CARNAP, TARSKI, AND QUINE'S YEAR TOGETHER: LOGIC, MATHEMATICS, AND SCIENCE

by

Gregory G. Frost-Arnold

A.B., University of Chicago, 1999

M.A. in philosophy, University of Pittsburgh, 2005

Submitted to the Graduate Faculty of

History & Philosophy of Science in partial fulfillment

Of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH

DEPARTMENT OF HISTORY & PHILOSOPHY OF SCIENCE

This dissertation was presented

by

Gregory G. Frost-Arnold

It was defended on

April 17, 2006

and approved by

Steven Awodey, Associate Professor, Philosophy, Carnegie Mellon University

Nuel Belnap, Professor, Philosophy, University of Pittsburgh

John Earman, Professor, History & Philosophy of Science, University of Pittsburgh

Dissertation Advisor: Laura Ruetsche, Associate Professor, Philosophy, University of Pittsburgh

CARNAP, TARSKI, AND QUINE'S YEAR TOGETHER:
LOGIC, SCIENCE AND MATHEMATICS

Greg Frost-Arnold, PhD

University of Pittsburgh, 2006

ABSTRACT

During the academic year 1940-1941, several giants of analytic philosophy congregated at Harvard: Russell, Tarski, Carnap, Quine, Hempel, and Goodman were all in residence. This group held both regular public meetings as well as private conversations. Carnap took detailed diction notes that give us an extensive record of the discussions at Harvard that year. Surprisingly, the most prominent question in these discussions is: if the number of physical items in the universe is finite (or possibly finite), what form should the logic and mathematics in science take? This question is closely connected to an abiding philosophical problem, one that is of central philosophical importance to the logical empiricists: what is the relationship between the logico-mathematical realm and the natural, material realm? This problem continues to be central to analytic philosophy of logic, mathematics, and science. My dissertation focuses on three issues connected with this problem that dominate the Harvard discussions: nominalism, the unity of science, and analyticity. I both reconstruct the lines of argument represented in Harvard discussions and relate them to contemporary treatments of these issues.

TABLE OF CONTENTS

CHAPTER I. HISTORICAL BACKGROUND & OVERVIEW OF HARVARD ACTIVITIES

1. INTRODUCTION: SETTING THE HISTORICAL STAGE

During the academic year 1940-1941, several giants of analytic philosophy—both established and budding—congregated at Harvard University. The list of philosophers is impressive. Bertrand Russell, who was only at Harvard during the Fall semester of 1940, originally emigrated from Britain in 1938. During 1940, he was embroiled in his infamous legal battles with the City College of New York. In the fall, he gave the William James Lectures at Harvard, a series of talks that presently became *An Inquiry into Meaning and Truth*.[1] Alfred Tarski arrived in the U.S. from Poland in August 1939 for the Fifth International Congress for the Unity of Science, held at Harvard. On September 1, the Nazis invaded Poland. Tarski received permission to stay in the U.S., though his family was stranded in Poland; he held a number of temporary positions (at Harvard, City College of New York, and the Institute for Advanced Study) over the following years, before becoming a professor at the University of California at Berkeley.[2]

Rudolf Carnap and Carl Gustav Hempel immigrated to the U.S. a few years earlier. In December 1935, Carnap moved from Prague to the University of Chicago, where he was eventually offered a permanent position (Carnap 1963, 34). Carnap was a visiting professor at

[1] Russell had also presented much of this material to a seminar he held at the University of Chicago during Winter Quarter 1939, which Carnap attended. The ASP has Carnap's notebook from this seminar.
[2] Anita Burdman Feferman has done significant work on Tarski's first years in the U.S.; see her "How the Unity of Science Saved Alfred Tarski" (1999), as well as the biography of Tarski co-authored with Solomon Feferman (2004).

Harvard during the academic year 1940-41. Hempel crossed the Atlantic after being invited by Carnap in 1937 to serve as his research associate (*ibid*, 35); he was with Carnap in Harvard as well. He had only published a few articles by 1940, most of which dealt with philosophical issues in probability theory. W. V. O. Quine had already taken up residence at Harvard (he was appointed to Instructor in 1936 and promoted to Assistant Professor in 1941). In 1940, the first edition of his *Mathematical Logic* was published. Nelson Goodman was awarded his Ph.D. at Harvard in 1941 after completing his dissertation, *A Study of Qualities* (submitted in November 1940), which later became *The Structure of Appearance*.[3] This group of philosophers held meetings under the heading of (what Carnap terms) 'Logic Group' regularly, and they had smaller, informal conversations as well.

Any student of the philosophy of logic, mathematics, and/or the natural sciences would like to know what these immensely influential and innovative thinkers discussed during their hours together. Such information would be valuable both for the light it could shed on the historical development of analytic philosophy, as well as for purely philosophical reasons: were interesting or compelling arguments made here that do not appear elsewhere? Fortunately, one *can* almost be a 'fly on the wall' for many of these conversations, both public and private: Carnap had the lifelong habit of taking very detailed discussion notes, and he often took such notes during his year at Harvard. These documents have been preserved and stored in the Rudolf Carnap Collection, part of the Archives for Scientific Philosophy (ASP) at the University of Pittsburgh. This dissertation takes these documents as its primary subject matter. Most of the

[3] (Goodman 1966, *ix*). Goodman also tells us that both Quine and Carnap "read *A Study of Qualities* with great care and made inumerable invaluable suggestions" (*ibid*., *x*). Also, he mentions that his dissertation was not nominalistic, as *Structure of Appearance* is (*xviii*); perhaps (part of) the spur to Goodman's conversion came from the conversations with Carnap, Tarski, and Quine in 1940-41. (However, his dissertation does discuss nominalism; see for example RCC 102-44-10, -11, which are Carnap's discussion notes for conversations with Goodman about his dissertation.

documents are records of discussions, but some are Carnap's own reflections on the topics, composed in private.

Several of the above-named philosophers also took part in a larger collaborative community, which was also founded in the Fall of 1940 at Harvard, called the 'Science of Science' dinner and discussion group. This group incorporated many prominent scientists, including many European emigres, as well as other philosophers. The Harvard psychologist S. S. Stevens, one of the champions of operationism in psychology, spearheaded the effort, apparently prompted by Carnap (Stevens 1974, 408). The mathematicians George David Birkhoff (and his son Garrett), Richard von Mises and Saunders MacLane, the sociologist Talcott Parsons, the economists Otto Morgenstern and J. A. Schumpeter, as well as Percy Bridgman, Herb Feigl, Philipp Frank, and C. I. Lewis were all invited to the first meeting; there were a total of forty-five invitations sent. Further details about the Science of Science group, including the text of that invitation and a list of invitees, can be found in (Hardcastle 2003). My focus here will be almost exclusively on the 'Logic group' and its participants, not the larger Science of Science group.[4]

2. OUTLINE OF THE PROJECT: A 'FINITIST-NOMINALIST' TOTAL LANGUAGE OF SCIENCE

As one might expect, Carnap's discussion notes from 1940-41 cover a wide range of topics. We have records concerning:

- the relations between metaphysics, magic, and theology (102-63-09),

[4] I have only found one document in the Carnap archives from this time period that mentions the Science of Science group: Carnap alludes briefly to von Mises' presentation, in the Science of Science group, of the Kolmogorov-Doob interpretation of probability (102-63-13). This allusion, however, shows that Hardcastle may be too hasty in concluding that what happened at these meetings "must be left to the historically informed imagination" (2003, 175).

- the concept of *proposition* (102-63-10, -11),

- the interpretation of the probability calculus (102-63-13),

- transfinite rules of inference (102-63-12),

- non-standard models of Peano arithmetic (102-63-08),

- comparisons of formal languages without types (e.g. set theory) over languages with types (exemplified by *Principia Mathematica*) (090-16-09, -02, -26),

- modalities (090-16-09, -25),

- Quine's recently published *Mathematical Logic* (090-16-02, -03, -26),

- the treatment of quotation-marks in formalized languages (090-16-13, -14),

- the possibility of a 'probabilistic' consequence relation (090-16-30),

- the relationship between the notions *state of affairs* and *model* in semantics (090-16-10, -11),

and other topics. Some of these are mentioned only briefly; others receive a more extended treatment.

However, a plurality of Carnap's discussion notes during the spring semester deal with what he and his collaborators call—most briefly—'finitism.' In these notes, Carnap calls this enterprise by a number of other names as well. The following are Carnap's section headings for entries related to this topic:

- "On Finitistic Syntax" (090-16-27)

- "Logical Finitism" (-24)

- "On the Formulation of Syntax in Finitistic Language" (-23)

- "Finitistic Language" (-06, -08)

- "The Language of Science, on a Finitistic Basis" (-12)

- "Finitistic Arithmetic" (-16)

- "Conversation about the Nucleus-Language" (-05)

However, this topic is *not* identical with the cluster of claims philosophers usually associate with the label 'finitism'—namely, the mathematical project associated with Hilbert and his school. Carnap, Tarski, and Quine were fully aware that Hilbertian finitism was long dead as a research program by 1940 (but see (Detlefsen 1986)), when these conversations began. Unlike Hilbert, they are not dealing with the foundations of mathematical inference (specifically, investigating which inferences employed in classical mathematics can be re-cast into a finitistically acceptable form). The notion of finitism first appears in these conversations when Tarski proposes rather strict, severe requirements a language must meet in order for it to qualify (in Tarski's eyes) as *verständlich*, that is, understandable or intelligible.

Tarski's proposal varies somewhat from meeting to meeting. Carnap records the first version of it as follows.

January 10, 1941.
Tarski, Finitism. Remark in the logic group.
Tarski: I understand at bottom only a language that fulfills the following conditions:
1. Finite number of individuals.
2. Reistic[5] (Kotarbinski): the individuals are physical things.[6]
3. Non-Platonic: Only variables for individuals (things) occur, not for universals (classes etc.)
(090-16-28)[7]

Three weeks later, Tarski offers a similar, though not identical, characterization of a language he finds completely understandable.

Finitism.
Tarski: I really understand only a finite language S_1:
only individual variables [identical to condition 3. above],
whose values are things [2. above];

[5] The transcription actually reads 'realistic,' but since Kotarbinski's fundamental philosophical view is called 'reism,' the emendation seems warranted.
[6] Tarski never shows much interest in specifying exactly what the 'physical things' are; Carnap (-22) usually takes them to be elementary physical particles (electrons etc.); at one point, Quine suggests 'quanta of energy' be used (-23).
[7] Tarski: Ich verstehe im Grunde nur eine Sprache die folgende Bedingungen erfüllt:
1. <u>Finite</u> Anzahl der Individuen.
2. <u>Realistisch</u> (Kotarbinski): Die Individuen sind physikalische Dinge;
3. <u>Nicht-platonisch</u>: Es kommen nur Variable für Individuen (Dinge) vor, nicht für Universalien (Klassen usw.)

whose number is not claimed to be infinite (but perhaps also not the opposite) [modified version of 1.].
Finitely many descriptive predicates.
(-25, p.2)[8]

Let us describe Tarski's proposed conditions for an intelligible language using the modern apparatus of model theory. We begin with the notion of an interpreted language *L* = <L, M, ρ>. L carries the syntactic information about the language: what the symbols of the language are, the grammatical category to which each symbol belongs, which strings of symbols qualify as grammatical formulae and which not, etc. The semantic scheme ρ determines the truth-values of a compound expression formed using logical connectives, given the truth-values of its constituents. M is an interpretation or model that fixes signification of the nonlogical constants of L. Specifically, M = <D, f>, where D is a nonempty set, and f is an interpretation function which assigns members of D to singular terms, assigns sets of ordered n-tuples $\subseteq D^n$ to n-ary relation symbols, and a member of $D^n \rightarrow D$ to each n-ary function symbol. Now let us use this apparatus to rephrase Tarski's idea more precisely and concisely. Tarski is describing a certain type of (interpreted) language *L* that has the following four characteristics, which I will henceforth refer to as Tarski's 'finitist-nominalist' conditions:

(FN 1) *L* is *first-order*.

In a fully understandable language, one cannot quantify over properties or classes, only over individuals. One might be tempted to interpret Tarski as claiming that any string that contains (the symbolic correlate of) 'For all properties Φ, …' is not a grammatical formula of L; 'being first-order' is a grammatical property. However, Tarski's proposal appears *not* to be a merely grammatical restriction. For immediately after the quotation above, Tarski explains that he is

[8] Tarski: Ich verstehe richtig nur ein endliche Sprache S_1: nur Individuumsvariable, ihre Werte sind Dinge; für deren Anzahl wird nicht Unendlichkeit behauptet (aber vielleicht auch nicht das Gegenteil). Endlich viele deskr[iptive] primitiven Prädikate.

perfectly willing to derive consequences using sentences containing higher-order variables according to the rules of a proof calculus—but he claims nonetheless that he does not truly *understand* these sentences. Specifically, Tarski says:

> I only 'understand' any other language in the way I 'understand' classical mathematics, namely, as a calculus; I know what I can derive from what… With any higher, 'Platonic' statements in a discussion, I interpret them to myself as statements that a fixed sentence is derivable (or derived) from other sentences. (RCC 090-16-25)

This notion of 'understanding,' which a 'calculus' alone cannot guarantee, will be discussed at length in II.2 below. Thus, (FN1) should be taken as (in some sense) a semantic restriction, not a grammatical one.[9]

(FN 2) D consists of 'physical objects' only.

In the elaboration and discussion of (FN 2), numbers are specifically disallowed from D. Furthermore, because of (FN 1), not even the Frege-Russell reconstruction of numbers as classes of classes (or concepts of concepts) is allowed in a finitist-nominalist language.

What, exactly, are the 'physical objects'? The discussants do not show great interest in settling upon a specific interpretation.[10] (Nelson Goodman, in *The Structure of Appearance* and "A World of Individuals," stresses that nominalism as such places no restrictions on what the individuals countenanced by the nominalist are (1966, 39; 1956, 17).) Three options considered are: (i.) elementary physical particles (electrons etc.), (ii.) mereological wholes composed of elementary particles (or quanta of energy) (Quine), so that the objects referred to by the names 'London' and 'Rudolf Carnap' will qualify as physical objects, and (iii.) spatial and/or temporal intervals (Carnap; this suggestion derives from *Logical Syntax*'s co-ordinate or position

[9] I do not know how best—or even whether—to couch this restriction in model-theoretic terms. One proposal would be to omit variable-assignments for higher-order variables and functions in the definition of satisfaction. Obviously, if we do this, then a proof calculus that allows higher-order variables will generate proofs of sentences that have no models—which would usually be characterized as a failure of soundness. I say 'usually' because it does not seem likely that Tarski considers understandability and model-theoretic satisfiability as coextensive (rather, the former is a proper subset of the latter).
[10] For an account and analysis of Quine's later published remarks on physical objects, see (Dalla Chiara and Di Francia 1995).

languages) (090-16-23). Quine, in 1947's "On Universals" (which could be viewed as a companion piece to his joint article with Goodman on nominalism) constructs a "logic of limited quantification over classes of concrete individuals," whose variables "admit only concrete objects as values." Quine then asks:

But concrete objects in what sense? Material objects, let us say, past, present, and future. Point-events, and spatio-temporally scattered totalities of point events. Now suppose physics shows these to be finite in number. (1947, 82)

(FN 3a: *restrictive version*) D contains a finite number of members;

or

(FN 3b: *liberal version*) No assumption is made about the cardinality of D.[11]

The restrictive policy is Tarski's initial proposal, but in later conversations, he clearly leans towards the liberal policy (see 090-16-04 and -05). Carnap, in his "Autobiography," attributes the restrictive version to Quine and the liberal version to Tarski and himself (1963, 79).

The last restriction Tarski proposes for a finitist-nominalist language can be couched as follows:

(FN 4) *L* contains only finitely many descriptive predicates.

Tarski offers no justification for (FN 4), and the participants never discuss it, so I will not dwell upon it further. Presumably, these four finitist-nominalist restrictions do not uniquely single out one language— several different interpreted languages could satisfy (FN 1)-(4). I will call the

---

[11] In 090-16-04, however, Tarski excludes interpreted languages whose domain is empty or has uncountably many members:

Tarski: I would like to have a system of arithmetic that makes no assumptions about the quantity of numbers at hand, or assumes at most one number (0). Let A_n be the system of those sentences of customary arithmetic which are valid only if there are numbers <n; so A_0 has no numbers, A_1 has only 0, and so forth. Let A_ω be the entirety of customary infinite arithmetic. For the purpose of simplification, we want to exclude A_0, so we assume the existence of at least one number.

My (i.e., Tarski's) system should contain all and only the sentences that are valid in each of the systems A_n (n=1, 2, … ω). Here belong all sentences of the following form, e.g.: no functors occur, all universal operations are not negated at the beginning, no existential operators.

above four conditions the ‘finitist-nominalist (FN) conditions’ (the first two are nominalist, the third and fourth finitist), and any language satisfying them a ‘finitist-nominalist’ language.

In other formulations of the group’s project, an additional condition is placed on the language(s) they are attempting to construct. They wish to formulate a finitist-nominalist language that is rich enough to conduct investigations into the logic of science within it, including metalinguistic investigations (note “syntactic” and “semantic” below) of classical analysis and set theory (they sometimes call such a language a “nucleus language”).

Jan. 31, 1941
Conversation with Tarski and Quine on Finitism
…
We together: So now a problem: What part S of M [the metalanguage of science and mathematics] can we take as a kind of nucleus, so that 1.) S is understood in a definite sense by us, and 2.) S suffices for the formulation of the syntax of all of M, so far as is necessary for science, in order to handle the syntax and semantics of the complete language of science. (-25)

Similar sentiments are expressed a few months later.

June 18, 1941
Final Conversation about the nucleus-language, with Tarski, Quine, Goodman, and Hempel;
June 6 ’41
Summary of what was said previously. The nucleus language should serve as the syntax-language for the construction [*Aufbau*] of the complete language of science (including classical mathematics, physics, etc.). The language of science thereby receives a piecewise interpretation, since the n.l. is assumed to be understandable… . (-05)

On the one hand, the finitist-nominalist conditions place restrictions on an interpreted language’s richness; this condition, on the other, restricts a language’s poverty. Carnap, Tarski and Quine realize it may not be possible to construct a language that simultaneously satisfies this criterion as well as (FN 1)-(4). For immediately following the first of the two quotations immediately above, we find:

1. It must be investigated, if and how far the poor nucleus (i.e. the finite language S_1) is sufficient here. If it is, then that would certainly be the happiest solution. If it is not, then two paths must be investigated:
2a. How can we justify the rich nucleus (i.e., infinite arithmetic S_2)? I.e., in what sense can we perhaps say that we really understand it? If we do, then we can certainly set up the rules of the calculus M with it.

2b. If S_1 does not suffice to reach classical mathematics, couldn't one perhaps nevertheless adopt S_1 and perhaps show that classical mathematics is not really necessary for the application of science in life? Perhaps we can set up, on the basis of S_1, a calculus for a fragment of mathematics that suffices for all practical purposes (i.e., not something just for everyday purposes, but also for the most complicated tasks of technology). (-25)

In short, they suspect that a metalinguistic analysis of classical mathematics and physics may require a richer language than those allowed by the finitist-nominalist criteria, and if that suspicion is borne out, then either such a richer language must be shown to be understandable, or the weaker mathematics sanctioned in finitist-nominalist languages must be shown to be sufficient to deal with all sophisticated practical applications.[12] We are not told (unfortunately) whether this new condition 'trumps' the finitist-nominalist conditions or not. That is, if infinite mathematics is ultimately not understandable, and finitist-nominalist-condoned mathematics is insufficient for practical purposes, then what should be discarded—the demand for a single metalanguage of science, or the finitist-nominalist strictures on intelligibility? Thus it is difficult for us to ascertain the relative importance Carnap, Tarski, and Quine attached to these competing conditions. However, none of the participants assert that we should disavow or eliminate those portions of set theory (e.g.) that fail to meet the finitist-nominalist criteria (FN 1)-(4). Set theory can progress, even if parts of it are not fully intelligible—Tarski suggests that set theory then becomes a formal (i.e. uninterpreted) calculus, which merely indicates which sentences can be derived from others (090-16-28). But that is not a barrier on proving theorems.

A published summary of the finitist-nominalist project Carnap, Tarski, and Quine undertake in 1941 can be found at the end of Carnap's *Intellectual Autobiography*, in the section

[12] This last alternative was not usual at the time. Frege stressed the importance of understanding the meaning of number words in 'everyday' contexts. And Wittgenstein followed this lead: "In life a mathematical proposition is never what we want. Rather, we use mathematical propositions *only* in order to infer sentences which do not belong to mathematics from others, which likewise do not belong to mathematics"(*Tractatus* 6.211).

entitled “The Theoretical Language.” Carnap’s conception of this project has not changed substantially during the intervening years.

> We [Carnap, Tarski, Quine, and Goodman] considered especially the question of which form the basic language, i.e., the observation language, must have in order to fulfill the requirement of complete understandability. We agreed that the language must be nominalistic, i.e., its terms must not refer to abstract entities but only to observable objects or events. Nevertheless, we wanted this language to contain at least an elementary form of arithmetic. … We further agreed that for the basic language the requirements of finitism and constructivism should be fulfilled in some sense. We examined various forms of finitism. Quine preferred a very strict form; the number of objects was assumed to be finite and consequently the numbers appearing in arithmetic could not exceed a certain maximum number. Tarski and I preferred a weaker form of finitism, which left it open whether the number of all objects is finite or infinite. … In order to fulfill the requirement of constructivism I proposed to use certain features of my Language I in my *Logical Syntax*. We planned to have the basic language serve, in addition, as an elementary syntax language for the formulation of the basic syntactical rules of the total language. The latter language was intended to be comprehensive enough to contain the whole of classical mathematics and physics, represented as syntactical systems. (1963, 79)

Several of the features mentioned earlier re-appear here: the aim of understandability, disallowing abstract entities, a finite universe of discourse (in both the liberal and restrictive variants), re-interpretation of arithmetic, and that the basic language should serve as the ‘syntax language’ (which is part of the ‘metalanguage’) for the total language of science.

However, there are at least two notable discrepancies between this later description and the discussion notes of 1941. First, the term ‘constructivism’ is not explicitly used in the original formulation of the project.[13] Carnap is correct, though, to recall that the basic strategy was to begin with Language I of *Logical Syntax* and consider amendments to it. Second, Carnap later recalls the basic language as being an ‘observation language,’ i.e., a language whose terms designated observable entities, properties, and relations. There is both a conceptual and a historical mistake here. The conceptual mistake is Carnap’s conflation of ‘nominalistic’ with ‘observable’: the concrete/ abstract distinction is not coextensive with the observable/ unobservable (or theoretical) distinction. The nominalist (usually) denies the existence or

[13] However, finitism is standardly taken to be a species of constructivism. (Thanks to Paolo Mancosu for reminding me of this.)

epistemic accessibility of abstracta,[14] but she is free to believe in (concrete) unobservables. For example, atoms are usually considered concrete but unobservable.[15] This conceptual mistake is closely related to Carnap's mis-remembering of the historical episode. The requirement that the nucleus language be an *observation* language is *not* mentioned in 1940-41; there, the domain of discourse is often (though not exclusively) taken to include the elementary particles, entities whose names are *not* part of an observation language.[16]

Let us summarize and take stock. Carnap, Tarski, and Quine (and occasionally Goodman and Hempel) are attempting to construct a formal language that simultaneously meets the stringent finitist-nominalist constraints (FN1)-(4) and is rich enough to serve as a metalanguage for all of science, including (at least the bulk of) mathematics. Since these two conditions pull in opposite directions, this is a difficult goal to reach. I will postpone discussion of detailed objections to the project until Chapter III, but I wish to note here that Carnap, virtually from beginning to end, is strongly suspicious of the fundamental assumptions of this project, criticizing the finitist-nominalist restrictions on various grounds. Although he is willing and able to play by the rules Tarski has laid down in (FN 1)-(4), Carnap questions these rules repeatedly during the course of 1941.[17] In very general terms, Carnap's objections attempt to show that

---

[14] Goodman, in his (mature) defense of nominalism, takes a slightly different line. He writes: "the line between what is ordinarily called 'abstract' and what is ordinarily called 'concrete' seems to me vague and capricious. Nominalism for me consists specifically in the refusal to recognize classes" (1956, 16). Goodman offers this as a modification/ clarification of the doctrine of the original 1947 nominalism paper co-authored with Quine, in which *abstracta* were rejected.

[15] I am perhaps being too hard on Carnap here. He may just be using the term 'abstract' for what we today would call 'unobservable' (which could include most abstracta); see his (1939, 203-205).

[16] However, as Paolo Mancosu pointed out to me, Quine and Goodman favored sense data predicates as the basic terms for the descriptive part of the language (090-16-05. (Carnap, Tarski, and Hempel demurred.) If one follows Quine and Goodman on this, then the finitist-nominalist language will contain no basic terms for unobservable items.

[17] Carnap's position, it seems to me, is very much like John Burgess' today. Burgess has developed formal systems satisfying various versions of nominalist criteria, but also writes papers with titles such as "Why I Am Not a Nominalist." Carnap's view of nominalism is perhaps not quite so dim, but like Burgess, he throws himself into working within the rules set by the nominalist, without fully accepting the legitimacy of those rules.

either the finitist-nominalist restrictions yield undesirable consequences in the domain of the formal sciences, or that higher mathematics *is* genuinely meaningful.

3. MATHEMATICS IN A FINITIST-NOMINALIST LANGUAGE

Carnap, Tarski, and Quine appear to realize from the outset that one of the most pressing and difficult obstacles facing any attempt to construct a finitist-nominalist *Gesamtwissenschaftssprache* will be the treatment of mathematics. Can a language simultaneously meet Tarski's criteria for intelligibility and contain (at least a substantial portion of) the claims of classical mathematics? A substantial portion of Carnap's notes on 'finitism' deals with how to answer this question. The discussants focus on the simplest case, *viz.* classical arithmetic. A number of potential pitfalls present themselves: first, what is the content of assertions about numbers? Can we assert anything about them at all, given that the only objects we can quantify over are physical ones? Second, what should be done with numerals that aim to refer to numbers that are larger than the number of concrete things in the universe? That is, suppose there are exactly one trillion physical things in the universe; what should we then make of the numeral '1,000,000,000,001' and sentences containing it? Finally, what theorems and proofs of classical arithmetic are lost (or possibly lost)? I shall deal with each of these questions in turn.

A. Number

As seen in the previous section, one of the requirements for an understandable language is that abstract entities are not allowed to serve as denotata of names. So in such a language, the numeral '7' cannot name a natural number, considered as a basic individual object[18]—for the natural numbers are excluded from the domain of discourse. And as mentioned above, since a FN language must also be first-order, the Frege-Russell construal of numerals as denoting classes of classes is forbidden as well. But Tarski, Carnap, and Quine want the language to include, at the very least, portions of arithmetic. So how do they re-interpret numerals under this linguistic regime?

Tarski's strategy for introducing ordinal numbers[19] is the following: "Numbers can be used in a finite realm, in that we think of the ordered things, and by the numerals we understand the corresponding things" (090-16-25, p.2). Virtually the same[20] proposal is outlined in Carnap's Autobiography:

> To reconcile arithmetic with the nominalistic requirement, we considered among others the method of representing natural numbers by the observable objects themselves, which were supposed to be ordered in a sequence; thus no abstract entities would be involved.
> (1963, 79)

Let us illustrate this idea with a concrete example. Suppose, in our domain of 'physical things' that have been 'ordered in a sequence,' Tom is the eighth thing, John is the fourth, and Harry the eleventh. (Assume the numeral '0' is assigned to the first thing.) Then the arithmetical assertion '7+3=10' is re-interpreted as 'Tom + John = Harry.' Put model-theoretically, the interpretation function f of an interpreted language meeting the finitist-nominalist requirements assigns to (at least some of) the numerals of L objects in D: f(7)=Tom, f(3)=John. (Arithmetical signs such as

[18] I am here assuming, *contra* the current school of structuralism in philosophy of mathematics, that the natural numbers are treated as individuals, not 'nodes in a structure' or however else the structuralist wishes to characterize numbers.

[19] The group discusses cardinal number *very* briefly in (090-16-25, p.2):
"One can also ascribe a cardinal number to a class.
Quine: E.g. by the introduction of '(∃3x)…' as an abbreviation for '(∃x)(∃y)(∃z)(~..=..&).'"
I presume that what Quine intends is the following:
$\exists 3x\phi x \equiv \exists xyz(x\neq y.y\neq z.z\neq x.\phi x.\phi y.\phi z)$.

[20] The only difference is that Carnap claims that the things are "observable." As I have mentioned above, this is almost certainly either a 'mis-remembering' by Carnap, and not part of the original proposal, or a discrepancy between Carnap's terminology and ours (and that of Tarski and Quine in 1941).

‘+’ are defined via the version of Peano Arithmetic in Carnap’s *Logical Syntax*; see §14 and §20.)

This heterodox view of ordinal numbers raises a number of serious questions. First, from whence does the sequential order of the physical objects spring? That is, what determines that Tom is ‘greater than’ John, and that Harry is ‘greater than’ them both? Must this ordering reflect the actual spatiotemporal positions of Tom, John, and Harry?—And if so, where do we ‘start counting,’ so to speak? Fortunately, I think such questions can be avoided for the most part. The ordering is imposed, it appears, by stipulation; Tarski says: “we want the (perhaps finitely many) things of the world ordered in some arbitrary way” (090-16-23). We may assign *any* member of the domain of physical things to the numeral ‘0,’ and we may choose any other member of the domain to be its successor, and be assigned to the numeral ‘1.’ The sentence ‘0+1=0’ will come out *false* under any such stipulation, regardless of which physical objects we choose to ‘stand in’ for 0 and 1 (assuming more than one thing exists). The relation *is a successor of* need not reflect anything ‘in the order of things,’ spatial, temporal, or otherwise. I should stress that the above claim, *viz.* the order of the objects is purely conventional, does not explicitly appear in Carnap’s dictation notes, beyond what Tarski says in the text been quoted. However, there is *no* discussion in the notes of how the order is fixed, and the proposal just suggested would allow Tarski, Quine and Carnap to avoid entangling themselves in thorny questions, so it is quite possible that they imagined the order fixed by arbitrary stipulation.

There is a second, perhaps more obvious worry about this proposal to interpret number-language under a finitist-nominalist regime. Let us suppose that the sentence ‘Tom has brown hair’ is true. Then, since the name ‘Tom’ and the numeral ‘7’ both name the same object (model-theoretically, the interpretation function assigns both individual constants the same value), it appears that the sentence ‘7 has brown hair’ will be true. And numbers cannot be brunettes. So this finitist-nominalist interpretation of numerals will make true many assertions about numbers that, intuitively, we do not want to come out true. Tarski, Quine and Carnap do not even consider this problem (at least, there is no record of it in Carnap's discussion notes). Perhaps technical refinements could avoid at least some of these unwanted truths.[21] Note also that an analogue of this problem appears in set-theoretic interpretations of arithmetic, such as Zermelo's and von Neumann's. For example, using von Neumann's set-theoretic construction of the natural numbers, '2 is contained in 3' is true—and that does not match up with ordinary usage of arithmetical language. This example shows that the type of problem Tarski faces is not peculiar to his proposal, but rather is likely to occur in any situation in which some portion of language is given a construal using structures from some other part of science.

[21] One such refinement is suggested in (Field, *Truth and the Absence of Fact*, 214-215). His basic idea, couched in our terms, is the following. Recall that the ordering of physical objects of D should (most likely) be viewed as arbitrary, and that alternative orderings of the elements of D are possible that would still respect the truths of classical arithmetic captured in the original model (e.g. '7+3=10'). Perhaps this fact could be finessed to eliminate unwanted truths: while '7' may be assigned to a brunette in one assignment of physical objects to numerals, it will be assigned to a blonde in another, and to various hairless physical objects in other assignments. However, under *all* these assignments, '7+3=10' is true. This suggests the following refinement to Carnap, Tarski, and Quine's proposal to re-interpret numerals in a finitist-nominalist language: a (mathematical) sentence ϕ is true (in L) if and only if ϕ is true for all assignments of physical-object-values to numerals (satisfying certain conditions: for example, we do not want to include the assignment in which all numerals are assigned to a single object in D). This characterization is only a first approximation, and I will not dwell on this possibility further; nonetheless, this line of thought shows that perhaps there is a way to interpret '7' that meets (FN1)-(4) and certifies substantial portions of arithmetic as true, without committing us to the truth of sentences such as '7 has brown hair.' Of course, if the number of objects in the physical universe is finite, then this proposal will not capture all of standard arithmetic. Also note that the proposed refinement will still class '7 is a physical object' as a truth, since that sentence is true on all assignments of physical objects to the numeral '7.' (From the point of view of someone who endorses (FN1-3), this will perhaps not be considered unfortunate.) One could class this suggestion as a 'nominalist-structuralist' account of mathematics, for it meshes nicely with Benacerraf's founding statement of structuralism:

Arithmetic is therefore the science that elaborates the abstract structure that all progressions have merely in virtue of their being progressions. It is not a science concerned with particular objects—the numbers. The search for which independently identifiable particular objects the numbers really are… is a misguided one.
(1965/1983, 291)

Because nominalists refuse to treat numbers as objects, they thereby make common cause with the structuralists.

B. Interpreting numerals that are 'too large'

Now we come to a problem concerning mathematics in finitist-nominalist languages that Tarski, Quine and Carnap *did* recognize themselves. Suppose there are only k items in the universe. Carnap poses the question: "How should we interpret [*deuten*]" the number expressions k+1, k+2, …, "for which there is no further thing there?" (-06) Initially, the group considers three options (employing the usual notation that x′ is the successor of x):

(a.) $k' = k'' = \ldots = k$
(b.) $k' = k'' = \ldots = 0$
(c.) $k' = 0, k'' = 0', \ldots$

In each of these three cases, at least one of the Peano axioms is violated—and thus so is one of the axioms of Carnap's Language I (PSI) in *Logical Syntax*. If (a.) is adopted, then there exist two numbers (where 'number' here is interpreted as *physical object*) will have the same successor (contravening PSI 10); if (b.) or (c.) is adopted, then the number assigned to '0' will be a successor (contravening PSI 9) (Carnap 1934/1937, 31).[22]

None of these options is especially palatable, since none captures the truths of classical arithmetic substantially better than the others. For example, suppose there are 1000 physical things in the universe. Then, the sentence '600+600=700+700' will come out true under proposals (a.) and (b.), for it reduces to '999=999' and '0=0,' respectively. And under (c.), the just-mentioned problem will be avoided, but '0=1000=2000' will be true. So regimes (a.)-(c.) will all certify as true several equations that are false in classical arithmetic. A surprisingly large portion of the discussion notes is devoted to working through proposed solutions to this problem. Strategies other than (a.)-(c.) are also considered, such as identifying numbers with *sequences* of objects instead of objects, so that there is no 'last element' forced upon us.

[22] In 090-16-23, Tarski suggests that, for finitist-nominalist purposes, the Peano axioms for arithmetic should be constructed such that the axiom of infinity is treated as an *axiom*, unlike in PSI. For then, when the finitist-nominalist omits the axiom of infinity, as little arithmetical power as possible is lost.

And that is not the only problem: as Tarski notes, under these conceptions of number "many propositions of arithmetic cannot be proved in this language, since we do not know how many numbers there are" (090-16-25). That is, suppose that we do not know how many physical objects there are in the material universe; this ignorance will be formally reflected in a refusal to allow any assumptions about the cardinality of the domain of *L*. Then, there will be arithmetical assertions we can prove under classical arithmetic, but are unprovable in a finitist-nominalist language. If we allow ourselves *no* assumption about the cardinality of the domain (or just the assumption that at least one object exists, as Tarski suggests), we cannot even prove that '1+1=0' is false.[23] So not only are 'intuitively true' arithmetical sentences declared false in this language, but chunks of previously provable assertions are no longer susceptible of proof. This issue will be treated at greater length below in Chapter III, which deals with Carnap's objections to the finitist-nominalist project.

There were other suggestions for how to deal with numbers that are 'too large' in a finitist-nominalist regime; however, none meet with substantially more approval from the other discussants. Interestingly, they never consider treating k′, k″, etc. as *denotationless*, i.e., analogous to 'Santa Claus' (put model-theoretically: f(k′) is undefined, or assigned the null set, or some other chosen object). This approach (which we today could carry out using free logic) would avoid certifying '600+600=700+700' and '0=1000' as true (both would lack a truth-value in 'neutral' free logics, and would be false in 'positive' (assuming a supervaluational semantics) and 'negative' free logics). However, this strategy would not recapture the classical arithmetical truths about numbers greater than k.

[23] For example, if we assume (b.) or (c.) (the discussants' eventual favored choice), then if D contains exactly 2 elements, then '1+1=0' will be true.

In his private notes at this time, Carnap actually performs some basic calculations on the question of how many physical things there are (090-16-22).[24] Starting from a conjecture of Eddington's, Carnap works out that the number of particles in the universe is approximately 10^{77}. Then, using Quine's proposed ontology, in which classes of particles are the things (since bodies are classes of particles), the maximum number of "things" in the universe is approximately 2^(10^{77}).[25] That Carnap actually works out how to apply this finitist-nominalist language to a realistic case shows, I believe, that Carnap did take this project at least somewhat seriously, and that for him it was neither a game of word-play nor a merely technical exercise.

4. PRE-HISTORY OF THE 1941 FINITIST-NOMINALIST PROJECT

The next chapter will address the justifications or rationales for undertaking the finitist-nominalist project. Here, a different question is posed: from what historical sources do FN 1-3 spring? Tarski himself cites Chwistek (090-16-09) and Kotarbinski (090-16-28) for certain of the ideas he presents, so I first briefly outline the claims of these two Polish philosophers most relevant to the finitist-nominalist project. Next, I examine possible indirect lines of influence Russell's ideas may have had on the formation of the FN project. Finally, I cite contemporaneous skeptical complaints about infinity from Wittgenstein, Neurath, and Hahn.

---

[24] Here is the entirety of that document, which is undated.
Number and Thing.
Following Eddington, there are 2^{256} particles=$2^{(10x25.6)}$=$(10^3)^{25.6}$=$10^{3x25.6}$=10^{77}.
Quine takes the things to be classes of particles: 2^(2^256) = 2^(10^77) = 2^(10x10^76) = (2^10)^10^76 = (10^3)^(10^76) = 10^(3x(10^76)).

[25] Quine explicitly accepts the consequence of this view that unassociated and spatially disconnected items (e.g., my left index finger and the planet Neptune) can be grouped together to count as a single thing. Nelson Goodman later accepts this consequence in *Structure of Appearance* and subsequent writings; this is, I believe, partially motivated by his desire to use what he calls the 'calculus of individuals,' which later became what is now called 'mereology,' i.e., the formal/logical study of the part-whole relation.

A. The Poles (Chwistek, Kotarbinski, Lesniewski)

The finitist-nominalist project is Tarski's proposal; thus, it is natural to look to the philosophical ideas he was exposed to in Poland to find his inspiration for the radical ideas expressed in the FN conditions. Tarski mentions two Polish philosophers, and their characteristic views, by name in the notes: Leon Chwistek's nominalism and Tadeusz Kotarbinski's realism. Chwistek worked in Krakow, which was not part of the Lvov-Warsaw School, to which Tarski, Kotarbinski, and other prominent Polish philosophers belonged.

(i) *Chwistek's 'nominalism'*

In May 1940, months before the finitist-nominalist project is proposed and explored, Tarski visited the University of Chicago, where he and Carnap have an extended and wide-ranging discussion. Carnap's notes record that Tarski says to Carnap:

> With the higher levels, Platonism begins. The tendencies of Chwistek and others ("Nominalism") to talk only about describable things [=*Bezeichenbarem*] are healthy. The only problem is finding a good execution. Perhaps roughly of this kind: in the first language numbers as individuals, as in language I, but perhaps with unrestricted operators; in the second language individuals that are identical or corresponding to the sentential functions in the first language, so properties of natural numbers expressible in the first language; in the third language, as individuals those properties expressible in the second language, and so forth. Then one has in each language only individual variables, though dealing with entities of different levels. (090-16-09)

Note that Tarski's proposal here is fundamentally different from the FN project. First, this proposal *does* allow 'higher levels,' and thus would qualify as Platonism under the criterion mentioned in the first sentence, as well as under FN2, which Tarski labeled the 'non-platonic' requirement. Also, Tarski's suggestion here to use numbers as the individuals in the universe of discourse (*prima facie*) violates the restriction of the universe to physical objects only. So it is not evident (a) in what sense 'nominalism' is meant here, or (b) how this Chwistekian nominalism relates to the later FN conditions.

What does Chwistek mean by the term ‘nominalism’? It does not correspond directly to any of Tarski’s finitist-nominalist conditions (though as we shall see, Chwistek harbors a suspicion of infinity). Chwistek’s nominalism—which Tarski is appealing to in the above quotation—corresponds more closely to a demand for *predicativity* found in Poincare’s preferred solution to the paradoxes. Concerning Poincare’s view of mathematical objects, Chwistek writes:

Poincare was a decided nominalist and could not be reconciled to the existence of indefinable objects, much less to the existence of infinite classes of such objects. Poincare regarded his belief as the fundamental postulate of a nominalistic logic. He formulated this postulate as follows: ‘Consider only objects which can be defined in a finite number of words.’ (1935/1949, 21)

In short, the Chwistekian nominalist refuses to countenance the existence of any mathematical object that cannot be finitely described.

It should be noted that Poincaré does not use the word ‘nominalism’ in this sense. Rather, he views nominalism negatively, condemning it as follows: “They have thought that… the whole of science was conventional. This paradoxical doctrine, which is called Nominalism, cannot stand examination.” (1902/1905, 138; cf. xxiii, 105). Furthermore, Poincare calls his own view, which Chwistek dubbed ‘nominalist,’ by a different label: ‘*pragmatist.*’ The pragmatists stand opposed to those Poincaré dubs ‘*Cantorians*’—a label which corresponds, in certain ways, to the cluster of commitments and attitudes currently associated with mathematical ‘Platonism.’ Poincaré writes:

Why do the Pragmatists refuse to admit objects which could not be defined in a finite number of words? Because they consider that an object exists only when it is thought, and that it is impossible to conceive an object which is thought without a thinking subject. … And since a thinking subject is a man, and is therefore a finite being, the infinite can have no other sense than the possibility, which has no limits, to create as many finite objects as one likes. (quoted in Bouveresse 2004, 66)

So the appellation of ‘nominalist,’ in Chwistek’s mouth, corresponds to Poincaré’s ‘pragmatist’; this view is often called ‘constructivism’ today. Terminological differences aside, Chwistek unequivocally endorses the views just expressed by Poincaré:

Jules Tannery inferred that there must exist real numbers which cannot be defined in a finite number of words. Such a conclusion is clearly metaphysical. It presupposes the ideal existence of numbers only some of which can be known.
(1935/1949, 78)

And for Chwistek, like many of his contemporaries, ‘metaphysical’ is a term of harsh disapproval. We find fundamentally the same argument from both Chwistek and Poincaré: if we cannot successfully describe a purported mathematical object, then we should not be committed to the existence of that object. And the finitude of a description, for both men, is a necessary condition for its success—a reasonable requirement, since the describ*er*s are limited creatures. Recent work by Jacques Bouveresse demonstrates that this way of dividing up the warring camps in philosophy of mathematics is not unique to Chwistek (and Poincaré) at the beginning of the twentieth century: in that age, “Platonism… is opposed to constructivism. It rests on the assumption that the objects of the (mathematical) theory constitute a given totality” (2005, 58).

Let us return to Chwistek’s conception of nominalism, for it goes beyond the inadmissibility of (finitely) indefinable mathematical objects. “The doctrines of the nominalists,” Chwistek writes, “depend upon the complete elimination of such objects as concepts and propositions.” (1935/ 1949, 43). Here we find a closer connection to Tarski’s FN project, for eliminating higher-order quantification and restricting the domain of discourse to physical objects only will rule out any realistic construal of concepts and propositions. How is this much stronger claim related to the rejection of indescribable objects? Chwistek holds that if one is committed to the existence of indefinable objects, then one is committed to some sort of concept-realism (*Begriffsrealismus*). This point is argued for concretely in Chwistek’s “The Nominalist Foundations of Mathematics,” published in *Erkenntnis*, in the issue following the symposium proceedings on the logicist, intuitionist, and formalist 'foundations of mathematics,' by Carnap, Heyting, and von Neumann, respectively. There, Chwistek proves (within the simple theory of types) that a certain propositional function φ exists, but also that "φ is unconstructible [*unkonstruierbar*], so we have proved the existence of an unconstructible function, which is of course a metaphysical result that contradicts nominalism in a radical way" (1932-33, 370). Maintaining the existence of such unconstructible entities 'contradicts nominalism,' according to Chwistek, because it represents a very strong realism about concepts, as least if propositional functions are (in some sense) *independent* of us and our cognitive activities of thinking, knowing, and describing. When Chwistek asserts that, under a nominalist regime, 'concepts and propositions' must be 'completely eliminated,' he thus presumably means that the nominalist must eliminate concepts and propositions, conceived of as existing independently of our constructive mathematical activities. Without this final qualification, *constructible* functions would qualify as a 'contradiction of nominalism'—and Chwistek clearly does not want to bring all mathematical activity to a halt.

