Carnap, Tarski, and Quine's Year Together: Conversations on Logic, Mathematics, and Science

Greg Frost-Arnold

September 26, 2008

Contents

Preface

Several years ago, I began combing through the Rudolf Carnap Collection at the University of Pittsburgh's Archives of Scientific Philosophy, searching for material about Alfred Tarski's theory of truth. I fortuitously came across a folder full of dictation notes Carnap had taken during conversations with Tarski and others, mostly from 1941. Although I did not fully recognize then what I had stumbled upon, those notes are the originating cause of this book. The first portion of this book consists of my attempts to come to grips with those documents, both in terms of their place in the history of analytic philosophy, and their often surprising philosophical content. As the subtitle of this book suggests, I have grouped this content into three interdependent sections: mathematical nominalism (mathematics), analytic truth (logic), and the unity of science. This choice of what to focus on in these wide-ranging notes certainly reflects my own, sometimes idiosyncratic interests in and perspective on the history of analytic philosophy; another, almost entirely different book could most likely be written about the same archival material. Because the documents would amply reward such study, an edited version of the archival material itself can be found in the Appendix, along with an English translation.

I do not feel that *I* wrote this book. The final product is very much the result of many people—I just happen to be the person who put the most time into the group's project. My inability to see obvious errors that others spotted easily, and others' insightful and creative suggestions for new directions, truly made this a collaborative effort. John Earman and Nuel Belnap gave me very helpful guidance as this material coalesced slowly into a dissertation. An audience at HOPOS 2004 helped improve the core of what is now the first chapter; Alan Richardson and Chris Pincock, in particular, both provided fruitful suggestions. A more intimate audience at the 2004 PSA helped iron out some of the deficiencies of the final chapter; Rick Creath provided especially useful feedback both then and later, as the larger project progressed. His *Dear Carnap, Dear Van* was not only an invaluable research tool for present purposes, but also served as an exemplar after which this book is patterned. Michael Friedman, Don Howard, and Thomas Ricketts each brought their substantial erudition to bear on various ideas presented here; their ability to see the historical 'big picture' was a very helpful corrective to my myopia and ignorance. Andre Carus gave me very helpful suggestions about the historical big picture as well, and Chapter 1 in particular was thereby much improved. Marion Ledwig, Greg Lavers, and Jon Tsou each undertook the task of reading the entire manuscript when it was close to completion; their fresh eyes provided wonderful new perspective on ideas that had been bouncing around in my head for years. Jon, in particular, engaged my arguments seriously, and (thankfully!) would not let me get away with anything that was too quick. I am certain I am leaving out several people who helped me recognize and rectify deficiencies in the ideas presented here, only because I've been overeager to discuss this material with anyone who would listen for the last five years or so. I apologize to you all.

I have saved this book's three greatest intellectual debts for last. First

is Paolo Mancosu, who found the records of these conversations at roughly the same time I did, and began working on them shortly thereafter. Writing about documents that no one has ever seen before is not something for which the usual historical training on Plato or Descartes really prepares you. I am extremely fortunate that a philosopher as able—and as intellectually generous—as Paolo began studying these works as well: his perspective kept me from being completely lost in the dark as I attempted to wade through Carnap's notes. More than he perhaps realizes, his published work and private comments helped iron out serious distortions and outright misunderstandings in my conception of what is going on in this material, and he helped me to recognize better what is central and what is peripheral.

Steve Awodey gave me copious and insightful feedback every step of the way, especially when the project was just beginning: he read very, very rough drafts with more care and respect than they deserved at the time. His ability to comment constructively on ideas that were highly underdeveloped—instead of simply saying 'think about this more,' the common and easy response—was a great boon. And Steve has shown excellent editorial touch as well, giving me all the freedom I could want. Plus, he prevented me from saying some very wrong-headed things about Russell's theory of types.

Laura Ruetsche's help has been essential from the very beginning of this work. She possesses the enviable knack of seeing to the core of an issue whose surface I'll only scratch for 20 pages or so. Laura has been a true mentor, not merely a very dedicated and astute reader: she struck the perfect balance of encouragement for work I had already done, while showing me how much improvement could yet be done. I am extremely fortunate that she supervised my dissertation.

My wife Karen has not only given me a wonderful life, but has been a fantastic philosophical rudder as well. Whenever I get lost in the philosophical trees, she shows me the forest again. More importantly, when I'm confronted with an apparently insuperable obstacle, she shows me that it can be overcome, and when I stumble, she helps me up.

Chapter 1

Historical Background and Overview of Harvard Activities

1.1 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 had 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 [20, 34]. Carnap was a visiting professor at Harvard during the academic year 1940–41.

[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 Archives of Scientific Philosophy has Carnap's notebook from this seminar.

[2] Significant historical work on Tarski's first years in the U.S., as well as the rest of his life, can be found in [35]; especially relevant for the present period under study is Chapter 5: "How the 'Unity of Science' Saved Alfred Tarski".

Hempel crossed the Atlantic after being invited by Carnap in 1937 to serve as his research associate [20, 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 a position 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, 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 (RCC), part of the Archives of Scientific Philosophy at the University of Pittsburgh. The present work takes these documents as its primary subject matter. Most of the notes are records of discussions, but some are Carnap's own reflections on the topics, composed in private. These documents have been previously studied in [62], which provides an excellent overview to this material; I hope here to build upon Mancosu's work.

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 included 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 [98, 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, Herbert 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

[3]Goodman also tells us that both Quine and Carnap "read *A Study of Qualities* with great care and made innumerable invaluable suggestions" [49, *x*]. Also, he mentions that his dissertation was not nominalistic, as *Structure of Appearance* is [49, *xvii*]; perhaps (part of) the spur to Goodman's change 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.

and a list of invitees, can be found in [52]. My focus here will be exclusively on the ‘Logic group’ and its participants, not the larger Science of Science group.[4]

1.2 Outline of the Primary Project: A ‘Finitist-Nominalist’ Total Language of Science

Carnap’s discussion notes from 1940–41 cover a wide range of topics. We have records concerning:

- the relations between metaphysics, magic, and theology (RCC 102–63–09),
- the concept *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) to languages with types (exemplified by *Principia Mathematica*) (090–16–09, –02, –26),
- modality (090–16–09, –25),[5]
- Quine’s recently published *Mathematical Logic* (090–16–02, –03, –26),
- the treatment of quotation-marks in formalized languages (090–16–13),
- 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 extended treatment.

However, the 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 refers to 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)

[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” [52, 175].

[5] An insightful treatment of this portion of the notes can be found in [62, 332–335].

- "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 today 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 the time these conversations began.[6] 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). Rather, 'finitism' appears in these conversations when Tarski proposes rather strict 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 (Kotarbinski): the individuals are physical things.
> 3. Non-Platonic: Only variables for individuals (things) occur, not for universals (classes etc.)
> (090-16-28)

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

> Finitism.
> Tarski: I really understand only a finite language S_1:
> only individual variables, [cf. condition 3. above]
> whose values are things, [2. above]
> whose number is not claimed to be infinite (but perhaps also not the opposite). [modified version of 1.]
> Finitely many descriptive predicates. [new requirement]
> (090–16–25)

Let us describe Tarski's proposed conditions for an intelligible language using the modern apparatus of model theory.[7] We begin with the notion of an interpreted language $\mathcal{L} = \langle L, M, \rho \rangle$. L carries the syntactic information about

[6]See [33] for a modern attempt to resuscitate parts of Hilbert's program.

[7]This is somewhat anachronistic, but it helps us today understand the discussants' activities.

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; that is, ρ encodes the truth-tables. M is an interpretation or model that fixes signification of the nonlogical constants of L. Specifically, $M = \langle D, f \rangle$, 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 $D_n \mapsto D$ to each n-ary function symbol. Now let us use this apparatus to rephrase Tarski's idea more precisely. Tarski is describing a certain type of (interpreted) language $\mathcal{L}$ that has the following four characteristics, which I will henceforth refer to as Tarski's 'finitist-nominalist' (FN) conditions:

(FN 1) $\mathcal{L}$ is first-order.

In a fully understandable language, variables range over individuals only, so one cannot quantify over properties or classes. One might be tempted to interpret Tarski as claiming that any string that contains (the symbolic correlate of) 'For all properties X, ...' is not a grammatical formula of L, since 'being first-order' is a grammatical property. However, we probably should not view Tarski's proposal as a purely grammatical restriction. For immediately following the quotation above, Tarski explains that he is perfectly willing to derive the consequences of sentences containing higher-order variables according to the rules of a proof calculus—and standardly, ungrammatical strings cannot be operated on within a proof calculus. Tarski's complaint is that he does not truly understand these higher-order sentences. Specifically, Tarski says:

> I only 'understand' any other language [i.e., a language that does not meet his restrictions—GF-A] 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 full or complete 'understanding,' which a mere 'calculus' alone cannot deliver, will be discussed at length in 2.2 below. But we can see that (FN 1) is not a restriction on which strings are grammatical, since Tarski does not consider "higher, 'Platonic' statements" ungrammatical—for a 'calculus' has a grammar.

(FN 2) All elements of D are "physical things."

In the elaboration and discussion of (FN 2), numbers are specifically disallowed from D. Furthermore, because of (FN 1), not even the usual 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 things'? The discussants do not show a great deal of interest in settling upon a specific interpretation.[8] Tarski never proposes exactly what he thinks the 'physical things' are. Three options the group considers are: (i.) elementary physical particles, such as electrons etc., (ii.) mereological wholes composed of elementary particles (or quanta of energy) (Quine)[9], so that e.g. 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 final suggestion is derived from the co-ordinate or position languages of Carnap's *Logical Syntax of Language* (090–16–23). Quine, in "On Universals" [80], published six years after these discussions, presents 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. [80, 82]

But in 1941, there was no consensus—or even an explicit desire for consensus—concerning what is to count as a physical object in (FN 2).

(FN 1) and (FN 2) place what we would now call *nominalist* requirements on an intelligible language; let us turn to the properly finitist requirement:

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

or

> (FN 3_L: liberal version) No assumption is made about the cardinality of D.[10]

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

[8]Nelson Goodman, in *The Structure of Appearance* [49, 39] and "A World of Individuals" [48, 17], goes so far as to say that nominalism as such places no restrictions on what the individuals countenanced by the nominalist are, so long as they are individuals and not classes.

[9]For an account and analysis of Quine's later published remarks on physical objects, see [31].

[10]In 090–16–04, however, Tarski proposes to exclude 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 $A_n (n = 1, 2, \dots \omega)$. 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.

his autobiographical recollections of these discussions, attributes the restrictive version to Quine and the liberal version to Tarski and himself [20, 79].

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

(FN 4) $\mathcal{L}$ contains only finitely many descriptive predicates.

Tarski offers no justification for (FN 4), and the participants never talk about it after its initial proposal, so I will not discuss it further. Presumably, these four finitist-nominalist restrictions do not single out one unique language—several different interpreted languages could satisfy (FN 1-4). In what follows, I will call the 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 constraint is placed on the language(s) they are attempting to construct. They wish to generate 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 [GF-A: 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.
> (090–16–25)

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 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 ....
> (090–16–05)

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).[11] For

[11] In a lecture dated September 8, 1939, Quine had already suggested this line of thought: "nominalism. Probably can't get classical mathematics. But enough mathematics for physical

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).
> (090–16–25)

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] Unfortunately, we are not told whether this new condition trumps the finitist-nominalist conditions or not. That is, if classical mathematics is ultimately not understandable, and the mathematics condoned by the FN conditions 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 attach to these competing conditions. However, none of the participants assert that we should completely discard or deny those portions of (e.g.) set theory that fail to meet the finitist-nominalist criteria (FN 1–4). Set theory can progress unimpeded by philosophical scruples, 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

science? If this could be established, good reason to consider the problem solved" (quoted in [64].

[12] The second 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" [106, 6.211]. Carnap maintains such a viewpoint as well: "The chief function of a logical calculus in its application to science is not to furnish logical theorems, but to guide the deduction of factual conclusions from factual premisses" [16, 177].

(090–16–28). But that is not a barrier to proving theorems.

A published summary of the finitist-nominalist project undertaken by Carnap, Tarski, and Quine in 1941 can be found at the end of Carnap's Intellectual Autobiography, in the section entitled "The Theoretical Language." Carnap's conception of this project has not changed substantially during the intervening years, though subtle shifts are present.

> 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.
> [20, 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 the desire 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—though finitism is standardly taken to be a species of constructivism. And Carnap is correct 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 non-logical terms designate observable entities, properties, and relations. As we saw in the discussion of (FN 2), this is not right. 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 epistemic accessibility of abstracta,[13] but she is free to believe in (concrete) unobservable entities. For example, protons are usually considered concrete but unobservable.[14] 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 discussed in 1940-41; there, as said before, 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.[15]

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 (FN 1–4) and is rich enough to serve as a metalanguage for 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 3, but will note here that Carnap, virtually from beginning to end, is strongly suspicious of (FN 1–4), and he criticizes 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.[16] In general, Carnap's objections attempt to show that either the finitist-nominalist restrictions yield unpalatable consequences in the domain of the formal sciences, or that higher mathematics is genuinely meaningful.

[13]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" [48, 16]. Goodman offers this as a modification or clarification of the doctrine of the original 1947 nominalism paper co-authored with Quine, in which abstracta were rejected.

[14]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 [16, 203-205]. Also, in 090–16–12, Carnap writes that in a "finitistic," "understood language," the "individuals" will be "certain *observable* things and their *observable* parts" (my emphasis). But the point remains that this is inconsistent with Tarski and Quine's understanding of 'physical objects' as including electrons and quanta of energy.

[15]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.

[16]It seems to me that Carnap's position is very much like that of 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" [6]. Carnap's view of nominalism is perhaps not quite so dim, but like Burgess, he is undertaking a project in which he works within the rules set by the nominalist, without fully accepting the legitimacy of those rules.

1.3 Mathematics in a Finitist-Nominalist Language

Carnap, Tarski, and Quine apparently realize from the outset that one of the most pressing and difficult obstacles facing any attempt to construct a finitist-nominalist language for the analysis of science 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 sizable 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 entities we can quantify over are physical ones in a language meeting Tarski's restrictions? Second, what should be done with numerals that purportedly 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? I shall deal with each of these questions in turn.

1.3.1 Number

As seen in the previous section, in a fully understandable language, names are not allowed to denote abstract entities. So in such a language, the numeral '7' cannot name a natural number, considered as a basic individual object[17]—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 they must re-interpret numerals. How do they do so, in such a restrictive linguistic regime?

Tarski's strategy for introducing ordinal numbers[18] 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). Virtually the same[19] 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

[17] This assumes, *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.

[18] The group discusses cardinal number very briefly in (090–16–25).

[19] 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 (as well as that of Tarski and Quine in 1941).

> ordered in a sequence; thus no abstract entities would be involved. [20, 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 '+' are defined via the version of Peano Arithmetic for PSI in Carnap's *Logical Syntax* (see §14 and §20).)

This heterodox view of ordinal numbers raises a number of pressing 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 somehow reflect the actual spatiotemporal positions of Tom, John, and Harry?—And if so, where do we 'start counting,' so to speak? Fortunately, it appears such questions can be avoided for the most part. The ordering is intended to be imposed, it seems, by stipulation: Tarski says "we want the (perhaps finitely many) things of the world ordered in some *arbitrary* way" (090–16–23, my emphasis). 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. There is no further discussion in the notes of how the order is fixed, but 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.[20]

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 'Tom' and '7' the same value), it appears that the sentence '7 has brown hair' will be true. Whatever else numbers cannot be, they certainly 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. There is no record of Tarski, Quine, and Carnap

[20] As James Woodbridge pointed out to me, one might object that the notion of sequence or order *presupposes* some concept the natural numbers, at least on the standard definition of sequence (where a sequence is any class that can be put in one-one correspondence with the natural numbers). So has Tarski just imported numbers into the system? Perhaps not: perhaps the notion of sequence does not presuppose the concept of the natural numbers, even if our standard definition today makes use of them. For example, the notion of a spatial or temporal sequence or order is presumably learned before the natural numbers are.

even considering this problem. Perhaps technical refinements could avoid at least some of these unwanted truths.[21] Note, however, 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, '$2 \in 3$' 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 an interpretation in some other part of scientific language.

1.3.2 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, and spent a fair amount of time and energy discussing. Suppose there are only k items in the universe. Carnap poses the question: "How should we interpret" the numeral expressions '$k + 1$,' '$k + 2$,' ..., "for which there is no further thing there?" (090–16–06) Initially, the group considers three options (employing the usual notation, where x' is the successor of x):

(a) $k' = k'' = \ldots = k$

[21] One such refinement is suggested in [38, 214-215]. Fields basic idea, couched in our terms, is the following. Recall that the ordering of physical objects of D is 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 $\mathcal{L}$) if and only if ϕ is true for all assignments of physical-object-values to numerals (satisfying certain intuitive conditions: for example, we want to rule out assignments in which all numerals are assigned to a single object in D). This 'supervaluational' characterization is only a rough pass, 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 (FN 1–4) and certifies substantial portions of arithmetic as true, without also committing us to the truth of sentences like '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 make '7 is a physical object' true, since that sentence is true on all assignments of physical objects to the numeral '7.' (From the point of view of someone who endorses (FN1–4), perhaps neither of the previous two consequences will 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.
> [3, 291]

By refusing to treat numbers as objects, nominalists make common cause with structuralists.

(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 (recall that 'number' will be interpreted here as some 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 of some number (contravening PSI 9) [11, 31].[22]

None of these three options is especially palatable, since none captures the truths of classical arithmetic substantially better than the others. For example, imagine that 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 is translatable into '$999 = 999$' and '$0 = 0$,' respectively. And under (c), the problem just mentioned will be avoided, but the equally counterintuitive '$0 = 1000 = 2000$' will be true. Presumably, the arithmetical consequences will be no less absurd if we take a more realistic (i.e., larger) estimate of the number of objects in the universe. So regimes (a)–(c) will all certify as true many 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 *simpliciter*, so that there is no 'last element' forced upon us.

And making these unpalatable equations true is not the only problem with (a)-(c): 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 $\mathcal{L}$. Then there will be arithmetical sentences that are provable under classical arithmetic (say, in primitive recursive 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 '$1 + 1 \neq 0$'. 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 3.4.2 and 4.1.

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., as analogous to 'Santa Claus' (put model-theoretically: $f(k')$ is undefined). This approach (which we today

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

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.

In his private notes at this time, Carnap actually performs some basic calculations on the question of how many physical things there are in our universe (090–16–22). Starting from a conjecture of Eddington's, Carnap computes that the number of particles in the universe is approximately 10^{77}. Then, using Quine's proposed ontology, in which classes of particles are things (since bodies are classes of particles), the maximum number of 'things' in the universe is approximately $2^{10^{77}}$. 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 an empty game of word play nor a merely technical exercise.

1.4 Pre-history of the 1941 Finitist-Nominalist Project

The next chapter will address the justifications or rationales for undertaking the finitist-nominalist project. First, I wish to consider here a different question: from what historical sources do (FN 1–4) spring?[23] 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 that are most relevant to the finitist-nominalist project. Next, I present possible indirect lines of influence that Russell's ideas may have had on the formation of the 1941 FN project. Finally, I briefly sample contemporaneous skeptical complaints about infinity from Wittgenstein and Neurath.

1.4.1 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 during his intellectual development 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 reism. Chwistek worked in Krakow, which was not part of the Lvov-Warsaw School to which Tarski, Kotarbinski, and many other prominent Polish philosophers belonged.

[23]An excellent account of the historical trajectory of Quine's shifting attitudes toward nominalism can be found in [64].

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 had an extended and wide-ranging discussion. Carnap's notes record that Tarski said:

> With the higher levels, Platonism begins. The tendencies of Chwistek and others ("Nominalism") to talk only about describable things 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 to or correspond 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 (FN 2), which Tarski labeled the 'non-Platonic' requirement. Also, Tarski's suggestion here to use properties as the individuals in the universe of discourse (*prima facie*) violates the restriction of the universe to physical objects only. So it is not immediately evident (a) in what sense 'nominalism' is meant here, or (b) how this view relates to the later FN conditions.

What does Chwistek mean by the term 'nominalism'? It does not directly correspond to any of Tarski's finitist-nominalist conditions (though Chwistek harbors a suspicion of infinity, as we shall see). Chwistek's nominalism—to which Tarski appeals in the above quotation—corresponds more closely to the predicativism of Poincaré's philosophy of mathematics. Concerning Poincaré's view of mathematical objects, Chwistek writes:

> Poincaré 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. Poincaré 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.'
> [23, 21]

In short, the Chwistekian nominalist follows Poincaré's refusal to countenance the existence of any mathematical object that cannot be finitely defined.[24]

It should be noted that Poincaré does not use the word 'nominalism' in this sense. Rather, he views nominalism negatively, claiming that certain people

[24] For much more detail on Poincaré's 'predicativist' philosophical position, see [40], especially Chapter 7.

