ORIGINAL PAPER

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

S. Awodey · A. W. Carus

Abstract   In Carnap's autobiography, he tells the story how one night in January 1931, "the whole theory of language structure" in all its ramifications "came to [him] like a vision". The shorthand manuscript he produced immediately thereafter, he says, "was the first version" of *Logical Syntax of Language*. This document, which has never been examined since Carnap's death, turns out not to resemble *Logical Syntax* at all, at least on the surface. Wherein, then, did the momentous insight of 21 January 1931 consist? We seek to answer this question by placing Carnap's shorthand manuscript in the context of his previous efforts to accommodate scientific theories and meta-linguistic claims within Wittgenstein's *Tractatus* theory of meaning. The breakthrough of January 1931 consists, from this viewpoint, in the rejection of the *Tractatus* theory in favor of the meta-mathematical perspective of Hilbert, Gödel, and Tarski. This was not yet the standpoint of the published *Logical Syntax*, as we show, but led naturally to the "principle of tolerance" and thus to Carnap's mature philosophy, in which the inconsistencies between this first view and the principle of tolerance, which survived into the published *Syntax*, were overcome.

Keywords   Rudolf Carnap · Kurt Gödel · Ludwig Wittgenstein · Logical Syntax of Language · Vienna Circle

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

S. Awodey (✉)
Philosophy Department, Carnegie Mellon University, 5000 Forbes Ave.,
Pittsburgh, PA 15208, USA
e-mail: awodey@cmu.edu

A. W. Carus
Department of Philosophy, University of Chicago, 1115 E. 58th str.,
Chicago, IL 60637, USA
e-mail: awcarus@mac.com

more recent chapter,[1] 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? What problems did it solve? And what were the "possible applications to philosophy" 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 apparently never been consulted to answer to these questions. However, it has been preserved, and provides important evidence.[2] On the surface, it does not in the least resemble the *Logical Syntax* of which it is supposedly the "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 conception. It is on reconstructing this pre-1931 view, accordingly, that we first concentrate.

An obstacle that has made this view difficult to appraise, or even state, is the large role played in it by certain ideas from Wittgenstein's *Tractatus*. From the beginning, this book has attracted more interpretive interest than has the Vienna Circle, whose understanding of the book has therefore—we think somewhat unfairly—been classified as either a pale shadow or as an outright misunderstanding of Wittgenstein's doctrines. We think this traditional view[3] misleading because it takes for granted that the Vienna Circle's first priority was a correct understanding of Wittgenstein's own intentions. In fact, of course, the Vienna Circle, including Carnap, had been led to something like Wittgenstein's ideas from very different starting points, and they used whatever came to hand to address *their* problems, which were not Wittgenstein's. But in arriving at a provisional solution to those problems in the 1920s, certain

---

[1] The *Logical Syntax* is usually regarded as the origin of the "linguistic turn" in analytic philosophy; see e.g. the editor's introduction in Rorty's collection *The Linguistic Turn* (Rorty, 1967).

[2] Of the original shorthand document, pp. 1–23 and 44 are preserved in the Carnap Papers (Manuscript Collection No. 1029) held by the Special Collections Department, Young Research Library, University of California at Los Angeles (henceforth abbreviated "UCLA/ RC1029"), Box 4, folder CM14, item 1.

[3] Which now takes the updated form of claiming that the *Tractatus* has been wrongly interpreted because it has been mistakenly *assimilated* to Carnap's different conception. James Conant, for instance, 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 view of Wittgenstein has been obscured, Conant says, by its assimilation to Carnap's very different 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 rather than with the mental states or intentions of speakers; Hacker (2003) thinks Carnap was right and Conant is wrong; see also Proops (2001).

ideas from the *Tractatus* were of great significance to them. So the Vienna Circle view described in the first part of this paper is neither Wittgenstein's nor not Wittgenstein's.[4]

Our first priority will be to describe this view and show how deeply and intensely the Circle grappled, despite its own very different agenda, with certain basic ideas of the *Tractatus*. The early Vienna Circle based its criterion of meaning on Wittgenstein's picture theory, on his conception of language as exhaustively determined by rules, and on a universal language of the kind suggested by the *Tractatus*. We will argue that it was precisely Carnap's attempt to expand the scope of such a 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 as having come to him "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* 3 years later. This initial syntax idea, which found expression in only one of Carnap's publications,[5] was expounded systematically in a series of unpublished manuscripts written in the spring of 1931,[6] 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. Here Carnap attempted a definition of *analyticity* that would preserve Wittgenstein's idea of tautology in the new syntactic environment. This attempt was, however, subjected to a probing criticism by Gödel, who pointed out a fundamental flaw in Carnap's exposition. In the late 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* (1934), the original syntax idea of the first draft and the principle of tolerance co-exist in an uneasy truce. This would not last; by the following year, Carnap had allowed that the metalanguage for science could 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. In Sect. 2, we 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* becomes more comprehensible. We also argue that Gödel's critique played an important role in motivating Carnap's radical step toward tolerance in late 1932. Finally, we discuss the tension between the original syntax idea of January 1931 and the principle of tolerance in the published book, and identify which elements of the original idea were preserved in Carnap's later development. We also suggest that the fundamental

---

[4] That it is not straightforwardly Wittgenstein's we think the voluminous literature on the *Tractatus* (see previous footnote) makes adequately clear. That it is not *not* Wittgenstein's is attested to by several members of the Vienna Circle themselves; Carnap, for instance, said in his autobiography, "For me personally, Wittgenstein was perhaps the philosopher who, besides Russell and Frege, had the greatest influence on my thinking." (Carnap, 1963, p. 25).

[5] 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); this original syntax view is also in the background of Carnap's (1931) critique of Heidegger, "Überwindung der Metaphysik durch logische Analyse der Sprache".