In the same article, Chwistek also argues against the existence of propositional functions on more general grounds. He maintains that they are not *logical* entities, as Russell and others would have it, for they do not (roughly, and in our terms) stay within the boundaries of syntax alone. He then infers directly from their not belonging to 'pure logic' that they must belong to 'idealistic metaphysics.'

The axiom of extensionality,[26] despite all the arguments of Wittgenstein, Russell, Carnap et al., has nothing to do with logic, since the metaphysical problem whether the propositional functions should count as something different from the expressions, or simply as certain expressions, cannot be decided within logic. From the semantic[27] standpoint the axiom of extensionality is

[26] For Chwistek, the axiom of extensionality is: "any two propositional functions that agree in extension are identical" (Chwistek 1935/1949, 133).
[27] Chwistek's characterization of semantics is non-standard: for him, semantics is "the study of the structural and constructional properties of expressions (primarily of mathematics)" (Chwistek 1935/1949, 83). This is much closer to what we would now consider *syntax*.

simply false, since e.g. the expression ψ{x}∨ψ{x} is clearly different from ψ{x}, although the equivalence of the two expressions holds for all x. If one nevertheless assumes the axiom of extensionality, then one clearly is not dealing with the foundations of pure logic. One is working much more // on a kind of idealistic metaphysics, which I would like to designate as 'concept-realism' [*Begriffsrealismus*], in analogy with certain medieval theories" (1932-33, 368-369)

The argument is simple, and is not restricted to propositional functions: the two sentences 'Jack is thin and Jack is thin' and 'Jack is thin' are syntactically different, but Russell et al. wish to say that, in some sense, they are the same—specifically, they express the same proposition. But character-sequences differ between the two sentences, so the sameness is not purely logical—in Chwistek's idiosyncratic sense of 'pure logic.' And if it is not purely logical, Chwistek infers, it is metaphysics (presumably because it cannot be plausibly construed as empirical). We today, along with many of his logical contemporaries, would deny Chwistek's conception of 'pure logic' as far too weak. If 'pure logic' is allowed a basic inferential apparatus—characterized either syntactically/ proof-theoretically or semantically—then the two sentences above can be considered, in certain respects, the same.

Chwistek makes other claims in "The Nominalist Foundations of Mathematics" that find an echo in Tarski's FN project. For example, Chwistek speaks favorably of Felix Kaufmann's *Das Unendliche in der Mathematik und ihre Ausschaltung*, which Chwistek hails as "announc[ing] the renaissance of nominalism in Germany" (387). "Kaufmann's fundamental idea," Chwistek writes, is "that the meaningful sentences about properties of properties of objects are reducible to sentences about properties of objects" (385). This would come as welcome philosophical news to any proponent of (FN 1), the view that only first-order sentences are fully meaningful. Elsewhere Chwistek states that if one introduces the axioms of infinity and of choice (which are necessary to recover certain mathematical theorems), "one must realize that one has gained certain merely formal relations between sentences, but not contentful results" (371). This echoes, almost exactly, Tarski's view of higher mathematics under a finitist-nominalist regime: namely, higher-order language in mathematics would be characterized as an uninterpreted or empty calculus. In short, Chwistek's influence on Tarski's FN project is perhaps best characterized as indirect, insofar as Tarski does inherit a basic skepticism about the existence of a mathematical reality independent of the material world and our cognitive practices within it, but he does not simply take over Chwistek's so-called nominalism as a fundamental postulate.

(ii) *Kotarbinski's 'reism'*

Tarski cites Tadeusz Kotarbinski's 'reism' as the source of (FN2), the requirement that the domain of discourse must contain physical objects only. Tarski also helped translate one of Kotarbinski's introductory articles on reism into English (1935/1955). What is reism? It is the view that everything is a *res*, a thing. This is not a terribly informative formulation (recall Quine's answer to 'What is there?' viz., 'Everything' (Quine 1953/1961, 1)). More revealingly, Kotarbinski also labels his view 'concretism,' the claim that everything is *concrete* (so no abstracta exist), as well as 'pansomatism': everything is a body. Sometimes Kotarbinski uses 'reism' to designate the weaker view that everything is a *res extensa* or *res cogitans* (see 1935/1955, 489); but then Kotarbinski notes that *his* view, pansomatism, is a particular species of reism (generated by adding the assumption that "every soul is a body" (1935/1955, 495). As mentioned above in section 2, there is disappointingly little discussion in the Harvard notes of what the participants mean by 'physical thing'; Kotarbinski, fortunately, offers brief hints as to what he counts as a *res*. He writes: "'Corporeal', in our sense, means the same as 'spatial, temporal, and resistant'"; thus, Kotarbinski counts an electromagnetic field *in vacuo* as a body, since it is resistant and spatiotemporal (1935/1955, 489).

Kotarbinski himself recognizes that his pansomatism is closely related to nominalism, for he writes:

Concretism… joined the current of nominalism, if by nominalism we mean the view that universals do not exist. … Not only do properties not exist, but neither do relations, states of things, or events, and the illusion of their existence has its source in the existence of certain nouns, which suggest the erroneous idea of the existence of such objects, in addition to things. (1929/1966, 430)

As the end of this quotation makes clear, pansomatism is has not only an ontological component but also a linguistic-semantic one. (Ajdukiewicz's response to Kotarbinski's initial formulations of reism prompted this distinction to be made explicit.) Kotarbinski's idea is that every sentence containing a grammatical subject or predicate that does not designate concrete object(s) can be re-phrased, without loss of content, into a sentence in which all grammatical subjects and predicates *do* designate concrete bodies only (1935/55, 490).[28] For example, the reist will transform 'Roundness is a property of spheres' into 'spheres are round,' and 'The number of dogs in my house is even' into 'The number of dogs in my house are pairwise numerous.' What motivates such a transformation? Kotarbinski's polemical answer is as follows:

Generally speaking, if every object is a thing, then we have to reject every utterance containing the words 'property', 'relation', fact', or their particularization, which implies the consequence that certain objects are properties, or relations, or facts. (1935/1955, 490)

Kotarbinski also explicitly rejects classes (1935/1955, 492). Thus we see that (FN1) (as well as (FN 2)) is a consequence of reism. Kotarbinski says that we could declare utterances containing such words either false or nonsensical (on grounds of their ungrammaticality). 'Roundness,' 'property,' etc. are called '*onomatoids*,' that is, merely apparent names, not genuine names; or if one chooses to call them 'names,' then they are *denotationless* names, like 'Pegasus.' In places (e.g. 1966, 432), Kotarbinski suggests that the reist's paraphrase is in fact what was *really* meant all along.

[28] 'But,' the modern reader will object, 'predicates do not designate concrete bodies. Kotarbinski is guilty of a category mistake (or another form of nonsense): only singular terms designate individuals.' This modern understanding of predicates is not shared by Kotarbinski, who holds the (ultimately medieval) view that singular terms name a single (concrete) thing, while predicates name several (concrete) things.

To understand Kotarbinski fully, Stanislaw Lesniewski's basic views must be outlined, since Kotarbinski adopts, in service of reism, the formal logic of Lesniewski (which Lesniewski calls 'ontology'). Lesniewski rejects the classical set-theoretic conception of classes, replacing it with the notion of a mereological whole (which he nonetheless called a 'class,' for he believed it was the salvageable remainder of the notion Cantor studied). Lesniewski bases his logic on the symbol 'ε,' which is intended to formalize the (ordinary language) copula. 'A ε B' can be given two readings in natural language (both of which are simultaneously possible on Lesniewski's formalism): 'A is a proper part of, or identical with, B' (the mereological conception) or 'A is one of the Bs.' One might think this latter smuggles in class-membership. However, the idea is not that there is a property of B-ness; rather, 'B' just names many (concrete) things—along the lines of the medieval nominalists' view. Interestingly, Quine espouses the very same understanding of predicates in a 1937 lecture on nominalism to the Harvard Philosophy Club (MS STOR 299, Box 11). Finally, Lesniewski takes a pansomatist view of logic itself, as Peter Simons explains:

[E]xpressions, their components and the wholes they constitute are one and all concrete entities: marks on paper, blackboards etc. … Lesniewski does not, as is common metalogical practice, assume there are infinitely many expressions of every category available. A system of logic for him is no less concrete than any other chunk of language" (Simons 1993, 220).

As strange as this view may sound, we will find Tarski and (to a lesser extent) Quine defending this conception of language in the notes (see II.4.B)—and Goodman and Quine publicly defend it in "Steps Toward a Constructive Nominalism."

B. Russellian influences

It may be an understatement to say that Russell towers over logically-informed and logically-inspired philosophy in the twentieth century, especially before 1940. Although he was in residence at Harvard during the fall of 1940, Carnap's notes do not record sustained involvement on Russell's part in these conversations. Nonetheless, his well-known views about numbers, classes, and abstracta presumably had some more general effect—even if only indirectly—on Carnap, Quine and Tarski's discussions. That is, even if they did not adopt his views completely, at least his output over the previous five decades presumably influenced their conception of what counts as a philosophically pressing question. Put otherwise, what Russell considered philosophically problematic and important partially determined the conceptual horizon for philosophers who followed in his wake.

In particular, Russell found numbers and classes philosophically troubling. He calls numbers—which, on his analysis, are classes of classes—'fictions of fictions':

> Numbers are classes of classes, and classes are logical fictions, so that numbers are, as it were, fictions at two removes, fictions of fictions. Therefore, you do not have as ultimate constituents of your world, these queer things that you are inclined to call numbers. (1918/1956, ???)

The fact that the greatest philosophical luminary of Carnap, Tarski and Quine's early careers called classes 'fictions' and declared numbers—the things introduced to and manipulated by five-year-old children—to be 'queer things' could play some indirect role in inclining Tarski, Quine and others to consider Tarski's refusal to allow numbers into the universe of discourse *prima facie* plausible or reasonable. Along similar lines, Russell describes the assertion that numbers exist as metaphysics.

> [S]o long as the cardinal number is inferred from the collections, not constructed in terms of them, its existence must remain in doubt, unless in virtue of a metaphysical postulate *ad hoc*. By defining the cardinal number of a given collection as the class of all equally numerous collections we avoid the necessity of this metaphysical postulate, and thereby remove a needless doubt from the philosophy of arithmetic.
> (1918, 156)

This shows that Russell considered taking the existence of numbers as a primitive assumption a metaphysical maneuver—and as we shall see in detail later (II.2 and III.2), part of the motivation for undertaking the finitist-nominalist project is to demonstrate that (at least a substantive chunk of) mathematics is not metaphysics, but is cognitively meaningful. Finally, Tarski's requirement that the universe of discourse contains physical objects only can perhaps be seen as a return to the Russellian conception of logic that the *Tractatus* aims to dismantle, namely, that “logic is concerned with the real world just as truly as zoology, though with its more abstract and general features” (1919, 169). (The fact that Carnap considered this Tractarian view a lynchpin of the logical empiricists’ epistemology of mathematics may explain his near-instinctual aversion to the fundamental assumptions of the FN project (see III.2-4).)

However, of course, the logic of *Principia Mathematica* directly violates (FN1), since it is higher-order. However, the axiom of reducibility effectively states that all formulae have first-order equivalents, so even in *PM*, there is a sense in which first-order logic is privileged. The situation is more subtle when it comes to (FN3). *Principia Mathematica* adopts an axiom of infinity, but Russell was not happy about the need to do so, and he explicitly considered such an axiom extra-logical. So he would certainly be sympathetic to the motivation behind (FN 3) as well. Also, the fact that Russell considered it worthwhile to eliminate classes from the system of logic in *Principia Mathematica* (the ‘no-class’ theory) could have conceivably exerted some indirect influence on Tarski’s refusal to countenance abstract entities, and on Quine and (to a lesser extent) Carnap’s willingness to consider the FN project worth pursuing.[29]

C. Logical empiricists skeptical of infinity

Finally, I wish to consider very briefly certain remarks made by the logical empiricists and their allies concerning infinity in the first part of the twentieth century. Their attitude is, in general, much more hostile than the prevailing sentiments at the beginning of the twenty-first; as an indirect result of this general mood, Tarski’s condition that no infinities be presupposed in the

[29] However, Russell does not assert that classes do not exist; he is an agnostic, instead of an atheist: “we avoid the need of assuming that there are classes without being compelled to make the opposite assumption that there are no classes. We merely abstain from both assumptions” (1919, 184).

language of science and mathematics (FN3) would likely appear more reasonable. As just remarked, Russell was forced to introduce the axiom of infinity into the logic of *PM* in order to capture certain basic results in mathematics, but he found this maneuver philosophically unsatisfying: he considered every proof of a theorem *p* of classical mathematics that appealed to the axiom of infinity to be better understood as a conditional proof of the form ‘If the axiom of infinity holds, then *p*.’

Wittgenstein, in the *Tractatus*, also rejects the axiom of infinity (5.535). He suggests that in a logically perfect language, each object will have exactly one name; thus, there will be infinitely many objects if and only if there are infinitely many names of objects. So the problem with the axiom of infinity, on this line of thought, is that if it is true, then in a logically perfect language what it intends to say is superfluous (though Wittgenstein apparently considers the axiom itself meaningless): the infinitely many names already captures (‘shows’) the infinity of objects. During Wittgenstein’s so-called ‘middle period,’ his antipathy towards the notion of infinity grows stronger. I will not attempt to analyze his complex pronouncements in detail here, but at the most basic level, one worry seems to be that for finite beings speaking a finite language and pursuing finite goals, the introduction of infinity seems ill-suited. It is during this period, for example, that we find Wittgenstein saying privately, and his associate Moritz Schlick saying publicly, that laws of nature are meaningless, on the grounds that they speak about an infinite number of occurrences.

Wittgenstein, of course, was not a fully- fledged member of the Vienna Circle, so one might think his skepticism about infinity might be seen as traditional philosophical quibbling by the Circle members, who would consider such skepticism the result of insufficient knowledge of, and/or respect for, scientific practice. However, this is not the case. Otto Neurath, who had very little patience for what he called Wittgenstein’s ‘metaphysics,’ was also hesitant to introduce the term ‘infinite’ and its cognates into the unified language of science. For Neurath, it seems that the problem is not merely that the axiom of infinity is extra-logical (synthetic), or even that it is false, but rather that the very concept of infinity is, in some sense, unacceptable for a committed anti-metaphysical empiricist.

> Perhaps there are theological residues also… in certain applications of the concept of infinity in mathematics. The attempts to make mathematics finite, especially in applications to concrete events, are certainly part of tidying up [the language of science]. Frequently we need only to give a finite meaning to statements with infinitesimal or transfinite expressions.
> (1983, 43)

Similar to the views of Russell just above, the concepts of classical mathematics are ‘theological’—the sibling of ‘metaphysical.’ Note also that the FN project is, in part, an attempt to fulfill the final sentence of this quotation—for it attempts to confer a (potentially) finite meaning on claims of classical mathematics involving infinity. Years later, Neurath’s worries about infinity were raised again, when discussion among scientific philosophers focused on the concept of probability: “There remains the difficulty to apply a calculus with an infinite collective to empiricist groups of items, to which the expressions ‘finite’ and ‘infinite’ can hardly be applied” (1946, 81). This objection is not fully fleshed out; I mention this only to show that Neurath’s skepticism about infinity continued over several years before and after the Harvard discussions, and stretched across different topics. Hans Hahn’s “Does the Infinite Exist?” expresses more moderate worries about the notion of infinity. Hahn does not answer the question in his title unequivocally, but his comparison of the infinities studied in Cantorian set theory with the possible infinities in the physical world lead him to the conclusion that the word ‘exists’ “means something entirely different” in mathematics and natural science (1934/1980).[30] Such a position—that ‘exists’ in mathematical contexts is merely homonymous with ‘exists’ in non-mathematical ones—is philosophically contorted and therefore unappealing. Perhaps seeing

[30] Hahn’s eventual criterion for mathematical existence is mere consistency.

such unsatisfactory attempts to integrate the logico-mathematical notions with empirical ones would also lead Tarski, Quine and Carnap to consider restricting the universe of discourse to physical objects only: the sense of ‘exists’ is thereby given a unified, and thus hopefully more natural treatment, and the relationship between the claims of mathematics and the claims of natural science is clarified. In general, the FN project’s renunciation of infinity comes with a high cost (in terms of sacrifice of classical mathematics), but it brings in its wake the benefits of avoiding convoluted and fragmented attempts, such as Hahn’s, to reconcile the mathematical realm with the natural one.

In 1940, when many of the greatest scientific philosophers of the twentieth century spent a year together, the plurality of their academic collaboration focused on the question: ‘What form should an intelligible language adequate for science take, if the number of physical things in the universe is possibly finite?’ And, as a corollary, ‘How will this force us to change arithmetic?’ In this chapter, the conditions Tarski proposes a language must meet in order to be intelligible were examined, and how these conditions might be compatible with mathematical discourse. For many twenty-first century students of philosophy, it will be somewhat surprising that such a question is at the center of their discussions—why did they not discuss issues (that we today perceive to be) more closely related to the core of their published, public views? This sense of surprise might make it appear to some that the finitist-nominalist endeavor is a peripheral side-project that is basically unrelated to the real areas of research for these philosophers.

However, this appearance is deceiving. For, as we shall see, several topics that are fundamental to scientific philosophers (both pre- and post-1940) are intimately involved in this particular—and peculiar—question. The most obvious and direct connection is to Quine’s (short-lived) and Goodman’s (long-lived) nominalism. Carnap, as we shall see in III.1, assimilates the finitist-nominalist endeavor to his work on the relation between observational and theoretical languages, begun in 1936's "Testability and Meaning" and continued well after 1941. But the 1941 conversations are fundamentally a discussion of the relation of mathematics to the world, an issue Carnap and other logical empiricists (especially Schlick and Hahn) considered of paramount importance throughout their careers—for the new mathematical logic held the promise of delivering truths independent of empirical facts about the world, and thus a tenable empiricist view of mathematics (i.e., one that did not fall into Mill's view) appeared possible. For this reason, the finitist-nominalist project involves issues of analyticity, the topic that perhaps looms largest in hindsight. Very closely intertwined with analyticity is the question of the best form for the emerging field of (formal) semantics, the immediate heir of the logical empiricists' concern with the notion of *meaning* (and *meaningfulness*) that frequently occupies center stage in the twenties and thirties. The differences of opinion between Carnap, Tarski and Quine on this matter are many and varied, so I will not attempt to summarize them here. The admissibility of modal (and other intensional) languages is drawn into the discussion of the finitist-nominalist project, in part because they want to set up the language such that it is *possible* that the number of physical things in the universe is finite, but Tarski and Quine are skeptical of intensional languages. Finally, insofar as the finitist-nominalist project aims to develop a single, unified language for all scientific discourse, it is also intertwined with the unity of science, an idea which had a heyday in the thirties, but lives on today in various reductionism debates. In each of the following chapters, I not only present and analyze the details of the 1940-41 discussion notes in their own terms, but will also show how they relate to the wider themes just mentioned.

CHAPTER II. JUSTIFICATIONS FOR FINITIST-NOMINALIST RESTRICTIONS

The official year of birth for modern Anglophone nominalism is generally taken to be 1947, with Quine and Goodman's *Journal of Symbolic Logic* article "Steps Toward a Constructive Nominalism." In a footnote in that article, the authors acknowledge that the initial impetus and strategy for their nominalist project was proposed in 1940 by Tarski, and discussed with the authors and Carnap (1947, 112). Thus, the discussions at Harvard in 1940-41 can be seen as an important wellspring of current nominalism. In the previous section, Tarski's finitist-nominalist criteria (FN1)-(4) were set out and explored in some detail. A pivotal question that was not addressed then, but shall be in this chapter, is the following: *What motivates or justifies these finitist-nominalist criteria?* First, I discuss the justifications presented by Tarski, Carnap and Quine for undertaking the finitist-nominalist project. I discern four kinds of rationales Tarski, Quine and Carnap consider for a finitist-nominalist project. I shall class these motivations under four headings: intelligibility, the anti-metaphysical drive, inferential safety, and natural science. The first three support FN1 and 2, and the fourth FN3. Finally, I will briefly outline the two primary current justifications for nominalism, and describe how they relate to those considered by Tarski, Quine, and Carnap.

1. FIRST JUSTIFICATION FOR THE FN CONDITIONS: *VERSTÄNDLICHKEIT*

A. Assertions meeting the finitist-nominalist criteria are verständlich

As we have already seen (in 090-16-25; -28), Tarski claims that a language must meet his finitist-nominalist restrictions in order to qualify as 'fully understandable' or intelligible. I think it is fair to interpret Tarski as suggesting that FN 1-3 are necessary (and perhaps sufficient) conditions for a language to be intelligible. We also saw above (I.2) that Carnap, in his Autobiography, mentions no other motive for this language-construction project besides the aim of understandability. In the 1940-41 notes themselves, Carnap clearly interprets the FN criteria as necessary conditions on the understandability of a language as well, for we find him writing the following:

[L]ogic and arithmetic also remain in a certain… sense finitistic, if they should really be understood. (090-16-24)

Is this talk of sequences whose length is greater than the number of things in the world at peace with the principle of finitism? I.e., is such a sentence understandable for the finitist? (090-16-27)

In both these quotations, which Carnap wrote to himself in private (i.e., they are not part of a conversation with Quine and Tarski), Carnap is saying: for the finitist, if a language is understandable, then it meets the finitistic criteria. (Terminological note: Carnap calls a 'finitist' someone who accepts *all* of FN 1-3, not just FN3 alone.)

Quine's view of the relation between intelligibility and nominalism is similar to Tarski and Carnap's, but he puts the point differently. In December 1940, before Tarski has proposed the language-construction project, Quine delivered a lecture on the topic of "the universal language of science" (RC 102-63-04). In it, he discussed philosophers' attempts to eliminate certain "problematic universals" from the language of science. "In each case" of eliminating universals, Quine writes, "we do it in order to reduce the obscure to the clearer." This is very similar to Tarski and Carnap's view of the aim of the FN restrictions described above, for aversion to universals is a classical characteristic of nominalism, and presumably, the 'obscure' is less understandable than the 'clear.' (However, Quine may intend 'clear' to have epistemological, instead of semantic, force here; I expand on this suggestion below.) So, in short, Tarski, Quine, and Carnap all hold that a central aim of undertaking this language construction project is that such a language would be maximally ‘intelligible’ or ‘clear.’

B. What does ‘understandable’ *[verständlich]* mean in the discussion notes?

The obvious question to ask next is: What do the participants mean by *verständlich*? Unfortunately, the discussion notes record very little. (What little there is will be discussed in the next paragraph.) It is frustrating that the discussion notes lack an explanation of what intelligibility is, since it is, according to all parties involved, the acknowledged central motivation for constructing a finitist-nominalist language. More specifically, the notes lack an explicit explanation of why a language violating any of Tarski’s finitist-nominalist criteria is not (fully) understandable; why, for example, would a sentence beginning with ‘There exists a property such that…’ be as unintelligible as an obviously ungrammatical string of English symbols, or even as Heidegger’s infamous metaphysical claim “*das Nichts nichtet*”? Furthermore, the participants do not completely agree (though their positions are not disjoint) among themselves about which particular assertions are fully understandable and which not: Carnap claims, *contra* Tarski and Quine, that classical, infinite arithmetic *is* intelligible, though he holds set theory is not (RC 090-16-25). The notes from the final day of collaborative work on the finitist-nominalist language highlight how the concept of *Verständlichkeit* is unclear and imprecise for the participants. Carnap writes:

> We agree the language should be as understandable as possible. But perhaps it is not clear what we properly mean by that. Should we perhaps ask children psychological questions, what the child learns first, or most easily? (RC 090-16-05)

So Carnap himself does not know what is meant by *verständlich*, even six months into the project, and the dictation notes show no response to his query from the other participants. However, this quotation does suggest that for Carnap, understandability is a *pragmatic* characteristic (in Carnap's sense) of a sentence or language, i.e., a property that depends on the language-user; I shall make use of this idea later in the chapter.

Our interpretive prospects are not hopelessly bleak, for there is *some* material in the discussion notes that provides insight into what *verständlich* means for Tarski, Quine, and Carnap. In particular, Tarski contrasts an intelligible language with an uninterpreted formal calculus.

Tarski: I fundamentally understand only a language that fulfills the following conditions: [Here are the three finitist-nominalist restrictions; see full quote above on p.xx]… I only "understand" any other language in the way I "understand" classical mathematics, namely, as a calculus; I know what I can derive from other [sentences] (rather, I have derived; "derivability" in general is already problematic). With any higher "platonic" statements in a discussion, I interpret them to myself as statements that a fixed sentence is derivable (or derived) from certain other sentences. (He actually believes the following: the assertion of a sentence is interpreted as signifying: this sentence holds in the fixed, presupposed system; and this means: it is derivable from certain foundational assumptions.) (RC 090-16-28)

The contrast between 'intelligible language' and 'uninterpreted calculus' also appears, albeit more briefly, elsewhere in the discussion notes (RC 090-16-25, -04, -05). Quine makes a similar point in 1943's "Notes on Existence and Necessity": "The nominalist, admitting only concrete objects, must either regard classical mathematics as discredited, or, at best, consider it a machine which is useful despite the fact that it uses ideograms of the forms of statements which involve a fictitious ontology" (1943, 125). Thus, merely knowing (to put the point in the terminology of Carnap's *Logical Syntax*) the formation and transformation rules of a calculus does *not* constitute genuine understanding of the language corresponding to that calculus, i.e., the language that that calculus is intended to model formally.[31] That is, understanding what a sentence *s* means is *not* merely knowing which sentences are deducible from *s* and which sentences *s* is deducible from. This basic view has found a modern expression in Searle's 'Chinese Room' thought-experiment, which purports to show that a computer cannot understand a language, because a computer's

[31] (Formation rules determine the well-formed formulae or grammatical strings of a calculus; transformation rules are commonly called 'rules of inference.')

operations are restricted to the realm of syntax. Regardless of what the detailed content of *Verständlichkeit* might be for Tarski and Carnap, it at least requires that a language be more than an uninterpreted calculus or 'empty formalism,' in addition to its being a pragmatic notion.

C. What does 'understandable'/ 'intelligible' mean in Carnap's publications?

The contrast between an understood language and an uninterpreted calculus also emerges clearly in Carnap's published remarks during this period. For when Carnap discusses the notion of *understanding*, he repeatedly connects it with *interpretation*. In both *Foundations of Logic and Mathematics* in 1939 and *Introduction to Semantics* in 1942, to 'understand' a sentence is to (know how to) *interpret* that sentence—and to interpret a sentence is to assign it truth-conditions via 'semantic rules.' Carnap writes in *Introduction to Semantics*:

By a **semantical system** (or interpreted system) we understand a system of rules…of such a kind that the rules determine a **truth-condition** for every sentence of the object language… In this way the sentences are interpreted by the rules, i.e. made *understandable*, because to understand a sentence, to know what is asserted by it, is the same as to know under what conditions it would be true. (1942, 22; italics mine)

We find a virtually identical claim three years earlier, in *Foundations of Logic and Mathematics*:

Therefore, we shall say that we *understand* a language system, or a sign, or an expression, or a sentence in a language system, if we know the semantical rules of the system. We shall also say that the semantical rules give an *interpretation* of the language system. (1939, 152-153)

Clearly, Carnap's publications both before and after the Harvard conversations of 1940-41 contain a conception of understanding that dovetails with the conception of understanding he and Tarski articulate during the Harvard conversations, for both the published and the unpublished remarks treat uninterpreted calculi as not understood.

We now have the materials necessary to make the crucial point of this section, one that connects the notion of *Verständlichkeit* to broader themes in Carnap's work and twentieth century philosophy more generally. Since an interpretation of a grammatical string of symbols gives its truth-condition, and for Carnap (and many others) at this time, a sentence's meaning is its truth-condition, an interpretation supplies meanings to (otherwise meaningless) symbols. This is why, in the discussion notes, 'uninterpreted calculus' is contrasted with 'intelligible language': an uninterpreted calculus has not yet had a meaning conferred upon it. And, recalling that *verständlich* and its cognates are pragmatic notions, it appears that 'intelligible' is the pragmatic (i.e., language user-dependent) correlate of the semantic notion 'meaning.' That is, a speaker understands a particular sentence if and only if she knows that sentence's meaning. (Meaningfulness differs from understandability because a sentence can be meaningful, even in my native language, even if I do not understand it—I might lack the requisite vocabulary.) The term '*Verständlichkeit*' is thus intimately connected to discussions of meaning and meaningfulness, notions which have occupied center stage in analytic philosophy throughout much of its history.

There is another, derivative sense of 'understanding' that Carnap offers both at this time in his career and later; a brief detour is needed to examine it. This second sense does not appear in Carnap's discussions of semantics in general or in the abstract, but rather in his treatment of the semantics of fundamental scientific theories. Carnap's basic idea, put simply, is that *incomplete interpretations* may provide understanding as well, if certain other conditions (to be spelled out shortly) are met. In *Foundations of Logic and Mathematics*, Carnap first notes that we have less and less "intuitive understanding" of the fundamental terms of modern science, such as Maxwell's electromagnetic field and, more strikingly, the wave-function in quantum mechanics (1939, 209). (Here, 'intuitive' perhaps carries Kantian overtones, if it is not intended to match precisely the Kantian characterization of intuition.) Carnap writes that "the physicist…cannot give us a translation into everyday language" of the symbol for the quantum-mechanical wave-function (211). Given Carnap's account above, in which understanding is achieved via interpretation, this appears to create a problem: how can modern physical theories be understood on Carnap's account, given that some of the fundamental (i.e. primitive (207)) terms do not admit of direct interpretation?

Carnap maintains that there is a sense in which a modern physicist "understands the symbol 'ψ' and the laws of quantum mechanics" (211). This seems reasonable: it would be Pickwickian to claim that Einstein does not understand the general theory of relativity. Carnap suggests that a physicist's understanding consists in using a physical theory—including the 'unintuitive' terms, which cannot be 'translated into ordinary language'—to explain previously observed phenomena and make new predictions. And *this* sort of understanding can be achieved via a partial or incomplete interpretation of a calculus, provided that the uninterpreted, 'unintuitive' terms are appropriately inferentially connected to the interpreted terms. Thus Carnap writes:

It is true a [physical] theory must not be a "mere calculus" but possess an interpretation, on the basis of which it can be applied to facts of nature. But it is sufficient… to make this interpretation explicit for elementary [= interpreted] terms; the interpretation of the other terms is then indirectly determined by the formulas of the calculus, either definitions or laws, connecting them with the elementary terms. … Thus we understand '*E*' [the symbol for Maxwell's electric field], if "understanding" of an expression, a sentence, or a theory means // capability of its use for the description of known facts or the prediction of new facts. An "intuitive understanding"… is neither necessary nor possible. (210-211)

Carnap maintains this view many years later, in his *Autobiography*.

[T]he interpretation of the theoretical terms supplied by the [semantic] rules is incomplete. But this incomplete interpretation is sufficient for an understanding of the theoretical system, if "understanding" means being able to use in practical applications; this application consists in making predictions of observable events, based on observed data, with the help of the theoretical system. (1963, 78)

Carnap's basic picture is clear: a partially interpreted calculus qualifies as understood, provided that that calculus—including its terms that are not directly interpreted—is substantively inferentially related to the unproblematically understood terms and sentences of 'everyday language' and thus is useable in practical applications, especially explanation and prediction.

The terms that are not directly interpreted (i.e., the 'theoretical' ones) become useful 'for making predictions of observable events' only if they are inferentially connected to the directly interpreted ones.[32] (Note that this liberalized version of 'understanding' includes Carnap's narrower, original version as a degenerate case.) In *Logical Syntax of Language* §84, Carnap also claims that practical application is one means of interpretation—although there, the calculus to be interpreted is drawn from pure mathematics, not natural science. Carnap writes: "the interpretation of mathematics is effected by means of the rules of application" of pure mathematics to synthetic sentences (1934/1937, 327). Carnap's use of partial interpretations will be discussed at more length below, in III.2.A.

But what is the *point* of producing such a partial interpretation? What good does it achieve? In *Foundations*, Carnap first introduces the issue of the understandability of a physical theory in the following terms: (how) can a *lay person* understand the content of the theory?

> Suppose that we intend to construct an interpreted system of physics—or the whole of science. We shall first lay down a calculus. Then we have to state semantical rules… For which terms, then, must we give rules, for the elementary or the abstract ones? We can, of course, state a rule for any term, no matter what its degree of abstractness… But suppose we have in mind the following purpose for our syntactical and semantical description of a system of physics: the description of a system shall teach a layman to understand it, i.e., to enable him to apply it to his observations in order to arrive at explanations and predictions. A layman is meant as one who does not know physics but has normal senses and understands a language in which observable properties of things can be described (e.g., a suitable part of everyday non-scientific English). (1939, 204)

Here again, we see evidence that Carnap considers the intelligibility of a language as a pragmatic matter: what is understandable to one language-user (e.g. a particle physicist) will likely be different from what is understandable to another (e.g. a bus driver). On a tangential but interesting note, this text points to an interesting historical fact about twentieth century

[32] Carnap's theory of meaning for scientific language is thus a hybrid of what Fodor and Lepore have called (1992) Old Testament (roughly, referential) semantics and New Testament (roughly, conceptual role) semantics. Carnap employs Old Testament semantics at the level of 'observable' or 'elementary' terms, and the New Testament (or inferential role) semantics is applied to the 'higher' reaches of scientific language.

philosophy of science. The distinction that Carnap draws between ‘elementary’ and ‘abstract’ terms is virtually identical to the infamous distinction between the observational and theoretical vocabularies. This latter distinction is often said to aim at isolating the ‘empirical content’ of a scientific theory. (For example, van Fraassen attacks attempts to identify empirical import using this distinction in (1980, 54).) But we see in the above quotation that in *Foundations* Carnap does *not* draw the distinction in order to isolate empirical content, but rather to make the theory understandable to a layperson. These two aims are clearly different. (However, we cannot be thoroughgoing revisionists: in “Testability and Meaning,” Carnap also endorses the goal traditionally ascribed to him, and attacked by van Fraassen.) Additionally, the purpose Carnap states here in *Foundations* dovetails nicely with one of Neurath’s goals for Unified Science (and his Encyclopedia): to democratize science by presenting scientific claims in a form comprehensible to everyone. And though *Foundations* does not articulate that vision explicitly, the work *is* a monograph in Neurath’s *Encyclopedia of Unified Science*.

D. What does ‘understandable’ mean for Quine?

So much for Carnap’s published remarks on understandability[33]; how does Quine conceive of *Verständlichkeit*? Quine, to the best of my knowledge, never explicitly affirms or denies Carnap’s conception of understanding as interpretation. In fact, I cannot find a characterization of (much less necessary and sufficient conditions for) intelligibility or understandability anywhere in Quine’s published corpus. This absence is all the more conspicuous, given that Quine frames his critique of analytic truth in terms of analyticity not being “intelligible” (Two

[33] In 1955’s “Meaning and Synonymy in Natural Languages,” Carnap offers yet another characterization of understandability: “the theory of intention of a given language *L* enables us to *understand* the sentences of *L*. On the other hand, we can apply the concepts of the theory of extension of *L* only if we have, in addition to the knowledge of the theory of the intension of *L*, also sufficient empirical knowledge of the relevant facts” (1955/1956, 234).

Dogmas, p.xx). However, we do find hints about the meaning Quine attaches to 'intelligibility' during this period in two letters he writes to Carnap in the 1940s. In a 1947 letter to Carnap, Quine writes that he considers an 'exclusively concrete ontology' intelligible:

I am not ready to say, though, that when we fix the basic features of our language… our guiding consideration is normally convenience exclusively. In my own predilection for an exclusively concrete ontology there is something which does not reduce in any obvious way to considerations of mere convenience; viz., some vague but seemingly ultimate standard of intelligibility or clarity (RC 102-63-01; =Creath 1990, 410).

Two points concerning this quotation are relevant for present purposes. First, note that for Quine, 'intelligibility' and 'clarity' are (at least roughly) synonymous. This closely echoes the language of Quine's 1940 lecture, briefly discussed above, in which 'universals are eliminated' in order to 'reduce the more obscure to the clearer.' Carnap does not, as far as I know, tie intelligibility to clarity. Second, at this point in Quine's career, intelligibility is not a merely pragmatic notion, i.e., the most intelligible apparatus is not necessarily the most 'convenient' one. Furthermore, the standard of intelligibility is 'ultimate' or (in philosopher-speak) brute—not only is it irreducible to 'mere convenience,' but it is not reducible to anything else. A very similar sentiment is expressed in the Goodman and Quine paper on nominalism: "Why do we refuse to admit the abstract objects that mathematics needs? Fundamentally this refusal is based on *a philosophical intuition that cannot be justified by appeal to anything more ultimate*" (1947, 105, italics mine). Readers of the 1947 paper have usually found this justification (if it can be called that) for nominalism extremely unsatisfying (Burgess and Rosen 1996, 205). I suggest that we could interpret this 'intuition' as a (transformed) version of the 1941 demand for intelligibility; if we do, then at least some small light is shed on what Goodman and Quine might have had in mind with this cryptic claim. The justification for this exegetical conjecture is that just as *Verständlichkeit* is both the primary motivation for the 1941 project and yet remains unclear and vague to the people using the term, the 1947 'intuition' is the 'fundamental' impetus for undertaking the nominalist constructions, and yet Goodman and Quine offer no explicit explanation of it. The primary difference between the 1941 and 1947 versions of nominalism is that the former is primarily (though probably not exclusively) a *semantic* notion (claims involving abstracta are, at best, empty formulae), the latter is an *ontological* one (abstracta do not exist).

The considerations of the previous paragraph lead one to suspect that Quine holds, at this point in his career, that there are epistemic virtues independent of pragmatic ones. Such a suspicion is borne out by a letter Quine writes to Carnap in 1943 regarding their Harvard discussions two years earlier.

> [T]he program of finitistic construction system [*sic*] on which the four of us talked at intervals in 1941… may indeed be essential to a satisfactory epistemology. The problem of epistemology is far from clear, as you have emphasized; and essential details of the aforementioned program must depend, as we have seen, on some increased clarification as to just what the epistemological question is. I am more hopeful than you of the eventual possibility of such a clarification; i.e., the possibility of eventually reducing to the form of clear questions the particular type of inarticulate intellectual dissatisfaction that once drove you to work out the theory of the *Aufbau*, and Goodman his related theory. …
>
> [I]n the course of… discussion it began to appear increasingly that the distinguishing feature of analytic truth, for you, was its epistemological immediacy in some sense. … Then we [Tarski and Quine] urged that the only logic to which we could attach any seeming epistemological immediacy would be some sort of finitistic logic. (Creath 1990, 294-295)

First note the final two sentences of this quotation. Given that, in 1941, the stated goal of the finitist-nominalist project was to construct a fully understandable language, it seems reasonable to infer that, for Quine, understandability is closely related to (if not identical with) 'epistemological immediacy.' Presumably, whatever is epistemologically immediate does not stand in need of further justification (for such knowledge is not 'mediated'); we might view this as an expression of some form of epistemological foundationalism. (This form of foundationalism could be moderate: all this commits Quine to is the existence of some unjustified justifiers.) Quine's notion of intelligibility in the above letters differs from that found in Carnap's published remarks. In the latter, *Verständlichkeit* appears primarily to be a *semantic-pragmatic* concept: a language is understandable to a particular person if that person knows the meaning of its sentences (i.e., can interpret the language's symbols). In 1943, however, Quine apparently thinks of *Verständlichkeit* primarily as an *epistemological* and *non-*pragmatic concept.[34]

Why has Quine apparently slid from one to the other? When philosophers in the early 20th century undertake projects of analysis, we can distinguish two distinct types: semantic analysis, which uncovers the 'real meaning' or logical form of a sentence often 'hidden' beneath the sentence's surface grammar, and epistemological analysis, which uncovers the grounds for the truth of a proposition. These two types of analysis can be tied together; the verification criterion of meaning is one way to achieve that association, and Carnap asserts in the *Aufbau* that a constitution system aims to exhibit not only the epistemic order of our knowledge, but also the *meanings* of our concepts (1928/1967, 246). However, semantic and epistemic analyses can be kept distinct, and at times they are: for example, Russell's analysis of definite descriptions in "On Denoting" is clearly a semantic analysis, not an epistemological one (though, of course, it has epistemological consequences, such as the happy fact that we need not be acquainted with nonexistent objects). But since semantic and epistemic analyses were often conflated by early analytic philosophers, it is in some sense natural that Quine would run them together in his reflections on the finitist-nominalist project of 1941.

I would like to draw out two further points from the above quote, focusing particularly on its first half. First, it provides evidence that Quine, like Carnap, felt the participants in the

[34] There is a serious equivocation here in the term 'pragmatic,' and it stems from Carnap's terminology. When discussing semantic issues, Carnap uses 'pragmatic' to refer to those aspects of language and meaning that are not independent of an individual speaker. On the other hand, when Carnap says that the choice between languages or linguistic frameworks is a *pragmatic* one, he means that the decision of which language to use is not made on the grounds of any facts of the matter (i.e., there is not one 'true language'), but rather on the grounds of convenience and utility. Quine's interpretation of 'intelligible' is clearly non-pragmatic in the second sense (for it 'does not reduce to mere convenience'), but it is unclear whether it is non-pragmatic in the first, i.e., whether it is independent of individual language speakers. (In any case, the epistemic vs. semantic distinction between Quine and Carnap holds nonetheless.)

Harvard discussions had not clearly fixed the meaning of 'understandable,' and that they recognized this fundamental unclarity. Second, the beginning of the above quotation reveals something interesting about the relative intellectual trajectories of Carnap and Quine. As is well known, Quine recants epistemological foundationalism in his later, post-nominalist work (for particularly strident examples, see "Posits and Reality" and "Epistemology Naturalized"). Carnap had already moved, many years before, to a version of the anti-foundationalist epistemological position that Quine later expounds (same genus, very different species). It is *Quine* who in 1943 was "hopeful of the possibility of clarification of the epistemological problem," whereas Carnap was not. Quine, not Carnap, still hoped in 1943 for some sort of 'epistemologically immediate' material, which could be used as an Archimedean point in the analysis of knowledge. Thus Quine's brief history of empiricism found in "Epistemology Naturalized" can perhaps be read as an *autobiographical* history leading up to 1969, not the story of empiricism from Hume through Carnap, finding its consummation in Quine. However, it should be noted that, within Carnap's finitist-nominalist discussion notes, *Verständlichkeit* is not an explicitly epistemological (as opposed to semantic) concept[35], and it is not enlisted there to serve traditional epistemological purposes, either by Quine or the others.

In response to the above letter, Carnap writes that, for himself, 'the distinguishing feature of analytic truth' is unequivocally *not* its epistemological immediacy, but rather its independence from any contingent facts (Creath 1990, 308). Quine, in reply, concedes the point (Creath 311, 336). However, Quine could have avoided his mistake if he had paid closer attention to what he had already recognized in the first letter: in 1943, Carnap does not think that there is any clear traditional 'problem of epistemology' to be solved, so Carnap would likely not be interested in

[35] The closest *verständlich* comes to having an epistemic aspect is in the final conversation, where Carnap suggests that perhaps the order of *learning* of concepts in children may reflect the order of *intelligibility* of those concepts.

attempting to identify 'epistemologically immediate' or otherwise foundational items of knowledge. In sum, though Quine and Carnap disagreed in the 1940s about the possibility of well-posed epistemological questions from the standpoint of scientific philosophy, both agreed that '*Verständlichkeit*' had not been given an exact characterization, despite the central role that term plays in their Harvard discussions. And while Quine (after the fact, at least) treats *Verständlichkeit* as (primarily) an epistemological concept, Carnap tends to think of it as semantic-pragmatic.