"have thought that ... the whole of science was conventional. This paradoxical doctrine, which is called Nominalism, cannot stand examination" [72, 138; cf. xxiii, 105]. So in Poincaré's mouth, the term 'nominalism' means what we today would call 'conventionalism.' Furthermore, Poincaré 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 important ways, to the cluster of commitments and attitudes currently associated with Platonism in the philosophy of mathematics. 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.
> [5, 66]

So the appellation of 'nominalist,' in Chwistek's mouth, corresponds to Poincaré's 'pragmatist'; elements of this view are 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.
> [23, 78]

And for Chwistek, like many of his contemporaries, 'metaphysical' is a term of harsh disapprobation. We find fundamentally the same argument in 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 any describer is a limited creature.[25] 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" [5, 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" [23, 43]. Here we find a closer

[25] Tarski places logical-philosophical weight on the finitude of human language in his *Wahrheitsbegriff* monograph [101, 253].

connection to Tarski's finitist-nominalist project, for eliminating higher-order quantification and restricting the domain of discourse to physical objects 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 in detail in Chwistek's "The Nominalist Foundations of Mathematics," published in the issue of *Erkenntnis* immediately following the famous symposium proceedings on the logicist, intuitionist, and formalist 'foundations of mathematics,' by Carnap, Heyting, and von Neumann respectively, as a kind of response. There, Chwistek proves (within the simple theory of types) that a certain propositional function ϕ exists such "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" [22, 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'—far too strong a result, for Chwistek clearly does not want to declare all of logic and mathematics metaphysical.

In the same article, Chwistek also argues against the existence of propositional functions on more general grounds. He maintains that they are not purely *logical* entities, as Russell and others would have it, for they do not (to put it roughly, and in current 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 simply false, since e.g. the expression '$\psi x \vee \psi x$' is clearly different from 'ψx', 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

[26]Chwistek understands the axiom of extensionality as follows: "any two propositional functions that agree in extension are identical" [23, 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)" [23, 83] This is much closer to what we (and most of Chwistek's contemporaries) would consider *syntax*.

> 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,' in analogy with certain medieval theories. [22, 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 symbol-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). Most philosophers and logicians today, along with many of Chwistek's contemporaries, would reject this conception of 'pure logic' as far too impoverished.

Chwistek makes other claims in "The Nominalist Foundations of Mathematics" that are very congenial to 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 "the renaissance of nominalism in Germany" [22, 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" [22, 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 the axioms of infinity and of choice (which are necessary to recover certain classical mathematical theorems) are introduced into a logic, then "one must realize that one has gained certain merely formal relations between sentences, but not contentful results" [22, 371]. This echoes, almost exactly, Tarski's view of higher mathematics under a finitist-nominalist regime, assuming that understandable [*verständlich*] expressions have content [are *inhaltlich*]: 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 shares 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 adopt Chwistek's so-called nominalism as a fundamental given.

Kotarbinski's reism

Tarski cites Tadeusz Kotarbinski's 'reism' as the source of (FN 2), the requirement that the domain of discourse must contain only physical objects. Tarski also helped translate one of Kotarbinski's introductory articles on reism into English [58]. What is reism? Most simply, 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' [83, 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* [58, 489]; but then Kotarbinski adds that his own view, pansomatism, is a particular species of reism (generated by adding the assumption that "every soul is a body" [58, 495]). As mentioned above in 1.2, there is disappointingly little discussion in the Harvard notes of what the participants mean by 'physical thing'; Kotarbinski, fortunately, hints at what he counts as a *res*. He writes: "'Corporeal', in our sense, means the same as 'spatial, temporal, and resistant"'; thus, Kotarbinski counts (e.g.) an electromagnetic field *in vacuo* as a body, since it is resistant and spatiotemporal [58, 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.
> [57, 430]

As the end of this quotation makes clear, pansomatism has not only an ontological component but also a linguistic or 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 any concrete object or objects can be re-phrased, without loss of content, into a sentence in which all grammatical subjects and predicates designate concrete bodies only [58, 490].[28] For example, the reist will transform 'Roundness is a property of spheres' into 'Spheres are round.' What motivates such a transformation or translation? Kotarbinski's 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.
> [58, 490]

Kotarbinski also explicitly rejects classes [58, 492]. Thus we see that (FN 1) is a consequence of reism, along with (FN 2). Kotarbinski says that we could declare utterances containing such words either false or nonsensical. Kotarbinski calls 'roundness,' 'property,' etc. 'onomatoids,' that is, merely apparent names, not

[28]'But,' the modern reader may object, 'predicates do not designate concrete bodies. Kotarbinski is guilty of a category mistake (or some other 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. Interestingly, Quine holds this same view in the late 1930s [64].

genuine names; or if one chooses to call them 'names,' then they must be thought of as denotationless names, like 'Pegasus.' In places (e.g. [57, 432]), Kotarbinski suggests that the reist's paraphrase is in fact what was *really* meant all along—a hermaneutic reconstruction of everyday language, instead of a revolutionary one, in the terminology of [7].

To understand Kotarbinski adequately, Stanislaw Lesniewski's basic views must also 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 in Lesniewski's system): '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 exactly the 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.
> [97, 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 Harvard notes (see 3.4.2); furthermore, Goodman and Quine defend it in "Steps Toward a Constructive Nominalism."

1.4.2 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 indicate that Russell had sustained involvement 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. Even if they did not adopt Russell's 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 the philosophers who followed in his wake.

In particular, Russell found numbers and classes philosophically troubling. He calls numbers—which, on his preferred 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.
> [94, 270]

The fact that the leading philosophical luminary of Carnap, Tarski, and Quines 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 role in inclining Tarski, Quine and others to consider the refusal to allow numbers into the universe of discourse prima facie plausible or reasonable. Along similar lines, Russell claims that postulating the existence of numbers is 'ad hoc 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.
> [92, 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 (2.2 and 3.2), 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" [93, 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 3).)

However, of course, the logic of *Principia Mathematica* directly violates (FN 1), since it is higher-order. Nonetheless, the axiom of reducibility (roughly) states that all formulae have first-order equivalents, so even in *Principia*, there is a sense in which first-order logic is privileged. The situation is slightly more subtle when it comes to (FN 3). *Principia* 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 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]

1.4.3 Logical Empiricists skeptical of infinity

Finally, I wish to consider very briefly certain remarks made by 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 language of science and mathematics (FN 3) would likely appear more reasonable. As just remarked, Russell was forced to introduce the axiom of infinity into the logic of the *Principia* in order to capture certain basic results in mathematics, but he found this maneuver philosophically unsatisfying: he considered every proof of a theorem ϕ 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 ϕ.'

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 language with a finite vocabulary and rules, and pursuing finite goals, the introduction of infinity seems ill-suited—a position similar to that of Chwistek and Poincaré described earlier.

Wittgenstein, of course, was not a fully-fledged member of the Vienna Circle, so one could think that Circle members would view his skepticism about infinity as traditional philosophical quibbling, due to Wittgenstein's 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

[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" [93, 184].

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.
> [70, 43]

Similar to the views of Russell just above, Neurath claims the concepts of classical mathematics are 'theological'—a close relative (if not a species) of 'metaphysical.' Note also that the FN project is, in part, an attempt to fulfill the final sentence of Neurath's quotation—for it attempts to confer a finite meaning upon claims of classical mathematics involving infinity. Years later, Neurath's worries about infinity resurfaced, 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" [69, 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.

In 1940–41, 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 analyzing 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, I have examined the conditions Tarski proposes a language must meet in order to be intelligible, 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 these philosophers' discussions—why did they not discuss the issues that we today consider more closely related to the core of their published, public views? This sense of surprise might make it appear to some today that the finitist-nominalist endeavor is a peripheral side project for these great minds, 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. Second, as we shall see in 3.1, Carnap assimilates the finitist-nominalist endeavor to his work on the relation between observational and theoretical languages, which began explicitly in 1936 with "Testability and Meaning" and continued well after 1941. But, even more importantly, the 1941 conversations are fundamentally a discussion of the relation of mathematics to the natural 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 (i.e., non-Millian) empiricist view of mathematics appeared possible. For this reason, the finitist-nominalist project involves the notion of analytic truth, the issue that perhaps looms largest in historical hindsight. 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 often occupied center stage for them during 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 the analysis of all scientific discourse, it is also intertwined with the issue of the unity of science, an idea whose heyday was in the 1930s, but whose grandchildren live 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 examine how they relate to the wider themes just mentioned.

Chapter 2

Justifications for the Finitist-Nominalist Conditions

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 to that article, the authors acknowledge that the initial impetus and strategy for their nominalist project was proposed in the 1940–41 academic year by Tarski, and discussed with the authors and Carnap [87, 112]. Thus, these discussions at Harvard can be seen as a, if not the, wellspring of current nominalism. In the previous chapter, Tarski's finitist-nominalist criteria (FN 1–4) were set out and explored in some detail. A crucial 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, summarized under the following headings: intelligibility, the anti-metaphysical drive, inferential safety, and natural science. The first three support (FN 1) and (FN 2), and the fourth (FN 3). 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.

2.1 First Justification for the FN Conditions: *Verständlichkeit*

2.1.1 Assertions meeting the FN criteria are *verständlich*

As we have already seen (in 090–16–25 and –28), Tarski claims that a language must meet his finitist-nominalist restrictions in order to qualify as 'fully understandable' or intelligible. We also saw above (1.2) 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 views 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 reminder: Carnap calls a 'finitist' someone who accepts all of (FN 1–3), not just (FN 3) alone.)

Quine's view of the relation between intelligibility and nominalism is similar to Tarski and Carnap's, but he couches it 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" (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 primarily 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.'

2.1.2 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. It is frustrating that the notes lack an explanation of what intelligibility is, since all parties involved acknowledge it as the 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 Heidegger's infamous metaphysical claim "*das Nichts nichtet*"? Furthermore, the participants do not completely agree (though their positions are not mutually exclusive) among themselves about which particular assertions should count, in an intuitive or pre-theoretic way, as fully understandable and which not: Carnap claims, contra Tarski and Quine, that classical, infinite arithmetic *is* intelligible, though he holds full set theory perhaps is not (090–16–25).[1] The notes from the final day of collaborative work on the finitist-nominalist language highlight how unclear and imprecise the concept of *Verständlichkeit* is 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?
> (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 this query from the other participants. However, this quotation does suggest that for Carnap, understandability is a pragmatic characteristic (in Carnap's sense) of a term, sentence, or language, i.e., a property that depends on the language-user; I shall return to this idea later in the chapter.

Despite these discouraging signs, 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 conditions]
> 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 as-

[1] In the same document, Carnap also says that he considers understandability a matter of degree.

> sumptions.)
> (090–16–28)

The contrast between 'intelligible language' and 'uninterpreted calculus' also appears, albeit more briefly, elsewhere in the discussion notes (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" [78, 125]. Thus, to put the point in the terminology of Carnap's *Logical Syntax*, merely knowing 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.[2] That is, understanding what a sentence s means is not merely knowing which sentences are provable from s and from which sets of sentences s is provable. Such a viewpoint 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 operations are restricted to the realm of syntax [96]. Regardless of what the detailed content of *Verständlichkeit* might be, it at least requires that a language be more than an uninterpreted calculus or 'empty formalism,' in addition to its being a pragmatic notion.

2.1.3 What does 'understandable' or '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.
> [17, 22; bold in original, my italics]

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

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

> 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. [16, 152-153; my emphasis]

Clearly, Carnap's publications both before and after the Harvard conversations of 1940–41 reveal a conception of understanding that dovetails with the conception that 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 to 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 captured by its truth-condition, an interpretation supplies meanings to (otherwise meaningless) characters. This is why, in the discussion notes, 'uninterpreted calculus' is contrasted with 'intelligible language': an uninterpreted calculus has not yet had 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, although I do not understand it—for example, 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, 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, as the history of science has progressed, we have less and less "intuitive understanding" of the foundational terms of modern science, such as Maxwell's electromagnetic field and, more strikingly, the wave-function in quantum mechanics [16, 209]. (Here, 'intuitive' perhaps carries Kantian overtones, even 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 'ψ' (the symbol for the quantum-mechanical wave-function) [16, 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" [16, 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" [16, 211]. This seems reasonable: it would be Pickwickian to claim that Einstein and Hawking do not understand the general theory of relativity. Carnap suggests that a physicist's understanding consists in using a physical theory—including the 'unintuitive' terms that 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. 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 [roughly, observational] 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.
> [16, 210-211]

Carnap still maintains this view several 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.
> [20, 78]

Carnap's basic picture is clear: a partially interpreted calculus qualifies as understood, provided that such a calculus—including its terms that are not directly interpreted—is inferentially related in a substantive way 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 such as the quantum mechanical wave-function) can be useful 'for making predictions of observable events' only if they are appropriately inferentially connected to the directly interpreted ones.[3] (Note that this liberalized version of 'understanding' includes

[3] Carnap's theory of meaning for scientific language is thus a hybrid of so-called 'referentialist'/ 'representationalist' and 'inferentialist'/ 'conceptual role' semantics, or what Fodor and Lepore call Old and New Testament semantics, respectively [39]. Carnap employs Old

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 [11, 327]. Carnap's use of partial interpretations will be discussed at greater length below, in 3.2.1.

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 layperson 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).
> [16, 204]

Here again, we see evidence that Carnap considers the intelligibility of a language to be 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. an auto mechanic), and vice-versa. On a tangential but substantive note, this text points to an interesting historical fact about twentieth century philosophy of science. The distinction that Carnap draws between 'elementary' and 'abstract' terms is virtually identical to the now-infamous distinction between the observational and theoretical vocabularies. This latter distinction is often said to have been introduced in order to isolate the 'empirical content' of a scientific theory. (For example, van Fraassen attacks the attempt to identify empirical import using this distinction [104, 54].) But we see in the above quotation that in *Foundations* Carnap does not draw the distinction simply in order to isolate empirical content, but rather to make a scientific theory understandable to a layperson. These two aims are clearly different. (However, we cannot be thoroughgoing revisionists about the aim of the observational-theoretical distinction: in "Testability and Meaning" citeCarnap:36, Carnap

---

Testament semantics at the level of 'observable' or 'elementary' terms, and the New Testament (or inferential role) semantics is applied to the 'higher' or theoretical reaches of scientific language.

also endorses the goal traditionally ascribed to him.) Additionally, the purpose Carnap states here dovetails nicely with one of Neurath's goals for the Unified Science movement (and his Encyclopedia, of which *Foundations* is an installment): to democratize science by presenting scientific claims in a form comprehensible to everyone.

2.1.4 What does 'understandable' or 'intelligible' mean for Quine?

So much for Carnap's published remarks on understandability; how does Quine conceive of *Verständlichkeit*? Quine, to the best of my knowledge, never explicitly affirms or denies Carnap's conception of understanding (c. 1940) as knowledge of truth-conditions or interpretation. In fact, I have not found an explicit 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 in "Two Dogmas" in terms of analyticity and/or synonymy not being "understandable" [81, 32] or "intelligible" [81, 26]. 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.
> (quoted in [28, 410]).

Two points concerning this quotation are relevant for present purposes. First, 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 'problematic 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, 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 Goodman and Quine's "Steps Toward a Constructive 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" [87, 105]. Readers of the 1947 paper have usually found this justification (if it can be called that) for nominalism extremely unsatisfying (e.g. [7, 205]). I suggest that we could interpret the 'intuition' of 1947 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 their cryptic published claim. This exegetical conjecture draws some support from the fact 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, so too the 'intuition' of 1947 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, meaningless strings), 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, i.e., considerations independent of 'convenience' and simplicity. 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 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. [28, 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 *verständlich* language, it seems reasonable to infer that for Quine *Verständlichkeit* is closely associated 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, intelligibility 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 the old goal of intelligibility (or often 'clarity' in Quine's idiolect) primarily as an epistemological, non-pragmatic concept.[4]

Why has Quine apparently run together semantics and epistemology? When we survey the projects of analysis that philosophers in the early twentieth century undertake, we can distinguish two distinct types of endeavor: semantic analysis, which uncovers the 'real meaning' or logical form of a sentence often 'hidden' beneath the sentence's surface structure, 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,[5] 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 [8, 246]. However, semantic and epistemic analyses can be kept distinct, and many times in the analytic tradition 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 given semantic and epistemic analyses were often thoroughly conflated by early analytic philosophers, it is less surprising that Quine would run them together in his reflections on the finitist-nominalist project of 1941. However, it should be noted that, within Carnap's finitist-nominalist discussion notes, *Verständlichkeit* is not an explicitly epistemological (as opposed to semantic) concept. The closest *verständlich* comes to assuming an epistemic aspect is in the final conversation, where (as we have seen) Carnap suggests that perhaps the order in which children learn concepts may reflect the order of intelligibility of those concepts.

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

[4]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 that are dependent on individual speakers. 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 (at least in part) on the grounds of convenience, simplicity, or utility. This is also the sense of 'pragmatism' that Quine embraces in "Two Dogmas." Quine's interpretation of 'intelligible' is clearly non-pragmatic in the second sense (for it 'does not reduce to mere convenience'), but it is uncertain 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 in 1943.

[5]The basic idea survives in the 'liberalized' meaning criteria found in Carnap's later "Testability and Meaning": "Two chief problems of the theory of knowledge are the question of meaning and the question of verification. . . . But, from the point of view of empiricism, there is a . . . closer connection between the two problems. In a certain sense, there is only one answer to the two questions" [14, 419].

nominalist work ("Posits and Reality" and "Epistemology Naturalized" are particularly strident examples). However, 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 still "hopeful of the possibility of clarification of the epistemological problem," whereas Carnap was not. Quine, not Carnap, still held in 1943 that there could be some sort of 'epistemologically immediate' material, which could be used as an Archimedean point in the analysis of knowledge. Quine recognized that Carnap is not 'hopeful' about such a project in 1943. Thus Quine's brief history of empiricism found in "Epistemology Naturalized" can perhaps be read as an *autobiographical* history leading up to 1969, instead of as the story of old-fashioned empiricism from Hume through Carnap, which finds its anti-foundationalist consummation in Quine.

In response to the above letter from Quine, 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 [28, 308]. Quine, in reply, concedes the point [28, 311, 336]. However, Quine could have perhaps 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 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.

2.2 Second Rationale for the FN Conditions: Overcoming Metaphysics

Another rationale for pursuing the finitist-nominalist project present—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) held a very negative view of metaphysics. The group of Polish philosophers from whcih Tarski came, the Lvov-Warsaw School, also shared this anti-metaphysical animus to some degree [97], [107], though as a group, they tended to be neither as fervently (*pace* Chwistek, who was not part of the Lvov-Warsaw school) nor as unanimously anti-metaphysical as their Viennese contemporaries. For example, Tarski's somewhat more relaxed attitude towards metaphysics around this time appears in a letter from Tarski to Neurath in 1936: "even if I [Tarski] do not underestimate your battle against metaphysics. . . , I personally do not live in a constant and panic fear of metaphysics" (quoted in [65]). The impetus to eliminate metaphysics was shared by many analytic philosophers in the early twentieth Century, but it took varying forms; I will detail some of this variety later, in 6.

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 the logical empiricists and their intellectual kin is the following: if a string of symbolic marks x is metaphysical, then x is meaningless.[6] (The converse does not hold: the string ')yPQ))', 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, since to say that A understands p is (as seen in the previous section) to say that A knows the meaning of p. 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, no conditional of the form 'If x is meaningless, then x is not understandable' ever 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 2.1.3 above—that in these notes, 'intelligible' should be understood as the pragmatic correlate of 'meaningful'—is correct, then the goal of constructing an intelligible language coincides with the goal of constructing a language free of objectionable 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 Harvard discussants.

But Carnap's notes from the discussions of 1940–41 contain more than implicit attacks on metaphysics. There are explicit references to (odious) metaphysical theses as well. Tarski and Quine hold that adopting (FN 1) and (FN 2) would prevent a pernicious slide into a certain kind of metaphysics, which they call 'Platonism'—after the grandfather of all metaphysicians. Recall that Tarski labels (FN 1) (the requirement that variables but be first-order) the 'non-Platonic' requirement in the first articulation of his proposal (090–16–28). The participants do not offer a detailed or precise characterization of Platonism; but it involves at least higher-order logic and/or (transfinite) set theory.[7] ((FN 1) rules out higher-order logic, and adding (FN 2) to it rules out (even first-order)

[6]Precisely this characterization is found in Carnap's "Overcoming Metaphysics through the Logical Analysis of Language," but the same idea is clearly set forth in the *Tractatus*, as well as in many of Schlick's and Neurath's writings. Philosophical differences arise in cashing out the content of 'meaningless.'