[6] These shorthand manuscripts are preserved in UCLA/RC1029, Box 3, folders CM10 and CM11.

nature of the break that occurred in January 1931 has important consequences for the interpretation of the later Carnap, which even some of his closest readers seem not yet to have fully appreciated.

1 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. "It really does seem on first sight as if the very existence of mathematics must mean the failure of pure empiricism—as if we had in mathematics a knowledge about the world that doesn't come from experience, as if we had a priori knowledge," Hans Hahn had said. "And this evident difficulty for empiricism is so plain, so brutal, that anyone who wants to hold a consistent empiricism has to face this difficulty..." (Hahn, 1929, pp. 55–56). Wittgenstein had solved this problem. Though he had adopted Frege's and Russell's all-encompassing conception of logic as universally applicable and inescapable, he rejected the view that the logical laws were something like the most general laws of nature, or the laws of thought. For Wittgenstein, logical laws were not laws *of* something, that pertained to a universe of objects and expressed transparently in language along with other sorts of facts. Instead, language itself was recognized as a *medium*, as not part of the world but as representing the world to us. It represented the world to us, Wittgenstein thought, by means of logical pictures isomorphic to facts in the world. Frege's conception of logic as a gap-free system of rules, then, was extended to the whole of language, and connected to the world by this picture theory of meaning (Ricketts, 1996a, pp. 69–80).

The truths of logic were a by-product of this representational function of language. They ceased to be part of what languages describes, and became instead an artifact of the representational capacity of language. As such, logical truths became tautological and empty, despite still being universal (ibid., pp. 59–64). Hahn, like the rest of the Vienna Circle, thought this idea of critical importance:

> If one wants to regard logic—as this has in fact been done—as the study of the most general qualities of objects, as the study of objects in general [*überhaupt*], then empiricism would in fact be confronted here with an impassable hurdle. In reality, though, logic says nothing whatever about objects. Logic is not something that is to be found in the world. Logic only arises, rather, when—by means of a symbolism—we *speak about the world*... The sentences of logic say nothing about the world. (Hahn, 1929, pp. 56–57)

Of course the Vienna Circle did not simply accept Wittgenstein's view as stated in the *Tractatus*. "We learned much by our discussions of the book," Carnap later wrote, "and accepted many views *as far as we could assimilate them to our basic conceptions*." (Carnap, 1963, pp. 24–25, our emphasis) In particular, the Circle gave the *Tractatus* view an epistemological and positivistic twist, by interpreting the "atomic sentences" as elementary observation sentences. And they also combined it with logicism, so that the empty and tautological status Wittgenstein accorded logic was thereby also transmitted to all of mathematics. Naturally, they did not distinguish their version of Wittgenstein from Wittgenstein's own; to them it was a single and interlocking complex of ideas.

But even with these modifications, Wittgenstein's solution to the old problems carried a high price tag for the Vienna Circle. Two costly by-products of this solution were particularly unpalatable. The first of these problems arose from the Circle's own proposal to extend Wittgenstein's truth-functional language structure to mathematics (by means of logicism) and thereby to make it the basis of their unified language of science. For it seemed to them that the *Tractatus* did not allow for unbounded quantification, and this left it insufficient for expressing even a fragment of actually existing science. Indeed, it seemed that most theoretical science, and most classical mathematics, could not be expressed, or had no meaning, within (their understanding of) the *Tractatus* framework.[7]

The second problem concerned the status of the fundamental doctrines themselves, which seemed to fall victim to their own consequences. This was much discussed at the meetings of the Vienna Circle in 1930–31. Gödel, for instance, is recorded in the minutes as raising it very simply and bluntly on one occasion:

> *Gödel* asked how the discussion about logical questions could be justified, as it involves the utterance not of any meaningful sentences but only of elucidations [*Erläuterungen*]. This raises the question how admissable elucidations are to be demarcated from metaphysical pseudo-sentences. (ASP/RC 081-07-11; 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 e.g. the verification principle *itself* verifiable? Is *any* theoretical statement about language (even the language of science) verifiable? And if not, why is their status not every bit as metaphysical as that of the philosophies the Vienna Circle was consigning to the dustbin?[8]

Wittgenstein himself had of course confronted this problem head on, taking the heroic (if somewhat obscure) position that his own statements were, in the light of

---

[7] There has been considerable speculation (Floyd 2002, Geach, 1983; Soames, 1983; Sundholm, 1992) how Wittgenstein intended his operator *N* to make quantification "truth-functional" without resorting to a "molecular" finitism of the kind discerned in the *Tractatus* by the Vienna Circle, and most interpreters now think this is how Wittgenstein should be understood. However, the Circle clearly did *not* understand him this way, as Carnap retrospectively makes explicit in *Testability and Meaning* (Carnap, 1936–7, §23); in the *Tractatus*, Carnap's prime example there of a "molecular language", he states unequivocally that "truth-operations... are conceived as not including general operators [i.e. quantifiers]" (ibid., p. 18). And in his conversations with the Circle, Wittgenstein said nothing to contradict this interpretation (Waismann, 1967, e.g. p. 188). Moreover, Carnap says that "I too accepted a molecular language. According to the positivistic principle of testability in its most radical form, I restricted the atomic sentences to sentences about actual experiences..." (ibid., p. 19)—as indeed Wittgenstein himself seems to have done at least sometimes during this period; in a conversation of 1930–31 he says that "object" in the *Tractatus* is "used for such things as a colour, a point in visual space, etc" Wittgenstein, 1980, p. 120). If a scientific theory is a truth function of observation sentences, then it *can* only be a statement about a finite number of instances, not a universal law. This was why the picture theory, combined with the Circle's empiricism, made theoretical science as ordinarily conceived impossible.