2. SECOND RATIONALE FOR THE FN RESTRICTIONS: OVERCOMING METAPHYSICS

Another rationale for pursuing the finitist-nominalist project that crops up—both implicitly and explicitly—in the Harvard discussions is the desire to purge (cognitively significant) discourse of metaphysics. It is well known that the logical empiricists and their allies (e.g. Russell and Wittgenstein) hold a very negative view of metaphysics. The group of Polish philosophers from which Tarski came, the Lvov-Warsaw School, also shared this anti-metaphysical animus to some degree (Simons 1993, Wolenski 1993), though as a group, they tended to be neither as fervently (*pace* Chwistek) nor as unanimously anti-metaphysical as their Viennese contemporaries.[36] The impetus to eliminate metaphysics was shared by many analytic philosophers in the early 20th Century, but it took varying forms; I will detail some of these forms later, in chapter VI.

The anti-metaphysical drive is closely connected to the notion of *Verständlichkeit* discussed in the previous section. One characterization of metaphysics that is widespread among

[36] Thanks to Paolo Mancosu for correcting my earlier misrepresentations of these Polish philosophers.

the logical empiricists and their intellectual kin is the following: if a string of symbolic marks *x* is metaphysical, then *x* is *meaningless*.[37] (The converse does not hold: the string ')*y*PQ))', which is meaningless in standard formalizations of predicate logic, is not metaphysics.) And presumably, if a given word or sentence is meaningless, then it is not intelligible, not understandable. The connection to the finitist-nominalist project is clear: by *modus tollens*, if every word and every sentence in an interpreted language is 'fully understandable,' then there are no metaphysical words or sentences in that language. This argument is never explicitly articulated in the discussion notes; in particular, the conditional 'If *x* is meaningless, then *x* is not understandable' never appears. Nonetheless, given that that conditional seems patently true (how could one understand nonsense?), it seems reasonable to connect Carnap, Tarski and Quine's discussions of intelligibility in this way to their shared aversion to metaphysics *qua* cognitively meaningless utterances and inscriptions. And if the central claim of section 2.C is correct, i.e. 'intelligible' should be understood as the pragmatic correlate of 'meaningful,' then the goal of constructing an intelligible language is identical to the goal of constructing language free of metaphysics. In short, given the unintelligibility of meaningless discourse, a fully intelligible language would also be a language free from metaphysical impurities—and it is not unreasonable to hold that such a connection was at least implicit in the minds of the Cambridge discussants.

But Carnap's notes from the discussions of 1940-41 contain more than implicit attacks on metaphysics. There are explicit references to (objectionable) metaphysical theses as well. Tarski and Quine hold that the adoption of FN1 and FN2 would prevent a pernicious slide into a certain kind of metaphysics, which they call 'Platonism,' after the grandfather of all

[37] Precisely this characterization is found in Carnap's "Overcoming Metaphysics through the Logical Analysis of Language,' but the same idea is very clearly set forth in the *Tractatus*, as well as in many of Schlick's and Neurath's writings.

metaphysicians. Recall that Tarski labels FN1 (the requirement that variables may only range over the individual domain, not over classes or properties) the 'non-Platonic' requirement in his first articulation of his proposal (090-16-28). The participants do not offer a detailed or precise characterization of this term; but it involves at least higher-order logic and/or (transfinite) set theory.[38] (FN1 rules out higher-order logic, and adding FN2 rules out (first-order) set theory.) For example, we find Tarski saying:

It would be a wish and a guess that the whole general set theory, as beautiful as it is, will disappear in the future. With the higher levels, Platonism begins. The tendencies of Chwistek and others ("nominalism") to talk only about designatable [*bezeichenbarem*] things are healthy. (090-16-09)

(We shall leave aside Tarski's provocative claim that set theory might be completely overthrown someday, since it is irrelevant to our present discussion of the meaning of 'Platonic' in the notes.) And even earlier, in a discussion with Russell and Carnap, Tarski asserts: "A *Platonism* underlies the higher functional calculus (and so the use of predicate variables, especially higher levels)" (102-63-09).

In a December 1940 lecture at Harvard (and thus before Tarski introduces FN1-3), Quine distinguishes mathematics from logic as follows: "'logic' = theory of joint denial and quantification," while "'mathematics' = (Logic +) theory of $\in$." Quine then goes on to say that "*mathematics is Platonic, logic is not*" (102-63-04). Why should the set-membership relation introduce Platonic commitments? Quine explains that "there are no logical predicates," but '$\in$' is a predicate. He then claims:

Predicates first bring *ontological* claims (not because they designate, for they are syncategorematic here, since variables never occur for them; rather:) because a predicate takes certain objects as values for the argument variable; so e.g. '$\in$' demands classes, universals; thus mathematics is Platonic, logic is not" (*ibid.*).

[38] Although it does not specifically address Carnap's, Tarski's, or Quine's conception of Platonism circa 1940, (Bouveresse 2005) provides an excellent treatment of the shifting meanings of the term 'Platonism' in the early part of the Twentieth century.

That is, if there are any true statements of the form '$P \in Q$,' then there must be at least one class (provided the set-membership symbol has the intended interpretation). For Quine, accepting the existence of at least one class is tantamount to accepting Platonism. This position is stronger than the one he published a year before ("Designation and Existence," 1939), for there Quine asserts that a nominalist could hold '$P \in Q$' to be true, *provided* the nominalist does not quantify (ineliminably) over the *Q*-position. And later in the same lecture, Quine asserts that higher mathematics is based on "a myth," for the axioms of set theory are "not univocally determined" by "familiar *common sense* results for finite classes, parallel to *common sense* laws about heaps" (*ibid.*)[39] (Quine's conception of the relationship between 'myth' and 'metaphysics' is not clear; at the very least, 'myth' is not a term of approbation, epistemic or otherwise.) And Quine harbored these suspicions of set theory even before Tarski proposes constructing a finitist-nominalist language. In short, (first-order) set theory is Platonic, along this line of thinking, because it forces us to admit the existence of classes.

So why do Tarski and Quine also suspect higher-order logic of being metaphysics—even when the domain of discourse consists solely of (concrete) individuals? In Quine's May 1943 letter to Carnap, reflecting on the Harvard discussions, we find:

> I argued, supported by Tarski, that there remains a kernel of technical meaning in the old controversy about [the] reality or irreality of universals, and that in this respect we find ourselves on the side of the Platonists insofar as we hold to the full non-finitistic logic. Such an orientation seems unsatisfactory as an end-point in philosophical analysis, given the hard-headed, anti-mystical temper which all of us share; … So here again we found ourselves envisaging a finitistic constitution system. (Creath 1990, 295)

Presumably, the 'kernel of technical meaning in the old controversy' is composed of two decisions: (i.) whether (contra FN1) to allow non-concrete individuals into the domain of quantification (as discussed in the previous paragraph), and (ii.) whether (contra FN2) to adopt a

[39] In his notes after this sentence, Carnap has inserted an exclamation-point within parentheses; usually, this expresses surprise on Carnap's part.

higher-order logic. For Quine, by this point in his career, a commitment to higher-order logics brings in its wake a commitment to the 'reality of universals.' Why? In his "Designation and Existence," published a year before the Cambridge discussions, Quine had written his famous dictum "To be is to be the value of a variable" (1939, 708). In that same article, he uses this dictum to characterize nominalism within the framework of modern logic: a language is nominalist if its variables do not range over any abstract objects.[40] And properties and relations, which are quantified over in second-order logics, are (for Quine and many others) paradigmatically abstract entities.[41] In short, a language is metaphysical if it quantifies over abstract entities; in first-order set theory, those abstracta are sets, and in higher-order logic, those abstracta are relations.[42]

Taking a wider historical view, Quine has transformed the old issue of nominalism into a form acceptable to logical empiricists and their allies: "The nominalist… claims that a language adequate to all scientific purposes can be framed in such a way that its variables admit only concrete objects, individuals, as values" (*ibid.*). That is, what was previously seen as a metaphysical question ('Are universals real?') is transformed into a logico-linguistic question: 'Is a certain type of formalized language rich enough to capture the content of scientific discourse?' This shows very clearly that whatever differences Quine might have with Carnap at this time, Quine is fully on board with the basic research program Carnap trumpets in the *Aufbau*, "Overcoming Metaphysics through the Logical Analysis of Language," and later works;

---

[40] Quine writes: "In realistic languages, variables admit abstract entities as values; in nominalistic languages they do not" (1939, 708).

[41] Quine does not tell us where or how to draw the line between concrete and abstract entities (1939, 708). Also, he does not appear in "Designation and Existence" to hold that all abstract entities correspond to relation symbols in a formalized language; that is, Quine appears to leave open the possibility that abstracta be part of the domain of individuals.

[42] For more published remarks on Quine's conception of nominalism, Platonism, classes, and relations, see "On Universals" *JSL* **12** 1947, 74-84, and "Notes on Existence and Necessity," *J. Phil.* **40** 1943, 113-127, esp. section 5.

viz., a central task of modern scientific philosophy is to transform metaphysical (pseudo-)questions into well-posed logico-linguistic ones.

What is missing from both the discussion notes of 1940-41, as well as from writings before and after that time, is an explanation of *why* admitting abstracta (whether they be sets, relations, or anything else) as values of variables constitutes objectionable Platonism for Tarski and Quine. This becomes more troubling when we note that labeling higher-order logic and/or set theory as metaphysics does not mesh well with the characterization of metaphysics offered elsewhere by the logical empiricists and their allies. Both the explicatum and the explanandum of the term 'metaphysics' vary over time and between different thinkers, of course; some of these differences will be detailed in chapter VI. Nonetheless, most logical empiricists, most of the time, strongly resist classifying logic and mathematics as metaphysical. (It is sign of this that special exceptions are made in their accounts of meaning and knowledge to account for logic and mathematics; e.g., Wittgenstein's distinction in the *Tractatus* between pseudo-propositions that are nonsense [*unsinning*] and those that are senseless [*sinnlos*] places logic and mathematics in a separate category from metaphysics, even though both 'say nothing about the world.') So not only do Tarski and Quine omit an explicit explanation of why classes and relations are metaphysical, but such a view appears to clash with the view of metaphysics presented by many of their philosophical peers. However, Russell is an exception. For him, classes are "fictions of fictions"; recall his talk (I.4.B) of "these queer things called numbers," etc.

Furthermore, it may very well be that there is no explanation to give, and that the proponents of this view even recognize that fact. Thus, as mentioned earlier, in Quine and Goodman's published paper that they acknowledge is an outgrowth of the 1941 discussions, Goodman and Quine admit that their "refusal [to countenance abstracta] is based on a philosophical intuition that cannot be justified by appeal to anything more ultimate" (1947, 105).

So, the rejection of abstracta is not based on a more fundamental (or even articulated) theory of knowledge and/ or meaning that declares abstracta unknowable and/ or meaningless. Of course, we should not assume they speak for Tarski as well, but if Tarski did think that he could 'justify' his rejection of abstracta 'by appeal to anything more ultimate,' that justification is not recorded in the Harvard discussion notes. And if he did provide such a justification during these discussions, it did not impress Goodman and Quine enough to include it in their article. We today, looking back, could impute to them a causal theory of knowledge or meaning—perhaps even just as an implicit, unarticulated assumption—but such an interpretation would be conjectural.

3. THIRD RATIONALE: INFERENTIAL SAFETY/ TAKING RUSSELL'S PARADOX SERIOUSLY

The next justification for the finitist-nominalist restrictions I will discuss does not appear in Carnap's dictation notes; however, Carnap as well as Quine and Goodman mention this justification elsewhere, so I consider it here. The basic idea is that Russell's paradox reveals that certain types of logics suffer serious problems, and therefore such logics should be avoided. Differences arise, however, over the scope of this class of logics: Quine and Goodman consider the class of suspicious logics to be wider than Carnap does. In the Goodman and Quine 1947 paper, quoted immediately above, we find an expression of their argument.

> Why do we refuse to admit the abstract objects that mathematics needs? … What seems to be the most natural principle for abstracting classes or properties leads to paradoxes. Escape from these paradoxes can apparently be effected only by recourse to alternative rules *whose artificiality and arbitrariness arouse suspicion that we are lost in a world of make-believe*. (1947, 105, my italics)

And presumably, the (supposed) inhabitants of a 'world of make-believe' simply do not exist. Their argument can, I think, be cast as follows: if we admit quantification over classes and/or relations into our logic, then we can have either a 'natural' logic that leads to inconsistencies, or an 'artificial' logic that avoids inconsistencies in an *ad hoc* manner.[43] But neither a natural but inconsistent logic nor an artificial but consistent logic is particularly desirable. So Goodman and Quine recommend we no longer allow classes and relations as the values of variables. A similar idea appears in a letter written to Carnap in 1947, in which Quine claims that Platonism is likely responsible for the logical paradoxes.

> I agree that the logical antinomies are symptoms of a fundamental unsoundness somewhere, but I suspect that this unsoundness lies in platonism itself – i.e., in the admission of abstract values of bindable variables. (Creath 1990, 409)

Here, again, Quine asserts that the real lesson of Russell's paradox is that we should give up quantifying over abstracta. Quine was not alone: Paul Bernays expounded a comparable view several years earlier.

> Several mathematicians and philosophers interpret the methods of Platonism in the sense of conceptual realism, postulating the existence of a world of ideal objects containing all the objects and relations of mathematics. It is this absolute Platonism which has been shown untenable by the antinomies. (1935/1983, 261)

(It should be noted that Bernays, unlike the Quine of 1947, believes that a "moderate Platonism" can survive the paradoxes.)

Carnap can muster some sympathy for this impulse, but his response to Russell's paradox is not nearly as drastic. His basic idea is that, *ceteris paribus*, if one language's rules of inference and/ or axioms are a proper subset of another language's, then the first is more likely

---

[43] In "On Universals" (1947), Quine reiterates the charge of *ad hoc*-ness against the type-theoretic formulation of mathematics current at his time (in particular, he points to (Tarski 1933/1983, 279-295): "It is as clear a formulation of the foundations of mathematics as we have. But it is platonistic. And it is an *ad hoc* structure which pretends to no intuitive basis. If any considerations were originally felt to justify the binding of schematic predicate letters, Russell's paradox was their *reductio ad absurdum.* The subsequent // superimposition of a theory of types is an artificial means of restituting the system in its main lines merely as a system, divorced from any consideration of intuitive foundation." (80-81)

than the second to be consistent. And for some purposes, inferential safety is of paramount value, trumping deductive and/or expressive power. In his autobiography, Carnap writes: "It is true that certain procedures, e.g. those admitted by constructivism or intuitionism, are safer than others. Therefore it is advisable to apply those procedures as far as possible," though we do lose inferential strength by restricting ourselves to those means alone (1963, 49). Thus we can interpret Carnap as attempting to discover, in the 1941 discussions, the limits of a language in which only (ultra-)constructivist procedures are applied.[44] But his view is clearly different from Quine and Goodman's. For immediately following the above quotation, Carnap writes:

> However, there are other forms and methods which, though less safe because we do not have a proof of their consistency, appear to be practically indispensable for physics. In such a case there seems to be no good reason for prohibiting these procedures so long as no contradictions have been found. (1963, 49)

So whereas for Quine, Russell's paradox casts doubt upon any logic that quantifies over abstracta, Carnap is willing to use any logic that has not been shown inconsistent. For languages constructed to avoid the logical paradoxes, Carnap must either not consider them to be 'artificial' as Quine does, or else he does not consider artificiality a fatal flaw of such languages. Actually, Carnap would probably agree with both disjuncts of the previous statement: Carnap explicitly claims that the type restrictions are natural,[45] and his Principle of Tolerance would allow languages that feel intuitively 'artificial' as theoretically acceptable. So, in short, Carnap recognizes that weaker languages enjoy the advantage of being safer, insofar as they are less

---

[44] In Carnap's autobiographical recollection of the Cambridge discussions, he writes: "We further agreed that for the basic language the requirements of finitism and constructivism should be fulfilled in some sense" (1963, 79).

[45] In a discussion with Tarski about the comparative advantages of *Principia Mathematica*-style logics and set theory, Carnap says: "The levels appear to me completely natural and understandable [*verständlich*]; and to a certain extent, stratification too" (RCC 090-16-26, p.3). And in his later logic textbook, Carnap writes: "A language with no type distinctions… seems unnatural with regard to non-logical sentences. For since in such a language a type-differentiation is also omitted for descriptive signs, formulas turn up that can claim admission into the language as meaningful sentences that have verbal counterparts as follows: 'The number 5 is blue,' 'The relation of friendship weighs three pounds'" (1954/1958, 84).

likely to engender contradictions or lead from true premises to false conclusions, but he does not think the logical antinomies cast aspersions on every language that quantifies (ineliminably) over classes or relations, as Quine does. An analogy may be helpful here. When a scientific theory encounters robust data at odds with that theory's predictions, two options present themselves: reject the theory, or make an *ad hoc* modification in order to save it. The anomalous data are analogous to the paradoxes; Quine's response is closer to the first option, Carnap's to the second.

4. FOURTH RATIONALE FOR THE FN RESTRICTIONS: NATURAL SCIENCE

Another justification that Tarski and Quine offer for pursuing the finitist-nominalist project could be called 'arguments from natural science.' The previous rationales all supported (FN 1-2) (viz. the language is first-order and its domain contains only physical objects), which we could consider support for *nominalism*; the following, however, is only a justification for *finitism* (FN 3). Tarski begins with a reasonable assertion: the number of individuals in our world "is perhaps in fact finite" (090-16-25). (Tarski is following one of Hilbert's justifications, in "On the Infinite," for his very different type of finitism provided) If the universe does only contain finitely many physical things, and if (FN 1-2) hold, then it follows that D has finitely many members—and this is the restrictive version of (FN 3). If we wish rather to leave open the possibility that a finite number of physical things exist, and we accept (FN 1-2), then the liberal version of (FN 3) follows. Note that if one does not accept (FN 1-2) then (FN 3) becomes much more contentious. As explained previously, (FN 1-2) prevent the two most common ways of introducing mathematical objects into a language, and mathematical infinities are usually paradigmatic examples of infinite totalities.

Carnap replies to Tarski's claim by suggesting that there *are* infinities. These come in two varieties: logico-mathematical and physical. The usual mathematical infinities will directly violate the nominalist criteria. As one, less obvious example of the logico-mathematical kind of infinity, Carnap points to "pure locations, instead of things" (090-16-25), which he introduced and used in *Logical Syntax* as an interpretation of PSI and PSII. But pure locations are strange beasts: for they are not *physical* positions (those are not introduced until §50 of *Logical Syntax*), but they are not merely numbers either, even though (according to Carnap in *Logical Syntax*) they satisfy the Peano axioms. Commentators today still cannot agree on what exactly 'pure locations' are and how best to understand them (Lavers 2004, 304-307). Returning to 1941, Tarski responds to Carnap's suggestion by saying that the idea of pure locations "in *Syntax* made a great impression on him at the time, but he thinks there are still difficulties with it" (090-16-25, p.2). What these 'difficulties' might be is not further articulated in the discussion notes; in any case, Carnap's suggestion fails to reverse Tarski's original finitist convictions.

As examples of empirical, physical infinities, Carnap offers space and, with more conviction, time. He claims:

> even if the number of subatomic particles is finite, nonetheless the number of events can be assumed to be infinite (not just the number of time-points…but the number of time-points a unit distance away from each other, in other words: infinite length of time.)" (-24).

Carnap's suggestion to use events or spatiotemporal intervals instead of physical objects for the domain of a language of science obviously violates the letter of the law of (FN 2), but Carnap likely believes it does not violate its spirit—for spatiotemporal events are still part of the natural, physical world, unlike numbers and their ilk. (Kotarbinski, whom Tarski invokes when he proposes (FN 2), however, explicitly denied that *events* are acceptable for the reist (1929/1966, 432).) So, Carnap is suggesting, if we expand (FN2) to allow the domain to be not just physical objects but rather any entity that is (broadly speaking) part of the physical world, then (FN 3) does not force itself upon us—provided there are an infinite number of events.

Tarski responds to Carnap's challenge in two related ways. The first engages Carnap on his own terms; the second suggests that Carnap's critique has missed the fundamental point of introducing (FN 3). First, Tarski replies directly to Carnap's suggestion that space and time will provide us with infinities, even if there are only a finite number of physical objects in the universe. Tarski asserts that space and time, contrary to appearances, may actually be finite: "perhaps quantum theory will give up continuity and density" for both space and time by quantizing both quantities. Furthermore, Tarski says, time and space could both be circular, in which case there would not be an infinite number of finite spatial or temporal intervals. In short, Tarski claims that developments in quantum and relativistic physics may in fact show that space and time are actually finite.

Second, Tarski suggests that arguing that there is in fact an actually infinite quantity somewhere in nature misunderstands the motivation behind (FN 3), at least in its liberal version. Presumably (though Tarski does not state this completely explicitly) we should not assume the number of physical things in the world is infinite, because this is presumably an *empirical* matter. What Tarski does say is the following: "we want to build the structure of the language so that this possibility [viz. that the number of things is finite] is not excluded from the beginning" (090-16-23). The basic idea is simple: the form of the language we use to describe the empirical world should not prejudge the number of entities in the universe, and Tarski's scheme leaves this question open. Put otherwise, 'How many spatial positions (or 'temporal intervals') are there?' is just as empirical a question as 'How many subatomic particles are there?' If one accepts (FN 2), and if one also wishes to incorporate (first-order) arithmetic into one's language (as e.g. Carnap does in Languages I and II in *Logical Syntax*), then one would be committed to an infinite number of physical objects. To put the matter in Carnapian terms: how many entities there are in the universe—as well as the topological structure of (actual) space and time—are *synthetic* matters, and Tarski's recommendation of (FN 3) prevents them from becoming analytic ones. That is, questions about the number of things in the universe or about the structure of space and time should be determined by the structure of the world, not by the structure of the language used for science.

But, one may ask on Carnap's behalf, how exactly would allowing 'infinite arithmetic' (S_2) exclude the possibility of circular time from the beginning? Why can't we have an infinite arithmetic, and simultaneously believe that time and space are circular (or otherwise finite)? I believe a Tarskian could reply as follows. If one accepts that

(i.) numerals must be interpreted by broadly physical entities of one sort or another, and

(ii.) finite temporal and spatial intervals are broadly physical entities, and

(iii.) the only live candidates for infinite collections of physical entities are the temporal or spatial intervals (so we assume e.g. that there are only finitely many particles),

then admitting infinite arithmetic *does* force one to admit that either time or space cannot be finite. So, if Carnap truly *needs* non-circular time or non-spherical space in order to make the axiom of infinity true, then positing the axiom of infinity as part of the basic language of science does 'rule out from the beginning' the possibility that both space and time are finite.

5. CURRENT JUSTIFICATIONS FOR NOMINALIST PROJECTS

Current justifications for undertaking nominalist projects usually take one of two forms: an argument from (some version of) a causal theory of knowledge and/ or reference, and a desire to refute the so-called 'indispensability argument' for mathematical entities and theorems. The first is a positive argument for nominalism, the second is a negative argument against a popular objection to nominalism.

A. The positive argument: from a causal theory of knowledge and/or reference

A (perhaps ‘The’) modern argument for nominalism can be cast as a simple syllogism, whose major premise is a concise statement of the causal theory of knowledge.

P1. We can only have knowledge of things causally related (or relatable) to us.

P2. Numbers and other abstracta are not causally related (or relatable) to us.

Therefore, we cannot have knowledge of numbers or other abstracta.

(This argument is neutral with respect to the ontological question of whether abstracta exist or not.) Both premises have been challenged by various philosophers. More criticism has been leveled at the first, presumably because many philosophers consider a defining feature of an abstract object to be its ‘standing outside’ the causal order. I will not discuss these objections; an excellent treatment of the objection-and-replies dialectic can be found in (Burgess and Rosen 1996, chapter 1).

Another variant of this syllogism replaces P1 with a statement of the causal theory of reference:

$P1_R$. We can only *successfully refer to* things causally related (or relatable) to us.

The conclusion is modified accordingly: we cannot refer to abstracta. And presumably we cannot say much of significance about items to which we cannot successfully refer. In general, both the causal and referential forms of this syllogism seem not to have won many converts to the nominalist cause—at least in part because causal theories of knowledge and reference are not terribly popular nowadays.

Causal theories of knowledge and reference did not appear in an explicit, fully-fledged form until the 1960s and 70s, so it is not at all surprising that Carnap's 1940-41 notes do not contain explicit statements of the views expressed in P1 and $P1_R$. However, Tarski et al. would likely not explicitly *deny* (in 1940) that we can only know about or refer to entities that are somehow causally connected or connectible with us. After all, if something is a physical object, then (with some exceptions[46]) it *is* causally connectible to us; and if something is an abstract object, then it is not causally connectible to us (unless one holds, with Maddy, that when we see a pack of cards, we are in causal contact with a set of cardinality 52). Though a causal theory of reference does not appear in Carnap's discussion notes, a related notion may be at work implicitly in the collaborators' minds. Semantics is customarily described as the study of the relationship between language and world, and this is certainly how Carnap and Tarski conceived of it. If a semantic theory is to live up to its 'job description,' then it makes good sense to use, as the individuals in the universe of discourse of an interpreted language, objects that unequivocally belong to the world. If one believes that mathematical truth is, at bottom, essentially truth in virtue of language—as Carnap certainly did—then allowing the domain of discourse to contain mathematical objects would create a semantics that did not study the language-world relationship, but rather a language-language relationship. And even if one resists the Carnapian view that mathematical truth is linguistic truth, the fact remains that it is not clear that a language whose domain of discourse is the natural numbers can be used to make claims 'about the world,' so the language-world relation would be absent.

[46] For example, the laws of physics prohibit me from making contact with events outside my past light-cone—so, on the causal theory of reference, I cannot (successfully) refer to my breakfast tomorrow.

B. The negative argument: Rebutting the 'indispensability argument'

Shortly after the 1947 *JSL* paper appeared, Quine rejected nominalism. (Goodman did not.) Hilary Putnam and the post-nominalist Quine argued for the existence of mathematical abstracta on the grounds that relinquishing such abstracta would force us to relinquish much of modern science. We should be unwilling to pay that price for maintaining nominalist scruples. Current nominalist projects, such as Hartry Field's seminal *Science without Numbers* (1980), usually consist of 'reconstructive' projects that attempt to rebut the indispensability argument. In such a nominalist project, a certain field of natural science is recast in a form that does not appeal to any 'abstract' entities. Field claims that if empirical science can be reconstructed nominalistically, then belief in mathematical objects becomes "unjustifiable dogma." (1980, 9) The literature on the indispensability argument is vast, and I will not comment upon its merits. The only point I wish to stress is that one common justification or motivation for undertaking a technical nominalistic project today is to rebut the indispensability argument. By constructing a scientific theory that does not quantify over numbers, the modern nominalist shows that numbers are, in principle if not in practice, dispensable for that theory.

Given that the indispensability argument did not appear as such until sometime in the late 1960s or early 1970s, an explicit desire to rebut it cannot be a motivation for undertaking the 1941 project. However, there *is* a precursor of the modern indispensability argument in Carnap's notes. We encountered it above in I.2, while discussing the lower bound on the poverty of a finitist-nominalist language's expressive power; the relevant section is reproduced here:

"If S_1 [the finitist-nominalist language] does not suffice to reach classical mathematics, couldn't one perhaps nevertheless adopt S_1 and perhaps show that classical mathematics is not really necessary for the application of science to life? Perhaps we can set up, on the basis of S_1, a calculus for a fragment of mathematics that suffices for all practical purposes (i.e. not something just for everyday purposes, but also for the most complicated tasks of technology). (090-16-25)

This is not precisely the program Hartry Field (re-)inaugurated, but it is similar: in both cases, the aim is to show that a proper subset of modern mathematics is sufficient for all applications of mathematics in science. This leads us to an interesting question: what is the relationship between the modern indispensability argument and the Tarski-Carnap-Quine demand that their 'understandable' language be sufficient to express (at least a substantial portion of) mathematics and natural science? To answer this question, we need an explicit statement of the indispensability argument. One current formulation, due to Colyvan, is the following:

1. We ought to have ontological commitment to all and only those entities that are indispensable to our best scientific theories.

2. Mathematical entities are indispensable to our best scientific theories.

Therefore:

3. We ought to have ontological commitment to mathematical entities.
(2001, 11)

The 1941 project differs from the modern one primarily in the first premise: there is no normative claim concerning ontological beliefs explicitly forwarded in the discussion notes. Nothing in the texts decisively rules out attributing this position to the participants as an implicit belief, but this (potentially anachronistic) interpretation is certainly not forced upon us, either. Instead, we can view Carnap et al. (or a proper subset of them) as replacing normative-*cum*-ontological issues with the goal of a unified language of science (see Chapter VI). Whether failure of a language to meet that aim, regardless of its other merits, automatically disqualifies it in the discussants' eyes is not, as mentioned above, discussed in Carnap's notes. We know that Quine, a decade after the Harvard discussions, opts for disqualification: his grounds for eventually repudiating nominalism are that we cannot recover a sufficient amount of mathematics to do science if we abide by nominalist strictures. Carnap's principle of tolerance puts him in a somewhat different position: he would not 'disqualify' any language outright—he could merely say that a particular language is inexpedient for this or that purpose. And the relative value or importance of various purposes is not something about which there is (or can be) a fact of the matter. Thus one might expect that Carnap would remain ‘agnostic,’ so to speak, about languages meeting the FN conditions. However, he does not: he resists the fundamental assumptions of the project from beginning to end. The details of Carnap’s resistance form the core of the following chapter.

CHAPTER III. CRITICAL RESPONSES TO THE FINITIST-NOMINALIST PROJECT

1. WHY DOES CARNAP PARTICIPATE IN THE FN PROJECT, GIVEN HIS RESERVATIONS?

As previous chapters have shown, Carnap is an active participant in the finitist-nominalist project, in cooperation with Tarski and Quine. However, those familiar with Carnap's fundamental philosophical views might suspect that he also harbors serious reservations about it. This chapter covers Carnap's main objections to the FN conditions, and adds a Carnap/ Frege-inspired one of my own. Given that Carnap is not convinced of the merits or value of Tarski's proposed restrictions, and resists accepting them wholeheartedly, the question naturally arises: why *does* Carnap engage in this project? Carnap's participation appears to be more than politely humoring his respected colleagues. For not only does he discuss this topic repeatedly and at length with them, taking notes throughout, but he also works on the problems privately, in his own time, working out the (often dreary) details of formal axiom systems aiming to satisfy the FN criteria. I see at least two reasons Carnap would participate in this project, despite his skepticism toward its fundamental assumptions: (i.) the principle of tolerance, and (ii.) the possibility of assimilating Tarski's FN project to Carnap's own investigations into the relation between observational and theoretical languages.

One likely reason Carnap participates in the FN project stems from one of the most far-reaching components of his intellectual stance, namely, the principle of tolerance. The version of the principle most relevant for the issue at hand is the following: there is no one single correct language (or logic), and as a practical corollary for the working logician, different—even incompatible—logics may be developed, and their properties investigated. Carnap applies this abstract principle concretely, for he is willing to investigate, in detail, the construction and consequences of languages whose philosophical motivations or underpinnings he does not fully endorse. For example, Carnap's Language I (PSI) in *Logical Syntax* is evidence of this willingness. PSI is intended to capture formally the intuitionist stance about mathematics, a stance which Carnap himself never embraced as his own. Yet he nonetheless devotes a large chunk of *Logical Syntax* to the axiomatic articulation of such a language, and an investigation of its logical properties.[47] I suggest that Carnap, in 1941, is undertaking the same kind of endeavor, and with the same rationale, as in *Logical Syntax*. That is, he is once again attempting to construct a formal language that meets requirements he does not fully endorse *in propria persona*, and he can justify this activity by appealing, as he did in *Logical Syntax*, to the principle of tolerance.

However, merely citing the principle of tolerance does not explain why Carnap was willing to investigate the *particular* language-form Tarski proposes. The principle of tolerance merely supplies permission to study formal languages satisfying the FN strictures, but that permission applies equally to an infinite number of languages that Carnap never investigates. So we would like to find some further rationale for Carnap's engagement in the FN project that explains why, out of the infinitely many languages a tolerant *Wissenschaftslogiker* is permitted to investigate, Carnap chose to devote his energies to this one. That rationale, I believe, can be found in Section 13 of his *Autobiography*, entitled "The Theoretical Language." There, Carnap closely ties ('assimilates' may not be too strong) the 1941 FN project to work he had already begun on the relationship between the observational and theoretical parts of a scientific theory.

[47] To give a more complete picture, it should be noted that there is (at least) one other reason Carnap discusses PSI at length: it is a language in which its own syntax can be formulated. This feature of PSI both is interesting in itself, and decisively refutes then-current claims to the contrary, advocated by followers of the *Tractatus*. According to them, syntax could only be 'shown,' not 'said': "The rules of logical syntax must go without saying" (3.334)

This work has its early roots in the discussions of protocol-sentences in the early 1930s, but it assumes a more familiar and canonical form in 1936-7's "Testability and Meaning" (the first Carnapian text in which the observational/ theoretical distinction explicitly appears and does significant philosophical work) and 1939's *Foundations of Logic and Mathematics*. Here is Carnap's autobiographical, *post facto* explanation of how the 1941 project connects to work he had already done.

> In *Foundations of Logic and Mathematics*, I showed how the system of science … can be constructed as a calculus whose axioms represent the laws of the theory in question. This calculus is not directly interpreted. It is rather constructed as a "freely floating system," i.e., as a network of primitive theoretical concepts which are connected with one another by the axioms. … Eventually, some of these [theoretical concepts] are closely related to observable properties and can be interpreted by semantical rules which connect them with observables. …
>
> In subsequent years I frequently considered the problem of the possible forms of constructing such a system, and I often discussed these problems with friends. I preferred a form of construction in which the total language consists of two parts: the observation language which is presupposed as being completely understood, and the theoretical language. …
>
> My thinking on these problems received fruitful stimulation from a series of conversations which I had with Tarski and Quine during the academic year 1940-41… We considered especially the question of which form the basic language… must have in order to fulfill the requirement of complete understandability. (1963, 78-79).

For Carnap, the observation language 'is presupposed as being completely understood'; we saw a virtually identical claim in II.1.C, in the quotation from *Foundations* where Carnap characterizes 'elementary' or observational language as that which is most comprehensible to a layperson. Thus when Tarski begins talking at Harvard about the only kind of languages he 'truly understands,' and tries to 'build' as much of scientific language as he can out of such languages, Carnap links this to his own previous work on the connections between theoretical language and observational language he had developed in "Testability and Meaning" (of course, these connections are not strict biconditionals for Carnap after 1936). Assuming Carnap's memory is to be trusted in this matter,[48] he viewed Tarski's 1941 project as closely related to a project he himself had already begun a few years earlier. It is plausible that Carnap would have considered the time and effort invested in the finitist-nominalist program well worth it, if he believed it to be building upon a line of inquiry in which he was already interested and involved. If Carnap considered Tarski's proposal as a continuation of one of his own investigations, then we can understand why Carnap is not only *permitted* to work on Tarski's project (namely, tolerance), but also why he is *willing* to do so.

As a word of caution, it should be noted that neither of these rationales for working on FN languages explicitly appear in Carnap's discussion notes. However, this omission is not terribly surprising: why would Carnap want to discuss them with Tarski and Quine (would Tarski need a reason to become more interested in his *own* project?), or bother making a note of his own motivations to himself? Also, one can reasonably doubt that Tarski and Quine saw the FN project as an instance or extension of Carnap's investigations into the relation between the observational and theoretical parts of a scientific theory. Their proposals to make the domain of discourse subatomic particles or quanta of energy show that they are not attempting to restrict the intelligible assertions to those expressible using *observational* vocabulary alone. By contrast, recall that in *Foundations* (see II.1.C), Carnap identifies the observational portion of scientific language as that part which is (most) understandable to a layperson. So Tarski and Quine are not likely to see their goal as giving a partial interpretation of scientific theories in an observational language, since for them, unlike Carnap, 'observational' and 'understandable' are not coextensive predicates. And, as I will explain in the next section, if a partial interpretation were sufficient to render the *whole* language meaningful, then Tarski (and Quine) would not dispute

[48] It is at least possible that Carnap imposes this synthesis of his own observational/ theoretical work with Tarski's FN project only much later, since his intellectual autobiography is not written until more than a decade after the Harvard conversations occurred.

the meaningfulness of higher mathematical language. Let us now turn to Carnap's objections to the finitist-nominalist project.

2. ANSWERING THE FN CLAIMS: HIGHER MATHEMATICAL LANGUAGE *IS* MEANINGFUL

As noted in I.2, Tarski maintains that sentences that fail to meet the finitist-nominalist conditions are nothing more than empty formalism. As such they are to be treated as part of a mere calculus, which can be manipulated according to a set of rules but never given a genuine, philosophically acceptable interpretation, i.e., a meaning. Parts of classical mathematics are thereby classed as meaningless; the participants usually call such statements 'higher' mathematics, and I will follow their terminology. Carnap's intuitions, however, run strongly against considering (at least some of) higher mathematics meaningless. (In the discussion notes, after the FN criteria are proposed, there is no record of Quine explicitly siding with Tarski; however, his 1947 paper with Goodman unequivocally endorses Tarski's position. Quine's December 1940 lecture is leaning toward Tarski's viewpoint, but the two are not identical.) The discussion notes do not contain much material in which Carnap defends the meaningfulness of higher mathematics, but he does offer at least two (unfortunately brief) rebuttals. The first argument for the intelligibility of non-finitist arithmetic appeals to an analogy between claims in higher mathematics and in theoretical physics; the second suggests that invoking a *potential* infinity could make classical arithmetic intelligible.

A. Analogy between higher mathematics and theoretical physics

As mentioned in II.1.A, in December 1940, Quine gives a far-reaching lecture entitled (in Carnap's notes) "Logic, Mathematics, Science." In it, he claims that higher mathematics is 'Platonic,' and that science more generally is full of "myths" (see II.3 above). Carnap responds[49] to this charge as follows (this is the entirety of the document containing Carnap's response):

Dec. 20, 1940.
Quine is discussed.
Can we perhaps conceive of the higher, non-finitistic parts of logic (mathematics) thus: its relation to the finitistic parts is analogous to the relation of the higher parts of physics to the observation sentences? Thereby non-finitistic logic (mathematics) would become non-metaphysical (like physics). Perhaps also light is thereby thrown on the question, whether a fundamental difference between logic-mathematics and physics exists.
(RCC 090-16-29)

What does this quotation, couched in terms of 'metaphysics,' have to do with defending the meaningfulness of mathematics? For logical empiricists and their allies at this time, the following equivalence holds: an apparently meaningful string of symbols is nonsense (or 'meaningless') if and only if that string is metaphysics. (This equivalence, clearly expressed in the *Tractatus* and elsewhere, will be discussed at much greater length in chapter VI.) Furthermore, if a given string of symbols is *not understandable*, then it seems reasonable to hold that that string is meaningless or nonsense—especially since Tarski declares calculi that cannot be given a meaningful interpretation (i.e., an interpretation meeting FN 1-4) unintelligible.[50] Thus if Carnap can successfully show that non-finitistic mathematics is 'non-metaphysical,' he thereby shows that it is meaningful, and hence understandable.

---

[49] It is not clear whether Carnap's response was public or private. The heading of the note reads "Quine is discussed," but no interlocutors appear in this note of Carnap's—not even Carnap himself. Usually, when recording discussions, Carnap attributes claims to one person or another; ordinarily, the only occasions in these notes in which Carnap does not mention speakers is when he writes private notes for himself.

[50] It is inconsequential for the present argument, but one might wonder whether the converse direction holds (If x is meaningful, then x is intelligible). The characterization of understanding in *Introduction to Semantics* (1942, 22; quoted in II.1.C above) suggests that it does, so there is no substantive difference between 'meaningful' and 'intelligible.' (There is, of course, a difference between 'meaningful' and 'understood,' since there are plenty of meaningful English expressions that I do not understand.)

So, given this connection between metaphysics, meaningfulness, and intelligibility, what is Carnap's argument for the intelligibility of mathematics? He offers an argument by analogy; in Carnap's words, the analogy is:

observation sentences : higher parts of physics :: finitistic math : higher, non-finitistic math

What, concretely, does this schematization express? I take Carnap as conjecturing that the relationship between observation sentences and (e.g.) Einstein's field equations is sufficiently similar in the relevant respects (whatever those might be) to the relation between the statements of elementary (finitist) arithmetic and Zermelo-Fraenkel set theory, or even classical arithmetic as expressed in Peano's axiomatization. That is,

report from Eddington's eclipse expedition : EFE :: 2+5=7 : Peano arithmetic (or ZF)

If Carnap's conjecture is correct (and if I have interpreted him correctly), then to reach his desired conclusion that higher mathematics is not metaphysics, Carnap would only need to make the assumption that Einstein's field equations are not metaphysical. What should we make of this argument? First, it is not clear that Quine and/ or Tarski would grant that Einstein's field equations and the other fundamental laws of physics are *completely* or *unequivocally* non-metaphysical—for Quine at least claims in his December lecture that "science is full of myth and hypostasis"; in that declaration, Quine appears to include the laws of fundamental physics.

Second, is the analogy a good one? There are obvious dissimilarities between the two cases. For example: what form of 'finitistic' mathematical statements would be most similar to observation sentences (for relevant purposes)? Is '2+5=7' sufficiently like the reports that came back from Eddington's eclipse expedition? The latter are spatiotemporally specific, the former are not: as philosophers since Plato have observed, we do not say '2+5=7 at 3:30 PM EST, January 21 2005.' So might '2+5=7' be more analogous to phenomenological laws of natural science? However, this difference (like several others we could point to)[51] does not appear to be relevant to Carnap's argument that higher mathematics is meaningful and hence non-metaphysical. Perhaps all that Carnap needs to draw from the case of physics is the following:

1. Observation sentences are uncontroversially meaningful.

2. Sentences expressing substantive scientific theories, such as Einstein's field equations, stand in substantive inferential relations (though not equivalences, post-"Testability and Meaning") with these meaningful observation sentences.

3. These inferential relations make Einstein's field equations 'meaningful by association.' (In general, if a symbolic string *s* stands in a non-trivial inference relationship with a meaningful sentence, then *s* is meaningful.)

From these three premises, Carnap infers that the Einstein field equations are meaningful. Then the question is: do analogues of the above three hold in the mathematical case? Each of the following would need to hold:

1′. Sentences like '2+5=7' are uncontroversially meaningful.[52]

2′. Higher arithmetical statements, such as the Peano axioms or ZF, stand in substantive inferential relations (possibly weaker than equivalence) with finitist ones.

3′. These relations make the higher, non-finitist statements meaningful by association.

It seems to me that 1′ would be questioned by Tarski and Quine for sufficiently large numerals, that 2′ is obviously true, but premise 3′ would be contested as well. For 3′ is contrary to Tarski's claims (seen in II.1.A) that any sentence not interpreted in accordance with the FN conditions is

---

[51] For example, as Carnap points out (102-63-15), to derive an observational prediction from physical laws, *initial conditions* are necessary; however, there appears to be no analogue to initial conditions in the mathematical case, in which particular arithmetical theorems are derived from the Peano axioms. However, I cannot see how this difference would matter, pro or contra, to Carnap's claim that higher mathematics is made meaningful by the meaningfulness of lower mathematics and the inferential relations between them.

[52] Where exactly we draw the line between sentences like '2+5=7' and sentences of 'higher mathematics' is not important for the present argument—so long as we draw a line somewhere. This mirrors the situation with the observational/ theoretical distinction in philosophy of science: many who use the distinction (e.g. Carnap and van Fraassen) admit that it is vague, but allow people to draw the boundary within the class of vague cases wherever they like.

not intelligible, and is no more than a counter in an empty calculus. And we see virtually the same assertion in the Goodman and Quine paper on nominalism:

> if it [‘∀n(n+n=2n)’] cannot be translated into nominalistic language, it will in one sense be meaningless for us. But, taking that formula as a string of marks, we can determine whether it is indeed a proper formula of our object language, and what consequence-relationships it has to other formulas. We can thus handle much of classical logic and mathematics without in any further sense understanding, or granting the truth of, the formulas we are dealing with.
> (1947, 122)

However, despite Tarski, Quine, and Goodman’s (implicit) rejection of 3′, it is nonetheless a guiding assumption behind many (later) logical empiricist attempts at a criterion of meaningfulness (as we shall see in Chapter VI). Carnap, from (at least)[53] *Foundations* in 1939, argues that theoretical physics is meaningful or non-metaphysical on the grounds that it can be given a *partial* interpretation in the observational language, whose exact form is left open, but is assumed to be meaningful (see II.1.C). But one could question Carnap’s claim that a partial interpretation confers full meaningfulness or intelligibility on the entire calculus. And it seems that Quine and Tarski must not have accepted such a view, for if they did, they would not dispute the meaningfulness of higher mathematics.

