Carnap's Dream: Gödel, Wittgenstein, and Logical Syntax

S. Awodey and A.W. Carus

Philosophy

Methodology

Logic

Carnegie Mellon

Pittsburgh, Pennsylvania 15213

Carnap's Dream: Gödel, Wittgenstein, and *Logical Syntax*

S. Awodey and A.W. Carus

DRAFT: 16 July 2003

At a meeting of the Vienna Circle in early 1931, Gödel raised very simply and bluntly a critical point that had been troubling everyone present for some time:

> *Gödel* fragte, wie die Diskussion über logische Fragen zu rechtfertigen sei, da man ja dabei keine sinnvollen Sätze sondern nur Erläuterungen ausspricht. Es ergibt sich nun die Frage, wie die zulässigen Erläuterungen von den metaphysischen Scheinsätzen abzugrenzen sind. (ASP/RC XXX; Stadler 1997, p. 288)

This brings down to bare bones a central question facing the Vienna Circle during this period: What protected its critique of traditional philosophy from *itself*? Is the verification principle *itself* verifiable? Is *any* meta-linguistic statement (such as other tools in the anti-metaphysical kit, e.g. Occam's Razor) verifiable? And if not, why is their status not every bit as metaphysical as that of the Kantian, Hegelian, and Heideggerian philosophies that the Vienna Circle was consigning to the dustbin?

Like so much of the Vienna Circle's agenda, this question had been raised by Wittgenstein. His influence on the Circle is of course well-known, though their grasp of his writings is often dismissed as jejune.[1] What has not been understood is the depth and intensity of the Vienna Circle's, and particularly Carnap's, engagement with the ideas of the *Tractatus*, especially in the period between the *Aufbau* (1928) and the formulation of his new position in *The Logical Syntax of Language*, in the early 1930's. A universal language of the kind described in the *Tractatus* became the basis for the Vienna Circle's criterion of meaning.

---

[1] There is a long tradition of innuendo against the Vienna Circle's ability to understand Wittgenstein; Baker & Hacker, etc. A more recent form of this view has been formulated by James Conant, who attributes to early Wittgenstein the idea that meaning is in the mind of the speaker or the writer, without whom the mere physical sign is not a symbol; the meaning relation (putting the symbol into the sign) is established only in the minds of human interlocutors (Conant 2001, pp. 24-28). Whether a sign or a sentence has meaning cannot be objectively determined, according to this view, i.e. cannot be determined outside the context of its use and the intentions of its users. This is of course very different from Carnap's conception of meaning: "Carnap seeks a method that will furnish criteria that permit one to establish that someone else is speaking nonsense, whereas Wittgenstein (both early and later) seeks a method that ultimately can only be practiced by someone on himself. Wittgenstein's method only permits the verdict that sense has not been spoken to be passed by the one who speaks." (ibid., p. 61) Carnap, of course, understood Wittgenstein as concerned also with objective meaning.

And we will argue that it was precisely Carnap's attempt to expand the scope of this language structure, and the tension between this effort and developments within mathematical logic, that set the scene for the "breakthrough" to the syntax idea that Carnap describes so dramatically in his autobiography as having come to him during a sleepless night in January 1931, "like a vision".

This breakthrough, however, as we describe in the following, led first to a rather different doctrine from that of the published *Syntax* three years later. This initial syntax idea, which found expression in only one of Carnap's publications[2], was expounded systematically in a series of unpublished manuscripts written in the spring of 1931, culminating in three lectures to the Vienna Circle in the summer of 1931, and the first draft of what we know as the *Logical Syntax*, written in the autumn of 1931 and early 1932. This first draft, however, was subjected to a probing criticism by Gödel, who pointed out a fundamental flaw in Carnap's exposition. In the autumn of 1932, Carnap responded to this criticism by abruptly altering the fundamental doctrine of the book. The leading idea now became the celebrated "principle of tolerance", which would henceforth become a principal theme of Carnap's thought. In the published *Syntax*, the original syntax idea of the first draft and the principle of tolerance co-exist in an uneasy truce. This would not last; by 1935, Carnap had accepted that the meta-language for science should be semantic as well as syntactic. So part of the specifically syntactic doctrine of the 1934 book, arrived at in the initial breakthrough of January 1931, would turn out to be ephemeral, while the principle of tolerance became the cornerstone of Carnap's mature philosophy.

But it was no accident that Carnap's formulation of the principle of tolerance grew out of the original syntax view. We focus, in the following, on attempting to formulate this original syntax view, using a number of previously unknown documents. They tell a clear story. Against this background, we then argue, Gödel's critique of the first draft of the *Syntax* (which has not been preserved — at least no copy is known to us) becomes more

---

[2] The famous paper "Die physikalische Sprache als Universalsprache der Wissenschaft" (1932), later translated into English, with a new preface, as the pamphlet *The Unity of Science* (1934).

comprehensible. We also argue that Gödel's critique played an important role in motivating Carnap's radical step toward tolerance in late 1932.

We turn first to a fuller description of the situation in the period *before* the breakthrough of January 1931. By understanding the difficulties Carnap encountered in his attempts to work within Wittgenstein's framework, as well as the internal tensions of that position vis-à-vis other constraints Carnap faced, the full significance to Carnap of the 1931 breakthrough becomes easier to appreciate.

I. A New Foundation of Logic

The *Tractatus* was the undisputed point of departure for the Vienna Circle because in their view, Wittgenstein had solved the age-old Platonic problem of the cognitive status of mathematics, which remained a grave threat to any form of empiricism. "Es scheint zunächst tatsächlich, als müßte am Bestehen der Mathematik der reine Empirismus scheitern, als hätten wir in der Mathematik ein Wissen über die Welt, das nicht aus der Erfahrung stammt, ein Wissen a priori," said Hans Hahn. "Und die hier für den Empirismus vorliegende Schwierigkeit ist so auffällig, so einschneidend, daß jeder, der konsequenten Empirismus vertreten will, zu dieser Schwierigkeit Stellung nehmen muß. . ." (Hahn 1929, p. XX) Wittgenstein had grasped and articulated the implications of Frege's and Russell's deductive systems, and on the basis of those systems he had conceived the idea of a universal language as a standard by which to measure whether a sentence has meaning. This conception became the basis for the Vienna Circle's idea of a single, unified language of science as a basis for rational reconstruction.

And Wittgenstein had done this without accepting Frege's and Russell's metaphysical ideas either about the foundation of logic (an ontology of logical objects, for Frege; synthetic a priori logical principles, for Russell), or about the scope of logic; he rejected their view that logic applies universally to all objects in the world, abstract and physical. Instead, logic becomes an artifact of language itself; it is tautological and empty. Hahn, like the rest of the Vienna Circle, thought this idea of critical importance:

Wollte man die Logik — wie dies geschehen ist — auffassen als die Lehre von den allgemeinsten Eigenschaften der Gegenstände, als die Lehre von den Gegenständen überhaupt, so stünde hier der Empirismus tatsächlich vor einer unüberwindlichen Schwierigkeit. In Wirklichkeit aber sagt die Logik überhaupt nichts über Gegenstände aus. Logik ist nicht etwas, das sich in der Welt vorfindet, Logik entsteht vielmehr erst dadurch, daß — vermöge einer Symbolik — *über die Welt gesprochen wird*. . . Die Sätze der Logik sagen über die Welt nichts aus. (Hahn "Empirismus, Mathematik, Logik", 1929)