[7]Although it does not specifically address Carnap's, Tarski's, or Quine's conception of Platonism circa 1940, [5] provides an excellent treatment of the shifting meanings of the term 'Platonism' in the early part of the twentieth century.

set theory.) For example, we find Tarski saying (as we have seen before):

> 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 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 (FN 1–3)) (102–63–04), 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". Why should the set-membership relation introduce Platonic commitments? Quine explains that "there are no logical predicates," while '$\in$' is a predicate. He then claims:

> Predicates 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.

That is, if there are any true statements of the form '$P \in Q$,' then there must be at least one class (provided '$\in$' is given the standard 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 (in "Designation and Existence" [76]), 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 1940 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." (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.
> [28, 295]

Presumably, the 'kernel of technical meaning in the old controversy' is composed of two decisions: (i.) whether (contra (FN 1)) to allow non-concrete individuals into the domain of quantification (as discussed in the previous paragraph), and (ii.) whether (contra (FN 2)) to adopt a 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 1939's "Designation and Existence," Quine had written his famous dictum "To be is to be the value of a variable" [76, 708]. In that article, he uses this dictum to characterize nominalism within the framework of modern logic: a language is nominalist if its variables do not ineliminably range over any abstracta.[8] And properties and relations, which are quantified over in second-order logics, are (for Quine and many others) paradigmatically abstract entities.[9] In short, a language is metaphysical if it quantifies (ineliminably) over abstract entities; in (first-order) set theory, those abstracta are sets, and in higher-order logic, those abstracta are properties and relations.[10]

Taking a wider historical view, it merits notice that Quine has transformed the old issue of nominalism into a form more congenial 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" [76, 708]. 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 had with Carnap at this time, Quine is fully on board with the basic research program Carnap espouses in the *Aufbau*, "Overcoming Metaphysics through the Logical Analysis of Language," *Logical Syntax*, and later works; viz., that a central task of modern scientific philosophy is to transform metaphysical (pseudo-)questions into well-posed logico-linguistic

[8] Quine writes: "In realistic languages, variables admit abstract entities as values; in nominalistic languages they do not" [76, 708].

[9] Quine does not tell us where or how to draw the line between concrete and abstract entities [76, 708]. Also, he does not appear in "Designation and Existence" to hold that all abstract entities correspond to predicates in a formalized language; that is, Quine appears to leave open the possibility that abstracta be part of the domain of individuals.

[10] For further published remarks on Quine's conception of nominalism, Platonism, classes, and relations, see "On Universals" [80] and "Notes on Existence and Necessity" [78], especially §5.

ones about the language of science. He and Carnap might disagree on which language forms are preferable or acceptable, but the general strategy is the same.

What is conspicuously absent 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 as values of variables constitutes objectionable, metaphysical Platonism for Tarski and Quine. A small hint about the attractions of nominalism for Quine can be found in a lecture from October 1937, where he describes the aims of nominalism:

> 1) To avoid metaphysical questions as to the connection between the realm ofuniversals and the realm of particulars; how universals enter into particulars, or particulars into universals.
> 2) To provide for reduction to statements ultimately about tangible things, matters of fact. This by way of keeping our feet on the ground—avoiding empty theorizing.
> (Quoted in [64])

The first point, which hearkens back to Socratic questions concerning how particulars 'participate' in the universals they instantiate, is clearly a metaphysical question—but it is far from clear that set theory and/or second-order logic engage this hoary issue in any substantive fashion. Quine's second point is not much more enlightening—it simply expresses the suspicion of abstracta and predilection for concreta, which characterize (if not define) the nominalist's position.

The silence on Quine and Tarski's part over what is fundamentally unpalatable about admitting abstracta as values of variables becomes more troubling when we note that labeling higher-order logic and/or set theory as metaphysics does not mesh well with the conception of metaphysics offered elsewhere by those with logical empiricist leanings. Both the explicatum and the explanandum of the term 'metaphysics' vary over time and between different thinkers, of course. Nonetheless, most logical empiricists and their allies, most of the time, strongly resist classifying most of 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. For example, 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 objectionably 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. Recall his talk 1.4.2 of "these queer things called numbers," which he considers 'fictions of fictions.'

Furthermore, it may very well be that there is no explanation to give here, and that the proponents of this view recognize that fact. 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" [87, 105]. So the rejection of abstracta is not based on any more fundamental (or even articulated) theory of knowledge and/ or meaning that declares abstracta unknowable and/ or signs for them 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 even if he did provide such a justification during these discussions that Carnap did not record, 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 reference (see 2.5.1 below)—perhaps even just as an implicit, unarticulated assumption—but such an interpretation would be conjectural. Paolo Mancosu has suggested that such a conjecture would be misguided, and that epistemological concerns are not at the heart of Tarski and Quine's predilection for nominalism:

> What is striking about Tarski and Quine in comparison to contemporary nominalism is the fact that the motivation for nominalism is not argued on epistemological grounds. Contemporary nominalism has been, by and large, an attempt to reply to Benacerraf's dilemma on how we can have access to abstract entities. Tarski and Quine['s]... anti-platonism originates from metaphysical qualms and from methodological commitments favoring paucity of postulated entities.
> [64]

Mancosu is certainly right to stress the near-total absence of epistemological rationales offered by Tarski and Quine, especially in contrast with current nominalists. However, a few caveats to his claim should be registered. First, as this section has emphasized, Tarski and Quine are somewhat short on explanations for what exactly is objectionable or unacceptable about the 'Platonic metaphysics' they think is embodied in set theory or higher-order logic. If they were pressed on the issue (e.g., they were explicitly asked 'Whats so horrible about this particular bit of metaphysics?'), it is conceivable that they would fall back on epistemological justifications. Second, as was quoted in the previous section, Quine writes to Carnap in 1943 that he (and Tarski) considered FN languages to enjoy an 'epistemological immediacy'—so issues concerning knowledge are not completely alien to Quine's reasons for favoring nominalism. Finally, a remark Tarski makes in his *Wahrheitsbegriff* about "the nature of language itself" can be construed as a kind of epistemic justification for finitism. "[L]anguage, which is a product of human activity, necessarily possesses a 'finitistic' character, and cannot serve as an adequate tool for the investigation of facts, or for the construction of concepts, of an eminently 'infinitistic' character" [101, 253]. If our investigations are conducted within a finite language, then there is some sort of barrier from our accessing and expressing any truths about the infinite—and it is not unreasonable to think of this barrier as epistemic in character: our tools for inquiry are not adequate for certain tasks. So while Mancosu is right to say that Tarski and Quine certainly do not emphasize any epistemological justifications for nominalism in or around 1940–1941, it may well be too strong to say that their motivations lacked any epistemic component.

2.3 Third Rationale: Inferential Safety, or Taking the Paradoxes 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. Very roughly, 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 size of this class of problematic logics: Quine and Goodman consider the class of suspicious logics to be wider than Carnap does. In the Goodman and Quine paper on nominalism we find an expression of this 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.*
> [87, 105; my emphasis]

And presumably, the (supposed) inhabitants of a 'world of make-believe' 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.[11] 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 suggests that Platonism is likely responsible for the logical paradoxes.

[11] In "On Universals," Quine reiterates the charge of ad hoc-ness against the type-theoretic formulation of mathematics current at his time (in particular, he points to the system of [101, 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, 80-81]

> 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.
> [28, 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.
> [4, 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 so drastic. One basic lesson Carnap takes from the paradoxes is that, *ceteris paribus*, if one logic's rules of inference and axioms are stronger than a second logic's, then the first is more likely to contain an inconsistency than the second. And for some purposes, inferential safety is of paramount value, trumping inferential 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 [20, 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. 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. [20, 49; my italics]

So whereas for Quine, Russell's paradox casts doubt upon any logic that quantifies over abstracta, Carnap is willing to use any useful 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. In fact, Carnap would likely agree with both disjuncts of the previous statement: Carnap explicitly claims that the type restrictions are natural,[12] and his Principle of

[12] In a discussion with Tarski about the comparative advantages of *Principia Mathematica*-style logics vs. set theory, Carnap says: "The levels appear to me completely natural and

Tolerance would allow languages that feel intuitively artificial to be theoretically acceptable. So, in short, Carnap recognizes that weaker languages enjoy the advantage of being safer, insofar as they are less 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 (roughly speaking) 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. But as the discovery of Uranus shows, introducing a hypothesis to save a theory from potential refutation is not always the worse course of action for a science to take.

2.4 Fourth Rationale for the FN Conditions: Natural Science

Another justification that Tarski and Quine offer for pursuing the finitist-nominalist project could be called the 'argument 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 argument from natural science, however, is only a justification for finitism (FN 3). This argument roughly follows one of Hilbert's justifications, in "On the Infinite," for his very different type of finitism. Tarski begins with a reasonable assertion: the number of individuals in our world "is perhaps in fact finite" (090–16–25). If the universe does only contain finitely many physical things, and if (FN 1–2) hold, then it follows that the domain D has finitely many members—and this is the restrictive version of (FN 3). If we wish rather to leave open the possibility that an infinite 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 spirit of the nominalist enterprise. As

understandable [*verständlich*]; and to a certain extent, stratification too" (090–16–26). 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.' [18, 84]

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.) (090–16–24).

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 (and temporal differences between them) are still manifestly 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 [57, 432].) So, Carnap is suggesting, if we expand (FN 2) to allow the domain to be not just physical objects or bodies 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 initial 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 infinite quantity somewhere in the actual material world misunderstands one motivation behind (FN 3), at least in its liberal version. Presumably, we should not assume the number of physical things in the world is infinite, because this is *prima facie* an empirical matter. Tarski says: "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 (the liberal version of) Tarski's scheme leaves this question open, as it should. 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 classical (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 intuitively or pre-theoretically 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)? If one is on board with the FN project, then this Carnapian challenge can be answered. 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.

2.5 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.

2.5.1 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. 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 dialectic of objections-and-replies can be found in [7, 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 very popular nowadays, at least in the simple form described above. 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., if directly asked in 1941, probably would not explicitly deny 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 entity, then (with some exceptions[13]) it is causally connectible to us; and if something is an abstract entity, then it is not causally connectible to us (unless one holds, with [60], that when we see a pack of playing cards, we are in causal contact with a set of cardinality 52).

2.5.2 The negative argument: rebutting the indispensibility argument

Shortly after the 1947 Quine-Goodman 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*, usually consist of 'reconstructive' projects that attempt to rebut this 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

[13] For example, the laws of physics prohibit me from ever being in causal contact with events outside my past and future light cones.

belief in mathematical objects becomes "unjustifiable dogma" [37, 9]. The literature on the indispensability argument is vast, and I will not comment upon the independent merits of the argument. 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 in its full-fledged form until around 1970, an explicit desire to rebut it cannot be a motivation for undertaking the 1941 project. However, there is a substantive precursor of the modern indispensability argument in Carnap's notes. We encountered it above in 1.2, when discussing the lower bound on the poverty of a finitist-nominalist languages expressive power; here is relevant section again:

> If S_1 [the finitist-nominalist language of arithmetic] does not suffice to reach classical mathematics, couldnt 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 Hartry Field's program, but it is similar: in both cases, the aim is to show that a proper subset of modern mathematics, which can be made nominalistically acceptable, is sufficient for all applications of mathematics in science. Tarski comes even closer to Field's program in a 1948 letter to Woodger:

> classical mathematics is at present an indispensible tool for scientific research in empirical science. The main problem for me is whether this tool can be interpreted or constructed nominalistically or replaced by another nominalistic tool which should be adequate for the same purposes.
> [62, 346]

This prompts 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 present-day indispensability argument. One current formulation, drawn from 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. [25, 11]

The 1941 project differs from the modern one primarily in the first premise: there is no normative claim concerning ontological commitments 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 a normative-cum-ontological claim with the goal of a unified language of science (see 6). Whether failure of a language to meet that aim automatically disqualifies it in the discussants eyes is not, as mentioned above in 1.2, discussed in Carnap's notes. We know that Quine, a decade after the Harvard discussions, opts for disqualification: his rationale for eventually repudiating nominalism is that we lack sufficient 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 categorically—rather, he would likely say that a particular language is merely inexpedient for this or that purpose. And the relative value or importance of various, potentially conflicting 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. An examination of the details of Carnap's resistance form the core of the following chapter.

Chapter 3

Critical Responses to the Finitist-Nominalist Project

3.1 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, cooperating with Tarski and Quine throughout the academic year. However, anyone familiar with Carnap's fundamental philosophical views should 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, which pre-date the 1941 discussions.

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.[1] 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 other 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. Evidence of that rationale, I believe, can be found in the section 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. 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*. 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

[1] 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).

> 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.
> [20, 78-79]

For Carnap, the observation language 'is presupposed as being completely understood'; we saw a virtually identical claim in 2.1.3, in the quotation from *Foundations* in which Carnap characterizes an 'elementary' or observational language as one which is fully 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." Assuming Carnap's memory is to be trusted in this matter,[2] 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 we suppose that Carnap considered Tarski's proposal to be a continuation of one of his own investigations, 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.

3.2 Answering the Finitist-Nominalist's Arguments: Higher Mathematics *is* Meaningful

As noted in 1.2, 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 here. 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

[2] It is 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.

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 render classical arithmetic intelligible.

3.2.1 An analogy between higher mathematics and theoretical physics

As mentioned earlier (2.1.1), in December 1940, Quine gives a wide-ranging lecture entitled "Logic, Mathematics, Science." In it, he claims that higher mathematics is 'Platonic,' and that science more generally is full of "myths" (see 2.3 above). Carnap responds[3] 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.
> (090–16–29)

What does this quotation, couched in terms of 'metaphysics,' have to do with defending the meaningfulness of mathematics? As mentioned above, 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. 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. Thus if Carnap can successfully show that non-finitistic mathematics is 'non-metaphysical,' he thereby shows that it is meaningful, and hence understandable.

So, given this connection between metaphysics, meaningfulness, and intelligibility, what is Carnap's argument for the intelligibility of mathematics? He

[3]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.

offers an argument by analogy; in Carnap's words, the analogy is:
observation sentences : higher parts of physics :: finitistic mathematics : higher, non-finitistic mathematics
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 'higher' 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 conjectured analogy captures a relevant similarity (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 as we saw, Quine at least claims in his December 1940 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)[4] does not appear to be relevant to Carnap's claim that higher mathematics is meaningful and hence non-metaphysical. It appears that 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 render 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.)

[4] 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 analog 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.

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.[5]

(2′) Higher arithmetical statements, such as the Peano axioms or ZF, stand in substantive inferential relations (possibly weaker than equivalence) to 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 in tension with Tarski's claims that any sentence not interpreted in accordance with the FN conditions is 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 [$\forall n(n+n = 2n)$] 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. [87, 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 6). Carnap, from (at latest)[6] *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 2.1.3). But one could question Carnap's claim that a partial interpretation confers full meaningfulness or intelligibility on the entire calculus.

---

[5] 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 is analogous to 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, so long as neither side of the distinction is empty.)

[6] 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.

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.

To fill in the details of the historical context, it should be noted that Carnap is not the only philosopher to suggest an analogy between variable-free formulas of arithmetic and observation reports in natural science.[7] We also find it in Poincaré'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. . . .
>
> 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. [72, 13-14]

Poincaré introduces the analogy to illustrate an important similarity between reasoning in mathematics and in physics. However, Poincaré holds that the justification for the ampliative 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 Poincaré 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. (088–30–03)

As in Poincaré'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 (even if neither Poincaré nor Gödel ever explicitly argued for that claim). And as Carnap explicitly states in "Testability and Meaning," for empiricists, the question 'How is a claim

[7] [62, 340] also discusses historical precursors to this argument. He includes Hilbert's finitistic project as well. See [61] for further details of the general conception attributed to Poincaré and Gödel here.

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 for 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" [46, 483-484]. I will merely point out that Gödel here 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 other remarks seem to suggest.

3.2.2 Potential infinity to the rescue?

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).
> (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 richer than PSI in *Logical Syntax*: 'negative universal sentences,' such as '$\neg\forall(x = y)$', are not expressible in PSI (because there are no symbols for quantifiers; free variables are used to express generality). However, S_2 and PSI would both contain theorems that the Tarskian finitist would either remain agnostic 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, even 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, entelechies, essences, and other traditional topics of metaphysics that, on the logical empiricists' view, are mere pseudo-concepts devoid of meaning.[8] 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. 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 objectionable 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 what constitutes sufficient evidence for or against a given arithmetical claim, in contrast with the perennial wrangling of the metaphysicians who repeatedly talk past one another (see especially [19, 218-219]).

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).[9] 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., the material necessary 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 thoroughly 'physicalized' conception of numbers; this will be discussed in detail in 3.4.1 below.

Before moving on to Carnap's other responses to the finitist-nominalist project, two further remarks should be made. First, Carnap offers no expli-

[8] In other words, Tarski subscribes to what could be called a Parmenidean account of meaning and meaninglessness: the meaning of a word is what it designates, so a word that designates nothing has no meaning, i.e., is meaningless (and therefore, according to logical empiricists, metaphysical).

[9] See [54, 192] and [100, 525].

cation of the distinction between a potential infinity and an actual one.[10] The distinction between potential [*potentiales*] and actual is perhaps different from the distinction between possible [*möglich*] and actual. Carnap seems to 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 our 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 [7]. 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. Mathematics then studies what is possible. But of course, many philosophers who are sympathetic to the strictures of nominalism are unsympathetic to modal notions—just as Quine and Tarski are in 1941, as seen in the above quotation.

3.3 Are There Any Infinities Compatible with Nominalism? (and a Detour through Finitist Syntax)

Carnap tries repeatedly to resurrect some type of infinity that is compatible with the spirit, if perhaps 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 to classical mathematics. Second, at the close of 2, 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

[10] Hailperin has characterized a 'potential-infinite domain' as, roughly and abstractly, a finite set of basic objects and a set of deterministic rules for generating further objects from the basic ones [51]. He explicitly avoids assuming that there exists a set containing all the objects so generated.

from numbers, classes, and properties—this would presumably require a modification to the letter of (FN 2) ('the domain of quantification includes physical *objects* only'), though 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; as we have seen, 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 the existence of an infinite number of individuals is not purely logical. He suggests using *sequences* of physical objects, instead of objects themselves, to construct an *ersatz* infinity and thereby avoid reaching the 'final number.' Then 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, including syntax. 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 Carnaps 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 initial discussion of Tarski's project. He attempts to meet Tarski halfway by relaxing Tarski's restrictions somewhat, while preserving the spirit of the finitist-nominalist program. This material can be found in the appendix (090–16–27 and –23); I here summarize the main thrust of Carnap's proposal.

(S1) Individual symbols in the object language 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.[11]

[11]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.

(S5) Such a syntax may suffice to 'build an unrestricted arithmetic.'

This summary of Carnap's proposal for finitistic syntax shows clearly the sense in which he is attempting to broker a compromise between Tarski's radical program and his own philosophical sensibilities. In Carnap's proposal for syntax, the basic, component symbols of the object language[12] 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 (and therefore the proofs) in the object-language consist of non-spatial sequences of these symbols—so there are genuine sentences of the object language that do not occur anywhere in the physical universe. (The class of sentences is defined as 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 (S3–4). So, in Carnap's finitist 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, viz. 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 later, I believe this difference looms large in the subsequent 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 5.3.1 below, 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 dis-

[12] Calling this a language may be contentious—it is not obvious that one is still talking about something that can reasonably be considered a language if one cannot say that 'bird' and 'bird' are the same word.

cussion of Carnap's syntax will appear in 3.4.2 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: $\langle a \rangle, \langle a, a \rangle, \langle a, a, a \rangle \ldots$ 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 real 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—$\langle a \rangle$ is assigned to '0', etc.)

Carnap recognizes that, in a universe with a finite number of objects, 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 appealing to abbreviations appears to postpone the problem without solving it.