On a tangential note, Carnap is not the only philosopher to suggest this analogy between variable-free formulas of arithmetic and observation reports. We also find it in Poincare’s *Science and Hypothesis*. He writes:

> We see successively that a theorem is true of the number 1, of the number 2, of the number 3, and so on—the law [which holds for an infinite number of cases] is manifest, we say, and it is so on the same ground that every physical law is true which is based on a very large but limited number of observations.
>
> It cannot escape our notice that here is a striking analogy with the usual [=natural scientific] processes of induction. // …

[53] The germ of this idea appears slightly earlier, in “Testability and Meaning.” There, Carnap claims that, for empiricists, ‘confirmable’ is closely tied to ‘(empirically) meaningful,’ and Carnap holds that incompletely confirmable claims should be admitted as scientifically acceptable. That is, we find an epistemological analogue of the semantic thesis of partial interpretation in “Testability and Meaning,” plus the view that confirmability and meaningfulness are coextensive for empiricists.

No doubt mathematical recurrent reasoning and physical inductive reasoning are based on different foundations, but they move in parallel lines and in the same direction—namely, from the particular to the general. (1905/1952, 13-14)

Poincare introduces the analogy to illustrate an important similarity between reasoning in mathematics and in physics. In both cases, we extrapolate a (infinite) generalization from a finite number of particular givens. However, Poincare holds that the justification for the amplicative inference is different in the two cases: in the physical case, we must assume that the physical world will continue to behave in the future as it has in the past, along with the rest of the skeptical worries about induction. In the mathematical case, however, we need only assume that a mathematical operation can be repeated indefinitely. Note that Poincare is here concerned with the discovery and justification of physical and mathematical laws, not (explicitly) with their meaningfulness or intelligibility, as Carnap is.

Gödel, in a conversation with Carnap on March 26 1948, also suggests that there is a substantive analogy between higher mathematics and theoretical physics.

He [Gödel] sees a strong analogy between theoretical physics and set theory. Physics is confirmed through sense-impressions; set theory is confirmed through its consequences in elementary arithmetic. The fundamental insights in arithmetic, which cannot be reduced to anything more simple, are analogous to sense-impressions.
(RCC 088-30-03)

Here again, as in Poincare's case, Gödel is concerned with the *justification* of set theory and physical theory—not with its meaningfulness per se. But presumably, if an assertion is justifiable, then it must be meaningful (though perhaps neither Poincare nor Gödel ever explicitly had this thought). And as Carnap explicitly states in "Testability and Meaning," *for empiricists*, the question 'How is a claim confirmed?' is to a first approximation the same as 'What is the meaning of a claim?' Whether Gödel would endorse this tenet of empiricism is another question, almost certainly to be answered in the negative. Finally, I will leave it to others to speculate on how the brief quotation above relates to Gödel's famous claim that "we do have something like a perception…of the objects of set theory" (1963/1983, 483-4). I will merely point out that Gödel says that the claims of elementary arithmetic are '*analogous* to sense-impressions,'—they are not a *type* of sense-impressions, and neither do they belong to a genus containing them and sense-impressions, as some of Gödel's remarks and his interpreters seem to suggest.

B. Potential infinity

On January 31 1941, after Tarski has set out his finitist-nominalist criteria for the second time, Carnap directly objects to declaring the sentences of classical arithmetic unintelligible. One rationale Carnap offers for his view relies on the concept of 'potential infinity.'

> I: It seems to me that, in a certain sense, I really understand infinite arithmetic. Let us call it language S_2: only variables for natural numbers, with operators (so also negative universal sentences) for the purpose of recursive definitions. To Tarski and Quine's question, as I take it, if the number of things is perhaps in fact finite: … I do not feel as averse toward the concept of possibility as Tarski and Quine. It seems to me that the possibility always exists of taking another step in forming the number series. Thus a potential, not an actual infinity (Tarski and Quine say: they do not understand this distinction.)
> (RCC 090-16-25)

Carnap goes on to say that he is not as convinced of the intelligibility of set theory as he is of higher arithmetic, though the difference is likely one of degree. From the very brief description above, it appears that S_2 is similar, if not identical, to PSI in *Logical Syntax*: roughly, the usual predicate calculus, but without quantifiers (free variables are used to express generality), plus (Carnap's version of) the Peano axioms. (Carnap's PSI is, fundamentally, identical to what is now called primitive recursive arithmetic.[54]) The only difference I can detect is in the choice of domain: in *Syntax*, the universe of discourse is the (pure) spatial positions, whereas in S_2 it is the natural numbers. Furthermore, that difference is not great, for Carnap takes the pure spatial positions of *Syntax* to obey the Peano axioms. At the very least, S_2 is close enough to PSI, so

[54] Jeremy Avigad pointed this out, in personal communication.

that both would contain theorems that the Tarskian finitist would either remain undecided about or deem false.

How, exactly, is Carnap's invocation of potential infinity intended to demonstrate that classical arithmetic is understandable? Perhaps Carnap has something like the following in mind: suppose we begin counting by pointing to objects in the world and 'marking them off,' one number for each object. Further suppose the number of objects in our world is some finite *n*, and in the process of our marking off objects, we arrive at the 'final' object. Carnap appears to be claiming (correctly, *prima facie*) that we could still count past the number *n*; i.e., even if we 'run out' of objects, we can continue counting unimpeded. In such a situation, nothing could stop us from proceeding to *n*+1 and beyond (with our eyes closed, perhaps): to any finite number, we could always add one and produce a new number. Our ability to understand the structure and properties of the natural numbers, Carnap believes, is independent of how many things happen to exist in our world. The understandable outruns the actual.

Such an idea has intuitive pull; how might Tarski's viewpoint be defended? First, a finitist-nominalist could suggest that any numbers that we generate greater than *n* should be regarded as similar to 'Pegasus,' 'unicorn,' and other non-denoting words. Or, less hospitably to Carnap, numbers greater than the number of things in the world should be considered as inhabiting the same philosophical (epistemological and/or semantic) boat as God, entelechy, essences, and other traditional topics of metaphysics that, on the logical empiricists' view, are mere pseudo-concepts devoid of meaning.[55] So a Tarskian could happily grant that we are able to produce numerals intended to pick out a number greater than the number of things in the universe, provided that such numerals are somehow not genuinely meaningful, e.g. by being

[55] In other words, Tarski subscribes to the Parmenidean theory of meaning: the meaning of a word is its referent, so a word without any referent has no meaning, i.e., is meaningless (and therefore, according to logical empiricists, metaphysical).

metaphysical (in a pejorative sense). The ability to generate a concept does not ensure its meaningfulness. So, the bare fact that we can generate the numbers is no guarantee of their meaningfulness or their epistemological respectability—for we can generate (in some sense) noxious metaphysical pseudo-concepts as well. How might Carnap respond? The claims of classical arithmetic, unlike metaphysical ones, (i.) are (ineliminably) used in science (see especially *Logical Syntax* §84) and (ii.) are governed by a set of rules such that there is a standard of 'checkability' for them, and as a result all competent mathematicians will agree on the truth-value of a given arithmetical claim, in contrast with the perennial wrangling of the metaphysicians (see especially 1950/1956, 218-9).

Let us consider a second finitist response to Carnap's suggestion that the notion of potential infinity will save classical arithmetic. There is a tradition in finitist writings, deriving from Hilbert, of conceiving numbers as inscribed vertical strokes (so ||| is identified with the number three).[56] Tarski, drawing on this idea, could respond to Carnap that it is *not* true that, to a finite number of inscribed strokes, we can always add another: at some point, if the material of the physical universe is finite, we will run out of 'ink,' i.e., material to draw the strokes. So if numbers simply are sets of inscribed strokes, and the 'ink' of the universe is finite, then every number no longer has a successor. Carnap would likely answer this rebuttal by denying such a 'physicalized' conception of numbers; this will be discussed in detail in section 4.A below.

Before moving on to Carnap's other responses to the finitist-nominalist project, two further remarks should be made. First, Carnap offers no explication of the distinction between a potential infinity and an actual one.[57] The distinction between *potential* [*potentiales*] and actual is likely different from the distinction between *possible* [*möglich*] and actual. Carnap seems to

[56] See (Hilbert 1925/1983, 192) and (Tait 1981, 525).

[57] Hailperin has characterized a 'potential-infinite domain' (1992) as, roughly and abstractly, a (finite) set of basic objects and a set of deterministic rules for generating further objects from the basic ones. He explicitly avoids assuming that there exists a set containing all the objects so generated.

be suggesting that in *this* world one can always take a further step in the number series—not that there could have been more subatomic particles (or whatever fundamental entity of choice) than there are in our actual world. However, Carnap's use of the term 'potential' is perhaps merely sloppiness, for just a few days later (Feb. 17, 090-16-23), he writes (in another diatribe against the Tarskian viewpoint) that "arithmetic… deals with possible, not actual, facts," and as we saw in the quotation from the notes that began the present discussion, Carnap says that he is 'not as averse to the concept of *possibility* as Tarski and Quine'—*not* the concept of *potentiality*.

Second, Carnap's appeal to possibility in order to appease the finitist-nominalist is an undeniable precursor of one of the major current attempts to reconcile nominalist scruples with modern scientific practice. Carnap's basic idea has been developed extensively, by Chihara and Hellman in particular; for an excellent survey of this work, see the chapters in (Burgess and Rosen 1996) dealing with 'modal strategies.' These viewpoints, generally speaking, adopt nominalist conditions of some kind, and also adopt modal concepts governed by some form of modal logic. Thus, for example, instead of being committed to the assertion that an infinity exists (either as an axiom or a derived theorem), a modal nominalist could content herself with the assertion that it is *possible* that an infinity exists. But of course, many philosophers who are sympathetic to the strictures of nominalism are *un*sympathetic to modal notions—just as Quine and Tarski are in 1941, as seen in the above quotation.

3. ARE THERE ANY INFINITIES COMPATIBLE WITH THE FINITIST-NOMINALIST IMPULSE?

Carnap tries repeatedly to resurrect some notion of infinity that is compatible with the spirit, if not the letter, of the finitist-nominalist conditions. Presumably, once such an infinity is at hand, we can avoid hobbling arithmetic with finitist conditions. We have already seen two instances of this: in the immediately preceding section, we found Carnap suggesting that the notion of a potential or possible infinity could perhaps be compatible with the FN criteria, and serve as a stepping stone into classical mathematics. Second, at the close of Chapter II, we saw Carnap argue that space and time are (actually) infinite, even if the number of objects in the physical universe is finite. Since spatiotemporal intervals are not (usually considered) abstract objects—they do seem different from numbers, classes, and properties—this would require a modification to the letter of FN 2 ('the domain of quantification includes physical *objects* only'), but probably not to its fundamental spirit. Tarski appears, in that section of the discussion notes, to allow that spatiotemporal intervals are sufficiently un-abstract to be considered potential candidates for the nominalist's ontology, but he denies that the number of spatiotemporal intervals must necessarily be infinite; he holds rather that the number of spatiotemporal intervals that exist is a contingent matter.

Carnap offers a third strategy for recovering infinity that attempts to respect the finitist's worry that assuming an infinite number of individuals exist is not a purely logical assumption. He suggests using *sequences* of physical objects, instead of objects themselves, to construct an *ersatz* infinity and avoid reaching the 'final number.' Thereby the dubious assumption that there are infinitely many physical things can be avoided. However, Carnap discusses this proposal within the context of a finitist-nominalist treatment of *syntax*, a topic that I have not yet addressed; correspondingly, the particular objects and sequences thereof that Carnap has in mind are *symbols* (of the object language). Thus far, I have focused primarily on the effects of adopting the FN conditions upon mathematics, and upon arithmetic in particular. However, these conditions will require restrictions in other areas as well. One such area is syntax, and Carnap, Tarski and Quine do occasionally discuss the implications of the FN criteria for a theory of syntax. I will first quote the sections of the discussion notes dealing with Carnap's proposal for syntax at length, and subsequently offer my interpretation of what Carnap is doing in them.

Carnap's proposal for a finitist-nominalist syntax first appears in private reflections on the first extended discussion about Tarski's project. He attempts to meet Tarski halfway by relaxing Tarski's restrictions somewhat, while preserving the spirit of the finitist-nominalist program.

Feb. 16, 1941
On finitistic syntax.
(Stimulated by the conversation with Tarski on finitism, Jan. 31, 1941)
Tarski thinks we ought to take as expressions, sentences, and proofs only actually written down items. But this is much too narrow. Then PM does not contain a single proof of a theorem. But we can make it finitistic nonetheless: we take as symbols only actual things, but as expressions and proofs not only certain actual spatial arrangements of these things, but rather (non-spatial) sequences of these things, designated either by the series of names of these things, separated by commas (elementary sequence expression), or designated by descriptions, e.g. as the union of two previously-described sequences, for which we have introduced abbreviations. (So sequences of things, not kinds of things; the symbols are thus only tokens…)

Example: Symbols in the object language: x y z P ( ) ~ v ∃
Their names in the metalanguage: a_1 a_2 a_3 a_4 a_5 a_6 a_7 a_8 a_9
The sequential expression 'a_4, a_5, a_1, a_6' then designates the sentence 'P(x)' even if this sentence never actually occurs (as a spatial series of four things of this type).
Problem: Is this talk of sequences whose length is greater than the number of things in the world at peace with the principle of finitism? I.e., is such a sentence understandable [*verständlich*] for the finitist? [G.F.-A.: These final two sentences appear to be later additions from Feb. 19, jotted at the bottom of the page.]
…
There are still only finitely many symbols of the object language. But we can speak of expressions whose length is greater than the number of things. E.g.
|| a_7, a_1, a_{100}, ………| 5 ||
Here I write in a sequential expression, which designates a certain very long expression of the object language. This should be the designation of the expression in the object language that 5 identical partial expressions of the written form stand next to each other.
(090-16-27)

The day after making these notes to himself, Carnap meets with Tarski and Quine and discusses his ideas about syntax with them.

Feb 17 1941
Conversation with Tarski and Quine, on finitism, II: 17.2.41

I: If we have only finitely many things, and thus finitely many names 'a,' 'b,' … 'Q,' then we can build arbitrarily long sequences:

R(a, a)
S(a, a, a)
T(a, a, a, a)
⋮

Naturally, in the same world, we cannot write down arbitrarily long sequences; but with the help of abbreviations, we can indeed talk about them. With these, we can build an unrestricted arithmetic.

…

Should we understand as expressions here only those actually written down in ink, or arbitrary conceivable [*denkbare*] sequences out of actually present things? (So that the alphabet will only need to be written down once, somewhere.)
(090-16-23)

With this extensive stretch of text under our belts, let us attempt a concise formulation of Carnap's central ideas about a finitist-nominalist theory of syntax:

(S1) Object language symbols are concrete things, i.e. tokens; (it is possible that) there are finitely many of them.

(S2) Formulae (and proofs) of the object language are ('non-spatial') sequences of object language symbols. Some formulae are physically instantiated, others are not.

(S3) Formulae that are not physically actualized in object language symbols can be referred to metalinguistically, using sequences of names of the object language symbols.

(S4) Formulae that are not and could never be physically actualized can be referred to via abbreviations.[58]

(Carnap's conjecture) Such a syntax would suffice to 'build an unrestricted arithmetic.'

This *precis* of Carnap's proposal for finitistic syntax shows clearly the sense in which he is attempting to broker a compromise with Tarski's radical program. In Carnap's proposal for syntax, the component symbols of the *object language*[59] will meet the FN criteria: they are physical things, they are tokens instead of types (and thus need not involve us in properties and/or classes), and they are finite in number. However, under Carnap's proposal, the formulae

[58] Whether the abbreviations appear in the object language, metalanguage, or both is not entirely clear. The default assumption is that the abbreviations are in the object language, but in the example Carnap gives, he uses the symbols for the metalinguistic names of object-language symbols.
[59] Calling this a 'language' may be contentious—it is not clear that one is still talking about a *language* if one cannot say 'bird' and 'bird' are the same word.

(and therefore the proofs) in the object-language consist of non-spatial sequences of these symbols—so there are sentences of the object language that do not occur anywhere in the physical universe. (The class of sentences is defined by taking all those sequences—actualized or not—that satisfy the formation rules for the language.) Nonetheless, we can refer to such non-actual sentences either by their names in the metalanguage, or by using abbreviations for them in the object language (S3-4). So, in Carnap's syntactic theory, when we speak of 'all sentences of a language,' we include items that are not concretely realized in the actual world. Furthermore, under certain circumstances, Carnap's syntax would even admit items that *could not* be concretely actualized: if the amount of material in the world is finite, there will be sentences of the language that are too long to inscribe anywhere.

If we step back to take a much wider historical view, I believe this highlights a substantial difference between Carnap's approach to the philosophical study of language and Quine's. Carnap is very far from not only the nominalist Quine of "Steps Toward a Constructive Nominalism," but also from the behavioral linguist Quine of *Word and Object*, when Carnap claims that our definition of sentence should include items that are physically impossible to actualize. In both his nominalist and post-nominalist phases, Quine holds adamantly to the assumption that our language is part of the same material world we inhabit and which we study in the sciences, and should be studied accordingly, using empirical methods. Carnap, on the other hand, studies language primarily as a mathematical object instead of a natural one; this fact is highlighted by these consequences of his characterization of formulae for a finitist-nominalist syntax. For reasons to be explained in a later chapter, I believe this difference plays a large part in the later disputes over analyticity. In brief, Quine thinks that which sentences count as analytic is an empirical (i.e. synthetic) matter, whereas Carnap thinks that which sentences count as analytic (in a formalized language) is itself an analytic question, and thus need not be accountable to the particular facts of the matter in our actual world. As we will see in V.3.A, one fundamental difference separating Quine and Carnap in 1941 is that Quine believes linguistic concepts (such as analyticity) should be treated as *empirical*, *descriptive* concepts, whereas Carnap treats them as logico-mathematical.

Let us return to February 17, 1941, and specifically to Carnap's claim that his proposal will allow us to 'build up an unrestricted arithmetic.' (Further discussion of Carnap's syntax will appear in Section 4.B below.) Carnap is arguing that one could admit that there are only a finite number of things (and hence symbols) in the world, but nonetheless maintain that this group of things could be used to construct—though not in a physically actual sense—an infinite sequence in the metalanguage. As an extreme example to make Carnap's idea vivid and simple, make the Parmenidean assumption that there is only one thing *a* in the physical universe, whose name is 'a'—we must assume that *a* symbolizes itself, since there is only one thing in the universe. Carnap's suggestion apparently implies that we could still generate (though not with 'pencil and paper,' so to speak, in that universe) an infinite sequence:

<a>, <a, a>, <a, a, a> …

This infinite sequence of finite sequences would serve as the interpretation of the natural numbers in the finitist's language, instead of things themselves (rather, 'thing itself'). The natural numbers would then have the usual properties ascribed to them in classical arithmetic. (This example is unrepresentative in that no mathematical language can be set up within this object language, since there is only one symbol in the object language. To make the example non-trivial, imagine instead that *a* is some object in our world; Carnap's proposal would then be that we could interpret all of our arithmetical language using the above sequence as domain—<a> is assigned to '0', etc.)

Carnap recognizes that, in a universe with finite material, we will reach a sequence length after which we cannot physically write down any further elements. In the extreme example just presented, we reach that length immediately. Carnap's response (as we have seen above) is that we can use *abbreviations*, such as $10^{3,000,000}$, to refer to such sequences that cannot be inscribed. But while the abbreviational strategy will give us expressions for more numbers than would be available without abbreviations, it seems that we will nonetheless eventually run up against a 'ceiling'— introducing abbreviations will raise that ceiling considerably, but there will still be gigantic numbers that we could not express with our limited material for symbol-tokens. (I am assuming that we cannot introduce an abbreviation for transfinite numbers, since Carnap never mentions this, and it would violate the FN criteria in an obvious and blatant way.) So the problem is postponed, but not solved.

However, one might wonder whether this is a serious problem: what is the ultimate import of our inability to write down a symbol for something in a given language? There are many familiar results concerning the inexpressibility of various concepts in a given language: Tarski's theorem about the undefinability of truth is perhaps the most famous; his proof that there are undefinable real numbers is another example. In neither of these two cases is the result (standardly) taken to mean that there is no such thing as truth-in-*L*, or that undefinable real numbers somehow do not exist. In both cases, the moral usually drawn is about the limitations of the *language* in question, *not* about the things it treats. However, it should be noted that the inexpressibility in the FN case is quite different from that of Tarski's theorems about truth and real numbers. In the case Carnap discusses, the inexpressibility is quite clearly a *physical* limitation—e.g., we simply run out of 'ink' at some point—whereas the other results are logico-mathematical limitations: even if there were an infinite amount of physical 'ink' in our universe, truth-in-*L* still could not be defined within *L*.[60]

Niceties concerning the number of physical symbols aside, Carnap's proposal to use sequences in place of objects appears to violate the spirit of nominalism in an obvious way, which has probably already occurred to the reader: sequences are simply classes with further mathematical structure,[61] and classes are paradigmatic *entia non grata* (to borrow Quine's phrase) for the modern nominalist in general and a proponent of Tarski's FN project in particular. To be precise, merely allowing expressions for classes into one's language would not contravene either FN1 (no higher-order variables) or FN2 (only physical objects in D). For example, the truth of '$a \in \{a, b, c\}$' is compatible with all the finitist-nominalist conditions; furthermore, provided that 'a' denotes a physical object, so is the truth of '$\exists x(x \in \{a, b, c\})$'. However, '$\exists x(a \in x)$' could never be true in a language satisfying FN2 (assuming the standard interpretation for the symbols), because if it were true, the universe of discourse would have to include a set.[62] Quine sees the situation clearly in his immediate response to Carnap's proposal for syntax, which replaces physical things with sequences:

<u>Quine</u>: The decisive question here is whether we introduce variables for these sequences. We must do so in order to make an unrestricted arithmetic. But then we are thereby making an ontological assumption, namely about the existence of sequences. But if we do this, then we can

[60] Kripke and Woodruff's fixed-point theory of truth as well as Gupta's revision theory of truth do define *true-in-L* within *L*; in both cases, certain classical assumptions about the structure of language are modified, and the resulting logic is non-classical (e.g. bivalence fails).

[61] Specifically, a sequence is usually defined as a class plus a function that takes the members of that class to the first *n* natural numbers. Tarski uses this definition (1931/1983, 121). Carnap himself characterizes sequences in *Introduction to Semantics* as follows. "A *sequence* with *n* members is, so to speak, an enumeration of the objects (at most *n*); it can be represented in two different ways: (1) by a predicate of degree 2 which designates a one-many relation between the objects and the ordinal numbers up to *n*, (2) by an argument expression containing *n* terms (in this case, the argument expression and the sequence designated are said to be of degree *n*)." (1942, 18) These definitions, which presuppose the natural numbers, would presumably be rejected by a finitist-nominalist.

[62] Note that if we take '$\exists x$' as a first-order quantifier, the finitist-nominalist conditions do not make '$\exists x(a \in x)$' meaningless or unintelligible, but rather simply false.

in the same way also assume classes, classes of classes etc.; with that we also obtain an unrestricted arithmetic. But with that we would give up reistic finitism.

In order to violate FN2, it is not sufficient simply to allow expressions for classes (or sequences) into the language; there must be circumstances under which variables may be substituted for class- or sequence-terms. Quine's response is the obvious reply to Carnap's proposal, given the ground rules of attempting to construct a finitist-nominalist language. The more interesting or difficult question is: given that sequences are so similar to classes, what was Carnap thinking in proposing sequences? I see at least three viable possibilities. First, he might have simply had a momentary lapse, and focused solely on meeting the condition that we not assume an infinite number of physical things. Second, he might think that sequences are relevantly different from classes and other abstracta; however, there is no evidence in the notes that he did (and I have not found any evidence elsewhere). Third, he might disagree with the second sentence in the quotation from Quine: Carnap might believe we can recapture classical arithmetic without introducing variables. Unfortunately for us, Carnap's notes do not contain a direct reply to Quine, nor does the other surrounding text make it clear which of these three (if any) explains Carnap's proposal. In any case, none of Carnap's three attempts to reintroduce infinity into a finitist-nominalist language—by spatiotemporal intervals, by potential or possible infinity, or by sequences—meet with approval from his interlocutors.

4. ATTACKING THE FN CONDITIONS

Thus far in this chapter, we have examined Carnap's attempts to *defend* classical arithmetic from the apparently destructive drives of nominalism, by arguing that a finitist-nominalist could countenance higher arithmetic as meaningful, or could accommodate some sort of infinity.

However, Carnap does not merely defend his own viewpoint from Tarski and Quine's criticisms; he also takes the offensive, and attacks the finitist-nominalist conditions directly. Carnap's two primary criticisms are, first, that Tarski's view of numbers as physical objects rests upon a 'mistaken conception of arithmetic,' and second, that adopting the finitist-nominalist viewpoint in syntax will lead to unacceptable consequences for logic. Before turning to those two relatively well-developed criticisms, I will mention a very brief remark Carnap makes about Tarski's general idea, even before Tarski explicitly lays out his three conditions for an understandable language. On March 4 1940, Tarski gives a lecture at the University of Chicago on the semantic conception of truth. During this visit, he and Carnap discuss several topics in logic and philosophy, includingwhat form a formalized language for the purposes of science should take. Tarski suggests (roughly) that such languages should be predicative. Carnap responds to Tarski as follows:

> I: This restriction … corresponds to finitism and intuitionism; the tendency (since [Poincare]) of this restriction is healthy and sympathetic; but didn't it turn out that *mathematics is thereby complicated intolerably, and that the restriction is arbitrary?*
> (*ibid*., italics mine)

Carnap objects to revisions of mathematics that create unnecessary complications, and he believes that the system Tarski is describing would do so. This sentiment is not an isolated occurrence: Carnap makes basically the same point in his autobiographical essay (1963, 49). Tarski, immediately thereafter, agrees with Carnap that intuitionist mathematics is problematic, but suggests that even though the intuitionists have failed thus far to construct an elegant system of mathematics, we need not conclude it cannot be done.

A. Arithmetic is distorted

Carnap thinks the FN conditions distort the nature of arithmetic. Near the conclusion of a long conversation with Tarski and Quine on finitism, Carnap writes (in what might be an exasperated tone):

> It seems to me that the entire proposal suffers from a mistaken conception of arithmetic: the numbers are reified; arithmetic is made dependent on contingent facts, while in reality it deals with conceptual connections; if one likes: with possible, not with actual facts.
> (RCC 090-16-23)

As this quotation makes clear, Carnap feels there is something fundamentally wrong with Tarski's proposal to interpret numbers as physical objects, and its corollary that some of arithmetic becomes empirical and contingent. But, we may ask, what sort of sin is this—what, exactly, does Carnap believe is fundamentally wrong here? Presumably, it cannot be that Tarski's proposal fails to capture the essence of number, for Carnap is temperamentally opposed to questions of essence. Carnap's resistance to the FN conditions appears even more problematic, when we remind ourselves that in 1941, Carnap has been explicitly committed to his principle of tolerance for several years.[63] According to this principle, "Everyone is at liberty to build up… his own form of language as he wishes" (1934/1937, 52). *Prima facie*, Carnap's attack on the FN conditions seems *in*tolerant: why not leave Tarski in peace to construct a language that meets his criteria? The same point can be put in slightly different terms. One formulation of the principle of tolerance is: which sentences are taken as analytic is an analytic matter, not a matter of fact. (This is why Carnap saw Quine's radical translation thought-experiment as vindication for his own views.) Thus, analyticity is *language-relative*; what is analytic in one language will not be in another. The apparent problem raised by Carnap's criticism of the FN project can now be phrased as follows: Carnap appears to hold that arithmetic is analytic (or a priori) *simpliciter*, as opposed to analytic with respect to particular languages.

[63] However, as we shall see shortly, Carnap's tolerance takes a more moderate form after 1939.

So our question now is: what fault, exactly, does Carnap find with Tarski's basic idea—and (how) can this fault-finding be compatible with Carnap's commitment to both the principle of tolerance and his aversion to questions of essence? The error Carnap sees in Tarski's ways, I will argue, can be conceived of as one of *explication*; that is, Carnap thinks Tarski's explicatum (interpreting numerals as physical objects) misses the target explicandum (arithmetic of the natural numbers). Carnap does not articulate this in the discussion notes, but this attitude comes through fairly clearly in the roughly contemporaneous *Foundations of Logic and Mathematics*. There, Carnap writes:

> For any given calculus there are, in general, many different possibilities of a true interpretation.[64] The practical situation, however, is such that for almost every calculus which is actually interpreted and applied in science, there is a certain interpretation or a certain kind of interpretation used in the great majority of cases of its practical application. This we will call the *customary interpretation* (or kind of interpretation) for the calculus. … The customary interpretation of the logical and mathematical calculi is a logical, L-determinate interpretation; that of the geometrical and physical calculi is descriptive and factual.
> (1939, 171)

Carnap's basic ideas are clear: (i) Every formal calculus intended to model inferences in the sciences has a particular interpretation (or family of interpretations), called the 'customary interpretation,' associated with it. Carnap apparently believes that this interpretation is determined by 'practical application.' (Carnap does not provide the details of how scientific practice fixes meanings; the question of how use can fix meaning is still open (e.g. Gupta 1997).) (ii) Interpretations can be logico-mathematical or descriptive: for example, an interpretation that takes the universe of discourse to be the natural numbers or Zermelo's hierarchy of sets is logico-mathematical, while an interpretation whose universe of discourse contains all and only the US

[64] 'True interpretation' corresponds roughly to 'model' in modern terminology. More specifically, Carnap defines a *true interpretation* S of a calculus C as an interpretation that fulfills all of the following three conditions: (i) if the proof calculus C permits the derivation of *q* from *p*, then either *p* is false in S or *q* is true in S; (ii) if there is a proof of *p* in C, then *p* is true in S; and (iii) if there is a proof of not-*p* in C, then *p* is false in S (1939, 163).

presidents (or any other set of physical objects) will be descriptive. (iii) The customary interpretation for the arithmetical calculus is a logico-empirical one.[65]

I take (iii) to be a more precise formulation of what Carnap finds repellent about the FN conditions. Tarski is ignoring the customary interpretation of arithmetic, and thereby ignoring the meanings of the symbols '1,' '+,' etc., a meaning which is fixed by the 'practical application' of these symbols in scientific activity. I think it is fair to characterize this as a failure of explication. An explication replaces an old concept (that is in some way inferior, e.g. it is vague) with a new one (which is, hopefully, in some way superior, e.g. it is precise). One necessary condition on an acceptable explication is that the new concept overlaps sufficiently with the old one ('sufficiently' is of course a vague standard); the precise mathematical concept of a Lie algebra is a worthless explication of the concept of causation, because the two notions are unrelated. Tarski's interpretation of mathematical calculi is, by Carnap's lights, insufficiently similar to the usual conception Tarski aims to replace. It simply does seem that part of the use of arithmetical concepts is that we can always add one to any finite number, and thereby generate a new number, and any explication that fails to capture this aspect of our use has simply missed its mark.[66]

---

[65] Carnap makes the same point a few pages later:

The question is frequently discussed whether arithmetic and geometry… have the same nature or not. … [T]he answer depends upon whether the calculi or the interpreted systems are meant. There is no fundamental difference between arithmetic and geometry as calculi, nor with respect to their *possible* interpretations; for either calculus there are both logical and descriptive interpretations. If, however, we take the systems with their *customary* interpretation—arithmetic as the theory of numbers and geometry as the theory of physical space—then we find an important difference: the propositions of arithmetic are logical, L-true, and without factual content; those of geometry are descriptive, factual, and empirical. (1939, 198)

[66] Some readers may think Carnap's point here is similar to an oft-cited claim of Parsons': "The empiricist view… in the work of Professor Quine, seems subject to the objection that it leaves unaccounted for precisely the *obviousness* of elementary mathematics" (1980, 151). However, I think Carnap's point is different: the epistemological status of arithmetic (i.e., whether it's obvious or not), is probably not part of its *use*, and thus need not be captured in any explication of Carnap's.

This interpretation of Carnap (viz., Tarski misses his target explicandum) also shows why Carnap has not violated his principle of tolerance: one can be tolerant and still point out errors in explication. A tolerant stance requires Carnap to allow Tarski to set up whatever formal language he wishes; however, tolerance does not mandate that every formal language be equally expedient for every purpose, or that every formal language model every inferential practice equally well. Such a view would be madness. In fact, Carnap makes this point explicitly in *Foundations*: his response to the question 'Is logic conventional?' is 'It depends'—it depends upon the method one chooses for constructing a logic. If one begins by laying out the proof calculus purely formally, i.e., without regard for the meanings of the marks used, then of course one may lay down any rules whatsoever, and logic is conventional (and thus arbitrary) in a very strong sense. However, if one begins *not* purely formally, but with marks having meaning—i.e., with genuine words—then one cannot set up any calculus whatsoever, if the calculus is intended as a formalized version of the original, meaningful language. Under this second method, logic is not completely conventional, for the meanings of the words impose constraints upon the rules of the proof calculus (though Carnap acknowledges that there could be more than one proof calculus adequate to an interpreted language, so that logic is still somewhat conventional, even under the second method) (1939, 168-171). This is a moderated kind of tolerance. For example, if we take '$\vee$' to have the meaning that 'or' usually has in English (as opposed to treating it as a meaningless mark subject to certain rules of inference), then any calculus that allows one to infer p from $p \vee q$ alone is a very poor one. Similarly, applying this principle to the FN project, any arithmetical calculus in which we cannot infer the existence of a (new and distinct) number $n+1$ from the existence of the number n is a rather poor calculus for the usual meanings of '+', '1', and the other marks that appear in arithmetical writings.

Tarski, of course, could respond that his aim is not to *explicate* arithmetic as it is currently practiced, but rather to *revise* it fundamentally.[67] That is, his goal is not to capture as much of usual arithmetic as possible, but rather to determine how much of usual arithmetic can be saved, given (what he considers to be) a philosophically and scientifically sane conception of what exists.[68] On this view, Tarski sees himself as rescuing arithmetic from the clutches of a very dubious Platonism. In modern terms, his proposal could be viewed as a scientific revolution in the Kuhnian sense, in which many older, customary ideas are discarded. Specifically, a proponent of the FN conditions could argue that the axiom of infinity in Peano arithmetic should be considered analogous to the parallel postulate in Euclidean geometry. (This suggestion is not explicitly raised in the notes.) We say that (e.g.) the Pythagorean theorem is *mathematically* true in Euclidean geometry, but the theorem is *empirically* false of physical space, since (in a sense) one of its axioms fails to hold in the physical world. Could we perhaps analogously maintain that 'There exist infinitely many odd numbers' is mathematically true of standard arithmetic, but empirically false? Carnap stresses, throughout his writings, a distinction between *physical* and *mathematical* geometry, made famous by Einstein. Could there perhaps be a similar distinction drawn between physical and mathematical *arithmetic*? We have seen Carnap assert above and in footnote 65 that there *could* be an empirical/ descriptive interpretation of the arithmetical calculus, but that the *customary* interpretation of arithmetic

---

[67] In Burgess and Rosen's terms, on this interpretation, Tarski takes the FN project to be attempting a "revolutionary" re-conceptualization of arithmetic, instead of a merely "hermeneutic" task, in which the re-conceptualization "is taken to be an analysis of what really 'deep down' the words of current theories have meant all along" (1996, 6).

[68] As Steve Awodey pointed out to me, Carnap also thinks explicata can be revisionary as well—so that cannot be the essential difference between him and Tarski on this matter. But Carnap's allowed revisions are different in kind from those envisaged by the finitist-nominalist project: Carnap's revisions tend to be formal/ linguistic in character. Carnap's explicata make a vague explicandum precise, distinguish separate senses for ambiguous terms (e.g., two senses of probability), and remove inconsistencies in usage (e.g. reforming 'true' in everyday language). Carnap, as a rule, attempts to preserve as much scientific content as possible, while 'sanitizing' the language in the above three ways. Tarski is not concerned with content preservation as much as Carnap—and therein lies the fundamental difference between them.

involves only logical objects. But, we may query Carnap, how is the effective *applicability* of arithmetic secured? That is, what guarantees that inferences from true empirical premises using the arithmetical calculus will yield true conclusions, if our interpretation lacks any empirical elements?

Carnap could respond to a Tarskian revolution as follows: one may propose any revision of arithmetic (or any other set of concepts and/or claims) one chooses; however, if one revises *too much*, then it is no longer clear one is still doing arithmetic at all. And without some requirement that the target explicandum be captured to some degree, we are engaged in an enterprise without substantive standards for success. Furthermore, in the case of revolutions in the natural sciences, radical revisions that appear to 'change the subject' can be justified on the grounds that they lead to better predictive success. It does not appear that Tarski's proposal could improve our predictive powers—though I would not wish to rule out creative scientists finding a way to do so. (Euclidean geometry is 'shown empirically false' by (*inter alia*) identifying 'straight line' in the Euclidean vocabulary with light rays and freely falling bodies in the world—without this identification or one like it, it makes no sense to say that the parallel postulate has been 'empirically disproved.') In the formal sciences, standards are somewhat different: no one would have accepted Frege's notion of a concept script if it failed to preserve standard mathematical inferences; similarly, Weierstrass's revision of the concept of limit would be rejected if it did not sufficiently match the usage in previous theorems involving limits. In short, an appeal to view Tarski's proposal as a revision or revolution comes to a plea for exemption from a primary standard of success for projects of his type, viz., conformity with existing usage (the other primary standard in formal-mathematical explication being simplicity or elegance, along with formal consistency, of course).

B. Problems with proofs

We now return to the topic of finitist-nominalist syntax, introduced in Section 3. In Goodman and Quine's 1947 "Steps toward a Constructive Nominalism," the issue of a finitistic syntax is front and center. They explain the problem facing the finitist-nominalist, which Carnap had raised years before, quite clearly:

> Classical syntax, like classical arithmetic, presupposes an infinite realm of objects; for it assumes that the expressions it treats of admit concatenation to form longer expressions without end. But if expressions must, like everything else, be found in the concrete world, then a limitless realm of expressions cannot be assumed. Indeed, expressions construed in the customary way as abstract typographical shapes do not exist at all in the concrete world; the language elements in the concrete world are rather inscriptions or marks, the shaped objects rather than the shapes. … Consequently, *we cannot say that in general, given any two inscriptions, there is an inscription long enough to be the concatenation of the two*.
> (106, my italics)

Serious consequences follow for logical inference. Recall the standard textbook definition of a proof of *p* (in *L*): a sequence $<p_1, p_2, \ldots p_n>$ of formulas (in *L*) such that p_n is *p*, and for each $i \leq n$, either p_i is an axiom or p_i follows from some of the preceding members of the sequence using a rule of inference. Now, suppose all the 'ink' in the physical universe is used up in writing down the formulas $<p_1, \ldots p_{n-1}>$. We then cannot give a proof of the conclusion p_n even if it follows from the previous *n*-1 formulas, since there will not be any material left to write down the final formula. As a result, all our usual rules of inference (unless we radically re-conceptualize inference rules) will admit exceptions, and would thereby be overturned. For example, we are no longer guaranteed that from *A* and *B* one can infer $A \wedge B$, since for large enough *A* and *B*, there will not be enough material (in a finite universe) to write down the conjunction of both, after writing down the first two. Similar reasoning holds for other rules of inference. Such a finitistic proof calculus would be radically incomplete (in the technical sense): every model that satisfies *A* and satisfies *B* will also satisfy $A \wedge B$—but the proof calculus cannot prove that. Prospects for finitist syntax will be even dimmer if we take Tarski's original suggestion that "we ought to take as expressions, sentences, and proofs only actually written down items" (-27). For, if that restriction is adopted, we could only infer sentences that are inscribed somewhere (or spoken sometime, etc.). As we saw in Section 3, Carnap recognized these problems, and considered them to be a serious defect in the FN conditions.

5. AN OBJECTION TO THE FN PROJECT NOT IN THE NOTES

Before concluding this chapter, I would like to consider a final objection to the FN project that does not appear in the notes. There are, of course, many criticisms one might level against the finitist-nominalist viewpoint that are not raised in Carnap's discussion notes; generations of anti-nominalist and anti-finitist philosophers have generated a small library of them. However, I will present only this objection, because it is not one of the usual objections to nominalism in general, it is based on a thesis that seems to enjoy (some) consensus among philosophers, and because I believe it is Carnapian in spirit—though I will not argue that final point in detail. The crux of the objection is this: an answer to the question 'What counts as an individual?' (or '…as a unit'), which determines (in part) how many 'things' there are, is not the kind of claim about which there is a fact of the matter. In Carnap's terms, it is an *analytic* issue, not a synthetic one; perhaps it is (partially) analogous to a choice of co-ordinate system in mechanics.

Something similar to this idea appears in *Republic* VII: Plato claims the unit is intelligible, not sensible. More importantly for present purposes, it also appears in Frege's *Grundlagen*: Frege asks us to consider a pack of playing cards. If someone points to the pack and asks you 'How many (things) are there?', the correct answer will be: it depends—if the questioner is asking about the number of spades, the answer is *thirteen*, if about Aces, the answer is *four*, and if about cards, the answer is *fifty-two*. Thus the question 'How many (things) are there?' is not well posed, because it admits of more than one answer, depending on further specifications. And what holds for the pack of cards also holds, presumably, for the entire material universe: there is no fact of the matter about how many things there are. Of course, once one specifies what is to count as an individual (e.g. *spades*), then it becomes a well-posed question with a univocal answer (*thirteen*). Without such a further specification, there is no fact of the matter about what the units are in the natural world.

How might a Tarskian respond to this challenge? Here is one straightforward reply: no matter what is taken as a unit (i.e., no matter what are taken to be the elements of the domain of quantification), whether it be quarks, spatiotemporal intervals, quanta of energy, etc., one will always come out with a (possibly) finite number of things—as long as the domain is restricted to physical entities. (Obviously, allowing variables to range over **N**, or **R**, or the set-theoretic hierarchy would automatically yield an infinity.) Thus, the initial lack of a well-posed formulation is rendered innocuous—the Tarskian will let you turn it into a well-posed formulation in whichever way you please, so long as the only things the variables range over are physical in one way or another.

In the end, none of Carnap's criticisms of the finitist-nominalist project made headway with Tarski and Quine. I wish to highlight here, in conclusion, that Carnap's failure to win converts is, in many cases, not a function of the quality of Carnap's arguments, but rather of the differing fundamental philosophical stances Tarski and Quine brought to the table in 1941. First, if Tarski and Quine had accepted Carnap's suggestion that a partially-interpreted calculus should also count as meaningful or intelligible *simpliciter*, then they would have been strongly inclined to view Peano arithmetic, and perhaps even set theory, as also intelligible. Second, if Quine and Tarski were not so averse to modal notions, perhaps they would have accepted Carnap's proposal to use the notion of a potential or possible infinity in lieu of an actual infinity in order to build up classical mathematics. More generally, if Tarski and Quine were comfortable with the notion of sense (as opposed to reference) and intensional languages, then they might not think of numerals that are ‘too large’ as being *meaningless*, but rather simply denotationless but nonetheless meaningful. Carnap’s willingness to allow for sense and/or intension explains why, for him, ‘the understandable outruns the actual.’ Finally, if Tarski and Quine did not consider the study of language to be a strictly empirical enterprise, they might be more deeply worried about the very real problems Carnap points out that arise with syntax and proof under a finitist-nominalist regime. But if syntax only studies empirical language, and thus only the physically possible inscriptions, then consequences that strike Carnap as intolerable (e.g. given two expressions, one cannot always form their conjunction) appear bearable, if not desirable. The differences between Carnap’s conception of analyticity, and the conceptions of Tarski and Quine, are the subject of the following two chapters.