But Wittgenstein's solution to the old problems carried a high price tag. Two costly by-products of this solution were particularly unpalatable to the Vienna Circle. First, Wittgenstein's universal language, which the Vienna Circle made the basis of their unified language of science, was insufficient for expressing even a fragment of actually existing science. Indeed, it seemed that most classical mathematics could not be expressed within the *Tractatus* framework. And second, as we have already seen — in the quotation from Gödel at the beginning of this paper — Wittgenstein excluded the possibility of meta-linguistic discourse. Apart from his perhaps somewhat obscure considerations about the impossibility of articulating the picturing relation in language, there was the simple problem that if language consists only of truth functions of the atomic sentences, then meta-linguistic sentences do not qualify. Hardly a sentence in the *Tractatus* itself (or any of the Vienna Circle publications) could reasonably be construed as a truth function of atomic sentences. And Wittgenstein himself had made this self-undermining conclusion explicit in the final sentences of his book.[3]

These consequences of the *Tractatus* were unacceptable to the Vienna Circle because they conflicted with their central project of rational reconstruction. If much of science fails to qualify as meaningful, and discourse about language is excluded in any case, then it becomes impossible even to compare different expressions regarding their precision or their usefulness for some practical purpose. It becomes impossible to say, for instance, that a rationally

---

[3] Here again, it is customary to assume that the Vienna Circle failed to grasp this aspect of the *Tractatus*. Hilary Putnam, for instance, thinks the logical empiricists never considered the question: "An obvious rejoinder [to the verification principle] was to say that the logical positivist criterion of significance was *self-refuting*: for the criterion itself is neither (a) 'analytic'. . . nor (b) empirically testable. Strangely enough this criticism had very little impact on the logical positivists. . . I believe that the neglect of this particular philosophical gambit was a great mistake; that the gambit is not only correct, but contains a deep lesson." (Putnam 1981, p. 106)

reconstructed concept is *more precise*, or *more useful*, than the concept to be reconstructed. This obstructs the Vienna Circle's practical critique of metaphysics and unclear thinking, and undermines its entire Enlightenment project. It was imperative, therefore, to address these two problems.

The approach taken especially by Carnap was to adopt a strictly positivistic interpretation of the *Tractatus*. The atomic sentences were taken to picture something like Mach's "elements" or the *Elementarerlebnisse* of the *Aufbau*. Attempts were made to understand classical mathematics in finitist terms, and to interpret theoretical physics as nothing but abbreviations for truth-functional concatenations of atomic sentences. On these fronts, the constitution system of the *Aufbau* had left a good deal of unfinished business. Two gaps in particular had Carnap's attention during this period. First, the *Aufbau* had been criticized (by Reichenbach and Eino Kaila[4]) for failing to accommodate certain modes of inference required in actual science, such as empirical induction, probability, and statistical inference. In a talk at Reichenbach's seminar in late 1929, Carnap had responded that, in fact, all such inferences could be assimilated to truth-functional inference like that described by Wittgenstein. We can regard any mode of inference, whether in mathematics or in empirical science, he said, formalistically, as a rule for transforming sentences of a certain specified form into sentences of a different form. We can even take axiomatic systems of infinitary mathematics and theoretical physics in this way.

Axiomatic systems, though, raised a more fundamental problem for the entire logicist structure in which Carnap had developed the *Aufbau* system. The explicit definitions in which he had (nominally, at least) attempted to construct the whole of knowledge could not accommodate the "implicit definitions" of concepts in axiomatic systems that Schlick had described in his *Allgemeine Erkenntnislehre*, and Einstein in his lecture *Geometrie und Erfahrung*. This problem Carnap had addressed in a large-scale project to reconcile axiomatic definitions with logicism, and transform implicit into explicit definitions. The result was a large, unfinished manuscript entitled *Investigations in General Axiomatics*. The

---

[4] Reichenbach (1929, pp. HH-JJ) and Kaila (1930, pp. XX-YY) claimed that a fundamental assumption of metaphysical realism was required as a basis for scientific inference. [. . .]

central theorem of the first part of this manuscript proves that an axiom system is categorical if and only if it is complete [*Entscheidungsdefinit*]. Thus in particular, arithmetic is complete, as the Peano axioms are categorical.[5]

One important feature, in the present context, of the system described in the *Axiomatics* was that axiomatic systems are not purely syntactic, but are given a fixed range of interpretions within a basic system, a *Grunddisziplin*, as Carnap called it, of arithmetic and set theory. This made it possible to regard axiomatic systems as having content, as long as it could be shown that the sentences of the *Grunddisziplin* itself had definite meanings. So not only is every sentence in the language of arithmetic decidable, but it has a definite meaning as well, since it is interpreted in the *Grunddisziplin*.

Thus Carnap could claim, e.g. in a lecture in Warsaw of December 1930, that there is only one rule of inference in science: We can transform a sentence however we like, but the conclusion is to have no more content than the premisses; it is to constrain the range of possibly true atomic sentences no less than the premisses; i.e. no new atomic sentences are recognized as true. All laws of logic, as well as all rules of inference in science, he maintains, follow from this principle.

Within his *Axiomatics* manuscript, though, there is no effort to work out such a conception of logic, nor had Carnap specified there how the *Grunddisziplin* acquired its fixed interpretation. These tasks Carnap attempted in a loose sketch he wrote down in Davos in April 1929, when he was attending the "Europäische Hochschultage" where Heidegger and Cassirer debated the legacy of Kant (Friedman 2000). The sketch was headed, ambitiously, "New Foundation for Logic" [*Neue Grundlegung der Logik*]. Its main idea is to erect a Hilbertian axiomatic superstructure on a Wittgensteinian basis. The atomic sentences are pictures of elementary

---

[5] This project is discussed in Awodey and Carus (2001), where we also give a more detailed account of the importance of the *Gabelbarkeitssatz* for Carnap's *Aufbau* project as well as the Vienna Circle's entire philosophy of logic and mathematics. Carnap's proof of the *Gabelbarkeitssatz* is actually correct, in his own terms, despite appearances. It does not, however, actually capture what he intended, as we discuss in detail in the above paper, and as Carnap himself realized in 1930, even before Gödel's incompleteness results later that year.

facts, as in the *Tractatus*. But other signs, not given a definite meaning in advance, may also be added and treated just like atomic sentences, as may "inference rules" governing the transformation of given sentence forms into other sentence forms. All sentences containing the meaningless signs still have a definite meaning, Carnap argues, as they confine the total space of possibilities to certain rows of the truth-table of a complete truth-functional state-description of the world (of the kind envisaged by Wittgenstein). The only requirement of a "logic" so constructed — evidently intended as a preliminary sketch for building a *Grunddisziplin*[6] — is that it not allow inference to any atomic sentence that is not already among the premisses. Axiom systems may then be framed within such a "logic", and all theorems resulting from them can likewise be assigned a definite meaning because they constrain the truth-table of the complete state-description of the world. This is the case even if they contain signs for infinite sets. These, Carnap says, are licensed within his system, though not purely "formalistically" as in Hilbert; they have a definite meaning, even if not a complete one:

> Wenn man nun, um das Unendliche einzuführen, hierzu noch "ideale Aussagen adjungiert" (Hilbert), d.h. Formeln aufstellt, die keine inhaltliche Bedeutung haben, aber die Mathematik des Unendlichen abzuleiten gestatten, so haben wir auch hier wieder die Bedeutung der bedeutungslos eingeführten Zeichen dadurch ermitteln können, daß wir untersuchen, für welche logische Konstanten die Formeln zu Tautologien werden würden. (Carnap, "Über die Konstitution des Nicht-Gegebenen" (1929), ULCA RC/1029/Box 4/CM13/section XX, p. 62)