[8] Despite the Circle's (and especially Carnap's) intense preoccupation with this question in 1930–31, it soon emerged as an all-purpose, unanswerable one-line *refutation* of logical empiricism. We still find it used this way, e.g. by Hilary Putnam: "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)

what they themselves asserted, strictly nonsense, and that one must "kick away the ladder" once things were seen correctly. This subtle and profound view is developed from a distinction between "saying" and "showing" arrived at from the impossibility of referring to the structure of language itself, in the 4.12's of the *Tractatus*, and then applied with full force in the final sentences of the book.

But these two problematic consequences of the *Tractatus* were unacceptable to the Vienna Circle. To begin with, they conflicted with the Circle's central project of rational reconstruction. If much of existing theoretical 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 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. On a more basic level, the sublime character of Wittgenstein's position (to the extent it was even recognized by the Vienna Circle) was fundamentally opposed to their Machean scientific and positivistic temper. That the basic principles of a theory should have their own meaninglessness as a consequence could only be regarded as a new and refined form of reductio ad absurdum.

Though the Circle discussed these two problems (the scope of the scientific language, the status of the theory itself) in Wittgensteinian terms, they approached them with their own agendas firmly in mind. Carnap had himself, in his *Aufbau* (largely written before his 1926 arrival in Vienna, when he was most intensely exposed to Wittgenstein's influence), arrived independently at a similar impasse to Wittgenstein's.[9] He had found himself caught between his basic requirement that all knowledge be "structural" and the need to give empirical or phenomenological validation to instances of the one remaining basic relation.[10] In attempting to overcome this problem by eliminating the basic relation, he encountered perplexities of the representation relation analogous to those Wittgenstein had highlighted with the distinction between "saying" and "showing". But Carnap thought that although Wittgenstein had *raised* a very interesting and important question, he had not fully captured the problems underlying it, and that a more satisfactory articulation of the problem would make the issue clearer and more precise than Wittgenstein had.[11]

---

[9] From the first conception of the *Aufbau* in the early 1920s, Carnap's conception of "structural" had grown steadily narrower, but it had not quite, by 1926, reached the completely non-epistemological view Wittgenstein had put forward. "The *Tractatus*," in the words of Michael Dummett, "is a pure essay in the theory of meaning, from which every trace of epistemological or psychological consideration has been purged as thoroughly as the house is purged of leaven before the Passover." (Dummett, 1981, p. 679) In fact it is unlikely that Carnap, with his eye firmly on the task of reconstructing knowledge, would ever have adopted such a standpoint. So it is unlikely that he would have solved the problem of the *status* of logic (and, for him, mathematics) without Wittgenstein's aid; this is why he (and Hahn; see above) acknowledged their debt to Wittgenstein, and adopted Wittgenstein's terminology.

[10] This tension is explicitly acknowledged in §153 of the *Aufbau*, where Carnap admits that the basic relation on which the entire system has been based is still non-logical, and that this is inconsistent with his principle, stated in §15–16, that "all scientific statements are purely structural statements." The solution offered in §154–155 is explicitly tentative and problematic.

[11] In the *Syntax* he thought he had attained this higher degree of clarity and precision. It is no accident, we think, that so many sentences from the *Tractatus* are among the "pseudo-object" sentences listed in the two-column passages in §§78–80 of the *Syntax*.

The further question how the actual "picturing" relation between facts and sentences is to enter into the constitution of knowledge was one that, as far as Carnap was concerned, Wittgenstein had not even addressed; the *Tractatus* is extremely vague on the precise nature of its "atomic sentences". Carnap thought he had made some progress here by proposing a definite form, though he was open to the idea that the atomic sentences could take forms other than those proposed in the *Aufbau*.[12] He was aware, though, that the system of the *Aufbau* had left significant gaps and needed further work, especially in the "constitution of the non-given".[13] Hans Reichenbach and Eino Kaila[14] had taken exception, for instance, to the apparent exclusion from the constitution system of certain modes of inference required in actual science, such as empirical induction, probability, and statistical inference.

An even more fundamental problem for the entire logicist procedure by which Carnap had constructed the *Aufbau* was raised by axiomatic systems. 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 *Geometry and Experience*. 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 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.[15]

One important feature 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 under this view, but it has a definite meaning as well, since it is interpreted in the *Grunddisziplin*.

Within his *Axiomatics* manuscript, though, there is no attempt to explain how the *Grunddisziplin* acquires its fixed interpretation. This task 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"

---

[12] In his review of Kaila's critique of the *Aufbau* (Kaila 1930; see note 14 below) Carnap says that "today we do not yet know what form and content the atomic sentences will [ultimately] have", and he allows that the form given in the *Aufbau* may have to be changed (Carnap, 1931, p. 77).

[13] This included, for Carnap, not only the constitution of the specific modes of inference mentioned in the following sentence above, but also of the constitution of the framework of constitution itself; see the discussion below on the "New Foundation of Logic".

[14] Kaila's critique (1930) was the first book-length assessment of the Vienna Circle; it focussed its criticisms almost exclusively on Carnap's *Aufbau*. It frequently invokes Reichenbach (1929), which argues (pp. 26ff.) in favor of realism and against positivism, though not explicitly against Carnap.

[15] This project is discussed in Awodey and Carus (2001), which also gives 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 discussed in detail in the above paper, and as Carnap himself realized in 1930, even before Gödel's incompleteness results later that year.

[*Neue Grundlegung der Logik*].[16] Its main idea is to erect a Hilbertian axiomatic superstructure on a Wittgensteinian basis. The atomic sentences are pictures of elementary 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*[17]—is that it not allow inference to any atomic sentence that is not already among the premises. 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.[18] 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:

> If now, to introduce the infinite, one "adjoins ideal propositions" (Hilbert), i.e. writes down formulas that have no contentful [*inhaltliche*] meaning, but permit us to derive the mathematics of the infinite, then we have once again been able to determine the meaning of the signs introduced as meaningless, by investigating for which logical constants the formulas would become tautologies. (ULCA/RC1029/Box 4/CM13, item 2, p. 62)

Unlike Hilbert, Carnap does not regard axiomatic systems as strings of 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. All these inferences are now at the same level. In a talk at Reichenbach's seminar in late 1929, Carnap said that 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.[19] In a lecture in Warsaw of December 1930, he said, along these same lines, that there is only one rule of inference in science: We can transform a sentence however we like, but the

---

[16] This document is in Carnap's papers at the Archive of Scientific Philosophy, Hillman Library, University of Pittsburgh (henceforth abbreviated ASP/RC), catalogue mark 089-64-01.