CHAPTER IV. THE FINITIST-NOMINALIST PROJECT AND ANALYTICITY

If a modern-day philosopher is engaged in a free-association session, and the analyst's prompt is 'Carnap and Quine,' then the response will almost certainly be 'analyticity' or its kin. Quine's attack on the notion of analytic truth is, by most philosophers' standards, one of the most influential and widely adopted 'big ideas' of twentieth century Anglophone philosophy. Among scholars working in the history of analytic philosophy, the disagreement between Quine and Carnap over the analytic/ synthetic distinction has been one of the most studied and discussed episodes. Thus, one might hope that during Carnap and Quine's academic year together, they would discuss their conflicting viewpoints on this issue at length and in detail. Quine, in his autobiography, leads his reader to believe as much:

> The fall term of 1940 is graven in my memory for more than just the writing of *Elementary Logic*. Russell, Carnap, and Tarski were all at hand. … [They read the *Intro to Semantics* MS] My misgivings over meaning had by this time issued in explicit doubts about the notion, crucial to Carnap's philosophy, of an *analytic* sentence: a sentence true purely by virtue of the meanings of its words. I voiced these doubts, joined by Tarski, before Carnap had finished reading us his first page. The controversy continued through subsequent sessions and without progress in the reading of Carnap's manuscript.
> (1985, 149-150)

Unfortunately, this tantalizing claim is misleading. First, it misleads us in a relatively insignificant way: Quine's claim that the group did not advance past the first page of Carnap's manuscript is demonstrably false. Carnap's notes record a discussion of the adequacy of a particular definition that appears in chapter seventeen of the manuscript of *Introduction to Semantics* (RCC 090-16-03),[69] definition 18-1 in the published version. Quine's representation

[69] "Dec. 9, 1940
Tarski and Quine, On general semantics.
On the definition of 'entity u is *covered* by system S' ((I) D17-1)

of the situation in the above quotation is inaccurate in a second, more significant way. Although there are several scattered remarks in Carnap's dictation notes dealing with analyticity (or, in his preferred terminology at the time, with 'L-truth'), there are disappointingly few sustained discussions of the issue. (Of course, it is possible that there were many more such conversations on the topic of analyticity, but Carnap failed to record them. I know of no evidence for such a supposition beyond Quine's claim above; and as we have just seen, Quine's reminisces about this time period are not always veridical.)

Interestingly, the discussant who appears to bear the most sustained animosity toward analyticity is not Quine but Tarski. However, we need not despair that the 1940-41 notes shed no light on the vexed concept of analyticity. Not only are there scattered instances in which the group does directly discuss analytic truth and kindred concepts, but the finitist-nominalist project also bears a clear (albeit indirect) relation to analyticity. This relationship is the focal point of the present chapter; the discussion notes directly addressing analyticity are taken up in the following chapter. This chapter has three parts: first, I flesh out the conceptual relationship between finitism-nominalism and analyticity by determining which portions of arithmetic would become synthetic under a Tarskian regime; in order to do this, a digression through Carnap's conception of semantics circa 1940 is necessary. Second, I offer a historical conjecture about the development of Quine's attack on analytic truth, and lastly, I consider certain philosophical issues prompted by the previous two sections, with particular focus on Paul Boghossian's recent claims about analyticity.

Tarski: One should also include: Elements of designated classes, elements of elements of classes, and so forth.
Quine: The definition must be elastic enough that it also holds for languages in which the universal is not expressed through variables, e.g. Schönfinkel's system."
This becomes definition 18-1 in the published version. This does not show that Tarski and Quine read all of the manuscript up to the seventeenth chapter, but it does show that they did discuss more than the first page.

1. UNDER A FINITIST-NOMINALIST REGIME, ARITHMETIC BECOMES SYNTHETIC

Though Carnap, Tarski, and Quine do not directly discuss analyticity a great deal during their academic year together, the finitist-nominalist project, which does occupy a large portion of their time and energy, bears indirectly on the notion of analytic truth. How? As Carnap unhappily notes, under Tarski's regime "arithmetic is made dependent on contingent facts," i.e., it becomes a synthetic enterprise (RCC 090-16-23). This would be disappointing for Carnap, for he thinks one of the genuine intellectual advances made by the logical empiricists was showing that arithmetic is both analytic (contra Kant and Poincaré) and a priori (contra Mill and others) without lapsing into some form of Platonic metaphysics. Tarski's proposal would appear to Carnap as regressing to a Millian, empiricist view of mathematics.

Given that Carnap considers arithmetic to be synthetic under the finitist-nominalist restrictions, we can ask the further question: which parts, exactly, become synthetic? The answer is: less than one might initially think. I will justify that answer presently, but first we must clarify what is meant by 'synthetic' here. First, 'synthetic' does not mean 'neither logically true nor logically false' in the modern sense, i.e. 'false in at least one model, but not in all'—for if it did, classical first-order Peano arithmetic would be synthetic, since its postulates are only true in models that have an infinite domain. Second, one might attempt to cash out 'arithmetic becomes synthetic' via Carnap's distinction between *descriptive* interpretations and *logical* ones (discussed in III.4.A). A descriptive interpretation of a set of sentences takes as its domain empirical objects, while the domain of a logical interpretation of a set of sentences consists of logical objects—so Tarski's conditions turn mathematics into a descriptive language. However, although Carnap (as seen in previous chapters) thinks interpreting mathematical language as descriptive is a mistake, simply assigning the numerals to physical objects instead of numbers (considered either as individuals or in the Frege-Russell way) does not, by itself, make arithmetic synthetic. For then ‘Rushmore=Rushmore,’ ‘Carnap wrote *Principia Mathematica* or Carnap did not write *Principia Mathematica*,’ and any other instance of a logical truth containing descriptive terms would count as synthetic—an unpalatable consequence for Carnap. (Put otherwise: though Carnap claims that every sentence given a logical interpretation is analytically true or false, the converse is not true (1939, 180).) So what *is* the sense of ‘synthetic’ here? At this stage in Carnap’s career, a sentence is analytic in an interpreted language *L* if and only the semantic rules of *L* determine the truth-value of that sentence. If the semantic rules do not suffice to determine a sentence’s truth-value, then that sentence is synthetic or factual (and conversely).[70] So, if Carnap is correct that arithmetic becomes factual under Tarski’s restrictions, it must be the case that there are arithmetical claims whose truth-value is determined by the semantic rules of classical arithmetic, but whose truth-value is left indeterminate by the semantic rules of finitist-nominalist arithmetic.

In order to determine which arithmetical sentences become synthetic, we must answer the question: how does Carnap conceive of semantic rules in 1941? His conception is, in some ways, close to modern formal semantics, but there are clear differences as well. The fundamental unit of study for semantics for Carnap is the *semantical system*, which he defines as “a system of rules, formulated in a metalanguage and referring to an object-language, of such a

[70] In 1939’s *Foundations*, Carnap writes:
We call a sentence of semantical system S [= truth-tables for connectives plus rules of designation] (logically true or) *L-true* if it is true in such a way that the semantical rules of S suffice for establishing its truth. … If a sentence is either L-true or L-false, it is called *L-determinate*, otherwise (L-indeterminate or) *factual*. (The terms ‘L-true’, ‘L-false’, and ‘factual’ correspond to the terms ‘analytic,’ ‘contradictory,’ and ‘synthetic’, as they are used in traditional terminology” (155). Essentially identical claims are found in (Carnap 1942, 140-142).

kind that the rules determine a **truth condition** for every sentence of the object language… the rules determine the *meaning* or *sense* of the sentences" (1942, 22). A semantic system consists of three kinds of rules: rules of formation, rules of designation, and rules of truth. (It is the achievement of Tarski's *Wahrheitsbegriff* to show that the third can be defined in terms of the second.) The rules of formation provide a recursive definition of 'sentence of *L*'. Rules of designation provide designata for the (non-logical) signs.[71] Specifically, it consists of sentences of the form '*b* designates *c*,' where

- if *b* is a name (individual constant), then *c* is an object;
- if *b* is a predicate (or relation letter), then *c* is a property (or relation).

As an example of the first kind, Carnap mentions (where German is the object-language and English the metalanguage) "'mond' designates the moon," and as an example of the second, "'kalt' designates the property of being cold" (1939, 151). Note that Carnap treats individual constants extensionally, like modern models, but treats predicate letters intensionally, unlike modern models. What is the status of these rules, according to Carnap? "[T]he rules of designation do not make factual assertions as to what are the designata of certain signs. There are no factual assertions in pure semantics" (1939, 25). In short, the rules of designation are analytic, as are the rules of truth (assuming we are not engaged in empirical, descriptive linguistics).

The rules of truth are almost identical to the ones familiar to us today: the truth-values of sentences containing logical connectives are given by the usual truth-tables, and the rule for the universal quantifier is more-or-less identical to the one current today. The only substantive difference of formulation between Carnapian rules of truth and modern ones appears at the level

[71] The distinction between logical and non-logical signs is part of the semantic system, according to Carnap. We can think of it as part of the rules of formation, or as a separate, fourth set of rules associated with the semantic system (see 1942, 24).

of atomic sentences, and is due to Carnap's interpreting predicates as properties. Carnap writes (where the '*n*' subscript means the expression is in the grammatical category of *noun*, and the '*p*' subscript indicates *property*): "A sentence of the form '$\ldots_n$ ist $---_p$' is true if and only if the thing designated by '$\ldots_n$' has the property designated by '$---_p$'"(*ibid*.). In modern model theory, a (classical) model specifies an extension for every predicate, not a property (some people construe properties as extensions of predicates, but Carnap, like many, does not do so). For Carnap at this time, a property is an extension in every state of affairs; that is, a first-order property assigns a set of individuals to every possible world.[72] Lastly, if the language under consideration is to contain variables, Carnap says we must introduce *rule(s) of values*, which specify a range of values for each kind of variable in the language, as well as (what we today would call) *rules of satisfaction* for open formulas. (Rules of values are analogous to rules of designation, and rules of satisfaction are analogous to the rules of truth.) The rule of values, which specifies the universe of discourse for a language, is also an analytic claim in Carnap's view (1939, 174). The domain can be specified via simple enumeration, or by specifying a condition something must meet to be a member of the set; Carnap's own examples include "all space-time points, or all physical things, or all events, or all human beings in general" (1942, 44). (Variables themselves, however, can be either logical signs or descriptive signs, depending on the universe of discourse (1942, 59).)

With this characterization of Carnapian semantics in hand, we can better understand one of Carnap's claims that sounds most strange to our modern ears. This will (hopefully) be interesting for its own sake, but it will also highlight differences between Carnapian semantics and our own. In his discussion of the axioms of infinity and of choice in his autobiography, Carnap writes:

---

[72] This Carnapian notion of property is thus fundamentally identical to Montague's notion of *intension*.

"we [the Vienna Circle members] realized that either a way of interpreting them as analytic must be found, or, if they are interpreted as non-analytic, they cannot be regarded as principles of mathematics. I was inclined towards analytic interpretations; … I found several possible interpretations of the axiom of infinity, different from Russell's interpretation, of such a kind that // they make this axiom analytic. The result is achieved, e.g., if not things but positions are taken as individuals" (1963, 47-8).

The idea that an interpretation can make an axiom analytic is perplexing for a modern reader. For we now characterize logical truth as truth under all interpretations. If we think analytic truths are logical truths (and Carnap does call analytic truth 'L-truth,' i.e., logical truth), then it seems that analytic truths should be true under *all* interpretations, *contra* Carnap's suggestion above. But Carnap would not characterize analytic truth as truth-in-all-interpretations. Analytic truth for Carnap, as we have seen, is truth in virtue of the semantic rules; and one of the semantic rules specifies the universe of discourse. Thus if the domain is taken to be an uncontrovertibly infinite collection (e.g., the natural numbers), then the semantic rules alone will determine the truth-value of the axiom of infinity to be *true*. Of course, there will be other interpretations under which the axiom of infinity becomes analytically false (e.g., let D={0, 1, 2}), and others under which it becomes synthetic (e.g., D={x : x is a physical object}).[73]

But then the modern reader might worry: if we are allowed to include that much information about the language to determine which sentences are 'true in virtue of meaning,' then will there be any sentences that are *not* true in virtue of meaning? For example, looking at the matter from the modern perspective, suppose we are given an interpreted language, and that the interpretation function f of this language is such that f(t)=a and f(P)={a, b, c} ('t' is an individual constant and 'P' a monadic predicate). Then the truth-value of 'Pt' can be 'calculated'

[73] At this point, someone might object (especially if she is sympathetic to Tarski's finitist-nominalist program) as follows. First, it seems that mathematics is thereby forced to take a specific subject matter, in Carnap's case, positions. So it is no longer clear that we can legitimately apply these 'mathematical truths' to any and all physical objects, since we have restricted the domain of quantification to positions—yet Carnap does think mathematical theorems can be used to infer one factual statement about physical objects to another, and not just about positions. Second, by analogous reasoning, 'Less than 100 things exist' can be made analytic, if living U.S. presidents are taken as the individuals in the domain. That appears to be a nearly worthless kind of analyticity.

from the information about the interpreted language alone, i.e., no empirical tests need to be run to determine its truth-value. *Every* sentence appears to be analytic, an obviously unacceptable consequence to everyone, especially Carnap. What happened? Carnapian semantics would not allow f(P)={a, b, c} as a semantic rule of the language. Instead, the semantic rule for predicates take the form: f(P)= the property (of being) *X*. And thus the language alone would not (in general) determine the truth-value of 'Pt,' for (on Carnap's picture) it is an empirical question whether the object denoted by 't' in fact has the property designated by 'P.'[74] In short, whereas the modern conception takes 'logically true in *L*' to mean 'true in all M=<D, f> of *L*,' Carnap (to put the matter in modern terminology) fixes D, and then takes analytic truth to be 'true for almost all f'. The 'almost' must be included, because Carnap would place certain restrictions on the interpretation function.[75] For example: if, in a particular f, the object-language predicate corresponding to the property of being a horse is assigned set S, then in that same f, the set assigned to the predicate corresponding to the property of being a stallion must be a proper subset of S.[76] It is in this sense that 'All bachelors are unmarried' is an analytic truth. (This basic idea appears in *LSL* §34c-d (though without the 'almost,' and with different terminology since Carnap has not yet entered his semantic period), where Carnap defines 'analytic-in-language-II.' There, Carnap specifies the elements of D once and for all (as the class of accented expressions), but then sets up the definition of analyticity such that a sentence will be analytic if it is true[77] for all (grammatically appropriate) assignments of values to its variables and non-logical constants.)

[74] During his semantic phase, Carnap identifies "extension"—as opposed to intension—with "contingent reference or denotation" (1963, 63). That is, in order to determine the extension of a word (unlike its intension), empirical, factual information is necessary.

[75] The notation here is anachronistic, but the underlying idea is in Carnap: see *Introduction to Semantics* §19.

[76] I assume that the language contains primitive predicate letters corresponding to 'horse' and 'stallion.' Interestingly, this situation is one of the primary reasons Carnap considered the transition from syntax to semantics to be necessary: syntax alone cannot capture this relation between 'horse' and 'stallion' (1942, 87).

[77] In *LSL*, Carnap eschews the notion of truth; so this actually reads 'analytic.'

Now we can ask: what would be the semantic rules for arithmetic—and more specifically, for finitist-nominalist arithmetic? Since there is no list of such rules in Carnap's discussion notes, the following proposal is somewhat conjectural.[78] First, the maximal allowable domain is the set of all physical objects. Perhaps we should include, as allowable domains, all (non-empty) proper subsets thereof: Tarski remarks, as we saw in chapter I, that he would like an arithmetic that makes no assumption about the number of things in the world. For arithmetic to get off the ground, the elements of D must be arranged in a sequence; Tarski seems to suggest that we impose the order arbitrarily upon the physical objects, but it does not fundamentally matter what the source of this structure is, as long as it exists. The semantic rules for designation must be such that the first *n* numerals (starting from '0') designate the first *n* objects in the sequence. That is, 'S(*a*)' designates *b* if and only if *b* immediately follows *a* in the sequence of physical objects. However, it does not matter which object is the beginning member of the sequence, or which objects come where in the sequence. Setting up our semantic rules such that a single sequence is picked out once and for all will lead us to the problem that some numbers will be brunettes, discussed at the end of Chapter I. Thus I propose that we do not include in our semantic rules any one particular interpretation of the numerals, but rather just make all interpretations subject to the above constraint.

Finally, we need semantic rules to deal with numerals whose intended reference outstrips the number of physical objects in the world. Recall from I.3.B that Carnap records three proposals for interpreting such numerals. Assuming there are *k* physical objects in the universe, the three proposals are:

---

[78] In "Foundations of Logic and Mathematics," Carnap describes what a (true) interpretation of the Peano postulates would be: We have… to choose any infinite class, to select one of its elements as the beginning member of a sequence, and to state a rule determining for any given member of a sequence its immediate successor. … '[0]' designates the beginning member of the sequence; if '…' designates a member of the sequence, then '…′' designates its immediate successor; 'N' designates the class of all members of the sequence that can be reached from the beginning member in a finite number of steps. (1939, 181)

(a.) $k' = k'' = \ldots = k$
(b.) $k' = k'' = \ldots = 0$
(c.) $k' = 0, k'' = 0', \ldots$

The first two both follow the spirit of Frege's 'chosen object' proposal for handling non-denoting expressions; they differ from one another in that (a.) makes the 'final' object in the universe the chosen one, whereas in (b.) the chosen object is assigned to '0'. Option (c.) can be intuitively conceived as a circle whose circumference one can trace an indefinite number of times as one writes down the numerals: two numerals are assigned to the same object if and only if the numbers they are intended to denote are identical modulo *k*. In each of these three cases, at least one of the Peano axioms is violated. If (a.) is adopted, then two distinct 'numbers'—those are scare-quotes—will have the same successor (namely, the objects denoted by '*k*-1' and by '*k*'). (Though, of course, under all three proposals, if the domain is finite, there will be cases in which two *numerals*, such that one is an ancestral-numeral of the other, denote the same object.) If (b.) or (c.) is adopted, then the object designated by '0' will be the successor of some number.

Now we are in a position to ask: what, specifically, *are* the arithmetical claims that are classically analytic, but synthetic under Tarski's restricted arithmetic? First, any sentence of the language that asserts (or denies) that there exist at least *n* distinct numbers will become synthetic. (Of course, the same holds for 'There exist *n* distinct physical things'.) What about variable-free formulae of arithmetic, such as '2+3=5': do they maintain their analytic status under a finitist-nominalist regime? Some do, and some do not. The sentence '2+3=5' will be true regardless of the cardinality of the domain, and this is the case under any of (a.)-(c.), so it is analytically true in all three semantics. And '2+3≠5' will be false in any domain under (a.)-(c.), so it can be considered analytically false. However, the same cannot be said of '1000=2000' or '2+3≠7': each of these will be false in certain domains but true in others. Under rule (a.) or (b.), '1000=2000' will be true for domains with cardinality less than or equal to 1001, false otherwise.

Under rule (c.), this sentence is true for domains in which 1000=2000 mod *k* (*k*, as before, is the number of elements in the domain), false otherwise. For similar reasons, in certain domains the classically true '2+3≠7' will be false. The preceding can be generalized as follows: for all variable-free arithmetical sentences, all atomic sentences (i.e., those of the form *n*=*m*) that are analytically true in classical arithmetic will be analytically true under a Tarskian regime (assuming we adopt one of (a.)-(c.))[79]. However, atomic sentences that are analytically false in classical arithmetic become synthetic under the finitist-nominalist reconstrual (e.g. '1000=2000'), as do their negations (e.g. '2+3•7'). In short: though all the classically analytically true arithmetical sentences are analytically true under the finitist-nominalist setting as well, all the classically analytically false arithmetical sentences become synthetic.

Which other sentences become synthetic depends upon which particular semantic rules are adopted. What further sentences that are classically analytic become synthetic under (a.)? If we adopt the 'liberal' version of FN3, so that we allow it as a possibility that the number of physical objects in the universe is infinite, then the assertion (or denial) of 'No two numbers have the same successor' becomes synthetic, along with all the sentences that imply it. The same holds for 'No number is its own successor.' Both of these are false if the domain is finite, but not if the domain is infinite; thus the semantic rules alone do not determine the truth-values of these sentences. If, however, we endorse the stricter version of FN3, and claim that the number of physical objects in the universe is finite, then the truth-value of both of these sentences (and those that entail them) can be computed from the semantic rules—though here they become analytically *false*, unlike the classical case. Under the (b.) and (c.) semantics, these two sentences would be analytically true, *if* we allowed ourselves, among our semantic rules, the anti-Parmenidean assumption that the universe does not contain exactly one thing; without it,

[79] For example, if we declared numerals that are 'too big' to be meaningless (perhaps on the grounds that they are denotationless), then all the classical atomic truths will not be analytically true.

these two sentences will be synthetic under (b.) and (c.) as well. I do not see any reason why this anti-Parmenidean assumption should be considered a semantic rule: although it is pretty clearly false that our universe contains only one physical object, the kinds of reasons adduced to support that conclusion are primarily empirical in character. (Of course we could *stipulate* the anti-Parmenidean assumption as a semantic rule, but that would be unmotivated by the language we actually speak—it would be analogous to declaring 'World War II ended in 1945' a semantic rule.)

Which other classically analytic arithmetical sentences become synthetic under the semantic rules (b.) and (c.)? The situation parallels that of (a.) above: if we adopt the liberal version of FN3, then any assertion that implies the sentence '0 is not the successor of any number' or its denial will become synthetic. The truth-value of this sentence cannot be calculated from the semantic rules, since it will be false if the domain is finite, but could be true under an infinite domain. Similarly, if we adopt the strict version of FN3, then we can calculate the truth-value of this sentence (and all those which imply it) from the semantic rules—but unlike the classical case, it is evaluated *false*. And if the domain of discourse is allowed to contain only one individual, then this sentence is synthetic under semantic rule (a.).

2. RADICALIZATION OF QUINE'S CRITIQUE OF ANALYTICITY: A HISTORICAL CONJECTURE

Richard Creath has argued (1987) that Quine's 1936 "Truth by Convention" should not be read as a full frontal assault on the very idea of analytic truth. Such a reading is anachronistic, and arises from the temptation to read the radical criticism of analyticity found in "Two Dogmas" into an article written fifteen years earlier. On this interpretation, the Quine of 1936 is viewed as “more nearly a request for clarification than an attack” (1987, 487) on the notion of analytic truth. Creath marshals published and unpublished textual support for the view that Quine was not convinced the concept and its kin are fundamentally incoherent until years later. In particular, he points out, first, that “Truth by Convention” grew out of three lectures Carnap gave to the Harvard Society of Fellows in 1934, and these lectures praise Carnapian views almost unequivocally. Second, at the 1937 APA, Quine gives a lecture entitled “Is Logic a Matter of Words?”—and in it, Quine argues for (what he later calls) the ‘linguistic doctrine of logical truth.’ So we have Quine defending Carnaps views both immediately before and immediately after he wrote “Truth by Convention.” In what follows, I assume Creath’s picture is basically correct. If he is right, then this raises a further historical question: what prompted the radicalization of Quine’s attack on analyticity? Specifically, could the conversations of 1940-41 have contributed significantly to the deepening of Quine’s criticisms? I will provide evidence for this claim at the close of section V.1.B. If these discussions *did* foment Quine’s radicalization, how did they do so? No archival material in the Carnap collection from this period could, in my opinion, be considered a ‘smoking gun,’ so the following suggestion is conjectural.

In “Truth by Convention,” one way in which Quine questions the analytic/ synthetic distinction is the following. We establish the conventionality of the truths of logic by simply assuming certain sentence-forms involving ‘and,’ ‘not,’ and ‘all’ to be true by “linguistic fiat.” Why, Quine asks, if we are allowed to declare certain sentences true simply by linguistic fiat, could we not continue expanding this list of conventional truths, and include Einstein’s field equations (for example) in our list of sentences true by convention as well? And there is no reason to stop with fundamental physical laws: as long as we can declare any sentence true we like, we could include ‘Our solar system has nine planets.’

If in describing logic and mathematics as true by convention what is meant is that the primitives *can* be circumscribed in such a fashion as to generate all and only the truths of logic and mathematics, the characterization is empty; … the same might be said of any other body of doctrine as well.[80]

(Quine 1936/ 1976, 102)

In effect, Quine questions the existence of a reasonable and motivated 'cut-off' point for statements considered true by convention that would prevent an indefinite expansion of such truths beyond the realm of logic and perhaps mathematics. He cannot see any special quality that the terms 'or' and 'not' possess that (e.g.) 'mass-energy density' lacks, such that sentences involving the former but not the latter can legitimately be simply stipulated true. In short, we can read Quine as claiming that once we permit one sentence to be true by linguistic fiat, there is no principled ground for stopping unlimited inflation of such truths. This line of thought takes a more exact form in an article that Quine penned jointly with Goodman in 1940: "Elimination of Extra-logical Postulates." This article shows provides a formal procedure for converting any system of postulates framed in a formal language into a postulate-free language that has, in an important sense, the same content as the original postulate system. The basic idea is to transform the postulates, which could very well be considered synthetic claims (hence the 'Extra-logical' in the title), into *definitions* in the language, which are considered paradigmatically analytic by proponents of analyticity. Quine improved this formal recipe further in "Implicit Definition Sustained" (1961/1976).

In the finitist-nominalist conversations of 1940-1, Quine is presented with the converse possibility. Instead of *expanding* the conventional, and thus analytic, truths from logic and mathematics into natural science, Tarski presents a philosophically motivated language-form in which the number of presumed conventional truths is *contracted*. When Quine sees that

[80] A few years later, Quine will not even allow that the truths of mathematics *can* be so circumscribed. Gödel's incompleteness results show that we "can't even formulate adequate, usable conv'ns afterward," since no logical system captures all the logico-mathematical truths (MS storage 299, Box 12, folder: Phil. 20m-1940).

arithmetical assertions can become synthetic under certain conditions, this shows him concretely that the boundary between the analytic and the synthetic can be considered porous in *both* directions. In "Truth by Convention," only one of the directions is considered, and the analytic status of fundamental logic is not in doubt.[81] Yet we see precisely this questioning of the analytic status of logic in "Two Dogmas":

> Even a statement very close to the periphery can be held true in the face of recalcitrant experience by pleading hallucination or by amending certain statements of the kind called logical laws. … Revision even of the logical law of the excluded middle has been proposed as a means of simplifying quantum mechanics. (1951/1965, 211)

I am not claiming that Quine's willingness in "Two Dogmas" to renounce the supposed analyticity of logic and mathematics definitely stems from the 1940-41 finitist-nominalist project, in which certain arithmetical claims become synthetic. However, these conversations with Tarski and Carnap certainly *could* have planted the idea in his head, or (as is perhaps more likely) cultivated the germ of an idea he had already entertained. Additionally, I am not claiming this radicalization of Quine's critique of analyticity—namely, from 'The corpus of analytic truths can be indefinitely expanded' to 'The corpus of analytic truths can be indefinitely expanded *or contracted*'—is the only conceptual step needed to move from the Quine of 1936 to the Quine of 1951. In particular, Quine is not yet profoundly skeptical of the very intelligibility of synonymy in "Truth by Convention" or (as we shall see) in the 1940-41 discussions with Tarski and Carnap (where Quine, apparently without hesitation or compunction, defines analytic truth using the notion of synonymy). The radicalization I have attempted to account for in this section is thus only a part of Quine's intellectual journey from "Truth by Convention" to "Two Dogmas." Another will be described in V.1.B, where we see how Quine's (antecendent) antipathy towards intensional languages is transformed into a criticism of Carnapian analyticity.

---

[81] There is a footnote in *Word and Object* in which Quine basically says that "Truth by Convention" did not claim that there are no analytic truths (1960, 65n).

3. LANGUAGES WITHOUT ANALYTICITY

As observed in the previous section, Quine complains in "Truth by Convention" and later works that the body of analytic truths could be expanded without limit. In the most extreme case, every sentence of a language could be declared analytically true; this will occur in a language whose postulates and inference rules are inconsistent, as Carnap discusses in *LSL* §59. But the finitist-nominalist discussions show that the range of analytic truths can be contracted in a philosophically motivated way as well, and this notion appears at the close of "Two Dogmas." The extreme case here, of course, would be a language in which there were no analytically true or analytically false sentences whatsoever—not even the sentences traditionally called 'logical truths' or '(logical) contradictions.' This case would perhaps be interesting to study for its own sake, on the grounds that it might exhibit unusual pathologies. However, it would also be of interest because it would represent what (the later) Quine—the Quine who supposedly vanquished Carnap in the debates over analyticity—actually takes language to be. That is, the later Quine claims that there are no analytic sentences. I will attempt to make this claim precise by formulating it using model-theoretic apparatus, and investigate what follows from such a conception of language. In particular, I will attempt to shed some light on recent philosophical claims concerning analyticity, especially ones forwarded by Paul Boghossian and adopted by other prominent philosophers.[82] The aim of this section is *not* primarily historical; rather, it is to couch Quine's proposal in a modern idiom, and study it for philosophical gain. (The aim of the present model-theoretic analysis is to understand Quine's ideas in the same way that a generally

[82] The notion of a 'language without logic' has appeared at least twice in the recent literature: see (Creath 1996) and, much more briefly, (Sober 2000, 257).

covariant spacetime formulation of Newtonian gravitation theory aims at understanding Newtonian ideas.)

Let me say immediately that I recognize that analytic truth is conceptually distinct from logical truth. The logical truths are usually considered to be a proper subset of the analytic truths. Furthermore, Quine (1960) claims that even though we can develop and deploy a notion of logical truth, those logical truths should not be understood as analytically true, i.e., as true solely in virtue of the meanings of the words within them. My point, which Carnap and Quine agree upon, is that the strongest indication or evidence there is for the existence of analytically true sentences is the existence of (what we call) the logical truths. Thus a language without any logical truths would lack the strongest candidates for analytic truths.

A. Strictly Quinean Languages

There are at least two ways to capture formally the notion of a language with no analytic truths: syntactically (proof-theoretically) or semantically (model-theoretically). The syntactic characterization of language, which finds clear expression in Carnap's *Logical Syntax*, represents a language by an ordered pair: the set of formation rules and the set of transformation rules (1934/1937, 4, 167). In a conversation with Carnap in 1940, Tarski offers a similar characterization.[83] Formation rules (F) contain the grammatical information about the language: which symbols belong to which grammatical category, which strings of symbols are grammatical, which strings are sentences, etc. The transformation rules (T) are usually called

[83] Carnap writes: "for him [Tarski], a calculus is an ordered pair, consisting of a class (of sentences) and a relation (consequence)" (RCC 090-16-09). This is not identical to Carnap's characterization, because (in view of Gödel's results) there can be sentences of a given language such that B is a consequence of A, but no finite set of inference rules (and postulates) will allow the derivation of B from A.

rules of inference today.[84] It is often expedient to replace some of the transformation rules with axioms or primitive sentences (1934/1937, 29); when a language is characterized thus, we can represent the language as an ordered triple containing the set of primitive sentences (P) as its third member. To fix notation, let us represent a syntactically characterized language as <F, T, P>. All the sentences derivable from the rules of inference alone (and the axioms, if there are any) constitute the *derivability-class* or theorems of the language. Now we can characterize syntactically what I will call a 'strictly Quinean' language, namely, one that has no analytic truths.

L_{SYNTAX}=<F, T, P> is *syntactically strictly Quinean* iff the derivability-class of L_{SYNTAX} is empty.

Though the formation rules of a strictly Quinean language can be as rich as that of any other language, intuitively speaking, such a language does not have 'enough' transformation rules to derive any analytic truths. Obviously, P=∅ is a necessary condition a language must meet in order to fulfill the above definition. However, T need not be completely empty for L to be strictly Quinean: for example, we could include in T the usual inference rules (in a natural deduction formulation, e.g.) for ∧-introduction and elimination, as well as ∨-introduction and elimination, without the derivability-class of the language becoming non-empty. So 'A⊃(A∨B)' will not be in the derivability-class of the language, since T does not contain a rule for ⊃-introduction.

Now let us turn to a semantic conception of a language, where an interpreted language *L*=<L, M, ρ> is defined as in section I.2: L encodes grammatical information (exactly as F

[84] These rules of inference need not be restricted to those found in, say, a natural deduction system of first-order logic. I have introduced the syntactic characterization of language not just to cover the Carnap of *Logical Syntax*, but also for those students of language who consider themselves advocates of 'conceptual role semantics' today. (Roughly, the syntactic characterization is what Fodor and Lepore (1992) call 'New Testament semantics' and the semantic characterization matches up with 'Old Testament Semantics.')

above), M is a model <D, f>, and ρ provides the truth-values of compound sentences containing logical connectives, given the truth-values of their components. From the semantic perspective, a sentence of *L* is standardly considered logically true if it is true in all models. This gives us an obvious definition for a strictly Quinean language:

L=<L, M, ρ> is *semantically strictly Quinean* iff there are no sentences of *L* that are true in every model.

Under what circumstances will a (semantically characterized) language be strictly Quinean? As is well-known, a necessary condition for any sentences to be true in all models is that the same interpretation be given to certain symbols in all models, specifically, what are traditionally called the 'logical constants.' For example: 'Fa∨¬Fa' is logically true because for every assignment of 'F' and 'a', the sentence is true. The sentence would *not* be true in every model if we simultaneously varied the interpretation of '∨' (I am supposing it would keep its grammatical status as a binary sentential connective). Thus, if *no* symbols of a language are given a fixed interpretation, then that language would be strictly Quinean. But the converse does not hold (note the parallel with rules of inference in the syntactic characterization). Some symbols (but not too many) could be given a fixed interpretation in a strictly Quinean language: for example, we can let ρ contain the usual truth tables for '∧' and '∨'. Note that ρ=∅ is not sufficient to guarantee a strictly Quinean language: for example, if we introduce '=' with its usual interpretation, 'a=a' is true in every model (where 'a' is an individual constant).

Now, if we have a strictly Quinean language, at least some of the usual logical connectives will not be given fixed interpretations. Under such circumstances, we will need to modify our official definition of a language, if we still want such connectives to have a (truth-functional) meaning. One way to do this is to include in our definition of an interpreted language a fourth term ρ* that provides *in that particular interpreted language* an interpretation of the ‘logical’ connectives not covered by ρ. If we then reframe our definition of an interpreted language as <L, M, ρ, ρ*>, then we must characterize a logical truth as a sentence which is true for all M *and* for all ρ*. (Let us call a combination of a particular M and a particular ρ* an *interpretation*.) This new definition of a language obviously contains the standard one as a limiting case, namely, where ρ* is empty.

Finally, consider the following question concerning strictly Quinean languages: are there any (interesting or unifying) properties shared by all of them? I can offer only a partial and provisional answer. First, none will have a formal correlate of ‘only if’: in a syntactically characterized language, one can infer ‘p⊃p’ from the ⊃-introduction rule alone (assuming a reiteration rule is available). Similarly for semantically characterized languages: if there are sufficient resources to define the usual truth table for ⊃, then ‘p⊃p’ will be true in all models. Second, if the language has a formal correlate of ‘not’, then the language cannot contain any of the usual binary logical connectives: ‘p∨¬p’ will be true in all models/ derivable, ‘p∧¬p’ will be false in all models/ its negation will be derivable[85], etc. Finally, note that a language is strictly Quinean if and only if it is *presuppositionless*.[86] This is simply a consequence of a strictly Quinean language having no logical truths.

<u>B. Remarks on strictly Quinean languages</u>

1. Should strictly Quinean languages be considered *languages* at all? The answer may differ for different species of strictly Quinean languages. Let us call a language *maximally Quinean* iff it is strictly Quinean, and

---

[85] Or, if we prefer, we can (as Carnap does in *Logical Syntax*) characterize ‘*p* is analytically false’ syntactically as ‘Every sentence in the language is derivable from *p*.’

[86] I am using the usual definition of ‘presupposition’ here: A presupposes B ≡ if A is either true or false, then B is true (A is neither true nor false if B is not true).

- if the language is characterized semantically, ρ is empty;

- if the language is characterized syntactically, T is empty.

Now one way to frame the original question is: should we frame our original two definitions of language such that ρ and T are required to be non-empty? If such a requirement is imposed, then maximally Quinean 'languages' will not count as languages at all; some people might, with good reason, hesitate to call them languages. In the spirit of Carnapian tolerance, I forbear making any firm pronouncements on this score. I will say only that a syntactically characterized 'language' in which T=P=∅ seems to have much less of a chance of being considered a language than a semantically characterized 'language' in which ρ=∅, because in the latter case, the *model* is still there to provide some substance over and above the grammatical information. In the syntactic case, a maximally Quinean language is nothing more than a grammar, and many would balk at considering that alone a language, but in the semantic case, every name is still assigned an individual object, every predicate an extension, and so on.

The case of maximally Quinean languages need not be faced here, for as we have seen, there are strictly Quinean languages with non-empty T and ρ. These meet the canonical definitions of language used by philosophical logicians in an unproblematic way. So the question the Quinean must face is: does the fact that strictly Quinean languages have such an impoverished inferential structure[87] (or lack of presuppositions) disqualify them from being genuine languages? Again, I lean towards tolerance, for I am not entirely sure that there is a well-posed question in the neighborhood; perhaps for some purposes they would serve, but not for others. The only point I wish to make is that if someone *did* hold that strictly Quinean 'languages' were not bona fide languages—and this is not an unreasonable position—then it will

[87] 'Inferential structure' is meant to cover both syntactic and semantic conceptions of what follows from what.

be difficult for her to endorse Quine's radical critique of analyticity (viz. 'There are no sentences whose truth-value is determined in virtue of language alone'). She would have to either commit herself to the obviously unpalatable claim that there are no languages at all,[88] or assert that my formal characterization of Quine's critique fails to capture Quine's proposal adequately.[89] I will treat this second option presently, as part of the next two subsections.

2. Even if we decide, at least for certain purposes, to count strictly Quinean languages as bona fide languages, there is still a further question: could they model *our* languages remotely well? One knee jerk reaction might be: of course not!—our languages permit an abundance of inferences, not just those involving a logical connective or two alone, as the strictly Quinean languages appear to allow. The Quinean has a clear reply, however: it is indubitable that we make all sorts of inferences in our language—but the sorts of inferences we make using principles or rules of inference traditionally deemed 'logical' are not different in kind from the sorts of inferences we make using physical principles or psychological principles. In particular, the type of *justification* is identical for the various types of principles: it is, in the end, empirical. So the 'laws of logic' are (epistemologically) no different from laws of nature; it is presumably in this spirit that Quine writes: "I see no higher or more austere necessity than natural necessity" (1963/1976, 76). So Quine's view allows us to make all the inferences we usually do, but the grounds for these inferences are different from those Carnap endorses: for Quine, unlike for Carnap, the grounds are empirical. So a Quinean can work with the usual instruments of philosophical logic and model theory, using interpreted languages that are just as rich as the

[88] Though, as with all obviously unpalatable philosophical claims, an extremely able philosopher has held this view. Donald Davidson's "A Nice Derangement of Epitaphs" (1986) concludes that there is no such thing as language.

[89] Just to be explicit, the argument I have in mind is the following:
(P1) No (natural) languages contain analytic (including logical) truths or presuppositions.
(P2) Anything without presuppositions or analytic truths is not a (natural) language.
Thus, there are no languages.

usual ones, but she would understand those languages in an unusual way. There will still be certain sentences that are 'true in virtue of language,' but for a Quinean, 'in virtue of language' is not exclusive of 'in virtue of physics, psychology, etc.' Of course, there could still be other reasons to maintain that the Quinean conception of language does not model languages (sufficiently) well. For example, such a language will be virtually useless for those who study artificial languages.

3. The existence of strictly Quinean languages, I believe, provides a counter-example to Boghossian's claim (1996, 1997) that one cannot give up analyticity without adopting the indeterminacy of meaning thesis. This thesis, according to Boghossian, maintains

> that meaning facts themselves are indeterminate—that there is, strictly speaking, no determinate fact of the matter as to what a given expression in a language means. This is the doctrine I have called the thesis of indeterminacy of meaning. (1996, 362)

Boghossian considers this a problem, because most philosophers today (he reports, following Bill Lycan) reject the notion of analytic truth, but do not accept the indeterminacy thesis. That strictly Quinean languages *prima facie* constitute a counter-example to Boghossian's claim is clear, I think, in the case of semantically characterized languages: every term in such a language is assigned a (referential) meaning, and this assignment is done in the same way as in languages containing analytic truths. There is a 'determinate fact of the matter' about the meaning of (e.g.) individual constants in a strictly Quinean language, for they are assigned elements in the domain—and this is exactly how individual constants are assigned meaning in a non-Quinean language as well. The only difference, to repeat, is that fewer terms' meanings are held constant across interpretations in a strictly Quinean language.

The point made in the previous paragraph, although basically correct, stands in need of refinement and elaboration. One problem is that Boghossian's description of the indeterminacy thesis is either somewhat misleadingly phrased, or simply not Quine's own.[90] The moral Quine draws from the predicament of the radical translator is *not* that there is 'no determinate fact of the matter as to what a given expression in a language means,' but rather that there is nothing (scientifically respectable) to what a given expression means *over and above what is preserved in an adequate translation*, where an adequate translation is one that facilitates smooth conversation with a native speaker. One necessary condition on such a translation is that it must preserve what Quine calls the 'stimulus meaning' of a sentence.[91] It is only when some meaning *beyond* what is preserved in an adequate translation is attributed to a sentence that there is no fact of the matter as to whether that term has that meaning or not. To put the matter somewhat roughly, on Quine's view, there is determinate meaning, but there is no determinate *super-empirical* meaning.[92] ("Indeterminacy and Translation Again" and "On the Reasons for Indeterminacy of Translation") Thus Boghossian's claims, if taken at face value, appear to attribute to Quine a much stronger claim than the one Quine considered his arguments to have established.

So, one might wonder, do strictly Quinean languages still count as a counter-example to Boghossian's claim if we take up Quine's actual version of the indeterminacy thesis instead of the one Boghossian (apparently) forwards? The answer is *yes*, for the meanings assigned to the terms of L by M could outstrip what is preserved by an adequate translation. The meaning of a

[90] Some of Boghossian's remarks following the above quotation indicate that the former is likely the case.
[91] The stimulus meaning of a sentence is defined as the ordered pair of the sentence's affirmative stimulus meaning and its negative stimulus meaning. The *affirmative [resp. negative] stimulus meaning of t* is "the class of all the stimulations … that would prompt [the native speaker's] assent" [resp. dissent] toward t (Quine 1960, 32).
[92] A useful analogy, which Quine himself deploys (1969, 48-49), is drawn from spacetime physics. Just as modern physics rejects the notion of absolute velocity, but maintains that of relative velocity, Quine holds that questions about meaning that outstrip the observable realm are illegitimate, but the weaker notion of meaning that is strictly empirical is maintained. That is, absolute velocity : relative velocity :: Carnapian meaning : Quinean meaning.

singular term ‘t’ assigned by translation manual T_1 (very roughly,[93] $f_1(t)$) could be different from the assignment of a translation manual T_2 ($f_2(t)$), yet make no difference to observable behavior of native speakers, Quine claims, if compensatory changes are made elsewhere in the two translation manuals (or models). Since the assignments to ‘t’ differ in the two translation manuals, in one sense the meaning of ‘t’ will differ in the two cases, but this will (if Quine’s indeterminacy of translation thesis is correct) make no difference to facilitating smooth interactions with a native speaker.[94] Thus, even if we do not take Boghossian at face value (‘meaning is indeterminate’), but as intending the actual Quinean indeterminacy thesis (there is no super-empirical meaning), the existence of strictly Quinean languages still shows that one can repudiate analyticity completely without committing oneself to the indeterminacy thesis.

Someone unsympathetic to the claim that we can relinquish analyticity without the indeterminacy thesis could level the following objection against my strictly Quinean languages. If no (or very few) terms of a language are assigned fixed interpretations, then that very fact makes it seem that meaning *is* indeterminate for all (or almost all) terms of that language. How else would one characterize meaning being indeterminate, if not by allowing terms to take on varying meanings? Note first that the plausibility of this objection plays upon an ambiguity in ‘fixed’ vs. ‘indeterminate’: *within* a given interpreted (semantically characterized) language, every term has a perfectly determinate meaning; it is only when the interpretation is varied that certain terms’ meanings become in any sense indeterminate. Additionally, at least three further circumstances tell against this objection. First, if one maintains that the only genuinely

---

[93] This is ‘very rough’ because a translation manual is a function (or relation) from words in one language to words in another, whereas an interpretation function takes words to objects, sets, etc. The point remains the same, however: two incompatible interpretation functions can (if Quine’s indeterminacy of translation thesis is correct) both account for all observable linguistic behavior of a group of speakers of a particular natural language.

[94] I recognize that this is a case of ‘inscrutability of reference,’ not a classic ‘indeterminacy of translation’ case, but the point holds nonetheless.

meaningful terms of a language that are those assigned fixed meanings, then the only meaningful terms of a language will be its logical constants. But that, presumably, is an intolerable consequence (are the only meaningful terms of *our* languages the logical connectives?), so we should reject the premise. Second, we saw that, in certain cases, we can have terms with fixed interpretations (so at least *some* determinacy) in a language without any analytic truths. Third, even in a (semantically characterized) strictly Quinean language, every sentence is true or false (assuming our underlying language is classical), so it is at least meaningful in that minimal sense.