Unlike Hilbert, Carnap admits no purely formal, uninterpreted signs. Despite this, he calls his idea "radical formalism" because it allows not only logical inferences, but any sort of scientific inference — including, relevantly, inductive inference in empirical science or statistical inference — to be employed as part of a "system of logic" in this way. Though this idea is not thought through, and is in many ways incomplete, it indicates how Carnap was attempting to extend a truth-functional Wittgensteinian language to one usable for

---

[6] Though there is no explicit provision for the quantifiers, Carnap may have intended to develop them axiomatically, as Hilbert and Ackermann (1928, pp. XX and YY) had for both the propositional and the predicate calculus. The terminology of the "New Foundation" coincides with Hilbert and Ackermann, where the quantifiers are introduced by "formal axioms", which are distinguished from the "inhaltliche" (material, contentful) rules of inference — the term also used by Carnap.

mathematics and science, though the kind of solution Carnap was considering saw for mathematics very much the role that Wittgenstein had envisaged for it in the *Tractatus*:

> Der Satz der Mathematik drückt keinen Gedanken aus. Im Leben ist es ja nie der mathematische Satz, den wir brauchen, sondern wir benützen den mathematischen Satz *nur*, um aus Sätzen, welche nicht der Mathematik angehören, auf andere zu schließen, welche gleichfalls nicht der Mathematik angehören. (Wittgenstein 1922, 6.21-6.211)

In the course of 1930, however, this somewhat shaky "New Foundation for Logic" collapsed. Three developments contributed to undermine it. First, the *Gabelbarkeitssatz* fell victim to Gödel's first incompleteness theorem. As Gödel indicated in the discussion following the famous symposium on the philosophy of mathematics in Königsberg in September 1930 (at which Carnap had been the spokesperson for logicism, Heyting for intuitionism, and von Neumann for formalism), there could be true arithmetic sentences that were not provable:

> Man kann (unter Voraussetzung der Widerspruchsfreiheit der klassischen Mathematik) sogar Beispiele für Sätze (und zwar solche von der Art des *Goldbach*schen oder *Fermat*schen) angeben, die zwar inhaltlich richtig, aber im formalen System der klassischen Mathematik unbeweisbar sind. Fügt man daher die Negation eines solchen Satzes zu den Axiomen der klassischen Mathematik hinzu, so erhält man ein widerspruchsfreies System, in dem ein inhaltlich falscher Satz beweisbar ist. (Hahn et al 1931, p. 148)

Second, the incompleness result had an even more fundamentally devastating effect on logicism itself, which the Vienna Circle had relied on to guarantee the tautological (and thus empty) character of mathematics and indeed of all reasoning. As Carnap had expressed it in "Die Alte und die Neue Logik", it was this tautological character of logic and mathematics that undermined the presumption of metaphysicians that conclusions about the real world could be reached by reasoning alone, without factual knowledge. But now it turned out that there could be sentences of arithmetic that, despite the logicist explicit definition of the numbers, were not decidable after all.

Third and finally, we return to the quotation from Gödel with which we began this paper. Wittgenstein had argued that the structure of language "shows itself" but cannot explicitly be represented in language. This was, as we saw, a fundamental barrier to the Vienna Circle's

larger goals, and they sought to overcome it. Now the new work in mathematical logic, especially by Hilbert, Gödel, and Tarski, made extensive use of the distinction between a language and its meta-language. This work appeared to be rigorous, indeed more rigorous than older logical work like Russell's. It thus seemed to represent a clear counterexample to Wittgenstein's argument. Still, there was the difficulty that "elucidations", meta-linguistic sentences of the kind Wittgenstein and the Vienna Circle themselves had used in their writings, seemed impossible to put in the kind of mathematical form that Gödel and Tarski employed. And the problem remained that it seemed impossible to imagine how sentences about language could be expressed as truth-functions of atomic sentences.

II. The Sleepless Night and the Original Syntax Idea

Descartes recorded the content of his famous three dreams, often regarded as the origin of a new chapter in philosophy, in some detail. Carnap's dream, the origin of a more recent chapter, has come down to us in more fragmentary form, and requires some detective work to reconstruct:

> After thinking about these problems for several years, the whole theory of language structure and its possible applications in philosophy came to me like a vision during a sleepless night in January 1931, when I was ill. On the following day, still in bed with a fever, I wrote down my ideas on forty-four pages under the title "Attempt at a Metalogic". These shorthand notes were the first version of my book *Logical Syntax of Language*. (Carnap 1963, p. 53)

Just what *was* Carnap's "vision" during the sleepless night of 21 January 1931? How did it solve the problems that had brought down the conception of the "New Foundation"? And what further consequences did he see in it that made the idea so exciting?

The actual "Attempt at a Metalogic" (*Versuch einer Metalogik*, henceforth *Versuch*) that Carnap wrote down the following day has never been consulted to arrive at an answer to these questions. However, it has been preserved, and provides important evidence.[7] On the surface, it does not in the least resemble the *Logical Syntax* of which it is supposedly the

---

[7] It is available at the Special Collections Dept of Young Research Library, UCLA.

"first version". Only when we compare it with the view Carnap had been struggling toward previously do we understand the magnitude of the change in his view. In the *Versuch*, we find a perspective that is indeed radically different from the Wittgensteinian one of the "New Foundation". Carnap has here adopted the fully formal, "metalogical" point of view of Gödel and Tarski, according to which the logical language is a system of uninterpreted marks rather than meaningful signs. In the perspective of the "New Foundation", the atomic sentences were pictures of atomic facts, which gave them their meaning. In the *Versuch*, an atomic sentence is a finite sequence of superscript dots, followed by the letter "f" with a finite sequence of subscript dots, followed by a left parenthesis, followed by the letter "a" with a finite sequence of subscript dots, followed by a right parenthesis, e.g.:

$$\cdots\!\!f\ldots\ldots(a\ldots)$$

In the "New Foundation", a sentence is a tautology because of what it says, or not, about the world. In the *Versuch*, being a tautology is a property of a string of marks that is defined entirely in terms of its outer form — the type and order of the marks occuring in it. No use is made of the "meaning", "designation", etc. of the marks [*Zeichen*] in defining the central notions of truth-value assignment, consequence, tautology, and the like. Carnap even mentions that the undefined notion "true" might be better to avoid entirely.[8]

From the viewpoint of modern logic, this, this idea may not seem particularly momentous. Even at the time, it represented no technical innovation; Hilbert and others had been writing on formalism in connection with axiomatics for decades, and the methods of Gödel and Tarski were essentially that. But though Carnap's first attempt to formulate his "metalogic" was in terms of a particular formal system, his aim was not merely the mathematical study of a given formal logical system. His new idea was precisely to *apply* the insights of Hilbert, Gödel, and Tarski to the entirety of human knowledge. As we saw above, he had previously accepted Wittgenstein's basic account of the logical language framework in which all science was to be expressed, as the basis for the project of rational reconstruction. In that context the

---

[8] Quotation from document, reference.

"metalogical" perspective of regarding language purely as a system of rules, without reference to anything outside itself, was indeed a revolutionary idea.

Wittgenstein's theory of language form had called attention, for the first time, to the essential role of language in the expression of thought and the acquisition and limitations of knowledge. But like a Kantian form of intuition, it was an inescapable straight-jacket. The very nature of language, in Wittgenstein's view, prevented us from stepping outside it. Before Wittgenstein, language had been regarded as an essentially transparent medium for the expression of thought. The laws of logic, for instance, were considered by Frege and Russell to be laws of thought, judgement, or perhaps nature — but certainly not of language. Wittgenstein had recognized the essential role of language in determining logical (and other kinds of) truth, but had inferred from this a quasi-mystical doctrine of the ineffability of language structure and the inexorability of tautological logical truth. As we have seen, this doctrine (henceforth "Wittgenstein's prison") had implications that were incompatible with the Enlightenment ambitions of the Vienna Circle.