[17] Though there is no explicit provision for the quantifiers, Carnap may have intended to develop them axiomatically, as Hilbert and Ackermann (1928, pp. 22–23 and 53–54) 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 terms also used by Carnap.

[18] This idea, too, seems to have been suggested by the *Tractatus*, which says "The truth or falsehood of every single sentence changes something in the general construction of the world. And the range [*Spielraum*] that is left its construction by the totality of atomic sentences is precisely that which the most general sentences delimit." (5.5262)

[19] Carnap's shorthand notes for this talk are preserved in UCLA/RC1029/Box 4/folder CM13, item 3.

conclusion is to have no more content than the premises; it is to constrain the range of possibly true atomic sentences no less than the premises; 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 (ASP/RC 110-07-35, p. 2).

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 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*:

> The sentence of mathematics expresses no thought. In life it is never the mathematical sentence we need. We use the mathematical sentence *only* to derive sentences that do not belong to mathematics from other sentences that also do not belong to mathematics. (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:

> One can even (given the consistency of classical mathematics) give examples of sentences (of the kind stated by *Goldbach* or *Fermat*) that are correct in their content, but not provable in the formal system of classical mathematics. By adding the negation of such a sentence to the axioms of classical mathematics, one obtains a consistent system in which a sentence whose content is false is provable. (Gödel, 1931b, p. 202)

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. The Circle had needed all reasoning, mathematical or not, to have this character so as to undermine the metaphysical idea that conclusions about the real world could be reached by reasoning alone, without factual knowledge (Carnap, 1930, p. 25). But now it turned out that there could be sentences of arithmetic that were not decidable after all. The logicist explicit definition of the numbers had failed to guarantee the tautological character of arithmetic reasoning.

Third and finally, we return to Gödel's question. The apparent incompatibility of meta-linguistic discourse with Wittgenstein's framework was, as we saw, a fundamental barrier to the Vienna Circle's larger goals, and they sought to overcome it. The new work in mathematical logic, especially by Hilbert, Gödel, and Tarski, made essential 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 what the Vienna Circle read into Wittgenstein's final sentences. 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 were using.

2 The sleepless night and the original syntax idea

This threefold crisis brings us to Carnap's sleepless night. We are now in a position to understand the significance of the *Versuch einer Metalogik* he wrote the following day. What we find is a perspective that is 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.:

$$\overset{\cdots}{} f_{\cdots} (a_{\cdots})$$

An atomic sentence was thus a certain finite string consisting of instances of finitely many basic marks [*Zeichen*]—the instances themselves being *physical* marks, having a particular location on the blackboard or on a page. These physical marks represent the *Zeichen* of the calculus just as a geometrical diagram represents relations in pure, uninterpreted mathematical geometry.

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.[20]

From the viewpoint of modern logic, this idea may not seem particularly momentous. Even at the time, it represented no technical innovation; Hilbert and others had been treating axiomatic systems formally for decades, and the methods of Gödel and Tarski did 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 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 "metalogical" perspective of regarding language purely as a system of rules, without reference to anything outside itself, was indeed a revolutionary idea.

Before Wittgenstein, language had been regarded as an essentially transparent medium for the expression of thought. The laws of logic were considered by Frege and Russell to be laws of thought, judgement, or perhaps nature—but certainly not of language. Wittgenstein had recognized that they were laws of language. But he had arrived at this idea via a theory of representation that forced language to conform always and everywhere to a particular system of rules, arising necessarily from the

---

[20] In the margin of p. 3 of the manuscript, Carnap has scrawled, "Regarding the undefined concept '*true*'. It is completely different from the other concepts of metalogic. Perhaps avoidable? [Perhaps] just define which atomic sentences are the "basis" of a sentence, and how. (?)"

representational function of language—the picture theory. The possibility of representation determined a particular form of linguistic intuition, so to speak. This elementary logic built into our form of representation was, like a Kantian form of intuition, an inescapable straight-jacket. The very nature of language, in Wittgenstein's view (at least as seen by the Vienna Circle), prevented us from stepping outside it. One could call this quasi-Kantian view "Wittgenstein's prison".

Under the suggestive influence of Hilbert's formal approach to axiomatic systems and its use by Gödel and Tarski, Carnap was able to escape from Wittgenstein's prison by taking Wittgenstein's own idea of language as a system of rules one step further. Carnap distinguished the representational or meaning function of language from its purely combinatorial one, and now took the *latter*, rather than the former, as his starting point. The metalogical methods developed in pursuit of the very mathematical results (such as the incompleteness theorem) that had led to the disintegration of his Wittgenstein-based position in the "New Foundation", it turned out, also showed a way of breaking out of 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, this new recognition that linguistic structure could itself be investigated opened up a whole new method for the unification and clarification of knowledge. Thus 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, Carnap 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.[21] 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 of m-many dots. 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,

---

[21] An addition of 7 February 1931 to the manuscript says, "the syntax of the rows of dots is arithmetic" (p. 1).

and that to this hierarchy of languages there may be no limit. (Russell, 1922, p. 286)

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[22] must have indeed seemed rather compelling, at first sight.

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".[23] 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 (Carnap, 1932a, p. 435).

