleasant surprise to us graduate students when Homer Mason of the philosophy department at Minnesota told us that Craig would like to meet us for lunch. We were all familiar with Craig's observation—then very much discussed by philosophers of science—regarding the axiomatizability of recursively enumerable sets of sentences in a sub-language of the theory of which they are a part, and the more advanced students knew his celebrated Interpolation Lemma. So, along with six or so of my fellow students, I went off to have lunch with Craig and Mason.

We were all struck by Craig's modesty and accessibility. I'm also sure that each of us not only remembered, but was caught short by his challenge to our enthusiasm for Feyerabend, when he remarked, simply and without a hint of intimidation, "But *Create more theories* is not a thesis—not a claim that *says* anything." This remark—which I remember more than 40 years after the fact—re-awakened a skepticism in me that I saw I was in danger of losing, a skepticism that I believe has served me well ever since.

**Acknowledgements** Thanks to Anil Gupta for comments that led to several significant clarifications, and to John Corcoran for advice concerning the "Personal note" with which this paper concludes. Support from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged.

## References

Ben-Menahem, Y. (2006). *Conventionalism: From Poincaré to Quine*. Cambridge University Press.

Carnap, R. (1950). Empiricism, semantics and ontology. *Revue internationale de philosophie, 11*, 20–40.

Carnap, R. (1963). Replies. In P. A. Schilpp (Ed.), *The philosophy Rudolf Carnap* (pp. 859–1013). LaSalle: Open Court.

Carnap, R. (1966). In M. Gardner (Ed.), *An introduction to the philosophy of Science*. New York: Basic Books.

Craig, J. W., & Vaught, R. L. (1958). Finite axiomatizability using additional predicates. *Journal of Symbolic Logic, 23*, 289–308.

Cruse, P. (2005a). Ramsey sentences, structural realism and trivial realization. *Studies in History and Philosophy of Science, 36*, 557–576.

Cruse, P. (2005b). Empiricism and Ramsey's account of theories. In H. Lillehammer & D. H. Mellor (Eds.), *Ramsey's legacy* (pp. 105–122). Oxford University Press.

Demopoulos, W. (2003a). On the rational reconstruction of our theoretical knowledge. *British Journal for the Philosophy of Science, 54*, 371–403.

Demopoulos, W. (2003b). Russell's structuralism and the absolute description of the world. In N. Griffin (Ed.), *The Cambridge companion to Russell* (pp. 392–419). Cambridge University Press.

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

Demopoulos, W., & Friedman, M. (1985). Russell's analysis of matter: Its historical context and contemporary interest. *Philosophy of Science, 52*, 621–639.

DiSalle, R. (2006). *Understanding space–time: The philosophical development of physics from Newton to Einstein*. Cambridge University Press.

English, J. (1973). Underdetermination: Craig and Ramsey. *Journal of Philosophy, 70*, 453–461.

Gupta, A. (2006). *Empiricism and experience*. Oxford University Press.

Hempel, C. G. (1958/1965). The theoretician's dilemma: A study in the logic of theory construction. In H. Feigl & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science* (Vol. 2, pp. 37–98). Minneapolis: University of Minnesota Press; references are to its reprinting (with minor alterations) in his *Aspects of scientific explanation*. New York: The Free Press, 1965, pp. 173–226.

Hempel, C. G. (1963). Implications of Carnap's work for the philosophy of science. In P. A. Schilpp (Ed.), *The philosophy Rudolf Carnap* (pp. 685–710). LaSalle: Open Court.

Hempel, C. G. (1975). Rudolf Carnap, logical empiricist. In Hintikka (1975) pp. 1–13.

Hilbert, D. (1899/1971). *Grundlagen der Geometrie*; revised and enlarged by Paul Bernays and reprinted with an English translation by Leo Unger: *Foundations of geometry*. La Salle: Open Court.

Hintikka, J. (Ed.). (1975). *Rudolf Carnap, logical empiricist*. Dordrecht and Boston: Reidel.

Hodges, W. (2004). What languages have Tarski truth definitions? *Annals of Pure and Applied Logic, 126*, 93–113.

Kant, I. (1787). *Kritik der reinen Vernunft*, English translation by Norman Kemp-Smith: *Critique of pure reason*. New York: St. Martins Press, 1965.

Ketland, J. (2004). Empirical adequacy and ramsification. *British Journal for the Philosophy of Science, 55*, 287–300.

Maxwell, G. (1968). Scientific methodology and the causal theory of perception. In I. Lakatos & A. Musgrave (Eds.), *Problems in the philosophy of science* (pp. 148–160). Amsterdam: North Holland.

Maxwell, G. (1970). Structural realism and the meaning of theoretical terms. In M. Radner & S. Winokur (Eds.), *Minnesota studies in the philosophy of science* (Vol. 4, pp. 3–34). Minneapolis: University of Minnesota Press.

Newman, M. H. A. (1928). Mr Russell's causal theory of perception. *Mind, 37*, 137–148.

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

Putnam, H. (1962). What theories are not. In E. Nagel, P. Suppes, & A. Tarski (Eds.), *Logic, methodology and philosophy of science* (pp. 215–227).

Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review, 60*, 20–43, reprinted in his *From a logical point of view*. Harvard University Press, 1953, pp. 20–46.

Ramsey, F. P. (1929). Theories. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays* (pp. 212–236). Paterson, NJ: Littlefield and Adams, 1960.

Robinson, A. (1974). *Introduction to model theory and the metamathematics of algebra*, second revised printing of the 1963 edition. Amsterdam: Elsevier/North-Holland.

Russell, B. (1919). *Introduction to mathematical philosophy*. London: Allen and Unwin.

Russell, B. (1927). *The analysis of matter*. New York: Dover, reprint of 1954.

Russell, B. (1931). Review of F.P. Ramsey's '*The foundations of mathematics and other logical essays*', R. B. Braithwaite (ed.). *Mind, 46*, 476–482.

van Bentham, J. (1978). Ramsey eliminability. *Studia logica, 37*, 321–336.

Winnie, J. (1970). Theoretical analyticity. In R. Cohen & M. Wartofsky (Eds.), *Boston studies in the philosophy of science* (Vol. 8, pp. 289–305). Dordrecht and Boston: Reidel; reprinted in Hintikka (1975), pp. 143–159.