Finally, before moving on from Boghossian's treatment of analyticity, I will lob two more parting shots, in part because his mistakes have been unqualifiedly endorsed and propagated by (Sober 2000) and (Margolis and Laurence 2001). Though Boghossian wants to rehabilitate a notion of analyticity, he considers the notion of analyticity he finds in the logical empiricists severely wanting. I will argue that his criticisms miss the mark, either because his arguments are misleading at best, or because they seriously distort the historical record.

Boghossian dubs the notion of analyticity that he finds in the logical empiricists, and which he wants to attack, 'metaphysical analyticity.' First, calling any of part of Carnap's conceptual apparatus 'metaphysical' is, at the least, not an actor's category—Carnap famously rejects metaphysics of all shapes and sizes. So what does Boghossian mean by 'metaphysical analyticity'? Under this conception a sentence is analytic if its truth "is fixed exclusively by its meaning and not by the facts." Boghossian considers such a notion untenable. Why? Boghossian asks:

"Isn't it in general true—indeed, isn't it a truism—that for any statement **S**,
 S is true iff for some **p**, **S** means that **p** and **p**?
How could the *mere* fact that **S** means that **p** make it the case that **S** is true?"

(Boghossian 1996, 364; cited in (Sober 2000, 252) and (Margolis & Laurence 2001, 294))

Though he does not say so in exactly these words, Boghossian appears to be claiming that the truth of ⌜for some **p**, **S** means that **p**⌝ is never sufficient for the truth of ⌜**S** is true⌝, for any choice of **S** and **p**—that would be a reasonable way of cashing out the notion of having the truth of a sentence being fixed exclusively by its meaning.

If we do understand Boghossian's view in this way, then his claim is misleading at best. Consider a biconditional of the form

p if and only if q

If such a biconditional is true, we usually say that q is a necessary and sufficient condition for p. But, as we teach undergraduates in Intro Logic classes, if this biconditional is true, then (within the classical propositional calculus) so is

p if and only if (q and (if r then r))

(Any other theorem of the propositional calculus could be substituted for 'if r then r.') If we just read off the surface structure of such a sentence, one might think that q was no longer sufficient for the truth of p—because there appears is a second condition that has to be met in order for p to be the case, namely *if r then r*. Of course, strictly speaking, this is true: every sentence of the propositional calculus presupposes the truth of all the logical truths of the propositional calculus. However, it would be misleading to say that the truth of *q* is not a sufficient condition for the truth of *p* in our original biconditional—that is simply not the way we usually talk about sufficient conditions.

Hopefully the parallel with Boghossian's claim is clear. I certainly agree that his 'truism' is true. However, when the **p** in his schema is a logical truth, then the truism will be analogous to the second biconditional

p if and only if (q and if r then r)

That is, in the usual sense of 'sufficient condition,' we will have a case in which (contra Boghossian) the formula ⌜for some **p**, **S** means that **p**⌝ *is* sufficient for ⌜**S** is true⌝. To say otherwise, we would have to give up either the material we teach in introduction to logic or the usual conception of 'sufficient condition.'

Because Boghossian thinks this first conception of metaphysical analyticity is a non-starter, he considers a seconds version. The truth of a sentence is fixed solely in virtue of its meaning if "in some appropriate sense, our meaning **p** by **S** *makes it the case that **p***" (1996, 365). My final objection to Boghossian's treatment of analyticity is that this is an extremely misleading way of describing what Carnap (and other friends of analyticity) took the notion to be (and Boghossian apparently thinks it does (*ibid*.)). Boghossian rejects (along with Sober as well as Margolis and Laurence) this conception of analyticity for obvious reasons:

"Are we really to suppose that, prior to our stipulating a meaning for the sentence 'Either snow is white or it isn't' *it wasn't the case that snow is white or it wasn't?* Isn't it overwhelmingly obvious that this claim was true *before* such an act of meaning, and that it would have been true even if no one had thought about it, or chosen it to be expressed by one of our sentences?" (*ibid*.).

This is clearly an untenable position, and Carnap did not hold it. Rather, Carnap could (and perhaps would) reply to Boghossian roughly as follows. If we are going to communicate linguistically, we must use some language or other to do so. If the rules governing our language are expressively and inferentially rich enough to assert and infer much of what we assert and infer in everyday language and scientific communication, then the truth-values[95] of certain sentences are determined simply by the rules of the language we use alone. (Recall from our discussion of Quinean languages that no languages are both inferentially rich and presuppositionless.) If we make the rules explicit, as we do in formal languages, then we can

[95] Putting things in terms of truth-values is of course the semantic way of putting the point; from a syntactic point of view, we would have to say that certain sentences are *derivable* from the rules of the language we use alone.

‘compute’ which sentences have their truth-values determined by the rules of the language—they are, of course, the logical truths of the language. (Or, if we have characterized our language proof-theoretically instead of model-theoretically, they are the theorems of the language.) This view is expounded in the *Tractatus*, and it is Carnap’s basic view as well. Contrary to Boghossian’s explication, the linguistic rules, whether implicit or explicit, syntactic or semantic, are of course incapable of affecting what is the case—but they nonetheless can determine the truth-values of certain sentences of the language we speak. This is perhaps most vividly illustrated in the case of a language whose rules of inference are inconsistent: every sentence of the language is derivable (and is thus *analytically* true), including sentences of the form ‘p and not p’, but that does not mean that contradictions somehow exist ‘out in the world.’

C. Moderately Quinean languages

Thus far, we have discussed languages whose inferential structure, whether syntactically or semantically characterized, does not suffice to determine the truth-value of *any* of its sentences. I consider such languages to be an adequate representation of Quine’s basic view of analyticity. However, there is another interpretation of Quine’s view extant today, which can be supported by certain remarks Quine makes in various texts. (I think this interpretation is suboptimal, and will explain why below; for now, I will just sketch it). This interpretation takes Quine to be forwarding a weaker claim: the usual or standard ‘logical truths’ (= theorems of first order logic with identity) are allowed to be ‘true in virtue of meaning’ (or something similar); all the other so-called analytic statements (e.g. ‘No bachelor is married’), however, are rejected. This is phrased today as: truth (or inference) in virtue of ‘logical’ or ‘structural’ meaning is acceptable, but truth (or inference) in virtue of 'lexical' meaning (i.e. individual word meanings) is not[96] (Lycan 1994, 235; Fodor and Lepore 1992, §3). Let us call any language that contains *no* sentences true in virtue of their lexical meanings alone, but does contain sentences true in virtue of their structural meaning, a *moderately Quinean language.* Such a language would perhaps avoid the worries engendered by the strictly Quinean languages, namely: are they languages at all, and if so, are they plausible or reasonable models of interesting and useful languages?

Now, I think the moderately Quinean position may not be tenable—at least not for someone who is committed to sufficiently many of Quine's other arguments and claims. That is, the position is logically consistent, but lacks a principled justification. At the most basic level, the problem is this: (most of) the criticisms Quine levels at analyticity appear to apply with equal force to both structural analyticity and lexical analyticity alike. So if one is persuaded by Quine's arguments to adopt the conclusion that there are neither pure black nor pure white fibers in the lore we inherited from our forefathers, structural analyticity will not receive an exemption from this rule—at least not from Quine (or a committed Quinean). For example, certain people consider the following to be a convincing indictment of analyticity: none of our beliefs are immune from revision in the light of new experimental evidence; given that both putatively analytic truths 'respond' to empirical evidence in this way, it seems Carnap is propounding a 'distinction without a difference.' What one ultimately makes of this argument is irrelevant to my present purposes; the only point I wish to make is that this argument applies equally well to structural analyticity as well as lexical analyticity.

In order to make the problem I purport to see more general, one would need a list of all of Quine's arguments against analyticity, and show that each of them (on each of their reasonable interpretations) applies to sentences true in virtue of structural meaning; I shall not undertake this

[96] Actually, if one refuses *all* truths-in-virtue-of-lexical-meaning, then the usual treatment of '=' will have to go too, since it is standardly considered an interpreted dyadic predicate.

task here, since the amount of set-up work needed would be too great. Bill Lycan has offered a second kind of argument against the tenability of a moderately Quinean view of analyticity: there is no sharp distinction between lexical meaning and logical-structural meaning, or as he puts it, "mere lexical implications seem to *shade off* into clearly logical entailments by fairly smooth degrees" (1994, 237). This objection is in addition to the more often heard worry that there is no principled way of isolating all and only the logical constants, i.e., the distinction between logical constants and other morphemes is arbitrary.

In short, the basic idea is this: if you find Quine's arguments compelling, you will be pushed in the direction of strictly Quinean languages. On the other hand, if you find strictly Quinean languages objectionable for one reason or another (e.g., their extreme inferential austerity) the proponent of analyticity then has a fairly effective wedge. For if we allow Quine some analytic truths, why begrudge Carnap a few more? Of course, there are places where there are sharp enough criteria where one *could* draw a line (e.g., at decidability, so that we would stop at propositional logic plus monadic predicate logic, or at completeness, so that we could stop at first-order logic with identity), but are any of these independently motivated beyond just a desire to banish lexical analyticity?

Let us return to properly historical matters for a moment: what did Quine himself think? Certain interpreters do take Quine to be attacking lexical analyticity only (Lycan 1994, 236) (Fodor and Lepore 1992). This interpretation is not unmotivated: in "Two Dogmas" and elsewhere, Quine apparently reserves the most vitriol for 'meaning postulates' (as opposed to theorems of the propositional calculus, e.g.). Furthermore, Quine makes assertions such as the following, from "Carnap and Logical Truth":

> Logical truth… is… well enough definable… But when we would supplement the logical truths by the rest of the so-called analytic truths, true by essential predication [e.g., 'all bachelors are unmarried men'], then we are no longer able even to say what we are talking about.
> (1963, 404).

Other apparent approbations of logical truth appear in that article.[97]

However, in a footnote in *Word and Object*, Quine makes it clear that he considers dubious the claim that even the traditional logical (i.e., 'structural') truths are analytic.

There is a small confusion that I should like to take this opportunity to resolve… We who are less clear on the notion of analyticity may…seize upon the generally conceded analyticity of the truths of logic as a partial extensional clarification of analyticity; *but to do this is not to embrace the analyticity of the truths of logic as an antecedently intelligible doctrine.* I have been misunderstood on this score by Gewirth and others. (1960, 65n; italics mine)

In other words, the philosopher who considers analyticity dubious can point to the theorems of first-order logic and say to a pro-analyticity opponent: 'I grasp that you wish to explain what makes those sentences true, but I find your explanation (viz., that they are true in virtue of meaning) unintelligible.' Put otherwise, ostending the class of theorems of first-order logic resolves the objection that analyticity qua explicandum is vague, but no others. This shows, I think, that Quine himself would embrace the strictly Quinean view of language, not the moderate one. Quine argues not just for the unintelligibility of truth in virtue of lexical meaning, but of truth in virtue of structural meaning as well. Quine does think there is such a thing as logical truth, but he does not think such truths are true in virtue of the language in which they are couched.

[97] "Our notion of logical vocabulary is precise. And so, derivatively, is our notion of logical truth" (1963, 402).

CHAPTER V. DIRECT DISCUSSIONS OF ANALYTICITY (L-TRUTH) IN 1940-41

This chapter is an exposition and analysis of the characterizations of analytic truth, and the arguments concerning it, that appear in the 1940-41 discussion notes. Carnap prefers to characterize linguistic concepts in general, and analyticity in particular, in semantic and intensional terms, Tarski in semantic but extensional approach, and Quine syntactic and intensional language. I briefly compare some of the relative merits of each approach before examining Tarski and Quine's objections in the notes to Carnap's notion of analyticity. Tarski's two most well-developed objections to the analytic/ synthetic distinction are reconstructed and evaluated. The first, a version of the 'Any sentence may be held true come what may' argument, is shown to either misunderstand Carnap's position, or not conflict with it. Tarski's second objection, which is not familiar from public debates over analyticity, is based upon Gödel's incompleteness results. This argument does not tell decisively against the analytic/ synthetic distinction either, unless we characterize language and meaning fundamentally syntactically.

Quine, unlike Tarski, does not articulate complete arguments against Carnapian analyticity here, rather, he simply voices disagreement with two of Carnap's commitments. Nonetheless, Quine's points of contention do allow us to characterize the philosophical differences between the two men cleanly, and thereby understand the historical grounds and development of Quine's critique of analyticity. The first difference between the two is that Carnap holds sentences of the form '*p* is analytic' to be analytic, whereas Quine considers them synthetic. Second, Quine considers Carnap's characterization of analyticity in modal terms fundamentally unclear. Motivations and arguments (where possible) for each side are reconstructed, drawing on published work as well. Finally, this material suggests another historical conjecture concerning the radicalization of Quine's critique of analyticity (see IV.2). In *Logical Syntax*, Carnap is explicitly committed to analyzing logico-linguistic concepts in syntactic and extensional terms—Quine's lifelong preferred method. In the mid-thirties, however, Carnap shifted toward semantic and intensional treatments of linguistic concepts, but Quine did not follow him. Thus Quine's break with Carnap is not simply a matter of Quine changing his views, but of Carnap's views changing as well.

1. WHAT IS ANALYTICITY (CIRCA 1940)?

The aim of this chapter is to examine and analyze the treatment of analyticity in the 1940-41 discussion notes and related texts. Before proceeding, a potential terminological difficulty must be dispelled. Neither the word '*analytisch*' nor its cognates appear in Carnap's discussion notes of 1940-41. The phrase that does appear, and which corresponds for Carnap to 'analytic' at this time, is '*logically true*' (abbreviated as 'L-true'). Carnap explicitly states in print that his notion of L-truth is intended to be a modern, scientific version of the older, traditional philosophical notion of analyticity.[98] What makes this terminology somewhat unfortunate for us is that the current notion of *logical truth* is not identical to *analytic truth*. For the later Quine, the notion of logical truth (*qua* theorem, i.e. provable sentence within a calculus) is intelligible, whereas analyticity traditionally conceived is not. Thus Quine will *not* maintain that logical truths are analytic, at least in the Carnapian sense of 'true in virtue of the meanings of the words alone.' However, to complicate matter further, Quine finds certain empirical characterizations of analyticity acceptable at various points in his career; e.g., *Word and Object* and *Roots of Reference* both contain reformed usages for 'analytic.'

[98] Carnap writes: "there is the concept of logical truth, truth for logical reasons in contradistinction to empirical, factual reasons. The traditional term for this concept is 'analytic'; we shall use the term '*L-true*' (1942, 60-61; cf. 251).

Thus, despite initial appearances, 'L-truth' in 1941 should be interpreted (into our modern idiom) *not* as 'logical truth' but rather as 'analyticity,' in the full-blooded Carnapian sense of 'truth in virtue of (the meanings of) language, independently of any matters of fact.' (Carnap regards this sense of L-true/ analytic as rough—it is the explicandum, not the explicans.) This notion, not modern logical truth, is both Carnap's grail and the later Quine's target. (See (Creath 1990, 303) where Carnap spells out clearly the terminological differences between himself and Quine.) In the 1941 notes, Quine proposes "a criterion for logically-true: either logically provable or transformable through synonyms into a logically-provable sentence" (102-63-03, Jan.20, 1941). The first disjunct corresponds to the notion of theorem (a sense of logical truth that Quine never abandons), whereas the second disjunct is precisely the notion of analyticity attacked in "Two Dogmas" and elsewhere. This quotation shows that 'logically true' or 'L-true' in these notes should not be taken in the sense of 'theorem of a proof calculus,' or even as 'true in all models,' but rather as 'analytic.' In "Truth by Convention," Quine writes: "an *analytic* statement is commonly explained merely as one which, on replacement of definienda by definientia, becomes a truth of logic" (1936/1976, 87)[99] (in a footnote he attributes this formulation to Frege); analytic statements are precisely those sentences that are 'true by convention' (*ibid.*, 77). And in *Mathematical Logic*, Quine proposes to use 'analytic' for any logico-mathematical claims that are considered to be "'merely verbal transformations'" or 'disguised repetitions'" (1940/1958, 55).

Before plunging into the details of Tarski, Quine, and Carnap's differing conceptions of analyticity, a very rough schema of their approaches may help us see the forest while examining the trees. Carnap thinks analyticity should be treated as fundamentally *semantic* and *intensional*, Tarski agrees with Carnap that it is semantic, but holds that our account of it should be couched

[99] Carnap does not accept this characterization of analyticity, for it admits of counterexamples: see (Creath 1990, 304-305).

in extensional language as in his *Wahrheitsbegriff*, while Quine holds that the concept of analyticity should be, at bottom, cashed out in syntactically and extensionally. Each of the three chooses his approach not because of any particular view he has about analyticity, but because of much more general views he holds on the proper way to analyze language scientifically. That is, Carnap (by the late thirties) considers semantics a powerful philosophical tool and has no aversion to intensional languages (as *Meaning and Necessity* makes abundantly clear); Tarski is a great apostle of semantic methods, but all his important work is done using extensional languages, as he himself stresse to Carnap (090-16-09); and Quine tells us that he developed a very strong preference for extensional languages even before he left Oberlin. So, in a very general way, each of the three philosophers attempts to analyze 'analytic' during 1940-41 along roughly the same lines he would analyze any other term in scientific philosophy.

A. Carnap and Tarski

Carnap's basic conception of analyticity in the early forties has already been outlined in IV.1: a sentence *s* of language *L* is analytic-in-*L* if and only if the semantic rules of *L* determine the truth-value of *s*. This characterization appears in the discussion notes (090-16-11, 102-63-03), though it marks a shift from the characterization in 1939's *Foundations*: "In the Encyclopedia brochure [*Foundations*], I took 'L-true' to be 'true on the basis of the meaning of logical signs alone.' In the new MS [*Introduction to Semantics*]: on the basis of all signs" (102-62-03). This generalization is necessary for 'All mares are horses' to count as analytically true in English, a consequence Carnap considers desirable. This characterization of analyticity or L-truth is not intended as a formal definition. In *Meaning and Necessity*, Carnap calls this semantic-rule characterization a "convention," "an informal formulation of a condition which any proposed definition of L-truth must fulfill in order to be adequate as an explication of our explicandum"

(1947, 10). And in *Introduction to Semantics,* Carnap points out that 'true-in-virtue-of-semantic-rules' cannot be a metalinguistic definition of L-truth, since '… is a semantic rule' belongs to the *meta*metalanguage.

Carnap presents another characterization of analyticity in the discussion notes, and in roughly contemporaneous print. Carnap (apparently in his own voice) considers the following two definitions of analytic truth (where 'S' abbreviates 'semantic system' and 'C' abbreviates 'formal calculus'):

a_i is L-true $=_{\mathrm{df}}$ 1. a_i is true in every state of affairs in S.
2. a_i is true for each model of C.
(090-16-11)

What does each of these two definitions amount to? Let us take them in order. We have already seen (in IV.1) what a semantic system S is: essentially, an assignment of individuals to names, of properties and relations to predicates and relation letters, and the usual rules for logical particles familiar from the recursive clauses of Tarski's definition of truth. But what is a 'state of affairs' for Carnap at this time?

state = assignment of primitive descriptive predicates of the corresponding language to the individuals (of the universe of discourse of the language). Then each $\mathbf{pr}^1$ [GF-A: monadic predicate] is coordinated with a class of individuals, each $\mathbf{pr}^2$ [GF-A: binary relation letter] is coordinated with a class of ordered pairs of individuals.
(090-16-11)

In order for such an assignment to qualify as a full-blown state of affairs, the assignment must be complete, in the sense that every n-ary relation letter must be assigned a class of ordered n-tuples. The intuitive justification for allowing these assignments to vary within a single semantic system is presumably that any 'bare' (i.e., property-less) individual can bear any logically possible property or relation. (And in a Carnapian semantic system, individuals are 'bare' in this sense: the only information the semantic system provides specifically about them is their names.)

Additional conditions are imposed on object languages whose primitive predicates express properties that are not 'logically independent,' so that not all such assignments are allowed as genuine states of affairs. For example, in any particular state of affairs, the class assigned to the predicate 'mare' must be a subset of the class assigned to the predicate 'horse,' since the property of being a mare is only instantiated by entities also having the property of being a horse. When all the primitive predicates of the object language designate 'logically independent' properties, however, there are no such additional constraints (090-16-11). This mirrors the characterization of L-state in *Introduction to Semantics* §19K-L. Also, Carnap uses this framework to characterize a notion of synonymy: two predicates are synonymous if they "have the same extension not only in the actual world, but rather in every possible world, thus in every total-state [*Gesamtzustand*] ('*state'* in Semantics (I))" (102-63-07). Finally, the characterization of 'L-true' as 'true in all states of affairs' shows, perhaps more perspicuously than the 'true in virtue of the semantic rules' formulation, why Carnap held L-truth to be coextensive with necessary truth (090-16-25).

Now let us consider the second definition of L-true above, which uses the concept *model* instead of *state of affairs*, and *calculus* instead of *semantic system*. This definition of L-truth corresponds to the current model-theoretic notion of logical truth. Tarski introduces and uses this framework, which he characterizes thus:

> Models. Tarski apparently refers to a partially interpreted calculus, namely, all logical symbols are interpreted; for the usual signs, it is only determined that they are descriptive; but their interpretation is left open. A model for this system = a sequence of *n* entities, which are coordinated (as designata) to *n* descriptive signs.
> (090-16-11)

This is similar, if not identical, to the framework for semantics that Tarski uses in his *Wahrheitsbegriff* monograph and "On the Concept of Logical Consequence" (1936/ 1983, 416-7). It is also close to Carnap's notion of state of affairs of a semantic system. The primary difference is that Tarski does not first interpret predicates and relation letters as properties and relations, as Carnap's semantic system does. As a result, the 'additional conditions' imposed upon states of affairs involving logically dependent properties (such as *mare* and *horse*) are *not* imposed on the models—under the model/ calculus framework, 'All mares are horses' will not come out as L-true. Actually, this requires qualification: it holds only if 'mare' and 'horse' are (treated as) *primitive* predicates in the language. In "On the Concept of Logical Consequence," Tarski makes provision for non-logical constants that are defined; thus if we have, as a part of our language, the definition 'mare $=_{df}$ female horse,' then 'All mares are horses' will be L-true if we demand that all non-primitive constants be eliminated before applying the test for L-truth (1936/1983, 415). We can still make out a difference between Carnap and Tarski, though, in that Carnap would still want 'All mares are horses' to be analytic, even if the object language did not explicitly contain a definition of 'mare' (Creath 1990, 305). However, as Carnap notes, as long as all the properties designated by terms in a semantic system are logically independent, the states of affairs/ semantic system apparatus differs from the model/ calculus one only terminologically. And if, within the Tarskian model/ calculus framework, we have appropriate definitions for all the predicates expressing logically dependent properties, then any substantive difference between the two approaches also disappears.

Carnap's discussion notes record little of Tarski's own positive view of L-truth. It is not clear from the notes whether it is Tarski or Carnap who first raises the possibility of defining analyticity in terms of models; however, Tarski had already given this definition in print: "a class of sentences can be called *analytical* if every sequence of objects is a model of it" (1936/1983, 418). The only two direct statements that Tarski makes about L-truth in the notes are the following: "Tarski: We only want to apply 'logically true' and 'logical consequence' when it holds for every meaning of the non-logical constants" (102-63-12). (Again, Tarski may mean here ‘every meaning of the *primitive* non-logical constants.’) This formulation is not especially interesting or novel, but it does highlight one fact worth noticing: he considers logical truth to be fundamentally matter of *meaning*, i.e. of semantics, not primarily a syntactic affair. In this, he differs from the Quine of even 1940, as we shall see in the following subsection.

The second comment Tarski makes about L-truth occurs in a discussion about how to introduce a term ‘T’ representing temperature into their language of science (the sentence ‘T(a, x_0, y_0, z_0, t_0)’ is the formal correlate of the colloquial assertion that the temperature at space-time point (x_0, y_0, z_0, t_0) is a).

Just as we define ‘descriptive’ through an ultimately arbitrary enumeration, in the same way we also define the further concepts (‘L-true$_2$’ or whatever) through an enumeration of sentences in S [the language] involving T, so that the logical consequences (‘L-implies$_1$’) are taken as L-true$_2$. These sentences signify, for example: [1] T only takes quintuples, and [2] for true quintuples of real numbers, no 2 quintuples differ only in their first element, and [3] that for every quadruple, there is a quintuple with an appropriate first element; but furthermore also: [4] the function should be continuous, should have a first derivative, perhaps also a second etc. (090-16-10)

Note how different this is from Carnap’s proposal to characterize L-truth and L-implication. By simply stipulating (in the metametalanuguage) which sentences of the form “‘..T..’ L-implies ‘…T…’” are L-true, Tarski completely bypasses the intensional notions *state of affairs* and *logical possibility* (as Quine happily notes immediately thereafter).

B. Quine: analysis of language and analyticity should be syntactic and extensional

As discussed in IV.2, the Quine of 1940 does not consider the notion of analyticity (or ‘logical truth’) unintelligible, but his conception is nonetheless different from Carnap’s contemporaneous one. The fundamental differences stem from Quine’s preference for syntactic and extensional analyses of language over Carnap’s semantic and intensional approach. Here is the full context of Quine’s characterization of L-truth in the notes:

Carnap: I am inclined to take the following sentence as L-true as well (it would be logically or mathematically true): '$(x)(P(x)\supset Q(x))$', where 'P' is interpreted as 'black table,' and 'Q' is 'black.'
Quine: Yes, you can arrive at that, when you state an interpretation via the definition of 'synonym,' as a relation between expressions of the object language and either the metalanguage or perhaps a richer object language. ('Synonym' is intended so that it holds only for L-equivalent predicates, not for F-equivalent ones.) The named sentence then corresponds to a sentence '$(x)(P_1(x).P_2(x)\supset P_1(x))$', which is logically true. So a criterion for logically-true: either logically provable or transformable through synonyms into a logically provable sentence. (102-63-02)

This characterization of L-true is exactly like the characterization of analyticity in "Truth by Convention," except that here synonymous expressions take the place of definitions (1936/1976, 87). This account of L-truth, relying on definitions and/or synonyms, is clearly different from Carnap's, as seen in the previous subsection: there is no mention of models or states of affairs, so Quine's conception does not use the semantic concepts Carnap and Tarski deploy.

Of course, '*p* is synonymous with *q*' can certainly be construed as semantic, since its intended interpretation is that *p* and *q* have the same *meaning*. (And Quine holds out hope for an eventual clarification of the notion of meaning throughout the 1940s.) However, two considerations tell against viewing Quine as subscribing wholeheartedly to a fundamentally semantic approach to language analysis. First, given that Quine's characterization of L-true in 1941 is identical to that in "Truth by Convention," save that synonyms replace definitions, Quine could be thinking of '… is synonymous with - - -' as '$\ldots =_{df}$ - - -', and thereby interchangeable with a thoroughly syntactic treatment, if one treats definitions purely syntactically—and it appears Quine may think of them, at least at times, in this way, though the matter is not clear (Creath 1990, 297). Furthermore, even if synonymy is not closely tied to definition in 1941, Quine sees this semantic term as an intermediary step, a means to an end: the concept of synonymy is used to convert a sentence into a form whose truth-value could potentially be determined syntactically or proof-theoretically. Thus for Quine, semantics is in some sense secondary or subsidiary to syntax, whereas the opposite is true for Carnap and Tarski—for example, Tarski's famous definition of logical consequence (Tarski 1936/1983) is framed entirely model-theoretically.[100]

This view of Quine receives strong support from a comment Quine makes in his December 1940 lecture, in which he declares the syntactic characterization of logical truth to be 'more elementary' than the semantic one.

> <u>'Logically true'</u> can be defined syntactically, and even <u>protosyntactically[101]</u> (following Gödel's completeness proof):
> Infinite sets of axioms of quantification (axiom schemata, as in M.L. [*Mathematical Logic*]) and modus ponens.
> This is more elementary than the semantic characterization with the help of 'true.'
> (102-63-04)

Thus for Quine, calling a sentence 'logically true' means, first and foremost, that sentence is a *theorem*, a provable sentence. The definition of logical truth suggested in this lecture follows that in Quine's *Mathematical Logic*,[102] a book with a thoroughly syntactic flavor. He there claims:

> Insofar as logical truth is discernible at all, standards of logical truth can be formulated in terms merely of more or less complex notational features of statements; and so for mathematics more generally.
> (1940/1958, 4)[103]

So the meaning conferred upon notational features need not be considered, when attempting to discern whether a given sentence is a logical truth. And the characterization of the logical truths Quine first offers in "Truth by Convention" (and repeats in *Mathematical Logic*) is also more

---

[100] And in his Autobiographical essay, in the section entitled "Semantics," Carnap writes: "I had given the first definition of // logical truth in my book on syntax. But I now recognized that *logical truth in its customary sense is a semantical concept*. The concept which I had defined [in *Logical Syntax*] was the syntactical counterpart of the semantical concept" (1963, 63-64, italics mine). And in *Introduction to Semantics*, in discussing the definitions of various logical concepts, Carnap states: "This change of the definition from a syntactical to a semantical one is an essential improvement" (1942, 87). This is exactly the converse of Quine's view of the matter, as the next quotation from Quine makes clear.

[101] For Quine at this time, 'protosyntax' refers to logic of *ML* without the set-membership relation, plus one syntactic primitive 'Mxyz,' whose intended interpretation is too convoluted to give here; see (Quine 1940/1958, 288).

[102] See (Quine 1940/1958, 305).

[103] Shortly thereafter, Quine writes: "standards of logical or mathematical truth are to be formulated in terms merely of the observable features of statements" (1940/1958, 5).

syntactic than that of Tarski and Quine: "A statement is logically true if it… remains true when all but its logical skeleton [composed of the propositional connectives and the quantifiers] is varied at will; in other words, if it is true and contains only logical expressions essentially, any others vacuously" (1940/1958, 28).[104] Thus, though Quine may use semantic notions in the forties and afterwards, that does not mean that he adopts a fundamentally semantic view of L-truth in particular or of language analysis in general. This is in direct contrast to Carnap in the 1940s, who explicitly says that the proof-theoretic view of language is secondary to the semantic one in many respects (Carnap 1942, §39). One may object that Quine, in "Notes on the Theory of Reference" (Quine 1953, 132-138), is willing to admit a truth-predicate. However, closer attention to the details of that article supports my contention: Quine admits 'is true' only because the schema associated with Tarski's convention T

'…' is true if and only if …

exists to prescribe the use of the truth-predicate. Without this schema, it appears that Quine would not admit semantic notions.[105] This again supports my claim that Quine, while willing to deploy semantic vocabulary, nonetheless used it only when he was absolutely certain it was unobjectionable.

---

[104] See (Quine 1936/1976, 81). "An expression will be said to occur *vacuously* in a given statement if its replacement therein by any and every other grammatically admissible expression leaves the truth or falsehood of the statement unchanged. Thus for any statement containing some expressions vacuously there is a class of statements, describable as *vacuous variants* of the given statement, which are alike it in terms of truth or falsehood, like it also in point of a certain symbolic make-up, but diverse in exhibiting all grammatically possible variations upon the vacuous constituents of the given statement. An expression will be said to occur *essentially* in a statement if it occurs in all the vacuouc variants of the statement, i.e., it forms part of the aforementioned skeletion." (*ibid.*, 80).

[105] "These [semantic] paradoxes seem to show that the most characteristic terms of the theory of reference, namely, 'true,' 'true of,' and 'naming' (or 'specifying'). Must be banned from the language on pain of contradiction. But this conclusion is hard to accept, for the three familiar terms in question seem to possess a peculiar clarity in view of these three paradigms:
(4) '---' *is true* if and only if ---
(5) '---' *is true of* every --- thing and nothing else
(6) '---' *names* --- and nothing else.
(1953, 134)

From Quine's posthumous papers at Harvard, we also know that his feelings towards semantics just after the 1940-41 academic year were mixed at best. On November 5, 1941, Quine gives a talk entitled "The Scope of Semantics" to the Philosophy Club at Boston University. His concluding remarks are as follows:

So I feel that semantical analysis is of crucial importance to philosophy. But at the same time I feel that many of the most prominent claims that have been made for semantics are as yet unwarranted.

I can't see that any really objective, scientific progress along semantic lines has been made in connection with such supposedly semantic topics as:
a) elementalistic, levels of abstr., multiordinal, &c.,
b) meaningfulness,
c) protocol sentences,
d) analytic vs. synthetic sentences,
e) indicative vs. expressive use of language.

Perhaps progress <u>will</u> be made on some of these topics, and I expect semantics may prove useful in these topics as elsewhere; but I can think of nothing that I would point to, in any of these connections, as a definitive semantical accomplishment.

(MS STOR 299, Box 11)

Clearly, Quine has not given up completely on semantics in 1941, but the seeds of his later discontent are present. Some commentators, including Donald Davidson[106] and Peter Hylton (2001), consider the crux of Quine's criticism of Carnap to be that the kinds of explanations Carnap offers are not genuine explanations or unscientific explanations. As Quine himself says, an explanation that appeals to meanings is a "*virtus dormitiva*" explanation, since it "engenders the illusion of having explained something" (1953, 48). We see that more radical Quinean attitude foreshadowed in this 1941 lecture.[107]

Virtually all of Quine's logical work was avowedly and proudly extensional. The aim of his dissertation, for example, was to capture the logic of *Principia Mathematica* without recourse

[106] "My objection to meanings in the theory of meaning is not that they are abstract or that their identity conditions are obscure, but that they have no demonstrated use" (1984, 21).

[107] How might a Carnap respond to the *virtus dormitiva* charge? One reply is: for someone hoping for an explanation in empirical or even causal terms, a logico-mathematical explanation *will* seem to be an empty explanation. (This is why Hartry Field, in (Field 1972), calls Tarski's reduction of truth a 'bogus' explanation: Field is expecting a causal story, while Tarski provides only a mathematical one.) And Carnap's account of analyticity is logico-mathematical.

to any intensional concepts—in particular, propositional functions. One of Quine's first publications, 1934's "Ontological Remarks on the Propositional Calculus," argued that the concept *proposition* did not deserve a place in the total conceptual edifice of science. In "Confessions of a Confirmed Extensionalist" (2001), Quine claims that he was already committed to extensional approaches during his college days at Oberlin. And in the same piece, Quine remembers being pleased during 1932, his traveling fellowship year, to find that the extensional approach was taken for granted by Carnap and the Poles whom he visited during that time.

These last facts suggest a historical conjecture about the development of Quine's critique of analyticity. Quine's view in 1940 about how language should be analyzed is quite close to Carnap's in *Logical Syntax* six years previous: in that book, Carnap's analysis of every logical characteristic (analyticity included) is extensional and syntactic,[108] which, as we have seen, is Quine's preferred approach. Quine's public and private view of how language should be analyzed around 1940 (as well as before and after) is fundamentally the same as Carnap's view in *Logical Syntax*; but it is rather different from the explicitly semantic and intensional viewpoint Carnap advocates from *Foundations* onward. Thus, Quine's break with Carnap over analyticity can be seen as due to *Carnap changing* his position as much as Quine changing his: Carnap

[108] Carnap writes: "But even those modern logicians who agree with us in our opinion that logic is concerned with sentences, are yet for the most part convinced that logic is equally concerned with the relations of meaning between sentences. They consider that, in contrast with the rules of syntax, the rules of logic are non-formal. In the following pages, in opposition to this standpoint, the view that logic too is concerned with the *formal* treatment of sentences will be presented and developed. We shall see that the logical characteristics of sentences (for instance, whether a sentence is analytic, synthetic, or contradictory; whether it is an existential sentence or not; and so on) and the logical relations between them… are solely dependent on the syntactical structure of the sentences." (1934/ 1937, 1-2) Carnap also, in *Logical Syntax*, subscribes to the "thesis of extensionality," which states that "a universal language of science may be extensional" (245). Furthermore, Carnap considered intensional language suspect, on the grounds that many sentences couched within it are quasi-syntactic (or in the 'material mode'), and thus misleading at best (246). Finally, Carnap's extensionalism predates the *Syntax*: in the *Aufbau*, he declares that the logical value of a sentence is its truth-value alone; Fregean 'sense' is has "psychological" or "epistemic" value only (1928/1967, 84).

moved towards a semantic and intensional approach to language analysis, while Quine retained the syntactic and extensional approach exhibited in *Logical Syntax*. For example, in the 1934 "Lectures on Carnap," in expositing Carnap's notion of quasi-syntactic utterances, Quine writes: "It is in sentences dealing with reference, mention, meaning, denotation that we must be on our guard; also in modal sentences, both logical and empirical" (Creath 1990, 101). (In *Logical Syntax*, Carnap specifically attacks "sentences about meaning" in §75, and explains how to translate suspicious, "quasi-syntactic" intensional language (including the language of modalities) into hygienic extensional language in §§68-70.) Quine's guard stays up, when it comes to semantic and intensional language, over the following decades, while Carnap relaxes his.

Interestingly, Quine hints at just such a development of the situation in his "Homage to Carnap":

> Carnap was my greatest teacher. I got to him in Prague 38 years ago [from 1970, so 1932], just a few months after I had finished my formal studies and received my Ph.D. I was very much his disciple for six years. In later years *his views went on evolving* and so did mine, in divergent ways." (Creath 1990, 464; my italics)

First, note that Quine says that Carnap's 'views went on evolving'—and part of that evolution, as we have seen, is a willingness to study semantics and intensional languages. The historical question here is: what happened at the end of the sixth year, that Quine no longer considered himself 'very much Carnap's disciple'? It is unclear which stretch of six years Quine has in mind: 1932-37 (inclusive) or 1933-38. (In either case, Quine still considered himself 'very much Carnap's disciple' when he wrote "Truth by Convention.") Later in his "Homage," matters become even murkier, because Quine says that Carnap came to Harvard as a visiting professor in 1939, though Carnap did not arrive until 1940. Thus it could even be that the Harvard discussions themselves played an important role in ending Quine's discipleship to Carnap. More generally, it is possible that Quine sees Carnap's move away from the *Syntax* program as the turning point—Quine read the manuscript of Carnap's *Introduction to Semantics* for the University of Chicago Press in 1940. In other words, the radical critique of analyticity that Quine advocates by 1951 is as much a product of Carnap changing his views (towards fundamentally semantic and intensional approaches, away from syntactic and extensional ones) as Quine changing his. Carnap's shift is likely another factor in the radicalization of Quine's critique of analyticity between 1936 and 1951 (an issue first addressed in IV.2): what appears as a somewhat suspicious notion in "Truth by Convention" has become fundamentally unacceptable. Quine does not connect worries about analyticity in "Truth by Convention" to his distaste for intensional (and especially modal) languages, and this is likely because in 1936 he is believes that Carnap's preferred explanation of analyticity will be extensional and syntactic. Quine clearly rejects modal and intensional language in the 1941 notes, but he does not appear to see (either in 1936 or in 1941) that synonymy is also an intensional notion for Carnap by this time. In the discussion notes, Carnap clearly characterizes synonymy as modal: as quoted earlier, two predicates are synonymous if and only if they have the same extension in all possible worlds (102-63-07). This would rankle Quine, who writes that 'necessary' is a more "obscure" term than 'analytic' (1947 (Modal Logic), 45), so that the notion of analyticity should be used to explain the notion of necessity, *not* the other way around (1943, 121). In 1939's *Foundations*, we find L-true as 'true in virtue of semantic (as opposed to syntactic) rules,' and "two expressions are… *synonymous*… if they have the same designatum by virtue of the rules of S [the semantical system]" (153), but no characterizations of analyticity that appeal to modal notions.

Finally, let us consider one way in which Quine's preference for the syntactic approach to language-analysis contributes to his rejection of analyticity. Quine claims, in both "Two Dogmas" and "Notes on the Theory of Reference," that definitions/ characterizations of analyticity can only be done language-by-particular-language, so that giving a "definition of 'analytic-in-*L*' for each *L* has seemed rather to be a project unto itself" (1953, 138; cf. 32-36). Now, this seems plausible from a syntactic, proof-theoretic point of view. For example, in Carnap's *Foundations*, in the section on the calculus B-C (as opposed to the semantical system B-S), there are a few axioms and two rules of inference, namely modus ponens and "R2. *Rule of synonymity*: The words 'titisee' and 'rumber' may be exchanged at any place (i.e., if S_2 is constructed out of S_1 by replacing one of those words at one place by the other one, then S_2 is directly derivable from S_1 in B-C" (1939, 161). And clearly, another rule of inference (or postulate) would have to be added to the calculus for each additional synonymy we wished to represent. However, if we begin from the (Carnapian) semantic point of view, synonymy does not appear so *ad hoc*: every individual constant in the object language is assigned an object by the semantic rules, and if two individual constants happen to be assigned the same object by the semantic rules, then they are synonymous.[109] Similarly, every predicate is assigned a property, and if two predicates are assigned to the same property, then they are synonymous—in both cases, synonymy 'falls out of' the characterization of the language in a more natural way. On the other hand, a syntactically characterized language must add to the axioms or rules of inference to accommodate instances of synonymy as Carnap does in *Foundations*, though a semantically characterized language requires no such addition.

[109] Some will urge the following well-known objection to such a view: if our semantic system assigns Venus to both 'Morning Star' and 'Evening Star,' then 'Morning Star = Evening Star' will be true in virtue of the semantic rules, i.e., analytic.

2. TARSKI'S OBJECTIONS TO ANALYTICITY

When philosophers today think of critics of Carnap's notion of analyticity, Quine comes immediately to mind. However, when Carnap mentions criticisms of analyticity in print, we are usually as likely to find Tarski's name as Quine's. For example, in *Introduction to Semantics*, Carnap writes (citing (Tarski 1936/1983)): "Tarski expresses, however, some doubt whether the distinction between… L- and F-truth is objective or more or less arbitrary" (1942, 87; cf. *vii*). We find Carnap stressing Tarski's role in his autobiography as well: "my emphasis on the fundamental distinction between logical and non-logical knowledge, … which I share with many contemporary philosophers, differs from that of some logicians like Tarski and Quine" (1963, 13; cf. 30, 36, 62, 64). Note Carnap's choice of words: those who agree with Carnap on this fundamentally philosophical issue are 'philosophers,' while those who disagree are 'logicians.' This likely stems from Carnap's claim that Tarski and Quine's position is the result of their working almost exclusively in formal languages as opposed to the languages of natural science (1963, 932).

One of the documents in RCC 090-16 predates the academic year at Harvard. Tarski gave a talk at the University of Chicago at the end of spring term 1940, and on June 3, he and Carnap had an extended private discussion on topics of shared interest (RCC 090-16-09). One of the issues they discuss at length is Tarski's suspicion, voiced at the end of 1936's "On the Concept of Logical Consequence," that the logical/descriptive (or /'factual') distinction is somehow vague, unprincipled or almost arbitrary. There, Tarski writes:

> Underlying our whole construction [of the definition of consequence] is the division of all terms of the language discussed into logical and extra-logical. This division is certainly not quite arbitrary. If, for example, we were to include among the extra-logical signs the implication sign, or the universal quantifier, then our definition of the concept of consequence would lead us to results that obviously contradict ordinary usage. On the other hand, no grounds are known to me which permit us to draw a sharp// boundary between the two groups of terms. It seems to be

possible to include among logical terms some which are usually regarded by logicians as extra-logical without running into consequences which stand in sharp contrast to everyday usage.
(1936/ 1983, 418-419)

In short, the standard that must be satisfied by any division of terms into logical and extra-logical is conformity with existing 'everyday' usage (this is why the division is 'not quite arbitrary'). But Tarski thinks equally good levels of conformity can be reached by different choices for the division between logical and extra-logical terms. In other words, the logical/ extra-logical division is underdetermined by the (observed? or observable?) linguistic evidence. From this supposition that different choices of the boundary for 'logical term' could all capture the relevant linguistic phenomena equally well, Tarski concludes:

Perhaps it will be possible to find important objective arguments which will be enable us to justify the traditional boundary between logical and extra-logical expressions. But I also consider it to be quite possible that investigations will bring no positive results in this direction, so that we shall be compelled to regard such concepts as 'logical consequence,' 'analytical statement,' and 'tautology' as relative concepts which must, on each occasion, be related to a definite, although in greater or less degree arbitrary, division of terms into logical and extra-logical. The fluctuation in the common usage of the concept of consequence would—in part at least—be quite naturally reflected in such a compulsory situation.
(1936/ 1983, 420)

This idea—that the truth-values of 'A is a logical consequence of B' and 'C is analytic' are relative to a more or less arbitrary distinction between logical and non-logical terms—sounds *very* similar to Carnap's Principle of Tolerance, since specifying which terms are logical (and hence are given a fixed meaning) is an essential part of specifying a language. If it were (contra Tarski) completely non-arbitrary which terms are logical and which not (and assuming the meanings of the logical terms are also determinate), then there would be One True Logic, which is anathema to the tolerant Carnap. The primary difference is that Carnapian tolerance is not especially beholden to ordinary linguistic usage (especially in the *Syntax* period), so that Tarski's 'tolerance' is weaker than Carnap's. Given this similarity of viewpoint between Tarski and Carnap, it may strike us as somewhat surprising that Carnap considered Tarski one of his great opponents on the issue of analyticity. In order to alleviate this perplexity somewhat, in this section I examine Tarski's two primary arguments against the notion of analyticity in June 1940, as well as Carnap's replies. Since this discussion pre-dates those on finitism, it is not tightly linked to the later project undertaken at Harvard. We shall see that in certain ways, Tarski misunderstands or 'talks past' Carnap; nonetheless, genuine differences between them can also be formulated.