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 its "elimination of verification by comparison with facts [*Ausschaltung der Verifikation durch Vergleich mit Sachverhalten*]" (ASP/RC [diary entry of 26 June 1931]). As Carnap explained in "Die physikalische Sprache," not only criterial definitions but also ostensive definitions can be regarded as intra-linguistic.[24]

---

[22] The *Versuch* ends with a summary in four points: "(1) The particular *natural numbers* occur as signs of *the language itself*. (2) The so-called "*properties of natural numbers*" are not proper properties, but syntactic (Wittgenstein: internal) ones, so are to be expressed in the *metalanguage*. (3) A particular *real number* is a property or sequence of natural numbers, so is also to be expressed in the *metalanguage*. (4) The *properties of real numbers* are not real properties, but syntactic properties (with respect to the syntax of the metalanguage), and thus *to be expressed in the meta-metalanguage*." (p. 44)

[23] 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)

[24] "Elephant", for instance, criterially defined as an animal with certain characteristics, might be ostensively defined as "an animal of the kind present at a certain space-time location" (Carnap, 1932a, p. 435–436).

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 (naturally occurring or metaphysically unavoidable) components. Wittgenstein's idea of language as exhaustively characterizable by rules was taken a step further. The rules were no longer to be *found*, they were no longer objectively determinate and discoverable artifacts of the nature of representation, as Wittgenstein had appeared to suggest. Instead, they were a matter of human decision, conventions by which we set up the language of science. 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 (Jeffrey, 1994) 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 aspects were closely entwined, and obstacles remained to be overcome on the logical front.

3 *Metalogik*: the first draft of *Logical Syntax*

> The view that the terms "analytic" and "contradictory" are purely formal and that analytic sentences are empty of 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, which gives expression to Wittgenstein's absolutist conception of language and leaves out the conventional factor in the construction of a language, misses the mark. (Carnap, 1934, p. 139)

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* 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., §19). 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.[25]

---

[25] The first systematic exposition of the new view was in a series of three lectures to the Vienna Circle in June and July of 1931. These fell into the period during which Rose Rand was taking minutes of the Circle meetings, so they are recorded, somewhat elliptically, in ASP/RC 081-07-17, 18, and 19 (with further discussion of these lectures in 081-07-20), and published in Stadler (1997, pp. 314–334).

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. The identification of the logical with the formal seemed to restrict its scope to only what can be expressed with very limited means. If, however, there were no intrinsic constraints on the sorts of formal properties of formulas that could be considered, then perhaps there could be a formal criterion for mathematical truth *different* from mere provability. Since Gödel had shown that provability was insufficient—there were "true" arithmetical statements not derivable from the axioms—the identification of such a criterion was 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. But now observe that 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.[26] Such a configuration represents an analytic sentence that has no proof. 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 have been able to determine) but its table of contents. This lists the notion *analytic* alongside *synthetic* and *contradictory* under the heading "IV.B. Theory of content of formulas" (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.

Footnote 25 continued

Carnap appears, from evidence in the file containing the *Versuch*, to have changed the system to an arithmetized one on 17 June, the day before the second lecture. He spells out the difference this makes to the scope of the system at the conclusion of the second lecture as follows: "The difference between arithmetic metalogic and the metalogic portrayed previously is this: arithmetic metalogic treats not the empirically available, but all possible configurations. Our previous metalogic is the descriptive theory of certain given configurations, it is the geography of language forms, while the arithmetized metalogic is the geometry of language forms." (Stadler, 1997, p. 325) Also noteworthy in these talks is the fact that they contain no definition of analyticity, and that they take the view that only a single language is required (something like the later Language I). In answer to the question "So are we to draw the inference that there is only a *single* language?", Carnap replies "Well, there are sentences of very different form …, but all of them, even the metalogical ones, are in a *single* language." (ibid., p. 329).

[26] For instance: all the black pawns in their starting positions and the row behind them empty except for black king in one corner, white rook in the other.

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".[27] 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. 346–348). Carnap had apparently tried to define the 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:

$$(\forall x)f(x) \text{is analytic} \Leftrightarrow_{df} f(a) \text{is analytic for all numerical constants } a$$

In his definition, Carnap had apparently tried to use the same 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:

$$(\forall X)X(0) \text{is analytic} \Leftrightarrow_{df} A(0) \text{is analytic for all predicate constants } A(x)$$

But here there is no restriction on what predicate constants (i.e. "open sentences") $A(x)$ 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 metalanguage, in which the concepts of "set" and "relation" are already given. 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, 1931a). 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

---

[27] From the table of contents (ASP/RC 110-04-07) 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 June 1931 lectures to the Vienna Circle (see above, footnote 17) 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".

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 (ibid., pp. 350–352).

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 $(\forall X)\ldots X\ldots$. 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 (ibid., p. 354). 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",[28] which seems likely to be the reason Gödel never worked out his own part II (ibid., p. 356).

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".[29] 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.[30] 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.

---

[28] Presumably he refers here to Tarski (1932), which however gives only a bare summary; Gödel may have known more details from Tarski directly.

[29] In a recent paper it is claimed that "Carnap's main task in *Syntax* was to provide a reconstruction of mathematical truth" (Lavers, 2004, p. 296; cf also p. 308). This was true, as we have seen, of the first draft (the *Metalogik*)—but no longer of the published book. Note that the passages discussed by Lavers in support of his claim (ibid., pp. 297–299) were not included in the book's original edition (1934).

[30] Despite their view that the "logical syntax" (in Wittgenstein's sense) of the language was partly a matter of convention, the Vienna Circle up to this point generally held that there were "correct" and "incorrect" languages. In Carnap's case this is explicit in his papers during this period (1932a,b; see above, p. [16]) as well as retrospective comments soon after ("I was wrong in thinking that the language I dealt with was *the* language, i.e. the only legitimate language... Consequently I made the mistake of formulating my epistemological view in the form of an assertion—as most philosophers do—instead of in the form of a suggestion concerning the form of language." (1936–7, p. 20) We suggest that this conviction was shaken above all by Carnap's realization at this point that, if the meta-language was not part of the the *same* language, then there was *no* inherently privileged meta-language that could obviously or "naturally" claim to be "correct". Quine (1963) section VII thinks that Carnap's move to semantics was largely driven by this situation; as argued in the next section, we believe that Carnap's response to it was more fundamental.