Carnap seems to have realized, though, under the suggestive influence of Hilbert's formalist approach and the technical work of Gödel and Tarski, that one could recognize the essential role of language without being trapped within it. The metalogical methods developed in pursuit of the very mathematical results, like the incompleteness theorem, that had led to the disintegration of his Wittgenstein-based position in the "New Foundation" also showed a way of stepping outside Wittgenstein's prison, and making the structure of language itself the object of logical study.

As opposed to the confinement of all possible knowledge within the absolute constraints imposed by a (naturally or metaphysically) fixed structure of our means of expression, the new recognition that linguistic structure could itself be investigated opened up a whole new method for the unification and clarification of knowledge. Carnap retained Wittgenstein's language-dependence of knowledge, but threw off the shackles of Wittgenstein's prison in favor of the logicians' metalogical perspective.

Armed with this new insight then, and in the rush of enthusiasm that accompanied it, he apparently hoped to be able to solve the other problems that had undermined the "New Foundation", particularly those afflicting logicism. Arithmetic, it was envisaged in the *Versuch*, could evidently somehow be read off from the syntax of the logical object language — as opposed to being expressed in that language.[9] Thus the numbers are not defined as higher-order concepts in the Frege-Russell logicist style, but "purely as figures" [*rein figurell*], on the basis of the dot sequences attached to the symbols. Arithmetical properties and statements then belong to the meta-language. Thus e.g. the commutativity of addition n + m = m + n was supposed to follow from the fact that n-many dots written to the left of m-many dots gives the same series of dots as writing them to the right. The question of the need for mathematical induction *in the meta-language* is considered, but dismissed with some optimism.

If arithmetic was to be formulated in the meta-language of logic, then analysis was to be formulated in its meta-meta-language. For real numbers are properties or series of natural numbers, and properties of them and statements about them properly belong one level up. Carnap may have been guided, in this idea, by Russell's suggestion, in his introduction to the *Tractatus*, that one could perhaps break out of Wittgenstein's prison by using a scheme involving a hierarchy of languages:

> These difficulties suggest to my mind some such possibility as this: that every language has, as Mr. Wittgenstein says, a structure concerning which, *in the language*, nothing can be said, but that there may be another language dealing with the structure of the first language, and having itself a new structure, and that to this hierarchy of languages there may be no limit. (Russell 1922, p. HH)

Having now found the mechanism for such a scheme in the form of "metalogic", applying it to achieve a hierarchy consisting of language, meta-language, meta-meta-language, and so on[10] must have indeed seemed rather compelling, at first sight.

---

[9] Quotation from document, with reference.
[10] Quotation from last page of Versuch, w. ref.

Carnap says in his autobiographical account that not only "the whole theory of language structure" came to him like a vision, but also "its possible applications in philosophy".[11] These were spelled out later that year in the paper "Die physikalische Sprache als Universalsprache der Wissenschaft", which was later published in English, with a new preface, as the pamphlet *Unity of Science*. This paper is mainly known for its advocacy of physicalism, and is thus taken to represent a watershed in Carnap's epistemological views from the phenomenalism of the *Aufbau* to a Neurath-inspired physicalism. This epistemological aspect is certainly present in the paper, and reflected in its title. But the new syntactical doctrine is equally in evidence and, indeed, motivates the paper's physicalistic conclusions. After three pages of introductory discussion about the idea that all objects and facts are of a single kind, we are told that these expressions are a concession to the customary "material" [*inhaltliche*] way of speaking. The "correct" way, Carnap says, speaks of words rather than "objects" and sentences rather than "facts", for a philosophical investigation is an analysis of *language*. In a footnote he indicates that a comprehensive, strictly formal theory of language forms, which he calls "metalogic", will soon be forthcoming, and will justify the "thesis of metalogic" here invoked, that "meaningful" [*sinnvolle*] philosophical sentences are the metalogical ones, i.e. those that speak only of the form of language.

This represents a radically different basis for the critique of metaphysics from the one Carnap had previously adopted from Wittgenstein, whereby meaningful sentences were those that derived their meaning from atomic sentences by truth-functional combinations. Atomic sentences, as pictures of atomic facts, no longer play any role in distinguishing meaningful from meaningless sentences. The new metalogical or syntactic viewpoint is significant, as Eino Kaila agreed after discussion with Carnap a few months later, because of the "Ausschaltung der Verifikation durch Vergleich mit Sachverhalten." (ASP [diary]) As

---

[11] In the Vienna Circle, he says, "the philosophical problems in which we were interested ended up with problems of the logical analysis of language," and since "in our view the issue in philosophical problems concerned the language, not the world", the Circle thought that "these problems should be formulated not in the object language but in the metalanguage." It was therefore "the *chief motivation* for my development of the syntactical method," (our emphasis) to develop a "suitable metalanguage" that would "essentially contribute toward greater clarity in the formulation of philosophical problems and greater fruitfulness in their discussions." (Carnap 1963, p. 55)

Carnap explained in "Die physikalische Sprache," not only criterial definitions but also ostensive definitions can be regarded as intra-linguistic.[12]

So Carnap had comprehensively and definitively turned his back on the picture theory of the *Tractatus* — and thus also on its foundationalism. Meaning was no longer built up from some basic (metaphysically unavoidable or naturally occurring) components, but was determined by rules. Rules can be determined by humans. The upshot of Carnap's dream, then, was a liberation from the manacles of a fixed structure imposed on the human mind by natural or metaphysical factors beyond human control. January 1931 was the turning point in Carnap's career where the voluntarism to which he had already tended previously[13] could finally find its proper scope and expression. With respect to Wittgenstein's prison, this was literally an overnight transformation from slave to master. But for Carnap, the logical and philosophical thought processes were closely entwined, and obstacles remained to be overcome on the logical front.

III. *Metalogik*: the First Draft of *Logical Syntax*

> The view that the terms "analytic" and "contradictory" are purely formal and that analytic sentences have null content was stated by Weyl ... Later, Wittgenstein made the same view the basis of his whole philosophy. "It is the characteristic mark of logical sentences that one can perceive from the symbol alone that they are true; and this fact contains in itself the whole philosophy of logic." ... "And so also it is one of the most important facts that the truth or falshood of non-logical sentences can *not* be recognized from the sentences alone." This statement, expressive of Wittgenstein's absolutist conception of language, which leaves out the conventional factor in language-construction, is not correct. (Carnap 1934, p.[186])

In rejecting Wittgenstein's absolutism in favor of the logicians' conception of metalogic, the problem of elucidations was solved, but the excited solution to the problems of logicism suggested by the new metalogical standpoint turned out not to work. The rather odd idea that arithmetic could be read off from the meta-language of logic in a sense turned out to be *too*

---

[12] "Elephant", for instance, criterially defined as an an animal with certain characteristics, might be ostensively defined as "an animal of the kind present at a certain space-time location" (Carnap 1932, p. 435-436).
[13] Jeffrey (1992), etc. Different possible bases in *Aufbau*, different philosophical "languages in early 1920's in autobiography, etc.

correct, in that some essential meta-logical concepts (notably *provablity*) required for their formulation a combinatorial theory that was every bit as complicated as arithmetic itself. Thus in the late spring of 1931, Carnap decided to move to a conventional axiomatic arithmetic in the *object* language, so that the axiomatized arithmetic could then be used to express the metalanguage, using Gödel's method of arithmetization (ibid., p. [50-1]) This move had the further advantage of collapsing the entire hierarchy of languages and meta-languages into itself, at least in principle, by iterating Gödel's method of arithmetizing the metalanguage in the object language. Thus it appeared (for a time at least) that one could now get by with only a single language after all.[14]

However well this seemed to work, there was a price to be paid for it. For the very thing that had made the "metalogical" solution possible — i.e. the precise definablity of the central metalogical notions and their expressibility in the object language — was also responsible for the essential incompleteness of the logical treatment of mathematics. If only rules were to establish meaning, and there were no intrinsic constraints on the sorts of formal properties of formulas that could be considered, then there could be a criterion for mathematical truth *different* from mere provability. And as Gödel had shown, provability was insufficient; there were "true" arithmetical statements not derivable from the axioms; this made such a criterion essential. Carnap seems to have developed such a criterion sometime in the latter part of 1931, in the form of the notion of *analyticity*. This was to be a stronger sort of logical truth than provability in a formal system, but was still to be determined strictly in terms of the formal character of the symbols.