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 (in certain types of languages) is perhaps the most famous; the 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-$\mathcal{L}$, 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 of which 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-$\mathcal{L}$ still could not be defined within $\mathcal{L}$ (provided $\mathcal{L}$ meets certain fairly natural conditions).[13]

Niceties concerning the number of physical symbols aside, Carnap's proposal to use sequences in place of objects apparently violates the spirit of nominalism in an obvious way, which may have already occurred to the reader: sequences, as customarily understood, are simply classes with further mathematical structure,[14] 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 ones language would not contravene either (FN 1) (no higher-order variables) or (FN 2) (only physical objects in D). For example, the truth of '$a \in \{a, b, c\}$' is compatible with all the finitist-nominalist conditions; furthermore, so is the truth of '$\exists x(x \in \{a, b, c\})$', provided that at least one of 'a, b, c' denotes a physical object. However, '$\exists x(a \in x)$' could never be true in a language satisfying (FN 2) (assuming the standard interpretation for the symbols), because if it were true, the universe of discourse would have to include a set.[15] Quine sees the situation clearly in his immediate response to Carnap's proposal to replace physical things with sequences in a FN theory of syntax:

> Quine: 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 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 (FN 2), 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 governing a finitist-nominalist language. The more interesting or difficult question

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

[14] 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 [101, 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)" [17, 18]. These definitions, which presuppose the natural numbers, would presumably have to be rejected by a finitist-nominalist.

[15] In a language meeting the finitist-nominalist conditions, '$\exists x(a \in x)$' is not meaningless or unintelligible, but rather simply false.

is: given that sequences are so conceptually close to classes, what was Carnap thinking when he made this sequence proposal? I see at least three viable possibilities. First, 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). Second, he might disagree with the second sentence in the quotation from Quine: Carnap might believe we can recapture classical arithmetic without introducing variables. Third, he may have recalled that Tarski introduced the assumption, discussed above in 1.3.2, that "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). Thus, if Tarski allows that sequence, perhaps he would allow others. However, there is a clear difference between the sequence Tarski introduces and the one Carnap suggests: all the items ordered in Tarski's case are concrete, whereas Carnap generates a sequence of sequences. 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 of concrete symbols—meet with approval from his interlocutors.

3.4 Attacking the FN Conditions

Thus far in this chapter, we have examined Carnap's attempts to defend classical arithmetic from the apparently destructive drive 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, attacking 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, including what 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 [Poincaré]) 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*?
> (090–16–09, emphasis 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 [20, 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.

3.4.1 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. (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 constitutionally opposed to questions of essence. Carnap's resistance to the FN conditions appears even more problematic when we recall that, in 1941, Carnap has been explicitly committed to his principle of tolerance for several years.[16] According to this principle, "Everyone is at liberty to build up . . . his own form of language as he wishes" [11, 52]. *Prima facie*, Carnap's attack on the FN conditions seems intolerant: 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 analytic is itself an analytic matter, i.e. there is no fact of the matter concerning which sentences are really analytic. (This explains why Carnap saw Quine's indeterminacy of translation thesis as vindication for his own views.) Thus, analyticity is language-relative; what is analytic in one language need 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 be pushing the view that arithmetic is analytic *simpliciter*, as opposed to analytic with respect to particular languages.

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 commitments to the principle of tolerance as well as his aversion to questions of

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

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 (taking numbers to be physical objects) misses the target explicandum (arithmetic of the natural numbers). Carnap does not articulate this explicitly in the discussion notes, but this attitude comes through fairly clearly in *Foundations*, which is roughly contemporaneous. There, Carnap writes:

> For any given calculus there are, in general, many different possibilities of a true interpretation.[17] 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.
> [16, 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 debated (e.g. [50]). (ii) Interpretations can be logico-mathematical or descriptive: for example, an interpretation that takes the universe of discourse to be the natural numbers or a hierarchy of sets is logico-mathematical, while an interpretation whose universe of discourse contains all and only the US presidents (or any other set of physical objects) will be descriptive. (iii) The customary interpretation for the arithmetical calculus is a logico-mathematical one.[18]

---

[17] '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 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 $\neg p$ in C, then p is false in S [16, 163].

[18] Carnap makes basically 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.
> [16, 198]

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 let Tarski set up whatever formal language he wishes; however, tolerance does not mandate that every formal language model every natural language equally well. Such a view would be madness. In fact, Carnap makes this point explicitly in *Foundations*: his answer to the question 'Is logic conventional?' is '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) [16, 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—at least as a model of English. 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 given to '+', '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.[19] 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.[20] On this view, Tarski sees himself as rescuing arithmetic from the clutches of a

[19] 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" [7, 6]. Mancosu agrees that Quine and Tarski's "approach lies squarely in the revolutionary tradition" [64].

[20] As Steve Awodey pointed out to me, Carnap also thinks explicata can be 'revisionary' instead of 'hermaneutic' 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 the use of 'true' in everyday language). Carnap, in general, attempts to preserve as much scientific content as possible, while 'sanitizing' the language in the above three ways. Tarski is apparently not as concerned with content preservation as much as Carnap—and therein lies the fundamental difference between them.

very dubious Platonism. In modern terms, his proposal could be viewed as a scientific revolution in something like 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 (on an natural physical interpretation of the calculus) 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 in classical (Peano) arithmetic, but empirically false? Carnap stresses, throughout his writings, Einstein's distinction between physical and mathematical geometry. Could there perhaps be a similar distinction drawn between physical and mathematical arithmetic? We have seen Carnap assert above that there could be an empirical/ descriptive interpretation of the arithmetical calculus, but that the customary interpretation of arithmetic involves only logical objects. But, we may query Carnap, how then is the applicability of arithmetic secured?

Carnap could respond to a Tarskian proposal for 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 mathematical vocabulary with light rays and freely falling masses in the world—without this identification or one like it, it makes no sense to say that the parallel postulate has been 'empirically disproved,' because the geometry does not 'hook up' to the empirical world in any way.[21] 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 standards in formal-mathematical explication being simplicity or elegance, along with formal consistency, of course).

[21]For an extended and insightful development of related ideas, see [44].

3.4.2 Problems with proofs

We now return to the topic of finitist-nominalist syntax, introduced above in 3.3. In Quine and Goodman'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.*
> [87, 106; my emphasis]

Serious consequences follow for logical inference. Recall the standard textbook definition of a proof of ϕ (in a calculus C): a sequence $\langle \phi_1, \phi_2, \ldots \phi_n \rangle$ of formulas (in C) such that ϕ_n is ϕ, and for each $i \leq n$, either ϕ_i is an axiom or ϕ_i follows from some of the preceding members of the sequence using a rule of inference of C. Now, suppose all the 'ink' in the physical universe is used up in writing down the formulas $\langle \phi_1, \ldots \phi_{n-1} \rangle$. We then cannot give a proof of the conclusion ϕ_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 not longer be 'rules' in the usual sense. 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 semantically incomplete: every model that satisfies A and satisfies B will also satisfy $A \wedge B$—but the proof calculus will not be able to prove $A \wedge B$ from A, B. 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" (090–16–27). For, if that restriction is adopted, we could only infer sentences that are inscribed somewhere (or spoken sometime, etc.). As we saw in 3.3, Carnap recognized these problems, and considered them to be a serious defect in the FN conditions.

3.5 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 (i) it is not one of the usual objections to nominalism in general, (ii) it is based on a thesis that seems to enjoy (some) consensus among philosophers, and because (iii) I believe it is Carnapian in spirit—though I will not argue for that final point. The crux of the objection is this: an answer to the question 'What counts as an individual?' (or '. . . as a unit,' or '. . . as a thing'), 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 physics.

Something similar to this idea appears in *Republic* VII: Plato claims the unit is intelligible, not sensible. More importantly in the present context, it also appears in Frege's *Grundlagen*: Frege asks us to consider a complete 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, so long as the domain is restricted to physical entities. (Obviously, allowing variables to range over $\mathbb{N}$, or $\mathbb{R}$, 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 much headway with Tarski or 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 bring 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 intensional languages, then they might not think of numerals that are 'too large' as being meaningless, but rather simply denotationless but nonetheless meaningful. Carnaps willingness to allow for intension explains why, for him, the understandable outruns the actual. Finally, if Tarski and Quine did not consider the study of language to be strictly about physical, empirical language, 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 tolerable, if not quite desirable. The differences between Carnap's conception of the distinction between the descriptive and the logical, and the competing conceptions of Tarski and Quine, are the subject of the following two chapters.

Chapter 4

The Finitist-Nominalist Project and Analyticity

If a modern-day philosopher is engaged in a free-association session, and the prompt is 'Carnap and Quine,' then the response will almost certainly be 'analyticity' or related notions. 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 is one of the most studied 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. . . . 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 [of his manuscript for *Introduction to Semantics*]. The controversy continued through subsequent sessions and without progress in the reading of Carnap's manuscript. [85, 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 Carnaps 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* (090–16–03),[1]

[1] This documentary evidence 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.

which becomes definition 18–1 in the published version. Quine's representation 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 manifests the most sustained and direct animosity toward analyticity is not Quine but Tarski.

Fortunately, 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 two parts: first, I flesh out the conceptual relationship between finitism-nominalism and analyticity by sketching 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 radicalization of Quine's attack on analytic truth.

4.1 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 (090–16–23). This would be disappointing for Carnap, for he thinks one of the genuine intellectual advances made by the logical empiricists consisted in 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 some models but not in all. Second, one might attempt to cash out 'arithmetic becomes synthetic' via Carnap's distinction between descriptive interpretations and logical ones (discussed in 3.4.1). 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—another unpalatable consequence for Carnap. Put otherwise: though Carnap claims that every sentence given a logical interpretation is analytically true or false, he does not hold the converse [16, 180]. So what *is* the sense of 'synthetic' here? At this stage in Carnap's career, a sentence is analytic in an interpreted language $\mathcal{L}$ if and only if the semantic rules of $\mathcal{L}$ 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).[2] 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 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" [17, 22]. A semantic system consists of three kinds of rules: rules of formation, rules of designation, and rules of truth. (In Carnap's estimation, the fundamental achievement of Tarski's *Wahrheitsbegriff* is showing that the third can be defined given the first two.) The rules of formation provide a recursive definition of 'sentence of L.'

[2] In *Foundations*, Carnap writes:

> We call a sentence of semantical system S (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.
> [16, 155]

Essentially identical claims are found in *Introduction to Semantics* [17, 140-142].

Rules of designation provide designata for the (non-logical) signs.[3] Specifically, it consists of sentences of the form '*b* designates *c*,' where (i) if *b* is a name (i.e. individual constant), then *c* is an object, and (ii) if *b* is a predicate (or relation letter), then *c* is a property (or relation). As an example of the first kind, Carnap offers (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" [16, 151]. Note that Carnap treats individual constants extensionally, as they are in the usual notion of model in current mathematical logic, 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" [17, 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 of atomic sentences, and results from Carnap's interpreting predicates as properties instead of sets. Carnap writes (where the 'n' subscript means the expression is in the grammatical category of noun, and the 'p' subscript indicates a property): "A sentence of the form '$\ldots_{\mathrm{n}}$ ist $—_{\mathrm{p}}$' is true if and only if the thing designated by '$\ldots_{\mathrm{n}}$' has the property designated by '$—_{\mathrm{p}}$'' [16, 151]. In modern model theory, a (classical) model assigns to each *n*-ary predicate a set of *n*-tuples, not a property (some people construe properties as extensions of predicates, but Carnap, like many, does not). For Carnap at this time, a property is an extension in every state of affairs; that is, a first-order (monadic) property is a function that assigns a set of individuals to every possible world. This is identical to what is often called an *intension* in the literature. 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 [16, 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" [17, 44]. (Variables themselves, however, can be either logical signs or descriptive signs, depending on whether the variables only range over logical objects or not [17, 59].)

With this characterization of Carnapian semantics in hand, we can better

[3]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 [17, 24].

understand one of Carnap's claims that sounds most strange to modern ears. In his discussion of the axioms of infinity and of choice in his autobiography, Carnap writes:

> 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.
> [20, 47-48]

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, i.e., all models. If we recognize that Carnap holds analytic truths to be logical truths (during this period, he calls 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. This difficulty is solved by recognizing that Carnap does 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 such as 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$\}$).[4]

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 $f(P) = \{a, b, c\}$ (where t is an individual constant and P a monadic predicate). Then the truth-value of Pt is determined by the information about the interpreted language alone, i.e., no empirical tests need to be run to determine its truth-value (since we don't need to make any observations to ascertain that '$a \in \{a, b, c\}$' is true).

[4] At this point, someone might object as follows (especially if she is sympathetic to Tarski's finitist-nominalist program). 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.

Every atomic sentence then appears to be analytic—an obviously unacceptable consequence, especially to Carnap. What has 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 an atomic sentence 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 in the actual world. During his semantic phase, Carnap identifies "extension"—as opposed to intension—with "*contingent* reference or denotation" [20, 63; my emphasis]. That is, in order to determine the extension of a word (unlike its intension), empirical, factual information is necessary.

To put the point in a rough and ready way, whereas the modern conception takes 'logically true in L' to be truth in all $M =< D, f >$ of $\mathcal{L}$, 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 places certain restrictions on the interpretation function.[5] For example: if, for 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.[6] It is in this sense that 'All bachelors are unmarried' is an analytic truth, for its truth is fixed by the language in which it is couched. This basic idea also appears in *Logical Syntax* §34c-d (though without the 'almost,' and with different terminology, since Carnap has not yet entered his semantic period), in the definition of '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[7] for all grammatically appropriate assignments of values to its variables and non-logical constants.

Now we can ask: what would be the semantic rules governing arithmetical language—and more specifically, arithmetical language meeting the finitist-nominalist conditions? Since there is no list of such rules in Carnap's discussion notes, the following proposal must be somewhat conjectural.[8] First, the max-

[5]The notation here is anachronistic, but the underlying idea is in Carnap's *Introduction to Semantics* §19.

[6]This is assuming 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: he believes the syntactic conception of language cannot correctly capture this relation between 'horse' and 'stallion' [17, 87].

[7]In *Logical Syntax*, Carnap eschews the notion of truth; so this actually reads 'analytic.'

[8]In *Foundations*, 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

imal 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 1.2, that he would like to construct 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 suggests that we impose the order arbitrarily upon the physical objects, but it does not fundamentally matter what the source of this ordering is. 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, where '$S(x)$' means 'successor of x,' '$S(a)$' 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 in 1.3.1. Thus I propose that we do not include in our semantic rules any one particular interpretation of the numerals, but rather just make all admissible interpretations subject to the above constraint.

Finally, we need rules of designation to deal with numerals whose intended reference outstrips the number of physical objects in the world. Recall from 1.3.2 that Carnap records three proposals for interpreting such numerals. Assuming that 'k' is the name of the 'final' physical object in the universe, the three proposals are:

(a) $k' = k'' = \ldots = k$

(b) $k' = k'' = \ldots = 0$

(c) $k' = 0, k'' = 0', \ldots$

(There are $k + 1$ total physical objects in such a universe, since the first object is assigned to '0'.) 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,[9] whereas in (b) it is the first object (i.e., the one 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 $k + 1$. 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, since numbers are understood as physical objects here—will have

steps.
[16, 181]

[9]Graham Priest has defended such a picture of the natural numbers [73]; he has also explored generalizing models of arithmetic that have the other forms Carnap, Tarski, and Quine consider [74]. However, Priest studies these models within the context of the paraconsistent logic *LP*, unlike the Harvard discussants, who wish to retain classical logic.

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: which arithmetical claims 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, since 'There exist n distinct physical things' is synthetic. 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 '$2 + 3 \neq 5$' 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 '$2 + 3 \neq 7$': 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 $1000 = 2000 \bmod k + 1$ ($k+1$, as before, is the number of elements in the domain), false otherwise. For similar reasons, in certain domains the classically true '$2 + 3 \neq 7$' will be false. The preceding can be generalized as follows: for all variable-free arithmetical sentences, all atomic sentences or their negations (i.e., those of the form $n = m$ or $n \neq m$) that are analytically true in classical arithmetic will be analytically true under a Tarskian regime (assuming we adopt one of (a)-(c)) . 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. '$2+3 \neq 7$'). In short: though all the classically analytically true variable-free atomic or negated atomic arithmetical sentences are analytically true under the finitist-nominalist setting as well, the classically analytically false atomic and negated atomic arithmetical sentences become synthetic—with the exception of logical falsehoods, such as sentences of the form '$n \neq n$.'[10]

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 (FN 3), so that we allow 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 (FN 3), 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

[10] Greg Lavers helped me to see that final point correctly.

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, 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 presumably empirical in character. (We could stipulate the anti-Parmenidean assumption as a semantic rule, but that would be unmotivated by the language we actually speak—such a stipulation would be analogous to declaring 'Adolf Hitler died 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 (FN 3), 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 (FN 3), 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).[11]

4.2 Radicalization of Quine's Critique of Analyticity: A Historical Conjecture

Richard Creath has argued that Quine's 1936 "Truth by Convention" should not be read as a full frontal assault on the very idea of analytic truth. He argues that 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 Creath's interpretation, the Quine of 1936 is viewed as "more nearly a request for clarification than an attack" [27, 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. He points out, first, that "Truth by Convention" grew out of three lectures Quine gave to the Harvard Society of Fellows in 1934, and these lectures praise Carnapian views almost unequivocally. So, if the document that "Truth by Convention" grew out of was extremely sympathetic to Carnap's position, then Quine's position in "Truth by Convention" itself is likely not diametrically opposed to Carnap's position. Second, at the 1937

[11]The question of how to formulate a theory of arithmetic in finite models which Tarski *et al.* occupied themselves with during 1941 is still being investigated today [59], [67]. The question of what happens to first-order logic when we restrict ourselves to *finite* models is a vibrant research area as well; see [91] for an overview.

American Philosophical Association Meeting, Quine gave 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,' associated with Carnap's position. So we have Quine defending Carnap's views both shortly before and immediately after he wrote "Truth by Convention." There is a third piece of historical evidence for viewing Quine's critique in "Truth by Convention" as less radical than that of "Two Dogmas," which Creath does not mention. In his 1970 "Homage to Carnap," the eulogy for Carnap at the 1970 Philosophy of Science Association meeting, Quine says he first contacted Carnap "in Prague 38 years ago," which would be Fall 1932, and that he, Quine, "was very much Carnap's disciple for six years" [84, 41]. In other words, "Truth by Convention" must have been composed during Quine's discipleship under Carnap—and thus probably should not be viewed as fundamentally rejecting one of Carnap's most cherished ideas.

There is a fourth and final type of evidence for Creath's view. During the greater part of the 1940s, in Quine's published [78, 120] and unpublished writings [28, 298 and 332], his attitude toward analyticity is one of growing skepticism, but not the total dismissal that we find in "Two Dogmas." For example, in "The Problem of Interpreting Modal Logic," Quine claims to give an "interpretation of pre-quantificational modal logic" in terms of analyticity, viz., "The result of prefixing 'Necessarily' to any statement is true if and only if the statement is analytic' [79, 45]. Presumably, one would not give an interpretation of modality in terms of analyticity if one considered analyticity to be thoroughly incomprehensible. And in the same article, Quine calls the suggestion (which he attributes to Goodman) that the analytic-synthetic distinction is merely a matter of degree a "dismal possibility" [79, fn.4]—so it seems that Quine had not yet abandoned all hope for analyticity, even if he felt certain it had not yet received a satisfactory explanation. And a similar attitude is echoed in a 1947 letter from Quine to Morton White: "It's bad that we have no criterion of intensional synonymy; still, this frankly and visibly defective basis of discussion offers far more hope of clarity and progress, far less danger of mediaeval futility, than does the appeal to attributes, propositions, and meanings" [105, 339-340]. In short, throughout the 1940's, Quine shows a reluctance to accept and endorse the notion of analyticity, but he had not yet reached the thoroughgoing rejection we find in "Two Dogmas."

Paolo Mancosu has challenged Creath's view of the trajectory of the Quinean attack on analyticity. After presenting historical evidence demonstrating Tarski had been challenging the analytic-synthetic boundary and the unrevisability of logic from 1930 onwards, Mancosu says: "[t]his, however, raises the question of when exactly Quine arrived at the criticism of the analytic-synthetic distinction" [62, 331]. To phrase the question in terms of 'arriving at the criticism' of analyticity may be somewhat prejudicial to Creath's view, since he wants to portray Quine's criticism as slowly developing and changing over time—so there is not a particular moment at which Quine 'arrives' at his critique, just as there is not a particular instant in which a person arrives at adulthood from childhood.[12] But this is likely terminological quibbling: we can re-phrase Mancosu's question as 'When did Quine arrive at the final, radical rejection of analyticity?' Creath claims that it was in the summer of 1947, in the three-way correspondence with Goodman and Morton White [28, 31]. Mancosu argues that it was earlier. He offers as evidence a letter from Quine to Woodger, dated May 2, 1942, in which Quine discusses the 1940–1941 academic year.