4 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, 1932b, 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"). (ibid., 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*:

> The range of possible language forms, and thus of different possible logical systems is … incomparably larger than the very narrow range in which modern logical investigations have so far operated. Up to now there have only been occasional small departures from the language form given by *Russell*, which has already become classical… The reason for not daring to depart further from this classical form would appear to lie in the widespread view that such departures must be "justified", i.e. it must be shown that the new language form is "correct", that it represents the "true logic". It is one of the main tasks of this book to eliminate this view as well as the pseudoproblems and pointless squabbles arising from it. (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. "But they were hobbled by a striving for 'correctness'." And Carnap concludes with the famous words: "But now this barrier is overcome: before us lies the open sea of free possibilities." (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:

> Once it is understood that all pro- and anti-intuitionist considerations are concerned with the form of a calculus, the question will no longer be asked in the form "What *is* the case?" but rather "How do we *want* to set this up in the language being constructed?"... And with that, the dogmatic frame of mind that often makes the discussion unfruitful is banished. (ibid., p. 42)

This "dogmatic frame of mind" results, in Carnap's view, from the reliance on inherently vague philosophical "considerations [*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:

> Our attitude to demands of this kind may be stated generally by the *principle of tolerance: we do not want to impose restrictions but to state conventions*... *In logic there are no morals*. Everyone can construct his logic, i.e. his language form, however he wants. If we wants to discuss it with us, though, he will have to make precise how we wants to set things up. He has to give syntactic rules rather than philosophical considerations. (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 after 1945.[31] And he is careful to apply the insistence on precision to his own

---

[31] The classical exposition of this project is in Chapter 1 of *Logical Foundations of Probability* (Carnap, 1950a); for further discussion, see Stein (1992), Awodey and Carus (2004), section III, and Carus (2004), section II.

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.

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.[32]

5 Tensions within the published *Logical Syntax*

We began in Sect. 1 with a conception of meaning whereby it is built up from atomic components. One could call this view, which was derived from a reading of Wittgenstein (though it may not have been his own), a "meaning foundationalism". The meanings of all sentences rest on the representation of atomic facts by atomic sentences. We interpreted the new view Carnap arrived at during his sleepless night in January 1931 as the replacement of this meaning foundationalism by an axiomatic approach to language as a whole, in which all workings of the language are exhaustively specified by explicit rules stated in a meta-language. In its original statement, this "syntax" view completely excluded the possibility of "meaning"—in its old sense of representational correspondence between configurations of linguistic objects and configurations of objects in the world. There seemed no way of capturing any such correspondence in explicit formation or transformation rules for a language. Only a year after the *Syntax* book was published, though, Tarski's definition of truth suggested to Carnap that such correspondences could, after all, be captured in meta-linguistic rules. This amounted to defining a new notion of meaning "from above", in contrast to the one built up "from below" in the meaning foundationalism Carnap had rejected in January 1931. In this new scheme, the language itself was constituted by a system of rules which in turn permitted the rigorous specification of an "interpretation", as opposed to regarding the rules as descriptive and determined by a more fundamental notion of meaning built up from atomic components.

In January 1931, the rejection of meaning foundationalism and its replacement by an axiomatic view was all of a piece. Seen from the later, semantical perspective, though, this original syntax view could be regarded, retrospectively, as having been composed of a number of different elements that would later turn out to be separable. First, there was (a) the requirement that a language be entirely specified by explicit rules. But the "syntactic" view that seemed to follow from this can in retrospect be seen as having consisted of two separable parts: (b) the distinction between a language (a calculus, a purely syntactic symbol system) and its interpretation, and

---

(c) the prohibition of reference to the latter, and the restriction of the (*wissenschafts-logische*) meta-language to consideration of the former.

Components (a) and (b) are necessary pre-conditions for the tolerance idea. Without the requirement that language be specified by explicit rules, the alternatives among which tolerance is recommended are not fully specified. And without distinguishing language from content, there are no clearly distinct alternatives to be tolerant among. These two retrospectively visible components survive unscathed and undiminished into Carnap's semantic period. (So it is rather misleading to call them "syntactic"; Carnap's original term "metalogical" might be more appropriate.) Component (c), on the other hand, the overreaction against Wittgensteinian "meaning" that accompanied the original insight, did not survive. In distinguishing between a language and its interpretation, Carnap's first (and, as we saw, understandable) response was to reject that imprecise notion of meaning entirely. But this restriction was loosened when he saw that, in virtue of the precise specification of the object language, interpretations could *also* be specified by explicit rules (governing satisfaction, designation, and truth), in accordance with component (a) of the original syntax idea.

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 account 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 (Wittgenstein's, in the Circle's view). 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.