Analyticity was apparently to take the place of provability as the generalized notion of tautology or logical truth. To understand how this was intended, consider the analogy of a chess game. Think of the starting position of the pieces as the axioms, the permitted moves as the rules of inference, and a sequence of moves ending in checkmate as a proof of a theorem. There are configurations of pieces on the board that constitute checkmate, but cannot be reached from the starting position by any sequence of permitted rules. Such a

---

[14] Brief discussion here of the Referate, the discussion of them in the circle, main points raised, absence of analyticity def. (Also refer to Spanish article.)

configuration represents an analytic sentence that cannot be proved. In this way, the definition of analytic sentence can be phrased entirely formally, in accordance with all the same rules of inference, and yet still be wider than provability. Thus the absolute, Wittgensteinian conception of tautology could be saved, and indeed finally extended beyond propositional logic in accordance with the Vienna Circle's original ambitions.

Such a notion of analyticity was apparently defined in the first draft of the *Logical Syntax*, entitled *Metalogik*, of which nothing has been preserved (as far as we know) but its table of contents. This lists the notion *analytic* alongside *synthetic* and *contradictory* under the heading "IV.B. Theory of content of fomulas" (corresponding roughly to IV.B(a) of *Logical Syntax*, which — in the English translation — gives the general definition of "analytic"). This is followed in section IV.C by a discussion of soundness, consistency, and completeness, including sections on the "antinomies" and "the incompleteness of all formal systems" which appear to correspond closely to IV.B(c) of the (English) *Logical Syntax*, where the Gödel incompleteness of arithmetic is discussed.

We don't know exactly how analyticity was originally defined, but from the evidence available it is clear that the definition was defective. As we shall explain presently in more detail, Gödel objected to its application to the "extended model language".[15] And furthermore, he points out, it will be *impossible* to give a correct definition of it in *any* meta-language that can be faithfully represented in the object language, e.g. by arithmetization. This fact has since become known as Tarski's theorem on the indefinability of truth. Thus it turns out that Carnap's single language approach will not work after all.

Gödel's objection to Carnap's original definition of analyticity is explained in a letter dated 11 September 1932 (Gödel 2003, pp. XX-YY). Carnap had apparently tried to define the

---

[15] From the table of contents (ASP/RC XXX) it seems clear that a single language (corresponding to the later language I) was developed as the "model language" [*Modellsprache*]. (In "Die physikalische Sprache", it had been called the "system language" [*Systemsprache*]. Just as in the "Referate" (see above, footnote X) held just before Carnap embarked on composing the first draft, it seems that the "model language" was regarded as the "proper language" [*eigentliche Sprache*], while the full resources of classical mathematics could be developed by using the "model language" as a meta-language for axiomatic formal systems, Hilbert-style; the model language together with these axiomatic extensions was then called the "extended model language".

notion "analytic sentence" inductively, using what we would now call a substitutional treatment of quantification. Thus e.g. given an arithmetical sentence of the form $(\forall x)f(x)$, with quantification over the numerical variable x and $f(x)$ a formula with at most x free, one could reasonably define:

$(\forall x)f(x)$ is analytic $\quad\Leftrightarrow\quad$ $f(a)$ is analytic for all numerical constants a

In his definition, Carnap had apparently tried to use the same for strategy for higher-order quantifiers, for example over all properties or sets, as in $(\forall X)f(X)$. Thus e.g. for $f(X)$ of the simple form $X(0)$ one would have:

$(\forall X)X(0)$ is analytic $\quad\Leftrightarrow\quad$ $A(0)$ is analytic for all predicate constants A

But here there is no restriction on what predicate constant $A(x)$ are to be substituted for X in testing for analyticity, so among the substitution instances is e.g. the predicate $(\forall X)X(x)$ itself. Thus the definition is circular, and so it does not succeed in specifying the desired notion. The problem here is in the so-called impredicativity of the higher-order quantifier. One could restrict the substitutions to predicates of lower "order", in a suitably defined sense, and this would result in a workable scheme, but it would only provide a definition for a system like ramified type theory, which is inadequate for classical mathematics.

In his letter, Gödel suggests instead using a notion of "all sets and relations whatever" [*alle Mengen und Relationen überhaupt*] in place of "all predicates". An interesting footnote indicates that this need not be interpreted as Platonism, as he only suggests formulating the definition of "analytic" in a particular language, in which the concepts of "set" and "relation" are already defined. He goes on to say that he intends to use this idea to give a truth definition in Part II of his paper (presumably the missing sequel to Gödel 1931). And, moreover, that he believes it can not be done otherwise, and that the higher functional calculus can*not* be treated "semantically" (i.e. according to Carnap's strictly formal conception of metalogic).

In his first reply, a desperate Carnap attempts to reconstruct Gödel's proposal — the difficulty lies in the idea of "all values" for a predicate of the object language L. How is this to be understood, even with respect to another language L' in which the values are to be taken? It will not suffice to use only the predicates definable in L'; one apparently needs instead all "arbitrary" ones. And this latter notion strikes him as rather questionable [*ziemlich bedenklich*]. He finally asks for help in finding the right definition, especially since, as he says, everything else in his book depends on it.

Judging from his note of a few days later, Carnap finally did work out the solution for himself. He realized that the notion of "all values" of a predicate could be rendered in the formal meta-language L' simply by using a universal quantifier $(\forall X)...X...$. The key new idea here is that the language L' in which the values are taken needs to be stronger than the one for which they are given. In his (delayed) reply, Gödel confirms that this is the idea, and remarks that one cannot give the definition of "analytic" in the same language, otherwise "contradictions will result". He also points out that, presumably in the meantime, Tarski has already published a "similar" definition of "analytic"[16], which seems likely to be the reason Gödel never worked out his own part II.

For Carnap, ultimately, the resulting definition of "analytic" — which had previously been so crucial — was not even deemed important enough to include in the first edition of the book; it was omitted "for reasons of space". The problem with it was that, as hinted by Gödel in the footnote about Platonism, the notion of analyticity it defined was not absolute, but rather in a certain sense, conventional. It gave a notion of "analytic in L", but only with respect to *another* language L', used for the interpretation of L. There might be a natural or conventional choice for L' — type theory of the next higher type, or axiomatic set theory — but it could hardly be claimed that any particular such choice is the *correct* notion of analytic for a given language. This *language relativity* of the central notions of metalogic turned out to be more important to Carnap than the particular metalogical definitions themselves. And this brings us to the final step in the story we have been telling.

---

[16] Presumably he refers here to Tarski 1932, which . . .