> Carnap, Tarski and I had many vigorous sessions together, joined also, in the first semester, by Russell. Mostly it was a matter of Tarski and me against Carnap, to this effect. (a) C's professedly fundamental cleavage between the analytic and the synthetic is an empty phrase (cf. my 'Truth by Convention'), and (b) consequently the concepts of logic and mathematics are as deserving of an empiricist or positivist critique as are those of physics. In particular, one cannot admit predicate variables (or class variables) primitively without insofar committing oneself to 'the reality of universals.' [62, 331]

Mancosu believes this letter shows that "already in 1940–1941 Quine had explicitly rejected the notion of analyticity, and in 1942, he considered that rejection to be already in his 1936 paper 'Truth by Convention"' [62, 331]. In short, Mancosu, unlike Creath, believes that Quine's rejection of analyticity was complete well before 1947—by 1942 at the latest. He does not commit to a position in [62] concerning whether this completion comes in 1936, 1940–1941, or some other particular time.

Which picture of Quine's path to the rejection of analyticity is correct, ä's or Mancosu's? The historical evidence, as we have seen, appears to pull in opposite directions—perhaps because it reflects Quine's own ambivalence towards analytic truth during the period. One could accept Creath's picture and try to explain away all the apparently disconfirming evidence (such as Quine's 1942 letter to Woodger), or adopt Mancosu's position and explain away why Quine continues to use the concept of analyticity (albeit reluctantly) throughout the 1940s. Alternatively, one could attempt to steer a middle course, which I will attempt here. Creath's view is probably too strong in places, but nonetheless there is evidence that "Quine had not quite yet given up hope on analyticity" entirely in the mid-1940s.

For example, Creath likely overstates his case when he claims that "what is new" in "Truth by Convention" that cannot be found in Quine's 1934 lectures on Carnap is *not* "an attack on Carnap's doctrine." Creath admits that "there are some new arguments that are sometimes viewed as such an attack, but Quine himself answers these arguments" [28, 30]. While it is true that Quine provides rebuttals to his objections to analyticity in "Truth by Convention,"[13] he closes with the following:

[12] "Quine arrived at that break [viz., his "reject[ion of] Carnap's doctrine that there are analytic truths"] ... *only by stages*" [28, 31; my emphasis].

[13] Also, in a typical article, most philosophers canvass possible objections to their argument, without accepting those objections.

> We may wonder what one adds to the bare statement that the truths of logic and mathematics are true a priori, or to the still barer behavioristic statement that they are firmly accepted, when he characterizes them as true by convention . . . [A]s to the larger thesis that mathematics and logic proceed wholly from linguistic convention, only further clarification can assure us that this means anything at all.
> [84, 106]

It is difficult to maintain, in the face of such a strongly-worded conclusion, that in 1936 Quine did not think "Truth by Convention" presents serious and pressing problems for the conception of analyticity Quine attributes to Carnap. And the letter from Quine to Woodger that Mancosu presents makes it clear that by 1942 at the latest, Quine did consider "Truth by Convention" to constitute a serious objection to 'Carnap's cleavage between analytic and synthetic truth.'

Nonetheless, even if we admit that Quine believed in 1936 that the objections in "Truth by Convention" carried philosophical weight, the fact remains (as mentioned just above) that Quine was still willing, in public and private, to consider using the notion of analyticity during the early to mid-1940s—despite the fact that he had real reservations about the notion. Perhaps what the letter to Woodger makes clear is that Quine had, by 1942, rejected *Carnap's* preferred explication of analyticity at the time, but that Quine still thought the notion might eventually be clarified along other lines. This would help explain the otherwise puzzling fact that Quine spends relatively little time in "Two Dogmas" dealing directly with Carnap's then-current view (especially in the original version of the article): Quine had abandoned Carnap's preferred characterization of analytic truth several years earlier, and thus focused the article on what he considered more promising or plausible alternatives, such as Frege-analyticity. (See also 5.3 below.) This gives us the 'middle course' between Creath and Mancosu's views: Quine had completely rejected Carnap's version of analytic truth by 1942 (at the latest), but he was not yet willing to write off the concept entirely until he began writing "Two Dogmas." However, certain important parts of Quine's eventual complete repudiation of analyticity are already present in "Truth by Convention," and for this reason it makes sense that Quine points to it later as providing reasons for dissent. Finally, even if Mancosu is completely correct that Quine had thoroughly rejected the analytic-synthetic distinction by 1942 (or 1936), it is nonetheless fairly clear that Quine's *rationales* for his skepticism toward analyticity changed from "Truth by Convention" to "Two Dogmas." The evolution and development of Quine's justification for his view would still remain an open historical question, even supposing Mancosu's thesis is granted *in toto*.

In what follows, I will assume that "Truth by Convention" is, at least in some sense and to some degree, a less radical challenge to the concept of analytic truth than "Two Dogmas." (That is, I deny that Quine has completely rejected the very notion of analyticity by 1936—or more importantly for present purposes, before the 1940–41 discussions at Harvard.) If this is correct, then this raises a further historical question: what prompted the radicalization of Quine's attack on analyticity? How did we get from "Truth by Convention" to "Two Dogmas"? Specifically, could the conversations of 1940-41 have contributed significantly to the deepening of Quine's criticisms? The 1942 letter from Quine to Woodger makes this more plausible, as does Tarski's letter to Morton White after reading "Two Dogmas" and White's "The Analytic and the Synthetic: An Untenable Dualism," in which Tarski says he "found once more that your and Van's views on the problem of the analytic and the synthetic are very close to my own" [105, 119]—views which Tarski had been pushing since 1930. If the joint Harvard discussions did foment Quine's development, how did they do so? No archival material in the Carnap collection from this period can, in my opinion, be considered a 'smoking gun,' so the following is conjectural. In the remainder of this chapter, I give one possible way the 1940–41 discussions could have furthered Quine's radicalization; I provide another in 5.1.2.

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 then 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 (for example) Einstein's field equations 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,' or 'Greg Frost-Arnold has brown hair.'

> 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.[14]
> [84, 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 essentially involving the former but not the latter can legitimately be simply stipulated true. In short, we can read Quine as making a slippery slope argument: 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

[14]A few years later, Quine will not even allow that the truths of mathematics can be so circumscribed. He takes Gödel's incompleteness results to 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).

Extra-logical Postulates." This article 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 be intuitively synthetic (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" [84].[15]

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 supposedly analytic, conventional truths is contracted. When Quine sees that 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 logic and mathematics is not in doubt.[16] After suggesting that the behavioristic sign of analyticity is being held true come what may, Quine writes:

> There are statements that we choose to surrender last, if at all, in the course of revamping our sciences in the face of new discoveries; and among these there are some 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 may have to say of their status in the course of a subsequent sophisticated philosophy.
> [84, 102; my emphasis]

That is, in "Truth by Convention," Quine still understands the theorems of logic and mathematics to be analytic (though not in a fully Carnapian sense of 'analytic').

However, in "Two Dogmas," we see Quine question the analytic status even of logic:

> Even a statement very close to the periphery can be held true in the face of recalcitrant experience by pleading hallucination or by

[15]Quine writes:

> Briefly, the point is that there is a mechanical routine whereby, given an assortment of interpreted undefined predicates 'F_1' ... 'F_n', governed by a true axiom or a finite list of such, we can switch to a new and equally economical set of undefined predicates and define 'F_1' ... 'F_n' in terms of them, plus auxiliary arithmetical notations, in such a way that the old axioms become true by arithmetic.
> [84, 133-134]

[16]In a footnote to *Word and Object*, Quine basically says that "Truth by Convention" did not claim that there are no analytic truths [82, 65n.].

> 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. [81, 40]

That is, Quine believes that certain developments in the empirical enterprise of quantum physics could lead to changes in the logical laws—and therefore, even logic can be considered to be in part a synthetic matter, since it is responsive to new discoveries about the empirical world.[17] Quine's suggestion is that the class of paradigmatically analytic sentences can be contracted, a suggestion we see clearly in Tarski's FN project—though in "Two Dogmas," the contraction appears even more severe than in the finitistic language construction project. However, it is not stronger than Tarski's 1935 claim, made in conversation, that "he had never uttered a sentence which he had not considered to be revisable" [62, 331], and he makes similar claims in his 1944 letter to Morton White. And it is very unlikely that Tarski never voiced those views about logic in Quine's presence during their year together at Harvard.

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, in which certain portions of arithmetic are considered dependent on empirical facts about the world, certainly could have planted the idea in his head, or perhaps more likely, cultivated the germ of an idea he had already entertained. Additionally, I am not suggesting 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 1950. In particular, Quine is not yet profoundly skeptical 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. And "Two Dogmas," of course, contains a sustained attack on the notion of synonymous expressions. The radicalization I have attempted to account for in this section is only a part of Quine's intellectual journey from "Truth by Convention" to "Two Dogmas." Another, perhaps more important, will be described in 5.1.2, where we see how Quine's (antecendent) antipathy towards intensional languages is transformed into a criticism of Carnapian analyticity and Carnap's characterization of synonymy in the early 1940s.

---

[17] Because I draw a distinction between inflating and contracting the class of analytic truth, I must demur from Mancosu's assertion that making certain apparently empirical sentences "unrevisable despite all observations ... is just the other side of the coin of claiming that logical propositions might be just as revisable as the physical ones" [62, 330]. 'The other side of the coin' is of course metaphorical, so Mancosu may not intend to say they are similar; but 'Synthetic sentences can be unrevisable' and 'Analytic sentences can be revisable' are not equivalent (they are opposite directions of a conditional, if we make the standard, Carnapian identification of 'synthetic' and 'not analytic').

Chapter 5

Direct Discussions of Analytic 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. First, I outline each primary participant's preferred characterization of logico-linguistic concepts in general, and of analyticity in particular. I then 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 familiar from "Two Dogmas," either misunderstands Carnap's position, or does 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 proof-theoretically.

Quine, unlike Tarski, does not articulate complete arguments against Carnapian analyticity in the notes; rather, he simply voices disagreement with two of Carnap's core commitments. Nonetheless, Quine's points of contention do allow us to characterize the philosophical differences between the two men cleanly, and thereby better 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 themselves analytic, whereas Quine considers them synthetic. Second, Quine considers Carnap's characterization of analyticity in modal terms fundamentally unclear. Motivations and arguments for each side are reconstructed, drawing on published work when possible. Finally, this material suggests another historical conjecture concerning the radicalization of Quine's critique of analyticity. 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 shifts toward semantic and intensional treatments of certain key linguistic concepts, but Quine does 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.

5.1 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. 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 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 sentences alone.' To complicate matters 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 propose reformed usages for 'analytic.'

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 language, independently of any empirical 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 [28, 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). The first disjunct corresponds to the notion of theorem (a notion Quine never abandons), whereas the second disjunct is precisely the notion of analyticity Quine later attacks 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,' 'truth in virtue of logical form' (whatever that might be), or as 'true in all models,' but rather as 'analytic.'

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 before 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 in extensional language (as in his *Wahrheitsbegriff*), while Quine holds that the concept of analyticity should be, at bottom, cashed out syntactically and extensionally. Each of the three chooses his approach not because of any particular view he has about analyticity, but because of 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 stresses to Carnap (090–16–09); and Quine tells us that he developed a very strong preference for extensional languages even before he finished college. 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 logico-linguistic term in scientific philosophy.

5.1.1 Carnap and Tarski on Analyticity

Carnap's basic conception of analyticity in the early forties has already been outlined in 4.1: 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 of Logic and Mathematics*: "In the Encyclopedia article [*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 characterization in terms of semantic rules 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" [19, 10]. And in *Introduction to Semantics*, Carnap points out that 'truth in virtue of semantic rules' cannot be a metalinguistic characterization of L-truth, since '. . . is a semantic rule' belongs to the metametalanguage.

Carnap presents another characterization of analyticity in the discussion notes (090–16–11), 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'):

$$a_i \text{ is L-true} =_{df} \begin{array}{l} 1.a_i \text{ is true in every state of affairs in S.} \\ 2.a_i \text{ is true for each model of C.} \end{array}$$

What does each of these two definitions amount to? Let us examine them in order. We have already seen (in 4.1) 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 connectives familiar from the recursive clauses of Tarski's definition of truth. But what is a 'state of affairs' for Carnap at this time? In the notes, he writes:

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

In order for such a set of values assigned to linguistic expressions 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 *etc.* 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 ('state' in Semantics (I) [The manuscript of *Introduction to Semantics*])" (102–63–07). Finally, the characterization of 'L-true' as 'true in all states of affairs' shows, more perspicuously than the 'true in virtue of the semantic rules' formulation, why Carnap held L-truth to be identical to necessary truth (090–16–25).

Now let us consider the second definition of L-truth 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" [101, 416-417]. It is also close to Carnap's notion of state of affairs of a semantic system. The primary difference with the latter 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 the specification of our language, the definition 'mare $=_{df}$ female horse,' then 'All mares are horses' will be L-true, provided we demand that all defined constants be eliminated before applying the test for L-truth [101, 415]. We can still make out a difference between Carnap and Tarski, though, in that Carnap would want 'All mares are horses' to be analytic, even if the object language did not explicitly contain a definition of 'mare' [28, 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 and the model/ calculus one will agree on the class of L-true sentences. 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 in 1936: "a class of sentences can be called *analytical* if every sequence of objects is a model of it" [101, 418].[1] 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: Tarski considers logical truth to be first and foremost a matter of meaning, i.e. of semantics, not primarily a syntactic affair. In this, he differs from the Quine of even 1940,[2] as we shall see in the following subsection.

[1]There is a substantial body of research discussing how much difference there is between Tarski's concept of logical consequence and our current one. In particular, there is disagreement over whether Tarski operates with a variable or fixed domain conception of model in his 1936 paper on logical consequence. John Etchemendy [34] first suggested that Tarski was working with a fixed-domain notion, instead of the modern, variable one. There have been a number of rebuttals to Etchemendy's view, culminating in [47]. Tim Bays [2] and Paolo Mancosu [63] defend the fixed-domain view; the latter shows that Tarski still holds this conception of model in 1940, by drawing on unpublished material from that year.

[2]Of course, to define the notion of logical *truth*, Quine must have a notion of truth available, and truth is obviously a semantic characteristic of sentences. But, as will be discussed presently, Quine identifies the logical truths by assuming a class of true sentences has been given, and then identifies which among those sentences are logical via proof-theoretic (or 'syntactic,' in his terminology) characteristics of the sentences alone [78, 43].

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

> 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 a unique 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 any model-theoretic proposal to characterize L-truth and L-implication. By merely stipulating (in the metametalanuguage) which sentences of the form "'..T..' L-implies 'T"' are L-true, Tarski's proposal can completely bypass the intensional notions of state of affairs and logical possibility—as Quine happily notes immediately following Tarski's claim.

5.1.2 Quine: analysis of language–including analyticity–should be syntactic and extensional

As discussed in 4.2, the Quine of 1940 does not consider the notion of analyticity hopelessly 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 willingness to embrace semantic and intensional approaches. 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 '$(x)(P_1(x) \wedge P_2(x) \supset P_1(x))$', 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-truth is like the characterization of analyticity in "Truth by Convention," except that here (i) synonymous expressions take the place of definitions, and (ii) Quine has substituted 'logically provable' in 1940 for the earlier 'logically true' [84, 87]—probably in part because Carnap is here using 'logically true' to mean what we mean by 'analytic.' This account of L-truth, which relies on definitions and/or synonyms and appeals to provability, is clearly different from that of Tarski and Carnap seen in the previous subsection: here, there is no mention of models, states of affairs, or even 'meanings of non-logical constants'; the semantic concepts Carnap and Tarski deploy do not appear in Quine's formulation.

Of course, 'p is synonymous with q' is usually construed as semantic, since its standard interpretation is that p and q have the same *meaning*. (And Quine holds out hope for an eventual clarification of the notion of synonymy throughout the 1940s, as we saw in 4.2.) However, two considerations tell against viewing Quine's appeal to synonymy here as a wholehearted endorsement of a fundamentally semantic approach to language analysis. First, given that Quine's characterization of L-truth in 1941 is similar to that in "Truth by Convention," with 'synonym' replacing 'definition,' Quine could be thinking of 'x is synonymous with y' as '$x =_{df} y$', and thereby admitting a thoroughly syntactic treatment, if one treats definitions purely syntactically. It appears Quine may think of definitions in this way at least sometimes, though the matter is not completely clear [28, 297].[3] Furthermore, even if synonymy is not so 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 that allows us to determine syntactically or proof-theoretically whether or not it is a logical truth (in the modern, non-Carnapian sense). So even if definitions are fundamentally semantic, nonetheless 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 is framed entirely model-theoretically.[4]

This view of Quine receives strong support from the quotation from Quine just above, in which he frames the standard for L-truth in terms of provability instead of truth. Further evidence of Quine's privileging syntax over semantics is found in a December 1940 lecture, where he declares the syntactic character-

[3]Carnap, in both his semantic and syntactic periods, thinks of definitions, within proof theory, in this syntactic manner: "The definitions in a calculus are, so to speak, additional rules of transformation" [16, 173]; cf. [11, 66].

[4]And in his autobiographical essay, 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" [20, 63-64]. 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" [17, 87]. This is exactly the converse of Quine's view of the matter, as the next quotation from Quine makes clear.

ization of logical truth to be 'more elementary' than the semantic one.

> 'Logically true' can be defined syntactically, and even protosyntactically[5] (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)

This shows that for Quine in 1940, calling a sentence 'logically true' means, first and foremost, that that sentence is a theorem, i.e. a provable sentence. Theorems are the paradigmatic cases of logical truths for Quine. The definition of logical truth suggested in this lecture follows that in Quine's *Mathematical Logic* [77, 305], 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. . . . [S]tandards of logical or mathematical truth are to be formulated in terms merely of the observable features of statements.
> [77, 4-5]

So the meanings conferred upon notational features need not be considered when attempting to discern whether a given sentence is a logical truth.[6]

Though this is excellent evidence that Quine strongly preferred proof-theoretic approaches to semantic ones, the completely syntactic formulation of his view we have just seen is somewhat anomalous. In "Truth by Convention" and "The Problem of Interpreting Modal Logic" ten years later, Quine's definition of 'logical truth' does not eschew truth entirely. (How could it, given that it is logical *truth*?) Rather, Quine's procedure for characterizing the logical truths of a language is as follows: given a set of true sentences, pick out the logical truths among them by inspection of their proof-theoretic aspects alone. Quine gives the following "definition of logical truth": "deducible by the logic of truth-functions and quantification from true statements containing only logical signs. . . . The 'deducibility' spoken of . . . can be expanded into purely syntactical terms" [79, 43].[7] This shows that although Quine's usual, published notion of 'logical truth' does not eliminate the concept of truth altogether, he nonetheless believes what

[5] For Quine at this time, 'protosyntax' refers to logic of *Mathematical Logic* without the set-membership relation, plus one syntactic primitive '$Mxyz$,' whose intended interpretation is too convoluted to merit recapitulation here; for details, see [77, 288].

[6] This view of logical truth is also found in the *Tractatus*: "It is the peculiar mark of logical propositions that one can recognize that they are true from the symbol alone" (6.113).

[7] Quine continues:

> The word 'true' . . . cannot indeed be expanded; no enumeration of axioms or axiom-schemata would serve the purpose here, because of Gödel's proof . . . We could desire otherwise, especially in view of the logical paradoxes which are known to be connected with the general concept of truth. Nevertheless [this definition] is not without explicative value, as marking out the special notion of

makes such truths logical is purely proof-theoretic—and in that respect, he differs from Carnap and Tarski. Thus, though Quine may use the semantic notions of truth and reference in the forties and afterwards, that does not mean that he adopts a fundamentally semantic view of logical truth in particular or of language analysis in general. This is in direct contrast to Carnap in the 1940s, who explicitly states that the proof-theoretic view of language is derivative from or secondary to the semantic one in many respects [17, §39].

From Quine's papers at Harvard, we also know that his feelings towards the general project of semantics just after the 1940–41 academic year were mixed at best. On November 5 1941, Quine gave 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, etc.,
> b) meaningfulness,
> c) protocol sentences,
> d) analytic vs. synthetic sentences,
> e) indicative vs. expressive use of language.
> Perhaps progress will 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 rejected the semantic program outright in 1941, but he is certainly not one of its proponents. Some commentators, including Donald Davidson [32] and Peter Hylton [55], consider the crux of Quine's criticism of Carnap to be that the kinds of explanations Carnap offers, in semantic terms, are not genuine explanations, or they are 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" [83, 48]. We see that more radical Quinean attitude foreshadowed in this 1941 lecture.[8]

> logical truth (such as it is) of truth.
> [79, 43]

Worth noting here are (i) Quine's explicit appeal to 'syntactical' (proof-theoretic) methods for isolating the logical truths, and (ii) his worrying about the notion of truth ('We could desire otherwise').