Of the components of the original syntax idea, then, the two "metalogical" ones—(a) and (b)—are consistent with tolerance, while the rejection of meaning—component (c)—is inconsistent with tolerance. This tension is nowhere close to the surface in the published *Syntax* book, largely because these components of the original view, tolerance and the rejection of meaning, occupy different sections of the book, and never meet. In the famous introduction that we quoted from above, tolerance is highlighted. In the opening chapters and in the exposition of Language I, the rejection of meaning comes to the fore, except in §§16–17, in which tolerance is again in the spotlight. The exposition of Language II is largely neutral and technical, though in a few sections (e.g. §38) certain provisions (e.g. the elimination of classes) are implicitly defended as "correct". Section 4 of the book, on general syntax, is again dominated by tolerance; the optional nature of many different kinds of logical and mathematical systems is stressed. And in Sect. 5, finally, on "Philosophy and Syntax", the emphasis is almost entirely on the exclusion of meaning (the restriction to the "formal mode of speech"); tolerance hardly makes an appearance. Since this last section was by far the most widely read, it captured the imagination of philosophers. The distinction between "formal" and "material" modes of speech, introduced in "Die physikalische Sprache" (*Unity of Science*) in 1932, is discussed in more detail and applied to a wide range of philosophical problems. This distinction, of course, relies on the exclusion of

meaning, and is thus incompatible with the principle of tolerance.[33] This absence of confrontation between tolerance and the rejection of meaning is not by design, but resulted straightforwardly from the order of composition; part V was largely written before the October 1932 turn to tolerance, while part IV was written in 1933.[34]

While the original syntax idea was a necessary precondition for tolerance, then, one component of it (and, in the book, the most visibly high-profile component) would soon be jettisoned. This gave the appearance that the "syntax" doctrine—identified with the exclusion of meaning—had first been embraced, and then rejected again. "Meaning" was proscribed in 1931, it seemed, and then became acceptable again in 1935. What most observers failed to notice in this sequence of events was the permanence of the "metalogical" (or "top-down") components and the principle of tolerance consequent upon them. The rejection of meaning foundationalism in January 1931—Quine's second "dogma of empiricism"—was permanent; as we saw above, the *main point* of January 1931 had been the "*elimination* of verification by comparison with facts" (Sect. 2 above, final paragraph).[35]

Seen from this perspective, in fact, it becomes apparent that Carnap's abandonment of Quine's *first* "dogma" followed soon after; as we saw above (Sect. 3), the principle of tolerance was suggested by Carnap's recognition that there could be no language-transcendent notion of analyticity of the kind he had attempted in the first draft of the *Syntax* book. The idea of a single, canonical definition of analyticity was just what he *gave up* in October 1932, after Gödel's critique. Thereafter, analyticity is strictly language-relative. And this step was also permanent. It would appear, then, that Quine's supposed "two dogmas" were precisely the two components of the Wittgensteinian "absolutist" conception that Carnap, under Gödel's influence, left behind, successively, in January 1931 and October 1932.[36]

---

[33] As Carnap himself recognized, though he played it down; see Carnap, 1942, p. 250, where he discusses the possibility of "supplementing" the translation of material-mode sentences into formal-mode ones with translations into semantic and pragmatic sentences. He never actually attempted anything like this, however, for reasons well diagnosed by Ricketts (1996b), p. 247 and Goldfarb (1997), pp. 63–65. Indeed, his statement that "The explanation of the dangers of the *material mode* of speech . . . in §§78–80 remains valid" (ibid.) seems absurd, on the surface, given that the "material mode of speech" is *defined* in *Syntax* as non-syntactic (Carnap, 1934, pp. 212–213); i.e. it includes semantic sentences. This statement becomes reasonable, though, if we look at §§78–80 and realize that the instances of the "material mode of speech" considered there are nearly all philosophical (specifically ontological) ones of the kind he would later call "external" questions (Carnap, 1950b, p. 206). The translations into the "formal mode" offered there can nearly all be regarded as plausible translations of *ontological* "external" questions into a form that lends itself to their transformation into "practical questions of language choice". But in 1934 they are not yet *stated* as explicitly practical questions; they are stated as purely syntactic. This is an obvious difference.

[34] A detailed chronology of the book's composition was drawn up by Carnap and is preserved in his papers, ASP/RC 110-04-08 and 110-04-09.

[35] The second dogma is explicitly identified with the "verification theory of meaning" in Quine (1951), p. 41. We hope to make this case in more detail—for *both* of Quine's "dogmas"—in a subsequent paper.

[36] We are grateful to the Philosophy Department of the University of Constance (Germany), and especially to the participants in the seminar "Carnap der Logizist" in the summer semester of 2003, during which the present paper was largely written. We also thank Thomas Ricketts and Wilfried Sieg for helpful comments on earlier drafts, as well as those who attended talks and seminars where versions of this paper were given, at Stanford, Purdue, Jena, Paris, and the 2003 ASL meeting in Chicago.

References

Awodey, S., & Carus, A. W. (2001). Carnap, completeness, and categoricity: The *Gabelbarkeitssatz* of 1928. *Erkenntnis, 54*, 145–172.

Awodey, S., & Carus, A. W. (2004). How Carnap could have replied to Gödel. In S. Awodey, & C. Klein (Eds.), *Carnap brought home* (pp. 179–200). LaSalle, IL: Open Court.

Carnap. R. (1930). Die alte und die neue Logik. *Erkenntnis, 1*, 12–26.

Carnap, R. (1931). Review of Kaila (1930). *Erkenntnis, 2*, 75–77.

Carnap, R. (1932a). Die physikalische Sprache als Universalsprache der Wissenschaft. *Erkenntnis, 2*, 432–465.

Carnap, R. (1932b). Über Protokollsätze. *Erkenntnis, 3*, 215–228.

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

Carnap, R. (1936–7). Testability and meaning. *Philosophy of Science, 3–4*, 419–471, 1–40.

Carnap, R. (1942). *Introduction to semantics*. Cambridge, MA: Harvard University Press.

Carnap, R. (1950a). *Logical foundations of probability*. Chicago: University of Chicago Press.

Carnap, R. (1950b). Empiricism, semantics, and ontology. Repr. in *Meaning and necessity* (2nd ed., pp. 205–221). Chicago: University of Chicago Press, 1956.

Carnap, R. (1963). Carnap's intellectual autobiography and "The philosopher replies". In P. Schilpp, (Ed.), *The philosophy of Rudolf Carnap* (pp. 3–84, 859–1013). LaSalle, IL: Open Court.