A. Tarski's first objection: any sign can be converted into a logical sign

To begin their conversation, Carnap first notes that his intuitions are stronger in the case of sentences than of terms: the distinction between logical and descriptive sentences is intuitively clearer to Carnap than the distinction between logical and descriptive terms. Nonetheless, Carnap is willing to address the issue on Tarski's preferred ground, so Carnap proposes to distinguish logical terms from non-logical ones as follows: "indicate the simplest logical constants in the customary systems, and declare everything definable from them is also logical." Tarski replies that he "has no such intuition; for one could equally well reckon 'temperature' as a logical term as well": simply fix the truth-values of all the (atomic) sentences involving the predicate 'Temp' (representing temperature), and maintain that assignment "in the face of all observations." In other words, any atomic statement could be stipulated to have the value *true* by a semantic rule. Note that this is a stronger claim than that found in Tarski's paper on logical consequence: he makes no mention in the discussion notes of the requirement that the division into logical and descriptive signs must respect existing usage. Tarski's objection appears to be a version of the often-heard claim (which supposedly challenges Carnap's position on analyticity) that which assertions are taken as unrevisable is arbitrary: the truth-value of 'The temperature at spacetime point p_0 is t_0' could, for some investigator and/or in some language, remain the same 'in the face of all observations,' i.e., held true come what may (to echo "Two Dogmas"). We encountered a form of this objection in the earlier discussion of "Truth by Convention" (IV.2): at what point do we stop declaring sentences true by convention? I say 'supposedly' because The Principle of Tolerance is exactly the claim that the choice of assertions taken as analytic is arbitrary; different choices yield different languages.

Furthermore, as is becoming better known among current philosophers,[110] Carnap clearly holds that a statement's being held true come-what-may neither implies nor is implied by that statement's being analytic. Analytic sentences need not be held true come what may, for we can change the language we are speaking; and Carnap thinks that this is exactly what happened when the scientific community made the transition from the Newtonian view of gravitation to the relativistic one. As a result of this language change, the metric tensor changed from a logical sign to a descriptive sign. Conversely, a sentence can be held true come what may and remain synthetic: someone need only be dogmatic enough about what is actually the case, without asserting that what is the case is true in virtue of the meanings of the words she is using, or true in all states of affairs. So where did the notion that analytic statements are exactly those held true come-what-may originate, if not with Carnap? Quine first introduces this unrevisability criterion of analyticity in "Truth by Convention" (quoted at length in 3.A below), and reiterates it in "Two Dogmas," as a 'behavioristic' correlate of the old philosophical notion of analyticity. His challenge seems to be that Carnap's notion of analyticity is insufficiently empirical, and thus

[110] Stathis Psillos puts the point clearly: "A common criticism against analyticity, made by both Quine and Hempel, is that there is no point in distinguishing between analytic and synthetic statements, because all statements in empirical science are revisable… But since, as Hempel said, there are no such truths… there is no point in characterizing analyticity. However, such criticisms have always misfired against Carnap. Carnap never thought analyticity was about invioable truth, 'sacrosant statements,' unrevisability or the like. … Already in 1937 [that should be 1934: §84 was not one of the sections of LSL that appeared only in English-GFA], Carnap noted that no statements (not even mathematical ones) were unrevisable. Anything can go in the light of recalcitrant evidence" (2000, 154). And in the Schilpp volume, Carnap explicitly writes: "the concept of an analytic statement which I take as an explicandum is not adequately characterized as 'held true come what may'" (1963, p.921).

fails to be a scientifically respectable concept. (See (Creath 2004) for basically the same interpretation of Quine's challenge.)

Carnap responds to Tarski that, in a language where the truth-values of atomic sentences containing 'Temp' are all fixed (via semantic rules), 'Temp (t, p)' *would* be a "mathematical function, a logical sign, and not the physical concept of temperature." So Carnap apparently grants Tarski the point that, when laying down the semantical rules for one's constructed language system, one can make any term in the language under construction a logico-mathematical one—but then that term necessarily becomes a non-physical, non-factual term in the constructed language, even if it is homophonic (or homographic) with a factual term of natural language. If we decide (on pragmatic grounds) that the language we are constructing should capture the logical/ descriptive distinction (to the extent that it exists) in everyday language, then we cannot construct a language of physics including Tarski's imagined 'Temp' predicate. In either case, there is no contradiction: if are not required to re-capture everyday language within our artificial language, then it does not matter that 'Temp' (or any other term) becomes logico-mathematical. On the other hand, if we do require our artificial language to save the linguistic phenomena of extant usage, then we cannot stipulate the truth-values of all atomic sentences of the form 'Temp(t, p).' Carnap claims the crucial difference between logical and physical terms is shown as follows: for closed sentences containing the *physical* temperature predicate (as opposed to Tarski's un-empirical predicate 'temperature'), "we cannot find the truth-value through mere calculation." Thus, as seen earlier, a sentence is *analytic* in *L* if its truth-value can be arrived at via calculation—in particular, during Carnap's semantic period, this means calculation ultimately from the semantic rules of *L*, as the following subsection will make clear.

B. Tarski's second objection: Gödel sentences

Tarski seizes upon Carnap's characterization of analyticity in terms of calculability to lodge a second objection against the logical/ descriptive distinction. Tarski immediately retorts to Carnap's quoted statement above that 'Temp' would qualify as a logico-mathematical term on Tarski's construal: "That proves nothing, since that is often not the case for mathematical functions either, since there are undecidable sentences" in mathematics. That is, in sufficiently rich formalizations of arithmetic, there are logico-mathematical claims, such as the Gödel sentence, whose truth-values *cannot* be calculated from the axioms and rules of inference of that calculus. From the fact of undecidability, Tarski concludes that there is "no fundamental difference between mathematical but undecidable sentences and factual sentences." This argument is enthymatic, so any detailed reconstruction requires some conjecture. Here is one attempt to spell out Tarski's argument:

(P1) If a sentence *s* in language *L* is logico-mathematical, then *s* or *not-s* can be justified via mere calculation in *L*.

(P2) If a sentence *s* can be justified via mere calculation in *L*, then the axioms and inference rules of *L* suffice to prove *s*.

(Thm) If the Gödel sentence *G* is expressible in *L*, then neither *G* nor ~*G* can be proved from the axioms and inference rules of *L*.

(C1) Thus, neither *G* nor *not-G* can be justified via mere calculation. So *G* is not logico-mathematical.

At this point, the reconstruction becomes especially hypothetical.

(P3) But *G* is logico-mathematical.

Then, by *modus tollens*, either (P1) or (P2) must be false, and Tarski places the blame on (P1): mere calculability fails to separate the mathematical sentences from the descriptive ones. Then, in order to reach his stronger conclusion that there is 'no difference between undecidable mathematical sentences and empirical ones,' Tarski will need a premise in the neighborhood of the following:

(P4) No criteria besides calculability can effectively separate logical sentences from factual ones.

That is, if calculability is the only viable or plausible candidate for drawing a sharp distinction between logico-mathematical truths and factual ones, and the above argument shows that calculability cannot draw the distinction in the (intuitively) correct place, then there is no criterion to underwrite or support the distinction.

Immediately following Tarski's claim that there is no fundamental difference between the Gödel sentence (and its kin) and factual sentences, Carnap merely replies "It seems to me that there is" [*Es scheint mir doch*] such a difference, and the conversation ends there. Although we do not have a fully articulated rebuttal from Carnap here, we can infer his response with some confidence from his published writings. Carnap, facing the same *modus tollens* after (P3), would choose to reject (P2) and maintain (P1). The Carnap of the semantic period (and thus of 1940) would replace (P2) above with:

($P2_C$) If a sentence *s* can be justified via mere calculation, then the semantic rules of *L* suffice to determine that *s* is true.

The Carnap of *Logical Syntax* would likewise reject (P2) as too narrow a notion of calculation—there, Carnap allows calculation to include the infinite hierarchy of metalanguages (with transfinite rules of inference) associated with a given object language. (In the terminology of *Logical Syntax*, c-rules (for 'consequence'), not d-rules (for 'derivation,' answering to modern term 'proof'), are used to determine whether a sentence is analytic or not.) And in these stronger languages, sentences that cannot be proved in the object language can be proved. On either option—the semantic or the syntactic—the inference to (C1) is blocked, and Tarski's argument would be defused.

Thus, this dispute between Tarski and Carnap (as I have reconstructed it) reduces to the question of whether Carnap is entitled to this wider notion of calculation or not. Is the replacement of (P2) with ($P2_C$) legitimate? Talk of entitlement and legitimacy is somewhat obscure and metaphorical, but it is also a sign that our question—'What is calculation?'—has a normative component, like virtually all (philosophically interesting) questions of explication. Thus, a 'clean' answer is not to be expected. With that caveat stated, we can appropriate one argument for Carnap from *Logical Syntax*: "The sentence [*G*], which is analytic but irresoluble in language II, is thus in II_d [the language "which results from II by limitation to the d-rules"] an indeterminate sentence." And "sentences that are indeterminate" are "designated by us as descriptive, although they are interpreted by their authors as logical." Carnap says that the same holds for the language of *Principia Mathematica*. And further, he writes: "*the universal operator*… is a proper universal operator in languages I and II, but in the usual languages—for instance, in [*Princ. Math.*]—it is an improper one…, because these languages contain only d-rules." Thus we come to the surprising conclusion that "*the universal operator in both* Principia Mathematica *and* II_d *is not logical but descriptive*" (1934/1937, 231).[111] In short: a language with only d-rules makes certain 'apparently logical' operators (like ∀x) and sentences (such as the Gödel sentence) descriptive, but a language with c-rules as well can classify such operators as logical and such sentences as analytic. For present purposes, c-rules are the *Syntax*-era analogues of semantic rules in 1940. And presumably, an explication that makes the universal operator logical is preferable to one that does not.

I do not know of any sections of the Tarskian corpus that could be marshaled to support the narrower, proof-theoretic notion of calculability that Tarski (in the above reconstruction) endorses. Nonetheless, one can adduce arguments in favor of (P2). The current, widely-

[111] Incidentally, this shows (Creath 1996, 261) to be false.

accepted notion of computability is closer to the narrower conception of calculability—though modern computability is perhaps too narrow to even be a viable candidate for distinguishing logical from descriptive sentences. A second consideration that might be introduced to support the Tarskian viewpoint is subtler, but likely *not* one Tarski himself would have articulated—in part because it is not, in the end, compelling. Every genuine sentence that is expressed is expressed in some language. Without some form of language to speak or write in, we could not assert anything. But any realistic language permits inferences from certain sentences to others, that is, one component of the language is a *consequence-relation*.[112] Logically-minded students of language can codify the relation formally in the form of rules of inference (and axioms, considered as consequences of the null set). So if someone wishes to make an assertion in a natural language *l*, *l*-speakers must in some sense accept (perhaps only tacitly) the consequence-relation of *l*—otherwise she would not make the assertion she aims to make, but a different one. If someone makes an assertion in *l*, that person is thereby committed implicitly or explicitly to the sentences made true by *l*'s consequence-relation, if any. And if one uses the logician's formalization of *l* (call it *l**), then one is committed to the theorems provable from the rules of inference and axioms of *l**. Now, these sentences made true by the consequence relation of *l* (in *l**, the theorems), are not justified in the same way that other assertions couched in *l* (or *l**) are justified:[113] the theorems and their natural language correlates are 'taken on board' by the very expression of a proposition in *l* or its formalization. The fact that I express my assertions in *l** constitutes the ultimate justification (if 'justification' can be meaningfully spoken of here) for the theorems of *l**. This is an instance of a transcendental justification: a claim in *l** could not express what it does in fact express, if the theorems of *l** were not true. This is one way of

---

[112] Of course, languages in which no sentence is a consequence of any other sentence are a theoretical possibility: recall the discussion of 'Strictly Quinean Languages' in IV.3.
[113] Unless the views of the later Quine on mathematics are correct.

explaining why, for Carnap and other logical empiricists, mathematics and logic are not susceptible of empirical justification.

Now we have reached the point where a Tarskian can lodge a complaint against Carnap's wider notion of truth in virtue of calculability; the complaint will be clearer applied to the Carnap of *Logical Syntax*, so he will be the target. There, Carnap does not identify the analytic truths (i.e., the truths in virtue of calculation) with the sentences derivable from the inference rules and axioms of *l**, but rather from the inference rules and axioms of a stronger metalanguage (and metametalanguage, etc.). Our previous rationale for sentences we consider 'true by calculation' has disappeared—for we express ourselves in the object language, not the metalanguage. If we confined ourselves to the inference rules and axioms of *l**, we would of course have exactly those sentences (reconstructed) Tarski considers 'true by calculation,' for the Gödel sentence (couched in *l**) is of course *not* a theorem of *l**.

However, we can forward a convincing rejoinder on Carnap's behalf. The treatment of sentences 'made true by the consequence relation' in the previous paragraphs had to be presented in a misleading way in order to make the argument for Tarski's viewpoint. Consequence is standardly taken to be a thoroughly *semantic* notion: 'A is a consequence of B' is usually taken to mean that whenever B is true, A is (or must be[114]) true. However, consequence was not treated semantically above, since 'sentence true in virtue of the consequence relation' was treated as equivalent to 'theorem'; this is inadequate for any incomplete proof calculus. Now, if we consider the language that we use not merely as a formal proof system, but as endowed with semantic properties, then Carnap's wider characterization of 'true in virtue of calculation' falls out of using a language. That is, if we accept that an *l*-user is in fact entitled to the sentences true in virtue of the consequence relation of *l* simply by expressing a proposition, then the *l*-speaker is

[114] John Etchemendy pushes this modal line in *On the Concept of Logical Consequence* and elsewhere.

entitled to the semantic resources of that language, not merely a structural-syntactic characterization of it in terms of formal axioms and inference rules. And allowing the *l*-speaker to draw on that semantic information corresponds to allowing *Carnap's* wider notion of calculation, viz. truth in virtue of the semantic rules of the language, not Tarski's narrower one. But it is here that Quine's indeterminacy of translation thesis, expressed in Chapter 2 of *Word and Object*, is salient: for he there denies that such semantic rules should be allowed as a part of the scientifically respectable picture of the world. If he is right, then Carnap's rebuttal to Tarski is unacceptable—though if Quine is right about semantic rules, then Tarski's complaint based on the Gödel sentence would be the least of Carnap's problems.

3. QUINE'S DISAGREEMENTS WITH CARNAP CIRCA 1940

A. Analyticity is an empirical concept, not a logical one

Quine's criticisms of Carnap had not reached their fully mature form by 1941. Nonetheless, the seeds had been planted, and discrepancies between the two philosophers were beginning to show. The first public appearance of significant differences between Quine and Carnap is "Truth by Convention." Its main argument, as we saw in IV.2, is that the domain of stipulated (conventional, analytic) truths can be expanded indefinitely, with no clear, principled cut-off point. But there is another suggestion, very briefly mentioned in "Truth by Convention," that reappears in a slightly different form in the Harvard discussion notes. This suggestion is developed in "Two Dogmas" (though still somewhat inchoate), and reaches its fully-fledged form in *Word and Object*. Here is the original passage from "Truth by Convention":

> there is the apparent contrast between logico-mathematical truths and others that the former are a priori, the latter a posteriori; … Viewed behavioristically and without reference to a metaphysical system, this contrast retains reality as a contrast between more and less firmly

accepted statements; and it obtains antecedently to any *post facto* fashioning of conventions. There are statements… which we will not surrender at all, so basic are they to our whole conceptual scheme. Among the latter are to be counted the so-called truths of logic and mathematics, regardless of what further we have to say of their status in the course of a subsequent sophisticated philosophy. (1936/1976, 102)

Note that if we do not have empirical ('behavioral') criteria for identifying the a priori (analytic) sentences, then 'a priori' and 'analytic' are metaphysical terms. That is, analyticity must be cashed out in empirical (synthetic) terms, or else it is just as semantically and epistemically objectionable as God, souls, and Heideggerian pronouncements. That is why Quine characterizes analytic truths as those 'held true come what may': language users' acceptance of claims is thought to be susceptible to empirical investigation.

In the Harvard discussion notes, we find a clear statement distinguishing Quine's position from Carnap's in roughly similar terms. The basic distinction can be stated as follows: consider a sentence of the form '*p* is analytic.' Carnap thinks such sentences are analytic, while Quine thinks they are synthetic, so their truth-value must be determined by empirical/ observational means. This appears in the context of the group's attempt to develop a notion of L-truth that is suitable not just for formal/ mathematical languages, but for the total language of science as well. As an example of an empirical term, they use 'T' to refer to temperature (at a spacetime point).

Tarski: The physicist chooses certain sentences as conditions that a proposed claim about T [the temperature relation-letter] must satisfy, in order to be assumed as (logically) correct, before experiments about the truth are made. [GF-A: Such sentences include: 'every atomic sentence involving 'T' must have exactly five arguments' (four spacetime coordinates and one scalar for the temperature), and 'Two atomic T-sentences cannot agree on all the spacetime coordinates and disagree on the fifth argument']
Quine: It is then the task of a behavioristic investigation to determine what conditions of this kind physicists set up.
I: No, that would give only the corresponding pragmatic concept. As with all other semantic (and syntactic) concepts, here also the pragmatic concept gives only a suggestion, and is not determined univocally. (090-16-10)

This quotation shows that Quine considered the question 'Which claims are analytic?' to be amenable to behavioristic answers. This, in turn, shows why Quine pursues accounts of analyticity and synonymy in *Word and Object* and elsewhere—a fact that might be perplexing, given Quine's reputation as the hero who has slain the analytic/ synthetic distinction. The point illustrated by the above quotation clears up this potential perplexity: the Quinean notions of 'stimulus synonymy' and 'stimulus analyticity' are explicitly and thoroughly *empirical* notions—thus they are scientifically respectable, and potentially useful for scientifically-inclined philosophers. As Quine writes a few years after the Harvard discussions:

the meaning of an expression is the class of all the expressions synonymous with it. … The relation of synonymity, in turn, calls for a definition or a criterion in psychological and linguistic terms. Such a definition, which up to the present has perhaps never even been sketched, would be a fundamental contribution at once to philology and philosophy.
(1943, 120)

Exactly such a psycho-linguistic criterion is, of course, spelled out at length in *Word and Object*.

Quine expresses a similar sentiment a few years later:

Synonymy, like other linguistic concepts, involves parameters of times and persons, suppressible in idealized treatment: the expression *x* is synonymous with the expression *y* for person *z* at time *t*. A satisfactory definition of this tetradic relation would no doubt be couched, like those of other general concepts of general linguistics, in behavioristic terms. I should like, as a service both to empirical semantics and to philosophy, to offer a satisfactory definition; but I have none. So long, however, as we persist in speaking of expressions as alike or unlike in meaning (and regardless of whether we countenance meanings themselves in any detached sense), we must suppose that there is an eventually formulable criterion of synonymy in some reasonable sense of the term.
(1947, 44)

Here again, we see that Quine thinks synonymy must at bottom (in non-'idealized' cases) be given a behavioristic account—just like other linguistic concepts.

Another way of couching this difference between Carnap and Quine starts from Carnap's characterization of analytic truth: it is 'true in virtue of meaning (of language), independent of empirical facts.' Quine can be interpreted as denying the equivalence of those two clauses: 'meaning' and its derivative terms, if they are to have a place in a total language of science, must be just as empirical as any other theoretical term of science. A final note about this difference between Carnap and Quine: Quine held Carnap's view in his 1934 "Lectures on Carnap": "Analytic propositions are true by linguistic convention. But… it is likewise a matter of linguistic convention *which* propositions we are to make analytic and which not" (Creath 1990, 64). So for a time in the mid-thirties, presumably ending with "Truth by Convention," Quine thought it perfectly acceptable to consider the sentence '*p* is analytic (true by convention)' to be itself analytic (true by convention).

While we now have a clear formulation of one difference between Quine and Carnap—a difference that persisted, it appears, for the rest of their careers—it is much less clear whether there is a well-posed question in the vicinity of this disagreement. To put the matter in a way favorable to Carnap, it is analogous to asking: 'Which is right: pure geometry or applied geometry?" (The Quinean would presumably respond to this description as follows: 'A given mathematical system wouldn't deserve the name 'geometry' at all, if it did not admit of an interpretation using spatiotemporal magnitudes in the empirical world.') Perhaps the best we can do is to indicate motivations and possible motivations lying behind each view. As a rationale for Carnap's view that analyticity should be treated as an analytic concept, we can point to his enthusiastic adoption of the formalizations of syntactic and semantic concepts. That is, Gödel showed how to formalize the predicate '…is provable (in *L*)' within number theory, thus showing that the concept of provability is just as logico-mathematical as addition, conjunction etc. Analyticity is of course not co-extensive with provability, but every statement provable in *L* is analytic in *L*, and the concept of analyticity (especially in *LSL*) is an extension of the notion of theorem. Thus, since '… is provable' is a logico-mathematical predicate, and analyticity is intended as a generalization of provability, this leads naturally to considering '…is analytic' to be a logico-mathematical predicate. This conclusion could only be bolstered by Tarski's demonstration how to define '…is true in *L*' in a purely logico-mathematical way, given the expressions of *L* and names for the expressions of *L*. (For one might wonder whether, even if provability is a logico-mathematical notion, truth might not be.) Since analyticity is a species of truth, Tarski's work lends plausibility to notion that it, like truth, can be treated as a logical concept by scientifically-minded philosophers.

What motivates Quine's view that analyticity should *not* be considered a logico-mathematical concept? One rationale appears in the quotation from "Truth by Convention" at the beginning of this subsection: the actual pattern of human acceptance of claims 'obtains antecedently to any *post facto* fashioning of conventions.' The same idea appears in the "Lectures on Carnap" as follows: "in any case, there are more and less firmly accepted sentences *prior to* any sophisticated system of thoroughgoing definition" (Creath 1990, 65; italics mine). If 'antecedently' and 'prior to' are construed temporally, then this is obviously true, since a theory always post-dates its subject: the behavior of falling apples and the Earth's tides obtained long before Newton proposed his law of universal gravitation. Thus Quine presumably has something akin to *conceptual* precedence or priority in mind. Conceptual priority can be a thorny matter, and it is made more difficult here since Quine does not spell out the sense of priority he has in mind; we can get some sense of it, however, by looking elsewhere. Quine, in both his "Lectures on Carnap" and "Truth by Convention," uses the explanatory fiction of having a list of all sentences currently accepted as true in front of us. Those sentences on this list that we accept so firmly that we would not reject them under any circumstances, Quine says, are those that can or should be declared true by convention (Creath 1990, 65). The way we select which sentences to elevate to analytic status via stipulation, for Quine, is by finding exactly those sentences that would never be abandoned. So the sense of conceptual priority at issue comes at least to this: if there are no irrevocable sentences, then there are no analytic sentences. And Quine suggests, in the closing section of "Two Dogmas," that the antecedent of that conditional is true.

Elsewhere in Quine's writings, the priority of firmness of sentence acceptance over analyticity is couched in even stronger terms. The notion of Carnapian analyticity is considered artificial, in a pejorative sense. In "The Problem of Interpreting Modal Logic," within the context of explaining how 'No spinster is married' can be taken as analytic, because it is a definitional abbreviation of a logical truth, Quine writes:

> I should prefer not to rest analyticity thus on an **unrealistic fiction** of there being standard definitions of extra-logical expressions in terms of a standard set of extra-logical primitives. What is rather in point, I think, is a relation of *synonymy*, or sameness of meaning, which holds between expressions of **real** language though there be no standard hierarchy of definitions.
> (1947, 44; my emphasis in bold, italics in original)

Quine had already privately made a similar point in a 1943 letter to Carnap:

> A common answer to this problem is to say that 'No spinster is married' is a definitional abbreviation of a logical truth, 'No woman not married is married'. *Here we come to the root of the difficulty: the assumption of a thoroughgoing constitution system*, with fixed primitives and fixed definitions of all other expressions, *despite the fact that no such constitution system exists*.
> (Creath 1990, 296; my emphasis)

Carnap responds directly to Quine's worry about the apparent need to set up an elaborate system of definitions in order to capture the notion of analyticity or L-truth. Carnap points out that, on his view of language, what is needed to capture the notion of an analytic truth is just a semantic system, that is, a Carnapian interpreted language (Creath 1990, 305). The semantic rules would suffice to guarantee the synonymy of 'spinster' and 'unmarried woman,' even if both were primitive predicates in the language. But Carnap still assumes the existence of a semantic system, which Quine would likely consider just as artificial and/ or unreal as a constitution system. At least, that is (one way of interpreting) the conclusion Quine draws from his indeterminacy of translation thesis, namely, that semantic rules of the sort Carnap favors have no place in our total science.

B. Modal and intensional languages unacceptable

There is a further reason why Quine declines to treat analyticity as a logico-mathematical concept. Quine holds that if a concept is to be a (scientifically respectable) logico-mathematical concept, then it must be truth-functional. Analyticity (and necessity) are not: 'World War II ended in 1945' and 'World War II ended in 1945 or it did not end in 1945' have the same truth-value, "'World War II ended in 1945' is analytic" is false, while "'World War II ended in 1945 or it did not end in 1945' is analytic" is true. In "Three Grades of Modal Involvement," for example, Quine writes: "In mathematical logic… a policy of *extensionality* is widely espoused: a policy of admitting statements within statements truth-functionally only" (1953/1976, 162). And Quine himself certainly espouses this view, as we saw in 1.B. A similar claim appears in "Notes on Existence and Necessity," published two years after the Harvard discussions:

> any intensional mode of statement composition… must be carefully examined in its relation to its susceptibility to quantification. Perhaps the only useful modes of statement composition susceptible to quantification are the extensional ones, reducible to '~' and '.'. Up to now there is no clear example to the contrary. It is known, in particular, that no intensional mode of statement composition needed in mathematics.
> (1943, 124-125)

The fact that mathematicians did not need intensional language, combined with the problems surrounding '9' and 'the number of planets' familiar from this article onwards, were sufficient rationales for Quine to reject all intensional language.

In the Harvard conversation notes, Quine approves of a particular explication of L-truth, by saying: "Thus we avoid "*state of affairs*," intensional language, and the unclear concept 'logically-possible'" (090-16-10). Though this is the only record in the 1940-41 notes of Quine's disapproval of these three interrelated concepts, he was hostile towards them his entire adult life (Quine 2001). This mention is important, because Carnap characterizes L-truth as truth in all states of affairs (where the set of all states of affairs is relativized to a semantically-characterized language), i.e., all logically possible (L-possible) worlds. For example, for a suitably formalized portion of ordinary English, there is no logically possible world or state of affairs in which the extensions of ‘Batchelor’ and of ‘unmarried man’ are not identical. And Carnap makes precisely this characterization of synonymy in the Harvard notes: two predicates (e.g.) are synonymous in *L* (i.e., they designate one and the same property) if and only if, in every possible world, the two predicates are co-extensive (102-63-07). Incidentally, this shows that Quine’s ‘No entity without identity’ complaint against properties (and intensional language more generally) is fundamentally based upon Quine’s rejection of the modal notion of a logically possible situation or state of affairs. For Carnap has provided, by 1941, an identity-condition for two properties: a property is an extension for each L-possible world, so two properties are identical if, in every L-possible world, they are coextensive. Once Carnap attempts to spell out ‘analytic’ in modal terms—as he does after taking his semantic turn—Quine’s hackles are raised, and the notion that seemed somewhat suspicious to Quine in “Truth by Convention” becomes, in this new form, fundamentally unacceptable. It is interesting that in “Two Dogmas,” Quine does not consider this newer, semantic characterization of analyticity, but sticks to the older one, which Carnap had rejected years earlier. One plausible explanation for Quine’s passing over Carnap’s characterization of analyticity in terms of L-possible worlds/ states of affairs is that Quine considered using such notions as the starting point of analysis irredeemably faulty—an attempt to explain the obscure by the more obscure. Quine says the latter in print:

> The notion of analyticity appears, at the present writing, to lack a satisfactory foundation. Even so, the notion is clearer to many of us, and obscurer surely to none, than the notions of modal logic; so we are still well advised to explain the latter notions in terms of it.
> (1947, 45)

This could explain why Quine would not attack Carnap’s preferred characterization of analyticity of the 1940s: after Carnap switches to an intensional analysis of analyticity, Quine regards all Carnap’s further forays non-starters. Once again, we see how the radicalization of Quine’s critique of Carnap is prompted, at least in part, by Carnap’s shift to intensional approaches to the study of language.

VI. UNITY OF SCIENCE AND THE REJECTION OF METAPHYSICS IN LOGICAL EMPIRICISM

A desire to unify human knowledge is both ancient and abiding. Plato, in the seventh book of the *Republic*, suggests that all knowledge is somehow derived from or based upon the Form of the Good. Two millenia later, Descartes' *Rules for the Direction of Mind* stresses the value and importance of developing a universal system of *scientia*. Related discussions continue today: reductionist and anti-reductionist philosophers working in various fields disagree about whether particular domains of knowledge can be unified in content, in method, or in their concepts. The logical empiricists also made the unity of science a central plank of their party platform. Another essential plank of their platform is the rejection of metaphysics; this, too, is neither unique to nor original with the logical empiricists. Hume, for example, famously recommended committing metaphysical writings to the flames, and the logical empiricists themselves explicitly acknowledge their historical predecessors in the struggle against metaphysics.[115] And the anti-metaphysical drive is not yet moribund: in much of van Fraassen's work over the last decade, especially *The Empirical Stance*, it is alive and well.

To most current philosophers, these two topics—the unity of science and the rejection of metaphysics—likely appear *prima facie* rather different. However, my central contention here is that these two ideas are intimately intertwined in the writings of many logical empiricists. Close attention to the writings of central logical empiricists on the unity of science and the elimination of metaphysics reveals that, metaphorically speaking, these goals are two sides of the

[115] Schlick writes: "the denial of metaphysics is an old attitude… for which we can in no way claim priority" (Schlick 1978, 492). Carnap echoes this sentiment: "Anti-metaphysical views have often been put forward in the past, especially by Hume and the Positivists" (Carnap 1934/1937, 280).

same coin. More prosaically, in different logical empiricists, from the 1920's through the 1950's, we find the following criterion (or an approximation thereof) at work for detecting metaphysics: an apparently meaningful, contentual utterance is metaphysical if and only if it cannot be incorporated into 'unified science' [*Einheitswissenschaft*]. I will focus on Carnap and Neurath, for they wrote most extensively on both the unity of science and the elimination of metaphysics, and their work is prominent among both their peers and modern scholars re-evaluating logical empiricism. Finally, I present an objection to this criterion for identifying metaphysics: there is no procedure to show that a given utterance cannot be incorporated into a unified language of science.

1. UNITY OF SCIENCE: UNITY OF LANGUAGE, NOT OF LAWS

The 'Unity of Science' movement, spearheaded by Otto Neurath and embraced by other logical empiricists, had its intellectual roots late 1920's Vienna. The philosophers associated with the official movement founded the *International Encyclopedia of Unified Science* and a series of international conferences, beginning in Paris in 1935. These philosophers were also directly responsible for the journal *Erkenntnis*, whose original English title was *Journal of Unified Science*. The ideas driving the official movement played a less direct role in *Synthese*'s early activities: the first sentence of the first issue of *Synthese* is "Ours is a time of synthesis," i.e., of unification. During the forties, *Synthese* regularly included articles under the title 'Unity of Science Forum.'

What did the logical empiricists mean by the phrase 'unity of science'? The unity that the logical empiricists speak of is *not unity of laws or theories, but rather unity of language*. This point is increasingly recognized in recent scholarship, e.g. (Creath 1996), so I will not attempt a complete substantiation of this claim. Nonetheless, I devote this section to an abbreviated elaboration and defense of this exegetical contention. First, to be explicit, Carnap, Neurath and others stress repeatedly that their thesis is *not* that the results of biology, psychology, sociology etc. can (or will) be ultimately derived from a single fundamental theory (presumably physics).[116] Thus the logical empiricists of the 1930's unequivocally do *not* endorse the kind of 'unity of science' found in (e.g.) Putnam and Oppenheim's "The Unity of Science as a Working Hypothesis" (1958). Rather, the logical empiricists' aim is to construct a *language* that can simultaneously express biological, psychological, social, and physical claims. Carnap emphasizes that the reduction of (e.g.) biological laws to chemical or physical laws is an open question: "there is at present no unity of laws… On the other hand, there is a unity of language in science, viz., a common reduction basis for the terms of all branches of science" (1938, 61). Neurath's views are similar. He does not demand a unity of laws: "Having a Universal Jargon [his term for his language of unified science] in common does not imply that the same scientific 'laws' have to be valid in the various fields of scientific research" (1946, 81). Neurath, the social scientist, stresses the autonomy of sociological laws: "Comprehensive sociological laws can be found without the need to be able… to build up these sociological laws from physical ones" (1983, 75).

Neurath is more antagonistic than Carnap to this unification of theories or laws. Neurath claims that a desire to fit all knowledge into a single Procrustean bed constitutes a fundamental error of Cartesian and Leibnizean rationalism, and he stresses that the model for unified science

---

[116] However, commentators nonetheless saddle logical empiricists in general, and Carnap in particular, with this view. Even Thomas Uebel, who usually provides helpful correctives to the stereotypical caricatures of the *Wienerkreis*, appears to succumb to this view of Carnap: "The second large-scale difference between Carnap and Neurath concerned the unity of science. Against *the hierarchy of reductively related theories*, Neurath put a much looser conception of unity … Neurath may well have felt that the supposition of a *reductive hierarchy of special sciences with physics at the base* was just a bit too counterfactual" (2001, 214-5, my italics). It is not the *theories* that are 'reductively related,' but rather the *languages*.

is not a *system*, but an *encyclopedia*: the claims of an encyclopedia, unlike the claims of a system, are not all derivable from a few precise axioms. For example, in the first article in the *International Encyclopedia of Unified Science*, Neurath (the *Encyclopedia*'s editor-in-chief) writes: "the great French Encyclopedia," whose work the new *Encyclopedia* continues, "was not a '*faute de mieux* encyclopedia' in place of a comprehensive system, but an alternative to systems" (1938, 7; cf. 2, 16, 20). This rejection of the single axiomatized system of knowledge in favor of a loosely connected encyclopedia is a *leitmotif* running throughout Neurath's corpus; it is expounded at length in his 1936 "Encyclopedia as 'Model'."

What the logical empiricists' unified science requires is not a unity of laws, but something weaker: unity of *language*. We saw Carnap explicitly state this immediately above. For Neurath as well, the crucial kind of unity is a unity of language: "We can use the everyday language which we use when we talk about cows and calves throughout our empiricist discussions. This was for me the main element of 'unity'" (1983, 233). Philipp Frank provides perhaps the simplest formulation of the unity of science thesis: "there is one and the same language in all fields" of science (1947, 165). Maria Kokoszynska, who visited the Vienna Circle from Lvov, offers a very similar characterization: "Every scientific sentence can be expressed in one and the same language" (1937/38, 326).[117] In *Logical Syntax*, Carnap offers the following more precise characterization of the thesis: every sub-language of science can be translated without loss of content into one language (1934/1937, 320).

---

[117] Kokoszynska's central contention in this article is an interesting challenge to this standard characterization. Tarski had shown in 1933 that, for a given language L (and given a number of relatively natural assumptions) 'true-in-L' is not definable in L, on pain of contradiction. 'True-in-L' is, however, precisely definable in the *metalanguage* of L. Kokoszynska thinks that the sentences of Tarski's metamathematics should count as scientific, so she sees his work as potentially disproving the unity of science thesis: not all scientific sentences *can* be incorporated into a single language. Neurath's reply can be found in (1983, 206-208).

The next question to ask is: which language or languages fit this description? Many logical empiricists agree that the *physicalist* language is one such. In *Logical Syntax*, Carnap states that the thesis of physicalism is precisely that the physicalist language can successfully serve as an overarching language for all of science (1934/1937, 320).[118] Carnap defends this thesis most extensively in "*Die Physikalische Sprache als Universalsprache der Wissenschaft*" (translated two years later under the title *Unity of Science*) (1932/1934). Neurath provides a detailed description of the physicalist language, which he also calls (at times) 'Universal Jargon.' It is *not* restricted to the vocabulary of physics. Neurath describes his Universal Jargon as "an everyday language that avoids certain phrases and is enriched by certain other phrases" (1983, 208); specifically, it 'avoids' metaphysical terms, and 'is enriched' by technical terms (1983, 91-92). Carnap characterizes the physicalist language as one whose sentences "in the last analysis… express properties (or relations) of space-time domains" (1934/1937, 151). Neurath makes similar statements: for someone who uses the physicalist language, "in his predictions he must always speak of entities in space and time" (1983, 75).

But the fact that the physicalist language can serve as the language for unified science does not imply that *no* other language could. For example, in the *Aufbau*, Carnap holds that both the phenomenal, 'autopsychological' language and the physical[119] language could function as languages for unified science. And Neurath writes: "We expect that it will be possible to replace each word of the physicalist ordinary language by terms of the scientific language—just as it is

[118] "The thesis of *physicalism* maintains that the physical language is a universal language of science—that is to say, that every language of any sub-domain of science can be equipollently translated into the physical language. From this it follows that science is a unitary system within which there are no fundamentally diverse object-domains, and consequently no gulf, for example, between natural and social sciences This is the thesis of the *unity of science*. … It is easy to see that both are theses of the syntax of the language of science." (1934/1937, 320)

[119] Carnap speaks of the 'physical language' in the *Aufbau*, *not* of the physical*ist* language. Neurath's term 'physicalist' was not known to Carnap at the time of the composition of the *Aufbau*. This terminological difference is of little consequence; for the 'physical language' in the *Aufbau* is either nearly identical to the (later) physicalist language, or a proper subset of the physicalist language.

also possible to formulate the terms of the scientific language with the help of the terms of ordinary language" (1983, 91); this shows that the 'physicalist ordinary language' is not unique.[120]

Finally, the logical empiricist unity of science thesis is not refuted by Suppes' observation (1978, 5) that the actual terminology used in various sub-disciplines of the sciences is increasingly divergent, with each subfield developing its own jargon. Other scholars (e.g. Creath 1996) have already noted that Neurath's and Carnap's unity of science theses do not claim to provide a descriptive account of extant scientific language and practice. In fact, Carnap explicitly agrees with Suppes' position in §41 of the *Aufbau*.[121] Carnap, in his most extended defense of the unity of science thesis (1932/1934), argues only that the various languages of science *could* be connected in principle, not that they *are* so connected in everyday scientific practice (if they were, there would be no work for the *Wissenschaftslogiker*). In sum, the logical positivists' unity of science thesis, especially as articulated and advocated by Carnap and Neurath, asserts that there exists a language in which all (scientific) knowledge can be couched, but not that this language is actually used, on a day-to-day basis, by scientists.

Finally, in order to avoid historical distortion, two alternative visions of the unity of science from logical empiricist sympathizers will be briefly mentioned. One of the most interesting expressions of a unity of science thesis can be found in J. H. Woodger's programmatic "Unity through Formalization."

---

[120] However, Neurath sometimes appears to privilege the physicalist language over others: "Unified science contains only physicalist formulations" (1983, 54); "Physicalism is the form work on unified science takes in our time" (1983, 56). Perhaps Neurath considers the physicalist language *best* for his purposes. And in 1932, the year after Neurath publishes these remarks, Carnap claims that the physicalist language is the only one currently known to suffice for this purpose.

[121] Carnap writes: "as far as the logical *meaning* of its statements is concerned, science is concerned with only one domain. … On the other hand, in its practical procedures, science does not always make use of this transformability [of statements into one domain] by actually transforming all its statements" (1928/1967, 70).

"some day all the major// branches of empirical science may be formalized… the several sciences would differ from one another only in the empirical constants which occur in them. … This, then, would be one way, and perhaps the only way, in which a real unity of science could be achieved; and an encyclopedia of the sciences would then consist of lists (with elucidations) of the fundamental constants with cross-references to the axioms in which they occur." (1937/1987, 164-5)

Woodger's basic idea, I take it, is to reformulate all the results of science within the language of *Principia Mathematica* (or another equally rich formal language), and then provide (empirical) interpretations for the individual and predicate constants. This is precisely what Woodger himself attempts to do for portions of biology and neurology in his *Axiomatic Method in Biology* (1937) and *Biology and Language* (1952). Note that even on Woodger's picture, we have a unity of language, but not a unity of laws: new empirical constants could be introduced at the level of biology, psychology, or sociology. However, Woodger's vision of a unified science differs from that of Carnap and Neurath outlined above in that Woodger does not explicitly place special stock in the physicalist langauge. He forwards criticisms of the physical (not physicalist) language, but does not consider them utterly decisive (1952, 278, 310). Hans Hahn offers a different view of the unity of science from those considered above, for he is not primarily concerned with a single language for science. He claims that the sciences are unified because they are all evaluated according to how well they are confirmed by observation; this is a version of what philosophers of science today call the 'methodological' unity of the sciences.

2. THE ELIMINATION OF METAPHYSICS

The logical empiricists are (in)famous for assuming an anti-metaphysical stance. All the major figures in the group, as well as most of their patron saints, railed against metaphysics. But how exactly did the logical empiricists purport to identify and excise perniciously metaphysical concepts and claims? This question becomes especially pressing if one agrees with Michael Friedman's assertion that "metaphysical neutrality rather than radical empiricism... is... the essence of Carnap's position" (1999, 110). Alan Richardson also puts this point strongly: "if there is one defining feature of Carnap's philosophy, it is the claim that both science and philosophy can be done in a way that is neutral with respect to the traditional issues of metaphysics" (1992, 45). Such claims need not be restricted to Carnap alone; metaphysical neutrality was a major, if not fundamental, goal for virtually all central logical positivists.

How do the logical empiricists purport to expunge metaphysics from science? The stereotypical view, promulgated in (Ayer 1959), is that the logical empiricists eliminate metaphysics via a comprehensive application of the verificationist criterion of meaning. This view has already been discounted somewhat in (Richardson 1992, 59) and, more indirectly, in (Creath 1982). As I hope to make clear, the verificationist criterion of meaning does play *some* role in *some* logical empiricist rejections of metaphysics—however, its role is often a subsidiary one, and exclusive focus upon it leads to a fundamentally incomplete and therefore distorted image of the logical empiricists' attack on metaphysics. A more complete picture of the logical empiricists' anti-metaphysical project requires keeping their unity of science thesis in view. Roughly put, one criterion separating meaningless metaphysics from cognitively significant discourse that holds over several decades for many logical empiricists is the following:

(M) An apparently declarative sentence or apparently descriptive term is *metaphysical* if and only if that (apparent) sentence or term *cannot be incorporated into a total language of science*.

Furthermore, for the logical empiricists, failures of incorporation into unified science often come in two varieties: a metaphysical claim is either (i.) *ungrammatical*, or (ii.) grammatical but *isolated*. Case (ii.) arises when a grammatical sentence contains metaphysical terms.

I must stress that (M) is an *idealization*. No formulation of its brevity can fully and accurately characterize the logical empiricists' views on metaphysics and unity of science, for the historical situation is fairly complex. Different logical empiricists hold somewhat different views, and a single thinker's ideas about metaphysics often shift over time. Furthermore, the *bi*conditional (M) usually does not appear in the texts as such. Rather, a given logical empiricist virtually always uses only *one* direction of implication at a time, even though that thinker is committed to both directions, and might even use the other direction elsewhere in the very same work. So, (M) should be understood as a slogan, from which actual formulations deviate to a greater or lesser degree, and not as a complete account of logical empiricists' views on the relation between metaphysics and unified science.

The next task, then, is to present a more complete and detailed account of the logical empiricists' rejection of metaphysics across several texts. By examining several variants of (M), we can determine to what extent (M) captures a basic element of logical empiricist thought, and also see what historical nuances and complexities (M) elides. In what follows, I focus on Carnap, for he, more than any other logical empiricist, works out detailed positions on both the unity of science and the rejection of metaphysics.[122] I then show that Neurath's texts support attributing (M) to him as well, though his expression of the rejection of metaphysics lacks the fine-grained particulars of Carnap's.