[8]How might a Carnapian 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. (I believe this can explain why Hartry Field, in [36],

One may object to my picture of Quine as preferring syntax to semantics on the grounds that Quine, in "Notes on the Theory of Reference" [83, 132-138] and elsewhere, 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 such a "peculiarly clear" schema, it appears that Quine would not admit semantic notions.[9] This again supports my claim that Quine, while willing to deploy semantic vocabulary, nonetheless uses it only when there are independent reasons for considering it unobjectionable.

Finally, 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 to any intensional concepts—in particular, to propositional functions. One of Quine's first publications, 1934's "Ontological Remarks on the Propositional Calculus," argues that the intensional concept *proposition* did not deserve a place in the total conceptual edifice of science. And in one of his last, "Confessions of a Confirmed Extensionalist" [86], Quine claims that he was already committed to extensional approaches during his college days at Oberlin. And in the same piece, Quine recalls 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.

5.2 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 them attributed to Tarski as to Quine. For example, in *Introduction to Semantics*, Carnap writes: "Tarski expresses, however, some doubt whether the distinction between . . . L- and F-truth is objective or more or less arbitrary" [17, 87; cf. vii]. We find Carnap stressing

calls Tarski's reduction of truth a 'bogus' explanation: Field is expecting a causal account of truth, while Tarski provides only a mathematical one.) And Carnap's account of analyticity is logico-mathematical.

[9] Quine writes:

> 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.
> [83, 134]

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" [20, 13; cf. 30, 36, 62, 64]. (Note Carnap's choice of words: those who agree with him 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 [20, 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 (090–16–09). One of the issues they discussed 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. In that article, 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.
> [101, 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,' and 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 available linguistic evidence; in 'ordinary usage,' the logical/descriptive boundary is vague. From this supposition that different choices of the boundary could 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.
> [101, 420]

In short, the truth-values of sentences of the form '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.

But this 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 for him. If it were (contra Tarski's suspicions) 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), and thus Tarski's position here is actually less 'tolerant' than Carnap's at the time. Given this similarity of viewpoint between Tarski and Carnap, it may strike us as somewhat surprising that Carnap considered Tarski one of his greatest opponents on the issue of analyticity. In order to reduce this perplexity, 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 the two can also be formulated.

5.2.1 Tarski's first objection: *any* sign can be converted into a logical sign

To begin their conversation concerning the tenability of the logical/descriptive distinction, Carnap proposes to distinguish logical terms from non-logical ones as follows: "indicate the simplest logical constants in the customary systems, and declare that 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," as follows: 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 this way, 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: in the discussion notes, he makes no mention of the 1936 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" (4.2): which sentences, exactly, are we allowed to declare true by convention, and which not? I say this 'supposedly' challenges Carnap's position because one form of the Principle of Tolerance is just that the choice of assertions taken as analytic is arbitrary;[10] different choices yield different languages.

Furthermore, as is becoming more fully recognized among current philosophers,[11] Carnap clearly holds that a statement's being held true come what may neither implies nor is implied by that statements being analytic: "the concept of an analytic statement which I take as an explicandum is not adequately characterized as 'held true come what may"' [20, 921]. Analytic sentences need not be held true come what may, for the language we are speaking can change; 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 one. 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 (for example, consider a theist of absolutely unshakeable convictions). So where did the notion that analytic statements are exactly those held true 'come what may' originate, if not with Carnap? It is Quine who first introduces unrevisability as a criterion of analyticity in "Truth by Convention" (quoted at length in 5.3.1 below), and reiterates it in "Two Dogmas," as a 'behavioristic' correlate of the old philosophical notion of analyticity. His fundamental challenge appears to be that Carnap's notion of analyticity is insufficiently empirical, and thus fails to be a scientifically respectable concept. ä [30] among others interprets Quine's challenge in basically the same way.

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." Here Carnap apparently grants Tarski the point that one can make any term in a constructed language a logico-mathematical one[12]—but

[10]For an explanation of why this is follows from the usual formulation of the principle of tolerance, see [42, 202] and [90].

[11]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 inviolable truth, 'sacrosant statements,' unrevisability or the like. . . . Already in [*Logical Syntax*], Carnap noted that no statements (not even mathematical ones) were unrevisable. Anything can go in the light of recalcitrant evidence.
> [75, 154]

[12]However, as we saw in 3.4.1, Carnap is clear in *Foundations* that if we are trying to model an extant language formally, then we cannot assign an arbitrary meaning to every term—for

then that term becomes a non-descriptive, 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 that includes Tarski's imagined 'Temp' predicate as the formal correlate of 'temperature' in the practicing scientist's parlance. In either case, there is no pressing problem for Carnap's view: 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), "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 we shall see in the following subsection.

5.2.2 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 mathematical claims, such as the Gödel sentence, which cannot be proved or disproved via the axioms and rules of inference of that calculus. From the fact of undecidability, Tarski immediately 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 ϕ in a (formal) language L is logico-mathematical, then ϕ or $\neg\phi$ can be justified via mere calculation in L.

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

(Thm) If a Gödel sentence G is expressible in L, then neither G nor $\neg G$ can be proved from the axioms and inference rules of L, if L is consistent.

then our formal language would not be an accurate model of our target language.

(C1) Thus, neither G nor $\neg G$ can be justified via mere calculation, so G is not logico-mathematical.

(P3) But G is logico-mathematical.

Since premises (C1) and (P3) are contradictory, at least one of (P1–P3) must be false. Both Carnap and Tarski accept (P3). 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, final 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 logico-mathematical 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 from (P1–3) 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" such a difference, and the conversation ends there. Although we do not have an articulated rebuttal from Carnap here, we can infer a more complete response with some confidence from his published writings. Carnap would most likely 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 ϕ can be justified via calculation, then the semantic rules of L suffice to determine that ϕ is true.

Put otherwise, calculability is identified with analyticity. But even the pre-semantic Carnap of *Logical Syntax* would reject (P2) as too narrow a notion of calculation—there, Carnap allows calculation to include an 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 what we now call '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—including, in particular, a Gödel sentence of the lowest-level language. On either option—appealing to semantic rules or to the hierarchy of metalanguages of *Logical Syntax*—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 can be somewhat obscure and metaphorical, but it is also a sign that this 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 [Gödel] sentence, 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" in such cases. 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 Carnap comes to the surprising conclusion that "the universal operator in both *Principia Mathematica* and II_d is not logical but descriptive" [11, 231].[13] In short: Carnap holds that a language with only d-rules makes certain 'apparently' logico-mathematical operators (like $\forall x$) and sentences (such as the Gödel sentence) descriptive or synthetic,[14] but a language with appropriate c-rules as well can classify such operators and sentences as logical or analytic.[15] So, Carnap could argue that his wider notion of calculation draws the line between logical and desciptive in the intuitively correct place, since it does not class the universal quantifier and the Gödel sentence as synthetic or descriptive.

After Carnap's syntax period, c-rules are replaced by semantic rules, and such rules play essentially the same role that c-rules did earlier. In a sufficiently rich formalization of arithmetic, a Gödel sentence that intuitively asserts 'This sentence is unprovable' is true in the standard model or intended interepretation of the language in which it is expressed (if the formalization is consistent)—and for Carnap, semantic rules fix a model or interpretation [16, 182]. There are non-standard models in which this sentence is false, but in such models the term 'provable' is assigned an interpretation in which it does not correspond to our usual notion of proof. Since Carnap's specification of a language during his semantic period includes fixing the intended interpretation, he can hold that the truth value of a Gödel sentence is determined by the specification of the language in which it is stated, i.e. it is analytic. In sum, Carnap could say in favor of his wider notion of calculation that an explication that makes the Gödel sentence synthetic fails to match up with the intuitive distinction between mathematical and empirical claims—though Tarski may deny that any such 'intuitive' boundary separates the Gödel sentence from the claims of physical theory.

[13] Incidentally, this shows ä's claim that Carnap would not allow any of the standard logical vocabulary to be descriptive [29, 261] is false.

[14] Neil Tennant has suggested that the Gödel sentence of a formal theory could be thought of as synthetic, for very similar reasons [102, 294]—though he is not fully committed to this suggestion.

[15] Carnap's case for a Gödel sentence being synthetic in proof-theoretic languages is straightforward, since neither it nor its negation is provable in the formal language. His argument that the universal quantifier is not logical is more involved, and it is not clear that it can be successful, given the completeness of first-order logic. (Carnap's argument would thereby be weakened, but the historical issue of how Carnap might have responded to Tarski would remain untouched.)

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-accepted notion of computability is closer to the narrower conception of calculability—though modern computability is perhaps too narrow even to be a plausible candidate for distinguishing logical from descriptive sentences. But the fact that there is a convergent and almost universally accepted explication of computability may count in favor of understanding calculation as computation—for at least the latter is sufficiently clear by virtually everyone's standards. 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.

Consider the following rough but natural line of thought: every genuine, meaningful sentence is couched in some language. Without some form of language to speak or write in, assertion as we know it would not be possible. But any language that is not massively impoverished has a consequence relation as one of its components: given that some set of well-formed formulae are true, the consequence relation indicates which further formulae of that language are also true.[16] Part of what makes a given language the particular language it is and not some other is its consequence relation: the difference between the language of the intuitionist and the classical logician can be clearly seen in their different consequence relations. Both languages contain grammatical strings of the form 'not-not-p,' but only in classical logic is p a consequence of this string. Logically minded students of language can codify this relation in the form of rules of inference (and axioms, if desired). We can formulate this basic point within the context of natural language users, as well: if someone makes an assertion in a natural language l, an l-speaker must in some sense accept (perhaps only tacitly) the consequence relation of l (or at least that portion of the consequence relation that bears on the assertion)—otherwise she would not make the assertion she does in fact make, but a different one; that is, she would not be speaking l. If someone speaks or writes in l, that person is thereby committed implicitly or explicitly to the sentences that are consequences in l of her utterances—otherwise she would be speaking or writing in some other language. And if one makes an assertion in the logician's proof-theoretic 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:[17] 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

[16]Of course, a language in which no sentence is a consequence of any other sentence is certainly conceptually possible, as a formal possibility.

[17]Unless Quine's later views on the empirical justification of logic and mathematics are correct.

is, in some sense, a transcendental justification: a claim in l^* could not express what it does in fact express, if the theorems of l^* did not hold. This is one way of explaining why, for Carnap and other logical empiricists, mathematics and logic are not susceptible of empirical justification: their warrant comes from being an unavoidable concomitant of the language we use to express anything about the world.

Now we have reached the point where a Tarskian can lodge a complaint against Carnap's wider notion of calculability; the complaint will be clearer applied to the Carnap of *Logical Syntax*, so he will be the immediate target. There, as mentioned above, Carnap does not identify the analytic truths (i.e., the truths in virtue of calculation) with the sentences provable 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, transcendental rationale for sentences we consider 'true by calculation' thereby disappears—for we express ourselves in the object language, not the metalanguage. If we confined ourselves to the inference rules and axioms of the object language l^*, then the only sentences 'true in virtue of calculation' will be the theorems of the object language—which is exactly Tarski's contention. For if a Gödel sentence can be constructed in l^*, then it is of course not a theorem of l^* (assuming l^* is consistent).

However, we can forward a forceful 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 somewhat 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 A is true whenever B is true. However, consequence was not treated semantically above, since 'sentence true in virtue of the consequence relation' was treated as equivalent to 'theorem'; but this is an inadequate characterization of consequence 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 meaningful language. Put in terms of language users, if we accept that an l-user is entitled or committed to the sentences true in virtue of the consequence relation of l simply by uttering a sentence in l, then the l-speaker is entitled to the semantic resources of that language, not merely a formal-syntactic characterization of it in terms of uninterpreted 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. Now, here Quine's indeterminacy of translation thesis 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' unscientific status, then Tarski's complaint based on the Gödel sentence is the least of Carnap's problems.

5.3 Quine's Disagreements with Carnap circa 1940

5.3.1 Analyticity is an empirical concept, not a logical one

As suggested above (in 4.2), Quine's criticisms of Carnap most likely had not reached fully mature form by 1940. Nonetheless, the seeds had been planted, and the fault lines between the two philosophers had begun to show. The first public appearance of significant differences between Quine and Carnap is "Truth by Convention." One of its main arguments, as described earlier, 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 at greater length in "Two Dogmas" (though it is still somewhat inchoate there), 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.
> [84, 102]

Quine here claims that if we lack empirical (specifically, 'behavioristic') 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, obscure Heideggerian dicta, and other bits of philosophy that the logical empiricists wished to avoid. That explains why Quine characterizes analytic truths in "Two Dogmas" as claims 'held true come what may': language users' acceptance of claims is thought to be susceptible to empirical investigation.

Recent commentators have suggested that Quine and Carnap 'talked past' one another a great deal in their debate over analytic truth. For example, as just mentioned, Carnap never accepts characterizing the class of analytic truths as the class of sentences held true come what may. However, in the Harvard discussion notes, we find a clear statement distinguishing Quine's position from Carnap's along the lines just discussed. 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 believes 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 suitable not just for formal/ mathematical languages, but for empirical 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 must satisfy, in order to be assumed as (logically) correct, before experiments about the truth are made. [Such sentences include: 'every atomic sentence involving 'T' must have exactly five arguments' (four spacetime coordinates and one scalar for the temperature)]
> 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 belong to empirical investigation (specifically, a 'behaviorist' study).

This, in turn, shows why Quine pursues accounts of analyticity and synonymy in *Word and Object* and elsewhere—a fact that might appear perplexing *prima facie*, 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' that we find in *Word and Object* and elsewhere 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.
> [78, 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.* . . . 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.
> [79, 44; my emphasis]

Here again, we see that Quine thinks synonymy must be at bottom (in non-'idealized' cases) 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, namely, 'truth 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 merit 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" [28, 64]. So for a time in the mid-thirties, presumably ending with "Truth by Convention" at the latest, Quine thought it perfectly acceptable to consider a sentence of the form 'p is analytic' (i.e., p is true by convention) to be itself analytic (true by convention). In short, Quine's view was once Carnap's.

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 over whether 'p is analytic' is itself analytic or not. Why? To put the matter in a Carnapian manner, it may be analogous to asking: 'Which is correct: pure geometry or applied geometry?' The Quinean could presumably respond to this description of the question as follows: 'A given mathematical system wouldn't deserve the name 'geometry' at all, if it did not admit of an interpretation involving spatiotemporal magnitudes in the empirical world.' Perhaps the best we can do here is to indicate possible reasons motivating 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' within number theory, thereby 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 a formal calculus is (to put the point in Carnap's terminology at the time) analytic in any interpretation that makes the axioms of the calculus true, and its rules of inference truth-preserving. Furthermore, for Carnap—and especially the Carnap of *Logical Syntax*—the concept of analyticity is an extension of the notion of theorem. Thus, since 'provable' is a logico-mathematical predicate, and analyticity is intended as a generalization of provability, this leads naturally to considering 'analytic' to be a logico-mathematical predicate. This conclusion could only be bolstered by Tarski's demonstrating how to define 'true in L' in a purely logico-mathematical way, given the expressions of L and names for the expressions of L.[18] (For one might wonder whether, even if provability

[18]For evidence and argument that Tarski considers his analysis of truth to be logico-mathematical instead of empirical, see [45, §3].

is a logico-mathematical notion, truth might not be.) Since analytic truths are a species of truth, Tarski's work lends further plausibility to notion that analyticity, 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" [28, 65]. If 'antecedently' and 'prior to' are construed temporally, then this is obviously true, since a theory always post-dates its subject: the physical behavior of falling apples and the Earth's tides occurred 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 notion, and it is made more difficult here since Quine does not spell out the sense of priority he is using; however, we can locate this sense to some degree by looking elsewhere. Quine, in both "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. The 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 [28, 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 the preceding conditional is true. (As mentioned earlier, Carnap denies Quine's claim that all analytic sentences are irrevocable; during a scientific revolution such as the transition from Newtonian to relativistic physics, the analytic truths of the language of physics change: the theorems of Euclidean geometry that make ineliminable use of the parallel postulate switch from being analytic to synthetic.)

Elsewhere in Quine's writings, the conceptual priority of firmness of 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," when explaining how 'No spinster is married' can be taken as analytic, on the grounds that 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.

> [79, 44; my emphasis]

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

> A common answer ... 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.*
> [28, 296; my emphasis]

In the letter he writes in reply, 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 [28, 305]. The semantic rules (assuming they aim to capture standard English) would suffice to guarantee the synonymy of 'spinster' and 'unmarried woman,' even if both were primitive predicates in the language, so definitions are unnecessary. 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 scientific picture.

5.3.2 Modal and intensional languages are 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 extensional. 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" [84, 162]. And Quine himself certainly espouses this view, as we saw in 5.1.2. By this standard, analyticity (and necessity) are clearly non-extensional: though '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. Quine draws on related facts to make a claim in favor of extensionality in "Notes on Existence and Necessity," published two years after the conclusion of 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 ... Up to now there is no clear

> example to the contrary. It is known, in particular, that no intensional mode of statement composition is needed in mathematics. [78, 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, ground Quine's rejection of 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 [86]. This mention is important, because Carnap characterizes L-truth as truth in all states of affairs, i.e., all logically possible (L-possible) worlds. For example, for a suitably formalized portion of ordinary English, there is no L-possible world or state of affairs in which the extensions of 'bachelor' and of 'unmarried man' are not identical. And Carnap forwards precisely this characterization of synonymy in the Harvard notes: two predicates (e.g.) are synonymous in a language (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 items more generally) relies essentially 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 each 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 intensional form, fundamentally unacceptable. It is interesting that in "Two Dogmas," Quine does not seriously consider Carnap's newer characterization of analyticity, but focuses primarily an older one, which Carnap had rejected years earlier. One plausible explanation for Quine's passing over Carnap's then-current 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 . . . is clearer to many of us, and obscurer surely to none, than the notions of modal logic" [79, 45]. This would explain why, in "Two Dogmas," Quine would not waste much ink attacking the characterization of analyticity Carnap prefers in the 1940s: after Carnap switches to an intensional analysis of analyticity, Quine regards all Carnap's further forays as non-starters. This suggests that 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.

5.3.3 A further step in Quine's radicalization

These facts suggest another 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,[19] which, as we have seen (recall 5.1.2), 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 1939's *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 moves towards a semantic and intensional approach to language analysis, while Quine retains the more 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" [28, 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 scientifically 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," quoted earlier:

> 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

---

[19] 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.
> [11, 1–2]

Carnap also, in *Logical Syntax*, subscribes to the "thesis of extensionality," which states: "a universal language of science may be extensional" [11, 245]. Furthermore, Carnap considered intensional language suspect, on the grounds that many sentences couched within it are quasi-syntactic (or the 'material mode'), and thus misleading at best [11, 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' has "psychological" or "epistemic" value only /cite[84]Carnap:28. Contrast this claim with what Carnap says to Nelson Goodman in 1940, in a conversation about Goodman's dissertation: "My objection against my *Aufbau*: . . . the extensional conception: definition of qualities etc. by enumeration" (102–44–11).

> 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.
> [28, 464; my emphasis]

First, note that Quine says that Carnap's 'views went on evolving'—and part of that evolution, as we have seen, is Carnap's willingness to pursue research in semantics and intensional languages. The historical question here, broached in the previous chapter, 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," chronological 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 be that in Quine's estimation the Harvard discussions played an important role in ending Quine's discipleship under Carnap. More generally, it is possible that Quine sees Carnap's move away from the *Syntax* program as a turning point—Quine read the manuscript of Carnap's *Introduction to Semantics* for the University of Chicago Press in 1940. In short, the radical critique of analyticity that Quine advocates by 1950 is perhaps as much a product of Carnap changing his views (towards fundamentally semantic and intensional approaches, away from exclusively 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 4.2): what appears as a somewhat suspicious notion in "Truth by Convention" has become fundamentally unacceptable.