Carus, A. W. (2004). Sellars, Carnap, and the logical space of reasons. In S. Awodey, & C. Klein (Eds.), *Carnap brought home: The view from Jena* (pp. 317–355). LaSalle, IL: Open Court.

Carus, A. W. (forthcoming). *Carnap and twentieth-century thought: Explication as enlightenment*. Cambridge: Cambridge University Press.

Conant, J. (2001). Two conceptions of *Die Überwindung der Metaphysik*: Carnap and early Wittgenstein. In T. G. McCarthy, & S. C. Stidd (Eds), *Wittgenstein in America* (pp. 13–61). Oxford: Oxford University Press.

Creath, R. (1990). The unimportance of semantics. *PSA1990, 2*, 405–415.

Dummett, M. (1981). *Frege: Philosophy of language* (2nd ed.). London: Duckworth.

Floyd, J. (2002). Number and ascriptions of number in Wittgenstein's *Tractatus*. In E. Reck (Ed.), *From Frege to Wittgenstein: Perspectives on early analytic philosophy* (pp. 308–352). Oxford: Oxford University Press.

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

Geach, P. (1983). Wittgenstein's Operator *N*. *Analysis, 41*, 573–589.

Gödel, K. (1931a). Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I. Repr. in K. Gödel, *Collected Works*, vol. 1. (pp. 144–195). Oxford: Oxford University Press, 1986.

Gödel, K. (1931b). Diskussion zur Grundlegung der Mathematik. Repr. in K. Gódel, *Collected Works*, vol. 1. (pp. 200–205). Oxford: Oxford University Press, 1986.

Gödel, K. (2003). *Collected works*, vol. 4; *Correspondence A-G*. Oxford: Oxford University Press.

Goldfarb, W. (1997). Semantics in Carnap: A Rejoinder to Alberto Coffa. *Philosophical Topics, 25*, 51–66.

Hacker, P. M. S. (2003). Wittgenstein, Carnap, and the New American Wittgensteinians. *Philosophical Quarterly, 53*, 1–23.

Hahn, H. (1929). Empirismmus, Mathematik, und Logik. Repr. in H. Hahn, *Empirismus, Logik, Mathematik* (pp. 55–58). Frankfurt/Main: Suhrkamp, 1988.

Hilbert, D., & Ackermann, W. (1928). *Grundzüge der theoretischen Logik*. Berlin: Springer.

Jeffrey, R. (1994). Carnap's Voluntarism. In D. Prawitz, B. Skyrms, & D. Westerståhl (Eds.), *Logic, methodology, and philosophy of science IX* (pp. 847–866). Amsterdam: Elsevier.

Lavers, G. (2004). Carnap, semantics, and ontology. *Erkenntnis, 60*, 295–316.

Proops, I. (2001a). The New Wittgenstein: A critique. *European Journal of Philosophy, 9*, 375–404.

Putnam, H. (1981). *Reason, truth, and history*. Cambridge: Cambridge University Press.

Quine, W. V. O. (1951). Two dogmas of empiricism. Repr. in Quine *From a logical point of view* (pp. 20–46). New York: Harper, 1953.

Quine, W. V. O. (1963). Carnap and logical truth. In P. Schilpp (Ed.), *The Philosophy of Rudolf Carnap* (pp. 385–406). LaSalle, IL: Open Court.

Reichenbach, H. (1929). Ziele und Wege der physikalischen Erkenntnis, in *Handbuch der Physik*, vol. 4: *Allgemeine Grundlagen der Physik*. (pp. 1–80). Berlin: Springer.

Ricketts, T. (1996a). Pictures, logic, and the limits of sense in Wittgenstein's *Tractatus*. In H. Sluga, & D. G. Stern (Eds.), *The Cambridge Companion to Wittgenstein* (pp. 59–99). Cambridge: Cambridge University Press.

Ricketts, T. (1996b). Carnap: From logical syntax to semantics. In R. Giere, & A. Richardson, (Eds.), *Origins of logical empiricism* (pp. 231–250). Minneapolis: University of Minnesota Press.

Rorty, R. M. (Ed.) (1967). *The linguistic turn: Essays in philosophical method*. Chicago: University of Chicago Press.

Russell, B. (1922). "Introduction" to Wittgenstein's *Tractatus*. Repr. in L. Wittgenstein *Logisch-philosophische Abhandlung* (pp. 258–286) ed. by B. McGuinness & J. Schulte, Frankfurt a. M. 1998: Suhrkamp.

Soames, S. (1983). Generality, truth functions, and expressive capacity. *Philosophical Review, 92*, 573–589.

Stadler, F. (1997). *Studien zum Wiener Kreis: Ursprung, Entwicklung und Wirkung des logischen Empirismus im Kontext*. Frankfurt/Main: Suhrkamp.

Stein, H. (1992). Was Carnap entirely wrong, after all? *Synthese, 93*, 275–295.

Sundholm, G. (1992). The general form of the operation in Wittgenstein's *Tractatus*. *Grazer Philosophische Studien, 42*, 57–76.

Tarski, A. (1932). Der Wahrheitsbegriff in den Sprachen der deduktiven Disziplinen. *Akademischer Anzeiger der Akademie der Wissenschaften in Wien, Mathematisch-naturwissenschaftliche Klasse, 69*, 23–25.

Waismann, F. (Ed.). (1967). *Wittgenstein und der Wiener Kreis*. Oxford: Blackwell.

Wittgenstein, L. (1922). *Tractatus Logic. Philosophicus*. Repr. in L. Wittgenstein *Logisch-philosophische Abhandlung* (pp. 1–178) ed. by B. McGuinness & J. Schulte, Frankfurt a. M. 1998: Suhrkamp.

Wittgenstein, L. (1980). *Wittgenstein's Lectures, Cambridge 1930–32*, from the notes of J. King & D. Lee, J. King, ed. Totowa NJ: Rowman & Littlefield.