IV. Tolerance

The first public signal that Carnap's thought had entered yet another radically new phase was the discussion contribution "Über Protokollsätze", a response to Neurath's paper (entitled "Protokollsätze") that had in turn responded to Carnap's "Die physikalische Sprache". This discussion contribution was written within a month or two after the above correspondence with Gödel, and Carnap is a changed man. A new tone has suddenly entered his writing, one he would stick with from then on, and that would become deeply characteristic: "In my view the issue here is not between two conceptions that contradict each other, but rather between *two methods for constructing the language of science, which are both possible and justified.*" (Carnap 1932, p. 215) And he spells out the grounds of this new pluralism:

> Not only the question whether the protocol sentences are inside or outside the syntax language, but also the further question regarding their precise specification, is to be answered, it seems to me, not by an assertion, but by a stipulation [*Festseztung*]. Though I earlier [in "Die physikalische Sprache"] left this question open . . . I now think that the different answers are not contradictory. They are to be taken as proposals for stipulations [*Vorschläge zu Festsetzungen*]; the task is to investigate these different possible stipulations as to their consequences and assess their usefulness. (ibid., p. 216)

To the best of our knowledge there is no record, either in Carnap's autobiography or in the form of a letter or note (or even a diary entry), recording the moment at which he embarked on this new direction. But the sense of discovery and enthusiasm is palpable in "Über Protokollsätze"; Carnap repeats the new message again and again. And he is very much aware that it represents an even more radical departure from his and the Vienna Circle's previous position:

> In all theories of knowledge to date there is a certain *absolutism*: in the realistic theories an absolutism of objects, in the idealistic ones (including phenomenology) an absolutism of the "given", of "experiences", of "immediate phenomena [*unmittelbare Phänomene*]". Even in positivism we find this residual idealistic absolutism; in the logical positivism of our circle -- in the works on the logic of science (epistemology) published to date by Wittgenstein, Schlick, Carnap -- it takes the more subtle form of an absolutism of primitive propositions ("elementary propositions", "atomic propositions"). (Carnap 1932, p. 228)

This sense of breakthrough is equally evident in the passages evincing the new "principle of tolerance" in the *Logical Syntax* itself. The preface to that book is every bit as messianic as the justly renowned preface to the *Aufbau*:

> Der Kreis der möglichen Sprachformen und damit der verschiedenen möglichen Logiksysteme ist . . . unvergleichlich viel größer als der sehr enge Kreis, in dem man sich In den bisherigen Untersuchungen der modernen Logik bewegt hat. Bisher ist man von der schon klassisch gewordenen Sprachform, die *Russell* gegeben hat, nur hin und wieder in einigen Punkten abgewichen. . . Der Grund dafür, daß man sich bisher nicht weiter von der klassischen Form zu entfernen wagt, liegt wohl in der weit verbreiteten Auffassung, man müsse die Abwiechungen "rechtfertigen", d.h. nachweisen, daß die neue Sprachform "richtig" sei, die "wahre Logik" wiedergebe. Diese Auffassung und die aus ihr entspringenden Scheinfragen und müßigen Streitigkeiten auszuschalten, ist eine der Hauptaufgaben dieses Buches.
> (Carnap 1934, p. v)

The first attempts to escape from the "classical" forms — which themselves only went back one or two generations! — were certainly daring, he says. "Aber sie waren gehemmt durch das Streben nach 'Richtigkeit'." And Carnap concludes with the famous words: "Nun aber ist die Hemmung überwunden; vor uns liegt der offene Ozean der freien Möglichkeiten." (ibid., p. vi)

The principle is stated, in the text of the *Logical Syntax* itself, in the context not of epistemology, as in its first application, but of philosophies of mathematics, particularly intuitionism. It is expressed as the exhortation to state meta-theoretic or *wissenschaftslogische* proposals in precise terms, as explicit rules or definitions, within the formation or transformation rules of a precisely defined language or calculus:

> Hat man sich einmal klar gemacht, daß alle pro- und kontra-intuitionistischen Erörterungen von der Form eines Kalküls handeln, so wird man die Frage nicht mehr in der Form Stellen "Wie *ist* das und das?" sondern "Wie *wollen* wir das in der aufzubauenden Sprache einrichten?". . . Damit verschwindet die dogmatische Einstellung, durch die die Diskussion häufig unfruchtbar wird. (Carnap 1934, p. 42)

This "dogmatische Einstellung" results, in Carnap's view, from the reliance on inherently vague philosophical "Erörterungen" rather than on precise statements of definitions and rules.

He indicates how he has tried, in Language I of the *Syntax*, to capture the philosophical concerns (expressed in various gradations of finitism or constructivism) voiced by Brouwer, Kaufmann, Wittgenstein, and others. But, he points out, there is no way of telling whether he has expressed *precisely* what they have in mind, as they have not expressed their views as proposed precise definitions and rules, but only in terms of vague "Erörterungen" that leave many specific questions open, when one gets down to the brass tacks of constructing an actual language. (ibid., p. 44) Or they impose restrictions and requirements that appear to be normative.

Carnap's most general statement of the principle of tolerance, therefore, addresses these tendencies directly, contrasting them with his own program of precise and explicit rules:

> Unsere Einstellung zu Forderungen dieser Art sei allgemein formuliert durch das *Toleranzprinzip: wir wollen nicht Verbote aufstellen, sondern Festsetzungen treffen. . .In der Logik gibt es keine Moral.* Jeder mag seine Logik, d.h. seine Sprachform, aufbauen wie er will. Nur muß er, wenn er mit uns diskutieren will, deutlich angeben, wie er es machen will, syntaktische Bestimmungen geben anstatt philosophischer Erörterungen. (ibid., p. 45)

Only by *replacing* the vague concept with a precise equivalent can the practical merits or drawbacks of a proposal be judged, for some defined purpose. And under the new regime of pluralism, where there can be no criterion of inherent "correctness", practical usefulness is the only criterion left for deciding whether a proposal should be pursued or left aside. The principle of tolerance fits well, then, into the project of "rational reconstruction" pursued by the earlier Vienna Circle, and sets the stage for the successor project of "explication", which Carnap would not formulate explicitly until the 1940's.[17] And he is careful to apply the insistence on precision to his own work as well. Attention and criticism should be focussed, he repeatedly insists, not on the "inexact" informal reflections in the text, but on the precise definitions given in terms of the proposed calculi. (refs.)

---

[17] The classical exposition of this project is in Ch. 1 of *Logical Foundations of Probability* (Carnap 1950a); for further discussion, see Stein 1992, Awodey and Carus 2003, section III, and Carus 2003, section II.

Unlike previous revolutions in Carnap's thought, this one was permanent; it became the basis of his thought for the remainder of his career. And it extended, as we have seen, far beyond logic itself into epistemology and such questions as the form of the observation language. It represents the final step away from the meaning foundationalism of the *Tractatus* to a kind of radical pragmatism, in which the only criterion for acceptance or rejection of a language form is its usefulness for a particular purpose. This conception was only partially spelled out in the *Logical Syntax* itself. Indeed, it was never fully articulated in detail by Carnap, but only indicated very generally in such later writings as "Empiricism, Semantics, and Ontology" and certain replies to critics in the Schilpp volume.