In "Truth by Convention," Quine does not connect worries about analyticity to his distaste for intensional (and especially modal) languages; this is likely because in 1936 he believes that Carnap's preferred explanation of analyticity will be extensional and (in part) syntactic. Quine clearly rejects intensional language in general, and modal language in particular, in the 1941 notes, but he does not appear to grasp fully that synonymy is also fundamentally 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 believes 'necessary' is a more "obscure" term than 'analytic' [79, 45], so that the notion of analyticity should be used to explain the notion of necessity, not the other way around [78, 121].

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 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" [83, 138; cf. 32-36]. Setting aside the question of whether this is a pressing objection (there are reasonable arguments that it is not[20]), this seems a more plausible complaint if our conception of language is a syntactic, proof-theoretic one. 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" [16, 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. This somewhat *ad hoc* treatment of synonymous expressions does appear to make the definition of analyticity in each language a 'project unto itself.' However, if we begin from the semantic (i.e. Carnapian) point of view, synonymy does not appear so *ad hoc*: every (referring) 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.[21] Similarly, every predicate is assigned a property, and if two predicates are assigned to the same property, then they are synonymous. In both the case of names and of predicates, synonymy 'falls out of' the characterization of the language in a natural way. On the other hand, a syntactically characterized language must 'tack on' to the axioms or rules of inference to accommodate each instance of synonymy (as Carnap says in *Foundations*), while a semantically characterized language requires no such addition.

[20]See [66, 44–45], which argues that Tarski's definition of 'truth,' which all parties agree is scientifically respectable, would suffer from the same supposed 'defect'—and the same holds for many syntactic concepts as well. Quine's reply is that truth is different, because Tarski's truth schema ('. . . ' is true if and only if . . .) grants an antecedent clarity to the notion of truth, thereby obviating the need for an explication [83, 134]—while 'analytic' lacks this clarity and thus needs to be explicated.

[21]One can 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. Examples of this sort push Carnap, in *Meaning and Necessity* and afterwards, to assign to each individual constant an individual concept, which is defined as an object in each possible world.

Chapter 6

Overcoming Metaphysics via the Unity of Science

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 their claims, methods, or concepts. The logical empiricists also made the unity of science a central plank of their party platform. A second essential plank of their platform is their antipathy toward metaphysics;[1] 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.[2] And the anti-metaphysical drive is not yet departed: in much of van Fraassens 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

[1]Michael Friedman [43] has shown convincingly that to say that Carnap and others 'reject' metaphysics is at best a very misleading way of characterizing the logical empiricists' enterprise–and not merely because they do not assert the falsity (instead of the meaninglessness) of traditional metaphysical theses. 'Overcoming,' Friedman argues, is a much more apt term.

[2]Schlick writes: "the denial of metaphysics is an old attitude . . . for which we can in no way claim priority" [95, 492]. Carnap echoes this sentiment: "Anti-metaphysical views have often been put forward in the past, especially by Hume and the Positivists" of the nineteenth century [11, 280].

reveals that, metaphorically speaking, these goals are two sides of the same coin. More prosaically, in different logical empiricists, from the 1920s through 1950, we find the following criterion (or an approximation thereof) at work for detecting metaphysics: an apparently meaningful 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. To conclude, I present an objection to this criterion for identifying metaphysics.

How, the reader may ask, is this related to previous chapters? Well, what is the *point* of undertaking Tarski's project of reconstructing arithmetic within a language of strictly empirical science (i.e., a language of science that does not countenance the existence of abstracta)? As suggested in 2.2, part of the motivation for pursuing this finitist-nominalist project is to purge any noxious metaphysical elements from the language used by scientists—including mathematical language. But why would revising mathematical language (and mathematics) achieve that anti-metaphysical end? Because, to a first approximation, for Carnap and his intellectual allies, the distinguishing mark of a metaphysical bit of language is that it cannot be incorporated into unified scientific language. It is the burden of the present chapter to argue for this last assertion.

6.1 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 in the Vienna of the late 1920s. 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 the early activities of *Synthese*: the first sentence of its first issue is "Ours is a time of synthesis," i.e., of unification. During the forties, *Synthese* regularly included articles under the series 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. [29], so I will not attempt a complete substantiation of this claim. Nonetheless, I devote this section to an abbreviated elaboration and defense of this 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).[3] Thus the logical empiricists of the 1930s unequivocally do

[3] However, commentators nonetheless saddle logical empiricists in general, and Carnap in

not endorse the kind of 'unity of science' found in (e.g.) Putnam and Oppenheim's "The Unity of Science as a Working Hypothesis" [71]. 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" [15, 61]. Neurath's views are similar. He does not demand a unity of laws; as a social scientist, he 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" [70, 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 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 his new Encyclopedia continues, "was not a '*faute de mieux* encyclopedia' in place of a comprehensive system, but an alternative to systems" [68, 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 linguistic: "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"' [70, 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 [41, 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" [56, 326].[4] In *Logical Syntax*, Carnap offers

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" [103, 214-215; my emphasis]. It is not the theories that are 'reductively related,' but rather the languages.

[4] Kokoszynska's central contention in this article is an interesting challenge to this standard characterization of the unity of science thesis. 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. However, 'true-in-L' is definable in the metalanguage of L. Kokoszynska's contention is 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

the following more precise characterization of the thesis: every sub-language of science can be translated without loss of content into one language [11, 320].

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.[5] Carnap defends this thesis most extensively in "*Die Physikalische Sprache als Universalsprache der Wissenschaft*" [The Physicalist Language as the Universal Language of Science], translated into English two years later under the non-literal but not misleading title *Unity of Science*, [9]. Neurath provides a detailed description of the physicalist language, which he 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" [70, 208]; specifically, it 'avoids' metaphysical terms, and 'is enriched' by technical terms of the special sciences [70, 91–92]. Carnap characterizes the physicalist language as one whose sentences "in the last analysis . . . express properties (or relations) of space-time domains" [11, 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" [70, 75]. But the 'properties' and 'entities' of biology, chemistry, geology, and much of psychology and sociology are spatiotemporal, so the language of physicalism is much richer than the language of physics alone. Note that the languages meeting Tarski's (FN 1–4) can be thought of as a kind of physicalist (perhaps 'hyper-physicalist'?) language, in that they completely reject allowing the existence of abstracta, entities usually conceived of as existing outside the spatio-temporal realm.

Although Carnap and Neurath hold that the physicalist language can serve as the language for unified science, they do not maintain that no other language could. For example, in the *Aufbau*, Carnap holds that both the phenomenal, 'autopsychological' language and the physical[6] language could function as languages for unified science. And Neurath writes: "We expect that it will be

---

found in [70, 206-208]. Much more historical detail and analysis concerning Kokoszynska's interactions with Neurath (and Carnap) concerning Tarski's theory of truth can be found in [65].

[5]Carnap writes:

> 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*.
> [11, 320]

[6]Carnap speaks of the 'physical language' in the *Aufbau*, not of the physical*ist* language. Neurath's term 'physicalist' apparently 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.

possible to replace each word of the physicalist ordinary language by terms of the scientific language—just as it is also possible to formulate the terms of the scientific language with the help of the terms of ordinary language" [70, 91]; this shows that the 'physicalist ordinary language' is not the only language that can serve to unify science, since he 'expects' that the technical 'scientific language' will be able to do so as well. However, it should be noted that Neurath sometimes appears to privilege the physicalist language over others: "Unified science contains only physicalist formulations" [70, 54]; "Physicalism is the form work on unified science takes in our time" [70, 56]. Perhaps Neurath considers the physicalist language best for his purposes, even if not unique. 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 of unifying scientific languages.

Additionally, the logical empiricists' unity of science thesis is not refuted by Suppes' observation [99, 5] that the actual terminology used in various subdisciplines of the sciences is increasingly divergent, with each subfield developing its own jargon. Other scholars (e.g. [29]) 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 accepts Suppes' position in §41 of the *Aufbau*: "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" [8, 70]. Carnap, in his most extended defense of the unity of science thesis [9], 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, one more vision of the unity of science from a logical empiricist sympathizer 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":

> 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.
> [108, 164–165]

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 language's constants. This is precisely what Woodger himself attempts to do for portions of biology and neurology in his *Axiomatic Method in Biology* [108] and *Biology and Language* [109]. 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 [109, 278, 310].

6.2 Overcoming Metaphysics

The logical empiricists are (in)famous for assuming an anti-metaphysical stance. All the major figures in the group, as well as many 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" [42, 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" [88, 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 [1], 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 [88, 59] and, less directly, in [26]. As I hope to show, 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.*

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 inferentially isolated from the rest of scientific language (in a sense to be elaborated presently).

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, fuller 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.[7] 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.

6.2.1 *Aufbau*

Let us begin with Carnap's treatment of metaphysics in the *Aufbau*. How does Carnap identify metaphysics there? 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 describes in the *Aufbau*. We can phrase Carnap's criterion for metaphysics in the *Aufbau* as follows:

($\mathbf{M}_{\text{Aufbau}}$) 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

[7] Schlick writes a good deal about the rejection of metaphysics, but does not discuss the unity of science in much detail; 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.

> 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.
> [8, 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.
> [8, 270–271]

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.[8] Other metaphysical concepts are shown to have the same property; none can be incorporated into a constructional system.

Finally, two significant differences between (M_{Aufbau}) and Carnap's later characterizations of metaphysics should be noted: first, in the *Aufbau*, Carnap thinks primarily in terms of concepts; sentences are secondary. Second, the *Aufbau* lacks the claim that many metaphysical sentences 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 in 1930.

---

[8] 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 'Concept C cannot be cashed out (defined) in terms of sense experience' will be equivalent to 'C cannot be incorporated into a constructional system.'

6.2.2 "Overcoming Metaphysics through Logical Analysis of Language"

Carnap's most focused attack on metaphysics is "Overcoming Metaphysics through the Logical Analysis of Language" [10]. 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, he claims that "only" with the "development of modern logic" can "the decisive step be taken" in this pursuit [10, 61]. Why? Carnap's justification is that an apparent sentence (even if it contains only meaningful words) is meaningless, i.e. metaphysical, if it cannot be expressed in a logical language of the form found in *Principia Mathematica*. This is why Carnap claims the 'development of modern logic' is essential to overcoming 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*.[9] 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 1930s "was in essentials that of Wittgenstein" [11, 322]; see also [70, 54].

We are now in a position to understand one of Carnap's criticisms of Descartes' 'I think, therefore I am.' Carnap claims that the statement 'I am' (or 'Greg is,' assuming 'Greg' is treated as an individual constant in the language of this sentence) cannot be put into the language of classical predicate logic: the concept of existence in the modern logic of Russell and Whitehead's *Principia* is not a predicate, but an operator that acts upon formulae. Thus Carnap claims it is impossible, given the resources of such a language, to express that an individual in the domain of discourse exists simpliciter: one can only say either '$\exists x Px$' or 'Pa' (in the usual notation), but the string of symbols '$\exists a$' is not an admissible sentence [10, 74]. Carnap concludes that Descartes' assertion is meaningless, since it cannot be expressed in the language of *Principia*.[10] Any 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 formal logic, 'nothing' is represented as a concatenation 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

[9] 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" [89, 26-27].

[10] I do not know why Carnap would not allow '$\exists x(x = \text{Greg})$ to express the colloquial 'Greg is' or 'Greg exists'; this is the usual way of expressing the existence of individuals in classical predicate logic. Perhaps in 1932 Carnap was under the spell of the Tractarian viewpoint that declares such an expression meaningless, and the identity-sign superfluous (5.531-5.534).

language. And, of course, one cannot (grammatically) place the string '$\neg\exists x$' into an object-variable position of a propositional function.

So much for Carnap's account of metaphysical sentences; when is an apparently meaningful 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?
> [10, 62]

Here again we find 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:[11] (1.) characterizes the network syntactically or proof-theoretically, (2.) semantically. The unified language of science provides this inferential network. Grammatical strings that cannot be placed within such a network of scientific claims (e.g. 'God is benevolent'), Carnap maintains, contain metaphysical terms. Which particular term(s) in such a sentence are the metaphysical ones? For a given term t (of any grammatical category), t is metaphysical if and only if no atomic sentence containing t can be incorporated into the unified language of science. For example, in 'God is male,' 'God' is metaphysical though 'male' is not, for no sentence of the form 'God is F' can be incorporated into the unified language of science, though sentences of the form 'x is male' can be. This view is very reminiscent of 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 meaningless words?

To solve this problem, Carnap appeals 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*?"[10, 62]. For Carnap in 1932, this question is answered by specifying the inferential relations between S and the "'observation sentences' or 'protocol sentences.' It is through this reduction that the word acquires its meaning" [10, 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" (as in Mach), "total experience and similarities between them" (as in the *Aufbau*), or simply "things" [10, 63]. Furthermore, two years later, in "*Über Protokolsätze*," Carnap states that which sentences are protocol sentences is a matter of decision [12].

[11] Logical implications must be excluded: 'God is benevolent' entails 'God is benevolent or water boils at 100 degrees Centigrade,' and is entailed by 'God is benevolent and mammals have hair'; but Carnap wants 'God is benevolent' to count as metaphysical.

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 up,' via inferential relations, to S. These two assumptions 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 of meaning for all sentences. We arrive at the full-fledged verification criterion of meaning (as well as the liberalized empiricist meaning criteria which appear later[12]) by adding to the two assumptions of semantic foundationalism a third: observation sentences (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 unified language of science. 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 unified language of science, by connecting the terms of the individual sciences to the semantically privileged sentences identified by the verification criterion, shows the sentences of physics, biology, and psychology to be meaningful.[13]

6.2.3 *Logical Syntax of Language*

As Carnap's philosophical views change over his career, so does his characterization of what is metaphysical. In 1934, *Logical Syntax* appears, and with it a

[12]The difference between the earlier, verificationist criterion of meaning and the later, liberalized ones (e.g. in "Testability and Meaning") is that the verification criterion requires that observation sentences entail every meaningful sentence, whereas later criteria allowed weaker logical relations to hold between the observation sentences and other meaningful sentences.

[13]The preceding two paragraphs are heavily indebted to Jon Tsou.

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 methods of the anti-metaphysical program must be altered somewhat in order to accommodate the principle of tolerance:

> 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.
> [11, 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 as well as those apparently descriptive sentences that cannot be connected with any language of empirical science are dismissed as metaphysical pseudo-sentences [11, 322]. 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 sentences labeled '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 illustrate and make plausible the claim that metaphysics could be considered language relative. Consider Languages I and II of *Logical Syntax*: Language I, intended to capture the mathematical intuitionist's point of view, 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 so-called unconstructable real numbers would qualify on Carnap's proposal as a metaphysical pseudo-sentence in I, but not in II. Now, 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" [53, 67]. As a second example, consider the relation between first-order and higher-order logics, much discussed in the 1940–41 conversations: any sentences of second-order logic containing higher-order predicates would, in first-order logic, be metaphysical on Carnap's criterion. And as we have seen above 2.2, philosophers who find second-order logic suspicious call its quantification over properties 'Platonism'—and Plato is one of the most notorious metaphysicians. Thus Carnap's suggestion that what one counts as metaphysics depends on the language one uses is borne out in these actual examples. In sum, in *Logical Syntax*, the means of identifying metaphysics is, at root, the same as that found in Carnap's earlier works, but modified to accommodate the principle of tolerance.

6.2.4 "Empiricism, Semantics, and Ontology"

In 1950's "Empiricism, Semantics, and Ontology," Carnap's basic idea for distinguishing metaphysics from acceptable forms of discourse is essentially the same as before. 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.*
> [19, 207; my emphasis]

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 formulation of their question in the common scientific language. Therefore ... they have not succeeded in giving the external question cognitive content.
> [19, 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 *Konstitutionsysteme*, but the basic strategy for identifying and eliminating metaphysics remains the same.

6.2.5 Neurath

So much for Carnap's views on metaphysics; what of Neurath's? Though he eschews Carnaps formal, precise languages in favor of his 'universal jargon' or 'universal slang,' which is based on everyday language and is modeled on the structure of an encyclopedia instead of an axiom system [68, 2, 7, 16, 20], he shares the fundamental idea we have seen in Carnap: an apparently meaningful sentence or term is metaphysical if and only if it cannot be incorporated into unified science. First, let us consider the 'only if' direction: "If it [a proposed scientific sentence] is . . . meaningless—i.e., metaphysical—then of course it falls outside the sphere of unified science" [70, 58].[14] 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" [70, 56]. Thus we find assertions such as the following: "If we systematically formulate everything we find in non-metaphysical formulations, we get nothing but physicalist formulations" [70, 73].

Neurath explicitly articulates the 'if' direction of the biconditional (M) as well:

> 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'.
> [70, 161]

Note that Neurath mentions the strictures against both ungrammatical and otherwise isolated apparent 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"; furthermore, such linguistic defects have no place in a language of unified science: "[a]n unblemished syntax is the foundation of an unblemished unified science" [70, 54]. Unfortunately, Neurath does not spell out general criteria for what counts as a 'blemish' in a language's syntax; he apparently defers to his more logically-inclined colleagues in this matter. In general, 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 (6.2.2): a sentence's meaningfulness is demonstrated by showing that it is connected via inferential relations to sentences whose meaningfulness is given antecedently. Carnap identifies these semantically privileged sentences as the protocol or observational ones (though, as we saw, he was willing in 1932 and after to leave the exact form of such sentences open). Neurath, it appears, takes a slightly different set of sentences as antecedently meaningful. Neurath repeatedly states that a language of unified science should take as its starting point everyday

[14] See also [70, 54, 57, 61, 73, 173].

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 most likely the best we can do. (There are other reasons Neurath starts with everyday language: he values the democratization and popularization of scientific knowledge,[15] and he is suspicious of any framework that aims to break free of our present historically given situation—which includes our language—and view the world *sub specie aeternitatis*.)

Neurath's writings make it clear that, for him, a central aim of unifying science is the demolition of barriers between the scientific study of nature [*Naturwissenschaften*] and of the mind [*Geisteswissenschaften*]. Thus, one might allege that my focus on the anti-metaphysical drive misses that aspect of his thought entirely. I fully concede, of course, that Neurath repeatedly and unequivocally urged the value of breaking down these disciplinary barriers. But, interestingly, Neurath claims that the motivation underlying the separation of the sciences is *metaphysical*. As 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.
> [70, 68][16]

So, according to Neurath, the assertions used to justify the existence of insuperable boundaries between *Geisteswissenschaft* and *Naturwissenschaft* are 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, Neurath says, it eliminates any theory that purports to deal with "a special sphere of the 'soul"' [70, 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' [13, 58-59]–though theology is often considered a branch of metaphysics. Unification of the sciences may be valuable for its own sake, but it also serves to eliminate metaphysics.

[15] "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" [70, 237-237].

[16] See also [70, 44, 50, 69].

6.3 A Difficulty: What *Cannot* be Incorporated into a Language of Science?

As I have argued, a concept or sentence 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 such an account of metaphysics is: how do we know which concepts and claims can be incorporated, and which cannot? The answer to that question will determine what is metaphysical and thus in need of excision from our scientifically respectable language. 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 offers more assertions than arguments.[17] For example, Carnap simply asserts that "essence," taken in its "metaphysical" sense, cannot be constructed in the autopsychological constitution system of the *Aufbau* or in any other constitution system [8, §161].

Carnap's treatment of the mind-body problem is similar, but it better illustrates the potential shortcomings of equating the metaphysical with the unconstructable. In a phenomenal or 'autopsychological' 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 corresponds to phenomenal entities instead. 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" [8, 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, one may wonder, what if Hempel and Oppenheim's groundbreaking "Studies in the Logic of Explanation" had been published not in 1947 but in 1922? The conception 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 par-

[17] Alberto Coffa expresses surprise at how scanty Carnap's argumentation is here [24, 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 [88, 60]. However, 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.

ticular counterfactual circumstance,[18] this points to a serious and fundamental difficulty. In any case where Carnap (or any other logical empiricist) asserts that a given term or sentence cannot be incorporated into any unified language of science, it is (epistemically) possible that another person could later show how that concept can, in fact, be so integrated. For example, Tarski showed how to define 'truth (in a language)' rigorously, a term that many logical empiricists previously considered the province of the speculative metaphysician [65]. 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, more properly philosophical, 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 in many cases of interest, such an incorporation, however unexpected, could conceivably be achieved tomorrow—given sufficient ingenuity.[19] Our limited technical creativity is not a demonstration of impossibility. I am not claiming that (M) fails to provide necessary and sufficient conditions for identifying metaphysical terms and claims; it may well be an extensionally adequate criterion. Rather, the problem is that, in many cases of interest, we cannot know whether those conditions have been met—and thus the criterion cannot be used to determine effectively whether a particular somewhat suspicious claim is objectionable metaphysics or not.