A. *Aufbau*

Let us Let us begin with Carnap's treatment of metaphysics in the *Aufbau*. How does Carnap there identify metaphysics? Carnap discusses the concepts of essence, reality, and the mind-body connection (among others), and concludes that each, if taken in their customary sense, is metaphysical. Each of these purported concepts is deemed metaphysical on the grounds that it cannot be incorporated into any 'constructional system' [*Konstitutionsystem*] of the sorts Carnap

[122] Schlick writes a good deal about the rejection of metaphysics, but does not meaningfully address the unity of science; Neurath writes a great deal on the unity of science, but his explanations or justifications for rejecting metaphysics are not as sustained as Carnap's.

describes in the *Aufbau*. We can phrase Carnap's criterion for metaphysics in the *Aufbau* as follows:

(M_{Aufbau}) An apparent sentence is metaphysical if and only if it contains concepts that cannot be constructed in a constructional system.

This connection between non-constructability and metaphysics is clear in Carnap's treatment of the metaphysical 'problem of reality':

The concept of reality (in the sense of independence from cognizing consciousness) does not belong within (rational) science, but within metaphysics. This is now to be demonstrated. For this purpose, we investigate whether this concept can be constructed, i.e., whether it can be expressed through objects of the most important types which we have already considered, namely, the autopsychological, the physical, the heteropsychological, and the cultural. (1928/1963, 282)

To show that a concept is metaphysical, it must be shown that that concept cannot be constructed from *any* basic objects—not just from phenomenal, 'autopsychological' ones, but also from physical, heteropsychological, and cultural basic objects. The mind-body problem (in Carnap's terms, the 'parallelism' between mental states and brain states) is similarly unconstructable:

The question for an *explanation of these findings* [viz., mental state tokens and brain state tokens can be placed in a one-to-one correspondence] *lies outside the range of science*; this already shows itself in the fact that this question cannot be expressed in concepts that can be constructed;… (This holds for any such constructional system and not only for a constructional system of our specific kind.) Rather, the quest for an explanation of that parallelism belongs within metaphysics. (Carnap 1928/1967, 270-1)

The above parenthetical remark indicates that, for Carnap, constructability is a more fundamental criterion than verifiability in determining whether a concept or claim is metaphysical, for presumably the 'specific kind' to which Carnap refers is the constructional system with autopsychological basis. That is, what makes a (pseudo-)concept metaphysical is not whether it can be cashed out in terms of certain first-person conscious experiences, but rather whether it can be incorporated within any constitution system—even one which takes physical or cultural objects as basic.[123] Other metaphysical concepts are shown to have the same property; none can be incorporated into a constructional system.

This non-conceptual nature of metaphysics is directly connected to the logical empiricists' well-known rejection of *intuition.* Carnap writes: "metaphysics does not wish to grasp its object by proceeding via concepts… but immediately through intuition", and the deliverances of intuition are ineffable (1928/1967, 295). The opposition between the conceptual and the intuitive is drawn from Kant's first *Critique*: intuitions are given to us via our sensibility, while the source of concepts is the understanding (Kant 1997, 155 A19/B33). The logical empiricists adopt this Kantian distinction, but insist that all (genuine) knowledge is *conceptual*. Schlick, in his "*Erkennen, Erleben, Metaphysik*" [translated as "Cognition, Experience, Metaphysics"] (1926/1979), declares metaphysics incoherent on the grounds that it attempts to have conceptual knowledge of the intuitive. Schlick there characterizes metaphysics as an attempt to have descriptive, conceptual knowledge of that which we can only have conscious awareness/ experiential knowledge (*Erkennen* instead of *Erleben*). And, says Schlick, conscious awareness [*Erlebnis*] is ineffable (for example, *what it is like to see red* cannot be explained to a blind person (*ibid.*, 99)), whereas knowledge by description [*Erkennen*] is, by Schlick's definition, communicable via symbols. Schlick says that "intuition" or "intuitive knowledge… is simply what we spoke of above as *Erlebnis*" (*ibid.*, 108). Thus Schlick's mutually exclusive categories of *Erlebnis* and *Erkenntnis* map nicely onto Kant's strict division between intuition and conceptual thought.[124] Metaphysics, on this view, is an attempt to describe the indescribable

[123] If one accepts Carnap's claims in the *Aufbau* that (1) everything that can be said in any construction system can be said in an autopsychological one, and that (2) all concepts can be *defined* in terms of the autopsychological basis, then 'C cannot be cashed out (defined) in terms of sense experience' will be equivalent to 'C cannot be incorporated into a constructional system.'

[124] There is an interesting issue here about the notion of metaphysics in Kant and in the Vienna Circle. Kant (usually: Aviii, Bxix) characterizes metaphysics as certain claims that "surpass the bounds of all (possible) experience," or "a wholly isolated speculative cognition of reason that elevates itself entirely above all instruction from experience" (Bxiv); but (1.) that is not Schlick's characterization here (?what

contents of consciousness, to communicate the incommunicable intuition—and thus metaphysics is a hopeless enterprise.[125] What is important for current purposes is that what is ineffable cannot, by definition, be incorporated into any language. Finally, there are two significant differences between (M_{Aufbau}) and (M): first, in the *Aufbau*, Carnap thinks primarily in terms of concepts; sentences are secondary. Second, the *Aufbau* lacks the claim that many sentences of metaphysics are ungrammatical. This idea, drawn from Wittgenstein's *Tractatus*, does not come into prominence in Carnap's writings until after the *Wienerkreis* reads the *Tractatus* intensively together.

B. "Overcoming Metaphysics"

Carnap's most focused attack on metaphysics is "Overcoming Metaphysics through the Logical Analysis of Language" (1932/1959). Here Carnap clearly draws the distinction, described above, between the two kinds of pseudo-sentences that cannot be incorporated into the language of science: (i.) ungrammatical strings of symbols, and (ii.) grammatical 'sentences' whose terms cannot be connected to the meaningful terms and sentences of the language. I shall deal with each in turn. Carnap begins "Overcoming Metaphysics" by noting that there have been several attempts throughout the centuries to abolish metaphysics from the intellectual landscape. However, "only" with the "development of modern logic" can "the decisive step be taken" in this pursuit (1932/1959, 61). Why? A sentence (even if it contains only meaningful words) is meaningless, i.e. metaphysical, if it cannot be expressed in a predicate calculus such as *Principia Mathematica*. This is why the 'development of modern logic' is so important to the elimination

about Carnap? Autopsychological basis vs. physical basis in the Aufbau?), and (2.) Kant's characterization sounds like the usual description of the logical empiricists' elimination of metaphysics. Furthermore, Kant characterizes metaphysics as a conceptual, not intuitive, enterprise, which thus belongs to reason, not sensibility (Bxiv).

[125] Schlick characterizes metaphysics differently in other writings. For example, in (1930/1959), he holds that metaphysics occurs when someone holds *propositions* when *actions* are appropriate instead. And in "Postivism and Realism," [REF?] he asserts that metaphysics occurs when a question about *meaning* is mistaken for a question of *fact*.

of metaphysics: we pick out metaphysical sentences by finding those strings of symbols which *appear* meaningful, but cannot be expressed in the logical language of the *Principia*.[126] This conception of metaphysics is fundamentally Tractarian: whatever cannot be expressed grammatically in the ideal symbolic language of the *Tractatus* is meaningless metaphysics. Carnap and Neurath explicitly state that their view on the elimination of metaphysics in the early 1930's "was in essentials that of Wittgenstein" (Carnap 1934/1937, 322; see also Neurath 1983, 54).

This constitutes one of Carnap's criticisms of Descartes' 'I think, therefore I am' (1932/1959, 74). Carnap claims that the statement 'I am' (or 'Greg is,' assuming 'Greg' is an individual constant in the language) cannot be put into the language of classical predicate logic: the concept of existence, in standard first-order logic, is not a first-order predicate, but an operator which acts upon open or closed formulae. Thus Carnap claims it is impossible, given the resources of the language, to express that an individual in the domain of discourse exists *simpliciter*; one can only say either '∃xPx' or 'Pa' (where 'P' denotes some property, and 'a' denotes an individual in the domain of discourse). The string of symbols '∃a' is not admissible as a sentence. Therefore, concludes Carnap, Descartes' assertion is meaningless, since it cannot be expressed in the logical language of *Principia*. (I do not know why Carnap would not allow '∃x(x=Greg)' to express 'Greg is'; this is the usual way of expressing the existence of individuals in classical predicate logic. Perhaps Carnap was under the spell of the *Tractatus* doctrine that declares such an expression meaningless, and the identity-sign superfluous.[127]) Every other sentence that cannot be expressed in Russell and Whitehead's logical symbolism is also declared meaningless, such as Heidegger's '*Das Nichts nichtet*': Carnap points out that in

[126] Alan Richardson has stressed this idea: "The universal applicability and expressive power of the new logic does all the serious work in the rejection of metaphysics" (Richardson 1998, 26-27).
[127] *TLP* 5.531-5.534

formal logic, ‘nothing’ is represented as a concatentation of the negation-sign and an existential quantifier, but Heidegger’s sentence treats it as a substantive, which would be represented as an individual constant in the language. And, of course, one cannot (grammatically) place ‘¬∃x’ into the ‘empty slot’ of a propositional function.

So much for Carnap’s account of metaphysical sentences; when is a *term* metaphysical, i.e., meaningless? Carnap takes us on a brief detour through sentences, for a term is shown to be meaningless by showing that atomic sentences containing that term are meaningless. He asserts that the question “What is the meaning of [an atomic sentence] S?” is equivalent to each of the following two questions:

> (1.) What sentences is S *deducible* from, and what sentences are deducible from S?
> (2.) Under what conditions is S supposed to be true, and under what conditions false?
> (1932/1959, 62)

Here again, we see a version of (M). In this instance, a sentence (and thereby its constituents) is shown to be meaningful by placing it within a larger ‘inferential network’[128] ((1.) captures the syntactic aspect of the network, (2.) the semantic). This inferential network is drawn from the language of science (Carnap’s example is: *x* is an arthropod if and only if *x* is an animal, has a segmented body, and has six legs). Grammatical strings that cannot be placed within such a network of scientific claims, Carnap maintains, contain metaphysical terms. This is very similar to the unconstructable concepts of the *Aufbau*. But, one may wonder, what guarantees that *any* sentences in the larger inferential network are meaningful?—Couldn’t we construct a network of nonsense words?

To answer this question, a logical empiricist would appeal to the verificationist criterion of meaning. Carnap states that ‘What is the meaning of S?’, and hence questions (1.) and (2.) above, are also equivalent to “(3.) How is S to be *verified*?” (*ibid.*). For Carnap in 1932, this question is answered by specifying the deducibility relations between S and the “ ‘observation

[128] Excluding purely logical implications: ‘God exists’ entails ‘God exists or water boils at 100 degrees Centigrade,’ and is entailed by ‘God exists and mammals have hair.’

sentences' or 'protocol sentences.' It is through this reduction that the word acquires its meaning" (1932/1959, 63). However, Carnap explains, the specific nature of the protocol sentences is irrelevant to the elimination of metaphysics: "For our purposes we may ignore entirely the question concerning the content and form of the primary sentences (protocol sentences)": they could deal with "the simplest qualities of sense" (Mach), "total experience and similarities between them" (the *Aufbau*), or simply "things" (*ibid.*). Furthermore, two years later, in "*Über Protokolsätze*," Carnap states that which sentences are protocol sentences is a matter of decision (1934/1987).

Carnap's claim that a word 'acquires its meaning' through its relation to the observation sentence indicates, I take it, that Carnap is making the following two assumptions. First, there exists some set of privileged sentences whose meaningfulness is uncontroversial, assumed, or somehow otherwise guaranteed (this set is the 'observation' or 'protocol sentences'). Second, an arbitrary sentence S is meaningful only if S is non-trivially inferentially related to this other set of sentences. Metaphorically, the meaningfulness of the semantically privileged sentences 'filters down,' via inferential relations, to S. This view about meaning might be called 'semantic foundationalism': just as an epistemic foundationalist holds that there are 'unjustified justifiers' that function as the ultimate source for all claims' justification, a semantic foundationalist holds that there are sentences and/or terms that function as the ultimate source for the meaning of all sentences. We arrive at the full-fledged verification criterion of meaning (as well as the liberalized empiricist meaning criteria which appear later[i]) by adding to the two assumptions of semantic foundationalism a third: observation sentences[ii] (and/or terms) are members of the set of semantically privileged sentences (and/or terms).

We can now see more clearly the respective roles empiricist meaning criteria and a unified language of science play in eliminating metaphysics. Verificationist meaning criteria sanction treating the observational sentences and terms as uncontroversially meaningful. Once we have that assumption, then to determine whether a given sentence is meaningful, we must determine whether it is properly inferentially related to the semantic foundation. But from where are these inferential relations drawn? They are supplied by the language of science, as Carnap's arthropod example above makes clear. If we have a total language of science in which the observational terms and sentences are properly inferentially related to rest of the scientific language, then all scientific claims are guaranteed to be meaningful. Furthermore, the assumption that certain sentences are uncontroversially meaningful offers a solution to the problem, mentioned above, of constructing an inferential network of meaningless strings. In short, Carnap needs both an empiricist criterion of meaning and a total language of science in order to eliminate all metaphysical claims while preserving all cognitively significant ones: the meaning criterion guarantees that the entire inferential network will not be meaningless, and the language of science, by exhibiting the inferential relations between the semantically privileged sentences and all the other scientific sentences, shows the sentences of physics, biology, and psychology to be meaningful.[129]

I will make a final remark about the verification criterion of meaning before returning to the main thread of the argument. As we have seen, for Carnap, the question 'How is S to be incorporated into a collection of knowledge claims (if at all)?' is intended to generate the same information (perhaps under a different description) as 'How is S verified (if at all)?' A similar idea appears in his *Philosophy and Logical Syntax*. There, Carnap asserts that two apparently different philosophical criteria for eliminating metaphysics will yield "the same result": (i.) accept only those claims that satisfy of the verification criterion, and (ii.) accept only those claims that have "the possibility of being placed in a certain system, in this case, the space-time-system of the physical world" (1935/1963, 429). This shows that the verification criterion does not *supplant* the metaphysical criterion based upon the unified language of science, but rather—at least in Carnap's mind—the two criteria are extensionally equivalent.

---

[129] For the previous two paragraphs, I am very indebted to Jon Tsou.

C. *Logical Syntax*

As Carnap's philosophical views change over hus career, so does his characterization of the metaphysical. In 1934, *Logical Syntax of Language* appears, and with it a slightly modified program for eliminating metaphysics. We find the same basic ideas as in "Overcoming Metaphysics," but with an added wrinkle: the principle of tolerance. In *Logical Syntax*, what counts as metaphysical becomes (to a degree) *language relative*, as follows:

(M_{LSL}) An apparently declarative sentence or apparently descriptive term is *metaphysical with respect to a language of science L* if and only if that (purported) sentence or term *cannot be incorporated into L*

where 'incorporation' is understood as before.

Carnap describes how the anti-metaphysical drive interacts with the principle of tolerance as follows:

The view here presented [in accordance with the principle of tolerance] allows great freedom in the introduction of new primitive concepts and new primitive sentences in the language of physics or the language of science in general; yet at the same time it retains the *possibility of differentiating pseudo-concepts and pseudo-sentences* from real scientific concepts and sentences, *and thus of eliminating the former*. [This elimination, however, is not so simple as it appeared to be on the basis of the earlier position of the Vienna Circle… On that view it was a question of "*the* language" in an absolute sense; it was thought possible to reject both concepts and sentences if they did not fit into *the* language.] (1934/1937, 322)

Carnap holds that we can still avoid metaphysical pseudo-concepts and pseudo-sentences, even if we adopt the Principle of Tolerance and thereby reject the notion that there is a single 'correct' language. As in "Overcoming Metaphysics," the 'sentences' that are ungrammatical, and those apparently descriptive sentences that cannot be connected with the language of empirical science are dismissed as pseudo-sentences, as metaphysics (*ibid.*). So while there might be more than one acceptable language of science, traditional metaphysical concepts will nonetheless still be excluded, for they will not occur in any language *of science* (even though they might appear in some other, non-scientific language).

Is it reasonable to hold, with Carnap, that what counts as metaphysics is language relative? If we think of metaphysics as *nonsense,* as the Vienna Circle and Wittgenstein do, then the label of ‘metaphysical’ *should* be indexed to a particular language—for what is meaningful in one language often simply will not be in another. Let us examine a Carnapian example to explore the possibility of the language-relativity of metaphysics. Consider Languages I and II of *Logical Syntax*: Language I, intended to capture the mathematical intuitionist’s viewpoint, is weaker than Language II, which is expressively rich enough to capture all of classical analysis. Thus, there are sentences that are grammatical in II, but ungrammatical in I, and hence metaphysical from the point of view of someone using Language I. (For example, a sentence about ‘unconstructable numbers’ would be a metaphysical pseudo-sentence in I, but not in II.) As a second example, consider the relation between first-order and higher-order logics: certain sentences of second-order logic would be, on Carnap’s criterion, metaphysical in first-order logic (namely, those involving higher-order predicates). Perhaps this relativization of metaphysics to languages reveals something insightful about the way the term ‘metaphysics’ is used. For intuitionists *do* find something suspect about the unconstructable numbers of classical mathematics, and some would be inclined to call claims about such entities ‘metaphysics.’ Heyting, expressing the intuitionist viewpoint, writes: “If ‘to exist’ does not mean ‘to be constructed,’ it must have some metaphysical meaning” ([1971] 1983, 67). Similarly, philosophers who find second-order logic suspicious call its quantification over properties ‘Platonism,’ after the grandfather of all metaphysicians. Thus Carnap’s suggestion, that what counts as metaphysics depends on the language one uses, is borne out in these examples. In sum, in *Logical Syntax*, the conception of metaphysics is, at root, the same as that found in Carnap’s earlier works, but modified to accommodate the principle of tolerance.

D. “Empiricism, Semantics and Ontology”

In 1950’s “Empiricism, Semantics, and Ontology,” Carnap’s basic idea for identifying metaphysics is essentially the same. However, the terminology has shifted: instead of speaking of constructional systems or languages, Carnap now speaks of linguistic frameworks. But here again, a claim is shown to be non-metaphysical by incorporating it into a (pragmatically) acceptable linguistic framework.

[T]he concept of reality…in internal questions is… [a] scientific, *non-metaphysical* concept. To recognize something as a real thing or event means to *succeed in incorporating it into the system of things…, according to the rules of the framework.* … (1950/1956, 207; my italics)

The importance of a shared scientific language for identifying metaphysics also recurs here. It is on precisely these grounds that Carnap criticizes philosophers who ask the 'external' question "Are there numbers?":

Unfortunately, these philosophers have not given a formulations of their question in the common scientific language. Therefore… they have not succeeded in giving the external question cognitive content. (1950/1956, 209)

And questions without 'cognitive content' are metaphysical. Thus, Carnap's attitude towards metaphysics in 1950 is very closely related to his view in the twenties; linguistic frameworks replace construction systems, but the basic strategy for identifying and eliminating metaphysics remains the same.

E. Neurath

So much for Carnap's views on metaphysics; what of Neurath? Though he eschews Carnap's formal, precise languages in favor of his 'universal jargon' or 'universal slang' based on everyday language with its sloppy *Ballungen*, he shares the fundamental idea found in Carnap: an apparently meaningful sentence or term is metaphysical if and only if it cannot be incorporated into unified science. First, the 'only if' direction:[130]

If it [a proposed scientific sentence] is… meaningless—i.e., metaphysical —then of course it falls outside the sphere of unified science. (1983, 58)

[130] See also (1983, 54, 57, 61, 73, 173). Neurath sometimes speaks of 'physicalism' instead of 'unified science,' but, for Neurath, "physicalism is the form work in unified science takes in our time" (1983, 56).

For Neurath, perhaps even more than for Carnap, unified science is identified with physicalism: "physicalism is the form work in unified science takes in our time" (1983, 56). Thus we find assertions such as the following, which make essentially the same point as the previous quote:

If we systematically formulate everything we find in non-metaphysical formulations, we get nothing but physicalist formulations. (1983, 73)

And the 'if' direction:

statements that through their structure or special grammar could not be placed within the language of the encyclopedia—in general 'isolated' statements, … are statements 'without meaning in a certain language'. For these statements the Vienna Circle has often used the term 'metaphysical statements'. (1983, 161)

Note that Neurath mentions the strictures against both ungrammatical and isolated 'sentences.' As an example of an ungrammatical (and hence metaphysical) assertion, Neurath offers Kant's categorical imperative. Neurath characterizes it as "a command without a commander," and thus as "a defect of language" (1983, 54). And, he elaborates, "[a]n unblemished syntax is the foundation of an unblemished unified science" (*ibid.*). Where Carnap employs a constitution system or a linguistic framework, Neurath uses an encyclopedic language based on everyday communication instead; but otherwise, their views are very close.

Recall the notion of 'semantic foundationalism,' mentioned above: a sentence's meaningfulness is demonstrated by showing that that sentence is connected via consequence-relations to sentences whose meaningfulness is given antecedently. Carnap identifies these semantically privileged sentences as the protocol, *observational* ones (though, as we saw, he was willing by 1932 to leave the exact form of such sentences open). Neurath suggests a different set of antecedently meaningful sentences. Neurath repeatedly states that unified science should begin from everyday language, with minor corrections. Why? One possible reason is that everyday language is meaningful if any language is; everyday language would be the most indisputable case of a meaningful language. We are more committed to the meaningfulness of everyday language than any other. Thus, if we have to pick a 'semantic foundation,' everyday language seems the best we can do. (There are other reasons Neurath starts with everyday language: he values the democratization and popularization of scientific knowledge,[131] and he is suspicious of any framework that aims to break loose of our present historically given situation.)

One might criticize my interpretation of Neurath's claims about the unity of science as follows: a central aim of work in unified science is demolition of the barriers between the scientific study of nature and of the mind; my interpretation misses that aspect entirely. I concede, of course, that Neurath repeatedly and unequivocally urged the value of breaking down disciplinary barriers. But, interestingly, Neurath claims that the motivation underlying the separation of the sciences is *metaphysical*. When his program is realized,

> each basic decomposition of unified science is eliminated…, for example, that into 'natural sciences' and 'mental sciences'… . The tenets with which we want to justify the division are… always of a metaphysical kind, that is, meaningless. (1983, 68)[132]

So, according to Neurath, any assertion used to justify a strict division of the sciences is metaphysical. If the various sciences were unified, then any such assertion would be ruled out. Thus, unified science, which shows disciplinary barriers are not insuperable, eliminates a certain kind of metaphysics—specifically, it eliminates any theory that purports to deal with "a special sphere of the 'soul'" (1983, 73), distinct from the remainder of the spatiotemporal world. Carnap makes a very similar point in "The Task of the Logic of Science," though he characterizes the mental/ material division as motivated by "mythological" and "divine" motives, and does not explicitly use the word 'metaphysical' (1934/1987, 58-9). Unification of the sciences may be valuable for its own sake, but it also serves to eliminate metaphysics.

---

[131] "A Universal Jargon… would be an advantage from the point of view of popularizing human knowledge, internationally and democratically. … [It] seems // to me something fundamentally anti-totalitarian" (1946/1983, 236-7). And children can easily be taught such a language: "But how does the elimination of metaphysics proceed in practice?… *Every child can in principle learn to apply the language of physicalism from the outset*" (1932/1987, 9).

[132] See also (1983, 44, 50, 69).

3. A DIFFICULTY: WHAT *CANNOT* BE INCORPORATED INTO A LANGUAGE OF SCIENCE?

A concept or sentence, then, is metaphysical if it cannot be integrated into any unified language adequate for science (or constitution system, or linguistic framework, etc.). The central and pressing problem for this account of metaphysics is: how do we know which concepts and claims can be incorporated, and which cannot? That will determine what is metaphysical and thus in need of excision. Let us focus first on the *Aufbau.* When Carnap gets down to the details of showing how essences and theses about the mind-body problem cannot be formulated in any constitutional system, he (unfortunately) offers more assertions than arguments.[133] For example, Carnap simply asserts that "essence," taken in its "metaphysical" sense, cannot be constructed in the autopsychological constitution system of the *Aufbau* or any other constitution system ([1928] 1963, §161).

Carnap's treatment of the mind-body problem is similar, but it better illustrates the potential pitfalls or shortcomings of equating the metaphysical with the unconstructable. In the constitution system, Carnap says, we can discern a "parallelism" between two "sequences," one of which corresponds to "the construction of physical objects" and a second which does not. The mind-body problem asks: "how can the occurrence of a parallelism of sequences of constituents be explained?" Carnap responds that this "question cannot be expressed in concepts that can be constructed; for the concept… 'explanation' … [does] not in this sense have any place in a constructional system of objects of cognition," and this holds for "*any* such constructional system"). Therefore, an "explanation of these findings lies outside the range of science" ([1928]

[133] Alberto Coffa expresses surprise at how scanty Carnap's argumentation is here (1991, 225). Richardson has suggested that this "lax" argumentation on Carnap's part is due to the fact that "Carnap takes it as a point of agreement between himself and the metaphysicians that metaphysical debates are not scientific debates," i.e., that metaphysical concepts are outside the ken of science (Richardson 1992, 60). If that were true, why would Carnap bother writing Part V ("Clarification of Some Philosophical Problems on the Basis of Construction Theory"), which reviews particular metaphysical concepts one by one, and argues that each is not constructable within the system? This indicates Carnap genuinely does intend to show (instead of simply assume) that certain concepts cannot be incorporated into a constitution system, thereby showing more specifically how metaphysical claims are not scientific claims.

1963, 270). In short, Carnap's position is that 'explanation' is an unconstructable concept, so the mind-body problem is unscientific metaphysics on the grounds that it requests an *explanation* of a certain parallelism between physical and phenomenal sequences.

But what if Hempel and Oppenheim's groundbreaking "Studies in the Logic of Explanation" were published not in 1947 but in 1922? The sense of 'explanation' offered in that article might be sufficiently precise, clear, and scientifically respectable for the Carnap of the *Aufbau* to think that a notion of explanation *could* be formulated within a constitution system. Regardless of what Carnap's reaction would have been under this particular counterfactual circumstance,[134] this points to a serious and fundamental difficulty. In every case where Carnap (or any other logical empiricist) asserts that a given term or sentence cannot be incorporated into any unified language of science, another person could later show how that concept can, in fact, be so integrated. For example, Tarski showed how to define 'truth' rigorously, a term that many logical empiricists previously considered the province of the speculative metaphysician. Claude Shannon gave the concept of information a mathematically tractable characterization, and spawned a fruitful sub-discipline of mathematics. In general, the regimentation of a sentence or term from pre-analyzed usage into a form acceptable for use in a unified language of science can be a difficult process, often requiring substantial intellectual creativity. Russell's struggles with descriptions provide another kind of example of this phenomenon. In short, (M) and its variations are problematic because there is no general procedure for determining the truth-value of sentences of the form 'Concept C cannot be incorporated into the unified language L' (much less 'into *any* unified language L'), because such an incorporation, however unexpected, could be achieved tomorrow given sufficient ingenuity. Our limited technical ability is not a demonstration of impossibility. I am not claiming that (M) fails to provide necessary and

[134] For example, the Carnap of the *Aufbau* might not have allowed that certain constructable sentences are somehow privileged by being 'laws of nature,' and the Hempel and Oppenheim analysis of explanation requires that we be able to identify such laws. (However, in (Carnap 1966), Carnap happily accepts and deploys the concept of a law of nature in his account of scientific explanation.)

sufficient conditions for identifying metaphysical terms and claims; rather, the problem is that, in many cases of interest, we cannot know whether those conditions have been met.

One might reply to this objection as follows: before Tarski's work, 'true' *was* a metaphysical term, and it only became part of cognitively significant discourse *after* the publication of *Wahrheitsbegriff*—and similarly for any other terms and sentences that are not now incorporated into a unified language of science. In effect, this reply suggests a friendly emendation of (M), by modifying the boundary marking off the metaphysical. Specifically, this reply endorses replacing (M) with

(M*) An apparently declarative sentence or apparently descriptive term is *metaphysical* if and only if that (apparent) sentence or term ***is*** *not incorporated into a total language of science*.

The only difference between (M) and (M*) is that the latter lacks the former's modal force. Adopting (M*) would constitute a departure from the logical empiricists' original conception of the link between metaphysical neutrality and unified science, but it would also defuse the objection raised in the previous paragraph.

However, it appears that (M*) creates a problem at least as severe as the one it solves: (M*) makes the line dividing metaphysics from cognitively significant discourse overly sensitive to the intellectual abilities and interests of the *Wissenschaftslogiker*. Suppose a new theory, employing a set of new terms, is introduced into the developmental psychology literature this year. If the people constructing a unified language of science are either underinformed or simply too dense to see how to connect these new terms with older, antecedently meaningful ones, then these novel terms will qualify as metaphysical under (M*). Even worse, under (M*) what qualifies as metaphysics will depend on the particular interests of the *Wissenshaftslogiker*. Suppose that no one in the group formulating a unified language has an interest in ecology; their efforts are focused instead upon incorporating (e.g.) chemical and psychological language into the unified language. Because time and resources are finite, the terms unique to ecology may not be incorporated into the unified language now (or ever), and thus large chunks of ecology would be classed as metaphysics by (M*), simply because no *Wissenschaftslogiker* had managed to fit that project into the schedule. The obvious remedy for this unacceptable delineation of the metaphysical is to hold that these new terms from developmental psychology and the terms unique to ecology may not be incorporated into a unified language of science yet, they nonetheless *could* be, and for that reason are not metaphysical. But that position is just the logical empiricists' (M).

Thus far I have argued that, in the writings of central logical empiricists, there is a close conceptual connection between the unity of science thesis and the elimination of metaphysics, and that this connection is captured, to a first approximation, by (M). In closing, I present one piece of evidence that this connection is not merely conceptual, but also *genealogical*. That is, the term 'unified science' [*Einheitswissenschaft*], suggested by Neurath, sprung directly out of the Vienna Circle's elimination of metaphysics. Neurath, recalling the Circle's discussion of the *Tractatus*, explains how he came to introduce the term.

> Eliminating 'meaningless' sentences became a kind of game… But I very soon felt uneasy, when members of our Vienna Circle suggested that we should drop the term 'philosophy' as a name for a set of sentences … but use it as a name for the activity engaged in improving given sentences by 'demetaphysicalizing' them … Thus I came to suggest as our object, the collection of material, which we could accept within the framework of scientific language; for this I thought the not-much-used term 'Unified Science' (*Einheitswissenschaft*…) a suitable one. (1983, 231)

Thus, the very term 'unified science' arose directly from a desire to re-name the anti-metaphysical goal of the *Wienerkreis*. The two goals are, metaphorically, two sides of the same coin: the elimination of metaphysics is the negative or destructive aspect, while the production of a unified scientific language constitutes its positive or constructive aspect.

BIBLIOGRAPHY

Benacerraf, Paul (1965/1983) "What Numbers Could Not Be," reprinted in *Philosophy of Mathematics: Selected readings*, second ed., Benacerraf and Putnam eds., Cambridge: Cambridge University Press.

Bernays, Paul (1935/1983). "On Platonism in Mathematics," reprinted in *Philosophy of Mathematics: Selected Readings*, second ed., Benacerraf and Putnam eds., Cambridge: Cambridge University Press.

Bouveresse, Jacques (2005) "On the Meaning of the Word 'Platonism' in the Phrase 'Mathematical Platonism'," *Proceedings of the Aristotelian Society*, 55-70.

Burgess, John (1983) "Why I Am Not a Nominalist," *Notre Dame Journal of Formal Logic* **24**, 93-105.

--- (1992) "How Foundational Work in Mathematics can be Relevant to Philosophy of Science" PSA 1992 vol. 2, pp. 433-441.

Burgess, John and Rosen, Gideon (1996) *A Subject with No Object*. Oxford: Oxford University Press.

Carnap, Rudolf (1928/1963). *The Logical Structure of the World*. Rolf George trans., Chicago: Open Court.

--- (1932/1934). *Unity of Science*. Max Black trans., London: Kegan Paul.

--- (1932/1959). "Overcoming Metaphysics through Logical Analysis of Language," in *Logical Positivism*, A. J. Ayer ed. Glencoe, IL: Free Press, 60-81.

--- (1934/1937). *Logical Syntax of Language*. Amethe Smeathon trans., Chicago: Open Court.

--- (1934/1987). "On Protocol Sentences," Creath and Nollan trans., *Nous* 21, 457-470.

--- (1936-37). "Testability and Meaning," *Philosophy of Science* 3-4

--- (1939). "Logical Foundations of the Unity of Science," in *International Encyclopedia of Unified Science*, Vol. I Number 1. Chicago: University of Chicago Press, 42-62.

--- (1942). *Introduction to Semantics*. Cambridge: Harvard University Press.

--- (1956) *Meaning and Necessity*, second ed. Chicago: University of Chicago Press

--- (1954/ 1958). *Introduction to Symbolic Logic and its Applications.* Meyer and Wilkinson, trans., New York: Dover Publications. (Originally published as *Einführung in die symbolische Logik*.)

--- (1963) “Intellectual Autobiography,” in *Philosophy of Rudolf Carnap*, Paul Schilpp, (ed.). La Salle: Open Court Press.

Chwistek, Leon (1932-33) “Die Nominalistische Grundlegung der Mathematik,” *Erkenntnis* **3**, 367-388.

--- (1935/1949) *The Limits of Science*. Helen Brodie and Arthur Coleman (trans.), London: Routledge and Kegan Paul.

Coffa, Alberto (1987) “Carnap, Tarski and the Search for Truth” *Nous* **21** (1987) pp. 547-572.

--- (1991) *The Semantic Tradition from Kant to Carnap.* Cambridge: Cambridge University Press.

Colyvan, Mark (2001) *The Indispensability of Mathematics*. Oxford: Oxford University Press.

Creath, Richard (1982). “Was Carnap a Complete Verificationist in the *Aufbau*?”, *PSA 1982*, 384-393.

--- (1987) “The Initial Reception of Carnap’s Doctrine of Analyticity,” *Nous* **21** 477-499.

--- (1990) Introduction to *Dear Carnap, Dear Van*

--- (1991) “Every Dogma Has Its Day” *Erkenntnis* **35**, pp. 347-389

--- (1992) “Carnap’s Conventionalism” *Synthese* **93** pp. 141-165

--- (1996a). “The Unity of Science: Carnap, Neurath, and Beyond,” in Galison and Stump eds., *The Disunity of Science: Boundaries, Contexts, and Power*. Stanford: Stanford University Press, 417-427.

--- (1996b) “Languages without Logic” in Giere and Richardson (eds.) *The Origins of Logical Empiricism* pp.251-265

--- (2004). “Quine on the Intelligibility and Relevance of Analyticity,” in *The Cambridge Companion to Quine*, Roger Gibson Jr. (ed.), Cambridge: Cambridge University Press.

Dalla Chiara, Maria Luisa and Giulano Toraldo di Francia (1995) “Quine on Physical Objects,” in *On Quine*, Paolo Leonardo and Marco Santambrogio eds., Cambridge: Cambridge University Press, 104-112.

David, Marian (1996) “Analyticity, Carnap, Quine, and Truth” *Nous* **30** pp. 281-296.

Davidson, Donald (1984). “Truth and Meaning,” in: Davidson D., *Inquiries into Truth and Interpretation*, Oxford: Oxford UP.

Detlefsen, Michael (1986) *Hilbert's Program: an Essay on Mathematical Instrumentalism*, Boston: D. Reidel.

Dummett, Michael (1956) "Nominalism" *Philosophical Review* **65** pp. 491-505.

Etchemendy, John (1990) *On the Concept of Logical Consequence*.

Feferman, Anita Buridan (1999). "How the Unity of Science Saved Alfred Tarski," in Wolenski and Köhler eds., pp.

Feferman, Solomon (1992) "Why a Little Bit Goes a Long Way: Logical Foundations of Scientifically Applicable Mathematics" *PSA Proceedings 1992* pp.442-455

Field, Hartry (1980) *Science without Numbers: A Defense of Nominalism*, Princeton: Princeton University Press.

--- (2001) "Quine and the Correspondence Theory," in *Truth and the Absence of Fact*, New York: Oxford University Press.

Frank, Philipp (1947) "The Institute for the Unity of Science: Its Background and Purpose" *Synthese* **6**, pp.160-167.

Friedman, Michael (1999) *Reconsidering Logical Positivism*, Cambridge: Cambridge University Press.

Goldfarb, Warren (1996). "The Philosophy of Mathematics in Early Positivism" in Giere and Richardson, pp. 213-230

Goodman, Nelson (1951/ 1966) *The Structure of Appearance*, 2nd ed. Indianapolis: Bobbs-Merrill.

Goodman, Nelson (1956) "A World of Individuals" in *The Problem of Universals: A Symposium*. Notre Dame: University of Notre Dame Press.

Hahn, Hans (1934/1980) "Does the Infinite Exist?" in *Empiricism, Mathematics, and Logic*, edited by B. McGuiness, Dordrecht: Reidel, 103-131.

Hardcastle, Gary (2003). "Debabelizing Science: The Harvard Science of Science Discussion Group, 1940-41," in *Logical Empiricism in North America: Minnesota Studies in the Philosophy of Science, vol. XVIII*, Gary Hardcastle and Alan Richardson, eds., 170-196. Minneapolis: University of Minnesota Press.

Hellman, Geoffrey (1992) "On the Scope and Force of Indispensability Arguments" *PSA 1992* vol. 2, pp. 456-464.

Heyting, Arend (1971/1983). "Disputation," in *Philosophy of Mathematics*, Benacerraf and Putnam eds., Cambridge, Cambridge University Press.

Hodges, Wilfred (1985/86) "Truth in a Structure" *Proceedings of the Aristotelian Society* **86**, 135-151.

Hylton, Peter (2001) "'The Defensible Province of Philosophy': Quine's 1934 Lectures on Carnap," in *Future Pasts: The Analytic Tradition in Twentieth-Century Philosophy*, J. Floyd and S. Shieh (eds.), Oxford: Oxford University Press.

Kotarbinski, Tadeuz (1929/1966) *Gnosiology. The Scientific Approach to the Theory of Knowledge*, O. Wojtasiewicz trans., Oxford: Pergamon.

--- (1935/1955) "The Fundamental Ideas of Pansomatism," D. Rynin and A. Tarski (trans.). *Mind* **64**, 488-500.

Lavers, Gregory (2004) "Carnap, Semantics, and Ontology" *Erkenntnis* **60**, 295-316.

Maddy, Penelope (1990) "Physicalist Platonism," in *Physicalism in Mathematics*. A. D. Irvine (ed.), Boston: Kluwer, pp. 259-290.

Neurath, Otto (1938). "Unified Science as Encyclopedic Integration," in *International Encyclopedia of Unified Science*, Vol. I Number 1. Chicago: University of Chicago Press, 1-27.

--- (1946), "Unity of Science Movement: After Six Years" *Synthese* 5, 77-82.

--- (1946b). "Wichtigkeit Vorsichtiger Bemerkungen," *Synthese* **5**, 81-82.

--- (1983), *Philosophical Papers*, R. S. Cohen and M. Neurath, trans. & eds., Dordrecht: Reidel.

Parsons, Charles (1980) "Mathematical Intuition" *Aristotelian Society Proceedings*, vol.80 (New Series) 1980.

Poincare, Henri (1902/1905) *Science and Hypothesis*. W. J. Greenstreet (trans.), New York: Dover.

Psillos, Stathis (2000) "Rudolf Carnap's 'Theoretical Concepts in Science'" *Studies in the History and Philosophy of Science* **31**(1), 151-172.

Quine, Willard Van Orman (1939) "Designation and Existence," *Journal of Philosophy*, **36**(26), 701-709.

--- (1940/1958) *Mathematical Logic*, revised edition. Cambridge: Harvard University Press.

--- (1943) "Notes on Existence and Necessity" *Journal of Philosophy* **40**(5), 113-127.

--- (1947) "On the Problem of Interpreting Modal Logic," *Journal of Symbolic Logic*, **12**(2), 43-48

--- (1953/1961) *From a Logical Point of View*, second edition. Cambridge: Harvard University Press.

--- (1966/1976) *The Ways of Paradox*, second edition. Cambridge: Harvard University Press.

--- (2001) "Confessions of a Confirmed Extensionalist," in *Future Pasts: The Analytic Tradition in Twentieth-Century Philosophy*, J. Floyd and S. Shieh eds., Oxford: Oxford University Press.

Quine, W.V. and Goodman, N. (1940) "Elimination of Extra-logical Postulates," *Journal of Symbolic Logic* **5**, 104-109

--- (1947) "Steps Toward a Constructive Nominalism," *Journal of Symbolic Logic* **12**, 105-122.

Richardson, Alan (1992), "Metaphysics and Idealism in the *Aufbau*," *Grazer Philosophische Studien* **43**, 45-72.

--- (1998), *Carnap's Construction of the World*. Cambridge: Cambridge University Press.

Ricketts, Thomas (1996). "Carnap: From Logical Syntax to Semantics" in Giere and Richardson (eds.) pp.231-250

Rosen, Eric (2002). "Some Aspects of Model Theory and Finite Structures" *Bulletin of Symbolic Logic* **8**, pp.380-403

Russell, Bertrand (1918) *Mysticism and Logic and other essays*. London: Longmans Green.

Russell, Bertrand (1918/1956) "The Philosophy of Logical Atomism," reprinted in *Logic and Knowledge*, Edited by R. C. Marsh (London: Allen and Unwin, 1956).

Russell, Bertrand (1919) *Introduction to Mathematical Philosophy*. London, Allen and Unwin.

Russell, Bertrand (1926) *Our Knowledge of the External World*, second edition (first ed. 1914). London, Allen and Unwin.

Schlick, M. (1926) "Experience, Cognition, and Metaphysics" in Mulder and van de Velde-Schlick (1978-9), *Moritz Schlick: Philosophical Papers*. Vol. 2, Dordrecht: Reidel, 99-111.

Shapiro, Stewart (1983) "Mathematics and Reality," *Philosophy of Science* **50**, 523-548.

Simons, Peter (1993) "Nominalism in Poland," in *Polish Scientific Philosophy: the Lvov-Warsaw School*, Francisco Coniglione, Roberto Poli, and Jan Wolenski eds., Amsterdam: Rodopi.

Stein, Howard (1992) "Was Carnap Entirely Wrong, After All?" *Synthese* **93**, 275-295.

Stevens, Stanley Smith (1974). "S. S. Stevens," in *A History of Psychology in Autobiography*, ed. G. Lindzey, 395-420. New York: Appleton, Century, Crofts.

Suppes, Patrick (1978), "The Plurality of Science," *PSA 1978*, Vol. II, 3-16.

Tait, W. W. (1981) "Finitism," *Journal of Philosophy* **78**, pp.524-556

--- (2000) "Remarks on Finitism," in *Reflections. A Festschrift Honoring Solomon Feferman*, W. Sieg, R. Sommer, and C. Talcott (eds.). Urbana: Association for Symbolic Logic, Lecture Notes in Logic, Volume 15 (2002): 407-16.

Tarski, Alfred (1944) "The Semantic Conception of Truth" *Philosophy and Phenomenological Research* **4**, pp.341-376

--- (1956, 1983 2nd ed.) *Logic, Semantics, Metamathematics*. Indianapolis: Hackett.

--- (1986) "What Are the Logical Notions?" *History and Philosophy of Logic* **7**, 143-154.

Uebel, Thomas (2001) "Carnap and Neurath in Exile: Can Their Disputes be Resolved?" *International Studies in the Philosophy of Science* **15**, pp.211-220.

Urquhart, Alasdair (1990) "The Logic of Physical Theory" in *Physicalism in Mathematics*. A. D. Irvine (ed.), Boston: Kluwer, pp. 145-154.

Wittgenstein, Ludwig (1921/1961) *Tractatus Logico-Philosophicus*, D. F. Pears and B. F. McGuinness (trans.). Atlantic Highlands, NJ: Humanities Press International.

Wolenski, Jan (1993) "Tarski as Philosopher," in *Polish Scientific Philosophy: the Lvov-Warsaw School*, Francisco Coniglione, Roberto Poli, and Jan Wolenski eds., Amsterdam: Rodopi.

Zach, Richard (2001) *Hilbert's Finitism: Historical, Philosophical, and Metamathematical Perspectives*. Unpublished dissertation; downloaded from `http://www.ucalgary.ca/~rzach/papers/hilbert.html`.