V. Semantics

To what extent was the breakthrough of the sleepless night, the original syntax idea, conceptually *necessary* for the subsequent move to tolerance, and to what extent are the two are *separable*? This question is of particular interest in the context of present discussions of the *Logical Syntax*[18], in which it is often argued that tolerance depends heavily on the syntax idea, and that the abandonment of the original syntax view significantly restricted the principle of tolerance, or made its limitations evident.[19]

The above discussion suggests an answer to this question. We saw in section II that the original syntax idea represented, above all, a rejection of *meaning* in Wittgenstein's sense. Meaning, in this "absolutist" view, as Carnap later called it, retained a certain arbitrariness or obscurity. It seemed obvious informally that some "meaningful" sentences (e.g. Newton's laws) failed to meet Wittgenstein's criteria. And what authority did those criteria claim? Only that of philosophical arguments, not that of precisely specified concepts embedded in a language framework defined by explicit rules. Before January 1931, Carnap and the Vienna Circle had been "in the grip of a picture", the picture of language deriving its meaning by truth-functional concatenation of atomic sentences representing atomic facts. But then

---

[18] In which Gödel himself has, once again, played a posthumous role through the publication in 1995 of his critique of Carnap's view (Gödel *1953/9); see Goldfarb (1995), Ricketts (1994), Friedman (1999), and our discussions (2003a, 2003b) of Gödel's argument.
[19] Goldfarb and Ricketts (1994); Ricketts (1996); Friedman (1999, ch. 9).

Carnap discovered that he could retain Wittgenstein's language-relativity of knowledge without this picture, by extending a Hilbertian or Tarskian formalist view from logic and mathematics to the whole of knowledge.

It seemed entirely reasonable, at that point, to conclude that it was the Wittgensteinian theory of "meaning" that had blocked the way to this insight. The response, accordingly, was a complete proscription of meaning: *nothing* extra-linguistic could constrain the acceptability of a precise meta-language for the use of *Wissenschaftslogik*. The shift away from the Wittgensteinian view had meant a corresponding shift from trying to *incorporate* "philosophical" (*wissenschaftslogische*, "elucidatory") discourse into the language of science itself to the construction of a precise *meta*-language for such discourse. This was a drastic change of perspective. The criterion of empiricism, for instance, now had to be reformulated as a constraint on the scientific object language (i.e. as a set of formation and transformation rules in the meta-language, an *internal* constraint, from "above") rather than a requirement of *meaning* (in terms of conditions for verification, an *external* constraint, from "below").[20] The critique of metaphysics, as we saw in section II, was similarly re-expressed, as a proscription of meaning in the meta-language.

But the new syntactic view led Carnap, as we saw in section III, directly to the principle of tolerance. The syntactic view required a definition of what it means to "result from the rules", and Gödel had shown that the traditional, intuitive definition — provability — did not suffice. A separate and richer meta-language was required for this definition, and there is no reason to single out any particular such meta-language as "correct". Returning to the chess analogy, there is no reason to prefer a statement of the conditions for "checkmate" in French to one in German — or to prefer one in physicalistic terms (as relative positions of certain carved figures of a certain color on a board of a certain pattern, etc.) to an abstract one in terms of procedural rules. The latter makes it possible to state chess as an algorithm, to

---

[20] There was an attempt, during this transitional period, to formulate as many constraints as possible internally in this way. Neurath (though never Carnap) occasionally was tempted to express this view in ways that bordered on a coherence theory. But at no point did this mean an abandonment of empiricism, as Russell (and later Coffa) suggested (Carus 1999, p. XX); it meant rather that the constraint of empiricism in the object language (the language of science) was to be expressed as a set of formation and transformation rules in the (*wissenschaftslogische*) meta-language.

discuss strategies for achieving checkmate, and so on, but these are mere practical advantages; they do not make the procedural language "correct" in any sense. And once this pluralism of meta-languages is admitted, there is no reason not to extend it to object languages. Why not change the rules specifying the object language as well, to optimize them for whatever practical purpose we developed the object language for? In chess, this might mean altering the rules to produce the longest possible games (in the limit, we could specify rules that make checkmate impossible — just as Carnap pointed out that we could of course construct languages that have no analytic sentences), or the "most interesting" possible games — opinions would diverge here how to define that further.

But this new pluralism of the principle of tolerance no longer supported the proscription of meaning. Under the new pluralism, the criteria for considering a concept or language are simply: (i) specifiability by explicit rules, and (ii) practical usefulness. Under (i), the semantic definitions of "designation" and "truth" qualify (whether they qualify under (ii) cannot be decided until various possible semantic languages have been tried out and applied to the problems we want to solve[21]). The proposal in *Meaning and Necessity* to employ these definitions in the "method of extension and intension" (Carnap 1954) can be regarded as an *explication* of — a replacement for — the previously vague, somewhat obscure and arbitrary, conception of "meaning".

This answers the question about the extent to which tolerance depends on a specifically "syntactic" approach. Tolerance depends on two components of the original insight of the sleepless night: (a) the distinction between a language (a calculus, a purely syntactic symbol system) and its interpretation; (b) the requirement that a language be entirely specified by explicit rules. These two components survive unscathed and undiminished into the semantic period. (So it is rather misleading to call them "syntactic"; Carnap's original term "metalogical" might be more appropriate.) What does not survive is the overreaction against "meaning" that accompanied the original insight. In distinguishing between a language and its interpretation, Carnap's first (and, as we saw, understandable) response was to restrict

---

[21] Carnap never claimed to have arrived at the definitive semantics; he presented his formulation as a first attempt, to get things started and as a basis for discussion: ". . ." (Carnap 1942)

interpretation to the object language (and there to one particular — physicalistic — interpretation), and dispense with it entirely in the elucidatory (*wissenschaftslogische*) meta-language. But this restriction was loosened when he saw that interpretation could be specified by explicit rules (governing satisfaction, designation, and truth), in accordance with requirement (b).

The original rejection of "meaning" had proscribed what seemed an occult property, just like the rejection by Lavoisier of the traditional explanation of burning as the *release* of a substance ("phlogiston", in Stahl's theory) into the surrounding air. The reinstatement of an explicated form of meaning reflected the realization that the informal idea of meaning had not *itself* been the culprit, but rather a particular, somewhat obscure conception of it. But the new explication of meaning met the standards by which the previous conception had been rejected. In the same way, the later reinstatement of the idea that burning (oxidation) involved the *release* of electrons by the substance being oxidized met the standards of the post-Lavoisier principle of the conservation of matter, by which all reactions are regarded as recombinations of indestructible atoms. The new explication of the informal concept of "meaning" has no more in common with the previous occult property than electrons do with phlogiston.

VI. An Application to a Particular Case of Remarkable Misunderstanding

This fundamental difference between the pre-1931 conception and Carnap's new explication of the concept of meaning was apparently not recognized by Quine, who in later years continued to regard the explicandum itself with suspicion.[22] If we look at the grounds of Quine's well-known critique in the light of the above development, though, we see more clearly that it is not really addressed to the theoretical concepts it purports to discuss, like analyticity and intensionality. It is addressed, rather, to the very foundation, to the principle of tolerance itself, and particularly to the qualifying requirement (b): the requirement that a

---

[22] The "theory of meaning" contrasted with "theory of reference" (Carnap 1955, p. 233). Analyticity, the truth of a statement based on the meanings of its words alone, thus came under attack as a special case of the attack on meaning.

language or concept be defined by explicit rules. For it emerges that Quine's rejection of "meaning" in Carnap's explications is based on a wholesale rejection of definitions and semantic rules. Definition, he says, "hinges on prior relations of synonymy" deriving from usage (Quine 1950, p. 27) and is thus a "will o' the wisp" (ibid., p. 32). He specifically includes Carnap's project of explication in this sweeping write-off (ibid., p. 25). An explicative (more precise) replacement for a vague explicandum must, he says, be synonymous with some subset of the explicandum's semantic field, and "such a definition still owes its explicative function ... to pre-existing synonymies." (ibid.) This not only misses Carnap's point in requiring that a language or concept be defined by explicit rules, but absurdly inflates the role of common usage in scientific progress. Does the explicative function of the electron really rest on its synonomy to phlogiston? Of course not; it is a mere coincidence that we can now explicate the informal notion of burning as substance release by putting electrons in the role that an earlier chemistry gave to a substance with occult properties. It is precisely by first ridding ourselves of such occult properties — by Lavoisier's overcoming of the intuitive notion of "substance release" — that progress was made. Carnap intended explication as a generalization of such scientific progress, in which the more precise concepts do not accommodate but *replace* the vaguer ones.