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: 'cannot' has been replaced by 'is not.' Adopting (M*) would

[18]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 [21], Carnap happily accepts and deploys the concept of a law of nature in his account of scientific explanation.)

[19]Of course, there are results in model theory concerning the definability or indefinability of certain notions in a particular language; e.g., 'finite' cannot be defined within standard first-order logic with identity. My remarks are obviously not intended to apply to cases in which indefinability can be demonstrated.

constitute a departure from the logical empiricists' original conception of the link between metaphysical neutrality and unified science,[20] but it would also defuse the objection raised in the previous paragraph.

However, (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 one 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' original (M).

Another defense of the logical empiricists' proposal for overcoming metaphysics could be drawn from what could be called an 'enlightened' (or perhaps simply 'greatly weakened') verification criterion of meaning.[21] Suppose we have some claim p that we suspect might be noxious metaphysics instead of the sanitary science its defenders claim it to be. Following the general spirit of the verification criterion (though certainly not its letter), one might propose:

> Given (the idealized fiction of) the total class of all current scientific claims, if adding p to that class generates empirical predictions (that need not be currently observable) that cannot be derived without p,[22] then p is not metaphysical.

In other words, if adopting a claim into the edifice of science allows us to make new predictions, then that claim is not objectionably metaphysical. Note that this, unlike all the variants of (M) considered thus far, is *not* a biconditional. In fact, the converse does not appear to hold. For example, there are a variety of formulations of classical mechanics; Hamiltonian and Lagrangian versions are perhaps the most used. If we 'begin' with e.g. Hamiltonian mechanics, 'adding' the Lagrangian formulation will not generate any new predictions (since the

[20]As Jon Tsou pointed out to me, (M*) also appears to run counter to Carnapian tolerance, since new language forms 'under construction' would apparently count as metaphysical.

[21]This suggestion is due to Jon Tsou.

[22]In other words, p is not *conservative* over the total class of current scientific claims.

two theories are empirically equivalent)—but we certainly do not want to count Lagrangian mechanics as unpalatable metaphysics. So while this new criterion for separating out the scientific wheat from the metaphysical chaff is going to be applicable in a number of cases (though it is not always obvious whether adding a sentence to the current stock of scientific knowledge does, in fact, generate new predictive content), this criterion will work only for fairly obvious cases, and will be of little use in more subtle or contentious cases. A more sustained attack or defense could be made for (M) (or its conceptual kin); however, I will not dwell on this matter further, for there is not much contemporary interest in separating out scientific elements in our discourse from metaphysical ones.

6.4 Conclusion: The Origin of the Term 'Unified Science'

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 approximately captured 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 program to overcome 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.
> [70, 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 part, while the production of a unified scientific language constitutes its positive or constructive aspect.

What does the unity of science movement have to do with the finitist-nominalist project pursued in 1940–41? The primary point of contact is the anti-metaphysical animus, which animates the unity of science as well as nominalism, both in 1941 and later. The initial suspicion towards classical mathematical language—which appears *prima facie* to commit us to strange abstract entities—is overcome by showing that a substantial portion of such suspicious language can be captured in a language that is thoroughly empirical: the range of its existential quantifiers are empirical, concrete things only. Mathematical discourse is shown to be meaningful—instead of metaphysical—by embedding it within paradigmatically meaningful discourse. And as this chapter endeavored to show, for Carnap and his intellectual allies, any concept or claim that can be incorporated into a unified language of empirical science is not metaphysical.

What, in the end, should we make of Carnap's 1940–41 Harvard notes as a whole? I must confess that when I began studying these documents, I was initially somewhat disappointed by the primary foci of the participants' labors. Although it was certainly surprising that Carnap, Tarski and Quine were discussing a seemingly strange set of questions about what arithmetic should look like if the number of things in the world should turn out to be finite, I did wonder whether these notes might simply be a historical curiosity—something like 'Isn't that funny, that these great minds spent the year discussing such a strange question that is tangential to their primary interests and achievements?' That is, I wished that Carnap, Tarski, and Quine had concentrated their efforts on issues that we consider to lie at the heart of their various logico-philosophical enterprises; for example, I wanted 75 pages of a *tête-à-tête-à-tête* over analytic truth, or new information that would decisively settle outstanding historical disputes among historians of analytic philosophy.

As my initial disappointment began to fade, and I began to think about the notes in more detail, it became clearer to me that these philosophers actually were still working on a set of issues not so far from those considered central to them—and to us today. For example, the relationship between mathematical theories and theories about the natural world is absolutely central for almost all of those caught up in the intellectual currents of logical empiricism—and for Carnap in particular. He saw Tarski's finitist-nominalist project, and Quine's support for it, as a retrogressive slide back into Mill's empiricist view of mathematics. Furthermore, Carnap viewed Tarski's program as closely related to his own long-standing project of investigating the relationship between what Carnap usually called the 'observational' and 'theoretical' parts of language; this again shows that the 1940–41 conversations should not be considered an 'outlier' irrelevant to Carnap's wider goals. The demand for intelligibility, as I argued in 2.1.3, is very closely tied to the issue of linguistic meaning, a central concern of both the logical empiricists and many philosophers who came after them. And despite my initial disappointment that the notes are not simply a sustained, direct confrontation over the analytic-synthetic distinction, the issue nonetheless makes prominent appearances. For example, in a FN language certain parts of arithmetic become synthetic, and light was hopefully thrown on the historical trajectory that led Quine to "Two Dogmas."

I firmly believe that much more of historical and conceptual worth can be mined from these dictation notes. I steadfastly hope more of value will be. With the growing interest in the history of analytic philosophy, and the history of science and logic in particular, there is—thankfully—good reason to believe this hope will be fulfilled.

Bibliography

[1] A. J. Ayer. *Logical Positivism.* Free Press, 1959.

[2] Timothy Bays. On tarski on models. *Journal of Symbolic Logic*, 66:1701–1726, 2001.

[3] Paul Benacerraf. What numbers could not be. In Paul Benacerraf and Hilary Putnam, editors, *Philosophy of Mathematics: Selected Readings.* Cambridge University Press, second edition, 1965/1983.

[4] Paul Bernays. On platonism in mathematics. In Paul Benacerraf and Hilary Putnam, editors, *Philosophy of Mathematics: Selected Readings.* Cambridge University Press, second edition, 1935/1983.

[5] Jacques Bouveresse. On the meaning of the word 'platonism' in the phrase 'mathematical platonism'. *Proceedings of the Aristotelian Society*, pages 55–70, 2005.

[6] John Burgess. Why i am not a nominalist. *Notre Dame Journal of Formal Logic*, 24:93–105, 1983.

[7] John Burgess and Gideon Rosen. *A Subject with no Object.* Oxford University Press, 1996.

[8] Rudolf Carnap. *The Logical Structure of the World.* Open Court, 1928/1963.

[9] Rudolf Carnap. *Unity of Science.* Kegan Paul, London, 1932/1934.

[10] Rudolf Carnap. Overcoming metaphysics through logical analysis of language. In A. J. Ayer, editor, *Logical Positivism.* Free Press, Glencoe, IL, 1932/1959.

[11] Rudolf Carnap. *Logical Syntax of Language.* Open Court, Chicago, 1934/1937.

[12] Rudolf Carnap. On protocol sentences. *Nous*, 21:457–470, 1934/1987.

[13] Rudolf Carnap. The task of the logic of science. In Brian McGuinness, editor, *Unified Science.* Reidel, 1934/1987.

[14] Rudolf Carnap. Testability and meaning. *Philosophy of Science*, 3-4:419–471 and 1–40, 1936-1937.

[15] Rudolf Carnap. Logical foundations of the unity of science. In Otto Neurath, editor, *The International Encyclopedia of Unified Science*, volume 1, pages 42–62. University of Chicago Press, 1938.

[16] Rudolf Carnap. Foundations of logic and mathematics. In *The International Encyclopedia of Unified Science*. University of Chicago Press, 1939.

[17] Rudolf Carnap. *Introduction to Semantics*. Harvard University Press, 1942.

[18] Rudolf Carnap. *Introduction to Symbolic Logic and its Applications*. Dover, 1954/1958.

[19] Rudolf Carnap. *Meaning and Necessity*. University of Chicago Press, second edition, 1956.

[20] Rudolf Carnap. Intellectual autobiography. In Paul Schilpp, editor, *Philosophy of Rudolf Carnap*. Open Court, La Salle, 1963.

[21] Rudolf Carnap. *Introduction to the Philosophy of Science*. Basic Books, New York, 1966.

[22] Leon Chwistek. Die nominalistische grundlegung der mathematik. *Erkentnis*, 3:367–388, 1932-33.

[23] Leon Chwistek. *The Limits of Science*. Routledge and Kegan Paul, 1935/1949.

[24] Alberto Coffa. *The Semantic Tradition from Kant to Carnap*. Cambridge University Press, 1991.

[25] Mark Colyvan. *The Indispensibility of Mathematics*. Oxford University Press, 2001.

[26] Richard Creath. Was carnap a complete verificationist in the aufbau? *Philosophy of Science Association (Proceedings 1982)*, pages 384–393, 1982.

[27] Richard Creath. The initial reception of carnap's doctrine of analyticity. *Nous*, 21:477–499, 1987.

[28] Richard Creath. *Dear Carnap, Dear Van*. University of California Press, 1990.

[29] Richard Creath. The unity of science: Carnap, neurath, and beyond. In Peter Galison and David Stump, editors, *The Disunity of Science: Boundaries, Contexts, and Power*, pages 417–427. Stanford University Press, 1996.

[30] Richard Creath. Quine on the intelligibility and relevance of analyticity. In Jr. Roger Gibson, editor, *The Cambridge Companion to Quine*. Cambridge University Press, 2004.

[31] Maria Luisa Dalla Chiara and Giulano di Francia. Quine on physical objects. In Paolo Leonardo and Marco Santambrogio, editors, *On Quine*, pages 104–112. Cambridge University Press, 1995.

[32] Donald Davidson. Truth and meaning. In *Inquiries into Truth and Interpretation*. Oxford University Press, 1984.

[33] Michael Detlefsen. *Hilbert's Program: An Essay on Mathematical Instrumentalism*. D. Reidel, Boston, 1986.

[34] John Etchemendy. Tarski on truth and logical consequence. *Journal of Symbolic Logic*, 53:51–79, 1988.

[35] Anita Burdman Feferman and Solomon Feferman. *Alfred Tarski: Life and Logic*. Cambridge University Press, 2004.

[36] Hartry Field. Tarski's theory of truth. *Journal of Philosophy*, 69:347–375, 1972.

[37] Hartry Field. *Science Without Numbers: A Defense of Nominalism*. Princeton University Press, 1980.

[38] Hartry Field. *Truth and the Absence of Fact*. Oxford University Press, 2001.

[39] Jerry Fodor and Ernest Lepore. Why meaning (probably) isn't conceptual role. *Mind and Language*, 6:328–343, 1991.

[40] Janet Folina. *Poincaré and the Philosophy of Mathematics*. St. Martin's Press, New York, 1992.

[41] Philipp Frank. The institute for the unity of science: Its background and purpose. *Synthese*, 6:160–167, 1947.

[42] Michael Friedman. *Reconsidering Logical Positivism*. Cambridge University Press, 1999.

[43] Michael Friedman. *A Parting of the Ways*. Open Court, 2000.

[44] Michael Friedman. *The Dynamics of Reason*. CSLI Press, Stanford, 2001.

[45] Greg Frost-Arnold. Was tarski's theory of truth motivated by physicalism? *History and Philosophy of Logic*, 25:265–280, 2004.

[46] Kurt Gödel. What is cantor's continuum problem? In Paul Benacerraf and Hilary Putnam, editors, *Philosophy of Mathematics: Selected Readings*. Cambridge University Press, 1963/1983.

[47] Mario Gomez-Torrente. Tarski on logical consequence. *Notre Dame Journal of Formal Logic*, 37:125–151, 1996.

[48] Nelson Goodman. A world of individuals. In *The Problem of Universals: A Symposium*. University of Notre Dame Press, 1956.

[49] Nelson Goodman. *The Structure of Appearance*. Bobbs-Merrill, Indianapolis, second edition, 1966.

[50] Anil Gupta. Deflationism, the problem of representation, and horwich's use theory of meaning. *Philosophy and Phenomenological Research*, 67:654–666, 2003.

[51] Theodore Hailperin. Herbrand semantics, the potential infinite, and ontology-free logic. *History and Philosophy of Logic*, 13:69–90, 1992.

[52] Gary Hardcastle. Debabelizing science: The harvard science of science discussion group, 1940-41. In Gary Hardcastle and Alan Richardson, editors, *Logical Empiricism in North America*, pages 170–196. University of Minnesota Press, 2003.

[53] Arend Heyting. Disputation. In Paul Benacerraf and Hilary Putnam, editors, *Philosophy of Mathematics: Selected Readings*. Cambridge University Press, second edition, 1971/1983.

[54] David Hilbert. On the infinite. In *Philosophy of Mathematics: Selected Readings*. Cambridge University Press, second edition, 1925/1983.

[55] Peter Hylton. 'the defensible province of philosophy': Quine's 1934 lectures on carnap. In Juliet Floyd and Sanford Shieh, editors, *Future Pasts: The Analytic Tradition in Twentieth-Century Philosophy*. Oxford University Press, 2001.

[56] Maria Kokoszynska. Bemerkungen über die einheitswissenschaft. *Erkenntnis*, 7:325–335, 1937-1938.

[57] Tadeuz Kotarbinski. *Gnosiology: The Scientific Approach to the Theory of Knowledge*. Pergamon, Oxford, 1929/1966.

[58] Tadeuz Kotarbinski. The fundamental ideas of pansomatism. *Mind*, 64:488–500, 1935/1955.

[59] Michal Krynicki and Konrad Zdanowski. Theories of arithmetics in finite models. *Journal of Symbolic Logic*, 70:1–28, 2005.

[60] Penelope Maddy. *Realism in Mathematics*. Oxford University Press, 1990.

[61] Paolo Mancosu. Mathematical explanation: Problems and prospects. *Topoi*, 20:97–117, 2001.

[62] Paolo Mancosu. Harvard 1940-1941: Tarski, carnap and quine on a finitistic language of mathematics for science. *History and Philosophy of Logic*, 26:327–357, 2005.

[63] Paolo Mancosu. Tarski on models and logical consequence. In J. Ferreiros and J. Gray, editors, *The Architecture of Modern Mathematics*. Oxford University Press, 2006.

[64] Paolo Mancosu. From harvard to amersfoort: Quine and tarski on nominalism. In Dean Zimmerman, editor, *Oxford Studies in Metaphysics*, volume 3. Oxford University Press, forthcoming a.

[65] Paolo Mancosu. Tarski, neurath, and kokoszynska on the semantic conception of truth. In Douglas Patterson, editor, *Alfred Tarski: Philosophical Background, Development, and Influence*. Oxford University Press, forthcoming b.

[66] Richard Martin. On 'analytic'. *Philosophical Studies*, 3:42–47, 1952.

[67] Marcin Mostowski. On representing concepts in finite models. *Mathematical Logic Quarterly*, 47:513–523, 2001.

[68] Otto Neurath. Unified science as encyclopedic integration. In *The International Encyclopedia of Unified Science*, volume I, pages 1–27. University of Chicago Press, 1938.

[69] Otto Neurath. ???? *????*, ?1946?

[70] Otto Neurath. *Philosophical Papers*. Reidel, 1983.

[71] Paul Oppenheim and Hilary Putnam. The unity of science as a working hypothesis. In Herbert Feigel, Michael Scriven, and Grover Maxwell, editors, *Concepts, Theories, and the Mind-Body Problem*, volume II of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, 1958.

[72] Henri Poincaré. *Science and Hypothesis*. Dover, 1902/1905.

[73] Graham Priest. Is arithmetic consistent? *Mind*, 103:337–349, 1994.

[74] Graham Priest. Inconsistent models of arithmetic i: Finite models. *Journal of Philosophical Logic*, 26:223–235, 1997.

[75] Stathis Psillos. Rudolf carnap's 'theoretical concepts in science'. *Studies in the History and Philosophy of Science*, 31:151–172, 2000.

[76] Willard Van Orman Quine. Designation and existence. *Journal of Philosophy*, 36:701–709, 1939.

[77] Willard Van Orman Quine. *Mathematical Logic*. Harvard University Press, revised edition, 1940/1958.

[78] Willard Van Orman Quine. Notes on existence and necessity. *Journal of Philosophy*, 40:113–127, 1943.

[79] Willard Van Orman Quine. On the problem of interpreting modal logic. *Journal of Symbolic Logic*, 12:43–48, 1943.

[80] Willard Van Orman Quine. On universals. *Journal of Symbolic Logic*, 12:74–84, 1947.

[81] Willard Van Orman Quine. Main trends in recent philosophy: Two dogmas of empiricism. *Philosophical Review*, 60:20–43, 1951.

[82] Willard Van Orman Quine. *Word and Object*. MIT Press, 1960.

[83] Willard Van Orman Quine. *From a Logical Point of View*. Harvard University Press, second edition, 1961.

[84] Willard Van Orman Quine. *The Ways of Paradox*. Harvard University Press, second edition, 1976.

[85] Willard Van Orman Quine. *The Time of My Life: An Autobiography*. The MIT Press, 1985.

[86] Willard Van Orman Quine. Confessions of a confirmed extensionalist. In Juliet Floyd and Sanford Shieh, editors, *Future Pasts: The Analytic Tradition in Twentieth-Century Philosophy*. Oxford University Press, 2001.

[87] Willard Van Orman Quine and Nelson Goodman. Steps toward a constructive nominalism. *Journal of Symbolic Logic*, 12:105–122, 1947.

[88] Alan Richardson. Metaphysics and idealism in the aufbau. *Grazer Philosophische Studien*, 43:45–72, 1992.

[89] Alan Richardson. *Carnap's Construction of the World*. Cambridge University Press, 1998.

[90] Thomas Ricketts. Carnap's principle of tolerance, empiricism, and conventionalism. In P. Clark and B. Hale, editors, *Reading Putnam*, pages 176–200. Blackwell, 1994.

[91] Eric Rosen. Some aspects of model theory and finite structures. *Bulletin of Symbolic Logic*, 8:380–403, 2002.

[92] Bertrand Russell. *Mysticism and Logic and Other Essays*. Longmans Green, London, 1918.

[93] Bertrand Russell. *Introduction to Mathematical Philosophy*. Allen and Unwin, London, 1919.

[94] Bertrand Russell. The philosophy of logical atomism. In R. C. Marsh, editor, *Logic and Knowledge*. Allen and Unwin, 1956.

[95] Moritz Schlick. *Moritz Schlick: Philosophical Papers*, volume 1. Reidel, 1978.

[96] John Searle. Minds, brains, and programs. *Behavioral and Brain Sciences*, 3:417–424, 1980.

[97] Peter Simons. Nominalism in poland. In Francisco Coniglione, Roberto Poli, and Jan Wolenski, editors, *Polish Scientific Philosophy: The Lvov-Warsaw School*. Rodopi, Amsterdam, 1993.

[98] Stanley S. Stevens. S. s. stevens. In *A History of Psychology in Autobiography*, pages 395–420. Appleton, Century, Crofts, New York, 1974.

[99] Patrick Suppes. The plurality of science. *Philosophy of Science Association (Proceedings 1978)*, 2:3–16, 1978.

[100] W. W. Tait. Finitism. *Journal of Philosophy*, 78:524–556, 1981.

[101] Alfred Tarski. *Logic, Semantics, Metamathematics*. Hackett, Indianapolis, second edition, 1983.

[102] Neil Tennant. *The Taming of the True*. Oxford University Press, 1997.

[103] Thomas Uebel. Carnap and neurath in exile: Can their disputes be resolved? *International Studies in the Philosophy of Science*, 15:211–220, 2001.

[104] Bas C. van Fraassen. *The Scientific Image*. Oxford University Press, 1980.

[105] Morton White. *A Philosopher's Story*. Pennsylvania State Press, University Park, PA, 1999.

[106] Ludwig Wittgenstein. *Tractatus Logico-Philosophicus*. Humanities Press International, Atlantic Highlands, NJ, 1921/1961.

[107] Jan Wolenski. Tarski as philosopher. In Francisco Coniglione, Roberto Poli, and Jan Wolenski, editors, *Polish Scientific Philosophy: The Lvov-Warsaw School*. Rodopi, Amsterdam, 1993.

[108] J. H. Woodger. *The Axiomatic Method in Biology*. Cambridge University Press, 1937.

[109] J. H. Woodger. *Biology and Language*. Cambridge University Press, 1952.