Quine seems on some occasions to have realized that the difference between himself and Carnap was not of a "cognitive" nature, and that it concerned more fundamental views about how to go about constructing the language of science.[23] But he seems nonetheless to have fallen victim to his own belief in an underlying continuity or partial synonymy between explicandum and explicatum. For it appears that the real target of his critique was not Carnap's *new*, post-1935, "pluralist" explication of "meaning", but the *previous*, pre-1931 "foundationalist" conception of meaning described in section I above. This is most clearly

---

[23] According to Howard Stein's recollection of the discussion after a colloquium talk Quine gave in Chicago in 1951, Carnap summarized the differences between himself and Quine (a summary with which Quine, at the time, concurred) approximately as follows: "Quine. . . and I really differ, not concerning any matter of fact, nor any question with cognitive content, but rather in our respective estimates of the most fruitful course for science to follow. Quine is impressed by the continuity between scientific thought and that of daily life — between scientific language and the language of ordinary discourse — and sees no philosophical gain, no gain either in clarity or fruitfulness, in the construction of distinct formalized languages for science. I concede the continuity, but, on the contrary, believe that very important gains in clarity and fruitfulness are to be had from the introduction of such formally constructed languages. . ." (Stein 1992, p. 279).

evident in the section V of "Two Dogmas", where the scene shifts, and we see the first dogma (that analytic sentences are distinguishable from synthetic ones) in terms of the second, that of reductionism ("the belief that each meaningful statement is equivalent to some logical construct upon terms which refer to immediate experience"; ibid. p. 20). Indeed, Quine specifically states that the "two dogmas are ... at root identical" (p. 41). The critique of meaning amounts, then, more specifically to a critique of the "verification theory of meaning", according to which "the meaning of a statement is the method of empirically confirming or infirming it" (ibid., p. 37). It is *this* conception of "meaning" which Quine evidently regards as lurking behind Carnap's position. A continuity (partial, pre-existent synonymy) of intended interpretation seems to be assumed. Quine takes Carnap's new explication of "meaning", in other words, to be still essentially wed to the foundationalism of the *Tractatus*.

But we have seen in section II that it was precisely this absolute, reductionist conception of "meaning" that Carnap left definitively *behind* after the sleepless night in 1931. In fact, as we also saw, he at first *overreacted* to this earlier conception by declaring *all* meaning off-limits — just as post-Lavoisier chemistry at first declared all notions of substance release during oxidation off-limits. Indeed, it was the "Ausschaltung der Verifikation durch Vergleich mit Sachverhalten" that Carnap stressed in 1931 as the main point of the new "metalogical" point of view. Only a dogmatic insistence that there must be a continuity of partial synonymy between proposed explicata for a single explicandum could arrive at the view that the semantic theory of Carnap's later works has any connection with reductionism to "the given".

The idea Carnap arrived at during the night of 21 January 1931 — the substance of Carnap's Dream, we might say — was no temporary aberration. Carnap did not abandon meaning and then return to it.[24] He temporarily abandoned an informal concept because his criteria for successful explication at first seemed impossible to meet in that case; immediately after the Dream, he thought the informal notion of "meaning" terminally misleading, and wanted it

---

[24] As Coffa (1991, p. XX) claims, for instance.

retired. Later, he saw it could be explicated according to the standards required for consideration under the principle of tolerance, so he relented. But his re-admission of the informal *explicandum* "meaning" was widely misunderstood as a continuity between the new explication and the old conception — which, it should now be clear, are as distinct from each other as the theory of ion exchange and that of phlogiston.[25]

References

Awodey, S. and A.W. Carus 2001 "Carnap, Completeness, and Categoricity: The *Gabelbarkeitssatz* of 1928", *Erkenntnis* 54, pp. 145-172.

Awodey, S. and A.W. Carus 2003 "How Carnap Could Have Replied to Gödel", in S. Awodey and C. Klein, eds. *Carnap Brought Home* (LaSalle, IL: Open Court).

Carnap, R. 1927 "Eigentliche und uneigentliche Begriffe", *Symposion* 1, pp. 355-74.

Carnap, R. 1932a "Die physikalische Sprache als Universalsprache der Wissenschaft", *Erkenntnis* 2, pp. 432-465.

Carnap, R. 1932b "Über Protokollsätze", *Erkenntnis* 3, pp. 215-228.

Carnap, R. 1934 *Logische Syntax der Sprache* (Vienna: Springer).

Carnap, R. 1955 "Meaning and Synonymy in Natural Languages", repr. in R. Carnap *Meaning and Necessity: A Study in Semantics and Modal Logic* (Chicago: University of Chicago Press, 2nd edition 1956), pp. 233-247.

Carnap, R. 1963 "Carnap's Intellectual Autobiography" and "The Philosopher Replies", in P. Schilpp, ed. *The Philosophy of Rudolf Carnap* (LaSalle, IL: Open Court), pp. 3-84, 859-1013.

Carus, A.W. 1999 "Carnap, Syntax, and Truth", in J. Peregrin, ed. *Truth and its Nature (If Any)* (Dordrecht: Kluwer), pp. 15-35.

Coffa, A. 1991 *The Semantic Tradition from Kant to Carnap: To the Vienna Station* (Cambridge: Cambridge University Press).

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

Friedman, M. 2000 *A Parting of the Ways: Carnap, Cassirer, Heidegger* (LaSalle, IL: Open Court).

Gödel, K. *1953/9, version III "Is Mathematics Syntax of Language?", in his *Collected Works*, vol. 3, pp. 334-355.

Goldfarb, W. and T. Ricketts 1992 "Carnap and the Philosophy of Mathematics", in D. Bell and W. Vossenkuhl, eds. *Science and Subjectivity; The Vienna Circle and Twentieth Century Philosophy* (Berlin: Akademie-Verlag), pp. 61-78.

Hahn, H. *et al* 1931 "Diskussion zur Grundlegung der Mathematik", *Erkenntnis* 2, pp. 135-49.

Jeffrey, R. 1994 "Carnap's Voluntarism", in D. Prawitz, B. Skyrms, and D. Westerståhl, eds. *Logic, Methodology, and Philosophy of Science IX* (Amsterdam: Elsevier), pp. 847-866.

Quine, W.V.O. 1951 "Two Dogmas of Empiricism", repr. in Quine *From a Logical Point of View* (New York: Harper 1953), pp. 20-46.

Ricketts, T. 1994 "Carnap's Principle of Tolerance, Empiricism, and Conventionalism", in P. Clark and B. Hale, eds., *Reading Putnam* (Oxford: Blackwell), pp. 176-200

Ricketts, T. 1996 "Carnap: From Logical Syntax to Semantics", in R. Giere and A. Richardson, eds., *Origins of Logical Empiricism* (Minneapolis: University of Minnesota Press), pp. 231-250.

Stein, H. 1992 "Was Carnap Entirely Wrong, After All", *Synthese* 93, pp. 275-95.