Vojtěch Kolman (ed.)

MISCELLANEA LOGICA VI

From Truth to Proof

Work on this volume was supported by the grant No. 401/06/0387 of the Grant Agency of the Czech Republic.

MISCELLANEA LOGICA

Typeset by Vojtěch Kolman & Marie Benediktová

ISBN 80-7308-145-8

Contents

The Past and Future of Miscellanea Logica

Miscellanea Logica came into being as an occasional bulletin of the Department of Logic of the Faculty of Philosophy and Arts of the Charles University in Prague. The first five issues, published between the years 1998 and 2003 (the first four of them in Czech), were of a truly miscellaneous nature: they contained different kinds of contributions to logic due mostly to the members and students of the department. (The are all available, in an electronic format, from `http://logika.ff.cuni.cz/artsys/index.php?req=group&id=24`.) The people who deserve most of the credit for bringing *Miscellanea Logica* to life and editing the first issues were Petr Jirků (by that time the head of the department), Vítězslav Švejdar and Kamila Bendová.

When I moved to the position of the head of the Department of Logic I started to feel that time has come for a step forward. The idea of a *new series* of *Miscellanea Logica*, the first issue of which you are reading, is that it should be no longer miscellaneous, but rather much more narrow-focused. Each volume will be devoted to a single topic and bring together papers addressing urgent problems of contemporary logic (including the logic/mathematics and logic/philosophy interfaces); hence the *new series* should be more of a *book series* than a journal. The list of contributors should also not be restricted to people directly related to the department — we hope that by timely issuing CFPs for future volumes we turn *Miscellanea Logica* into a truly international enterprise.

Jaroslav Peregrin

Call for Papers

Next volume of *Miscellanea Logica* ‘Consequence, Inference, Structure’, with the expected deadline in March 2007, welcomes mathematically oriented papers on topics related to nonclassical logics and consequence relations (preferably substructural logics) and their applications (e.g. in metamathematics and proof complexity). Submissions should be send jointly to: `marta.bilkova@gmail.com` and `peregrin@ff.cuni.cz`.

Marta Bílková

Introduction

The well-known fact of the existence of undecided arithmetical propositions, such as Goldbach's conjecture, has led to the expounding of two extremist theories of arithmetical truth, both of which consider this fact to be their vindication. Mathematical *mentalism* bases arithmetical truth on mental constructions, or intuitions, as opposed to the conventions of language, which Brouwer claims have nothing to do with mathematics proper. On the other hand, mathematical *Platonism* stresses the independence of arithmetic not only of any particular human subject (in which it has a point), but of the whole mankind. Whereas according to the first doctrine there is no need to distinguish effective means of proving truth from the truth itself, the second doctrine extrapolates the common and useful distinction between truths we already know and truths we have yet to recognize in the direction of true, yet unrecognizable reality on the one hand and of the fallible nature of human knowledge on the other. Thus, neither intuitionism nor Platonism is able to provide us with a satisfactory analysis of how *we* can know that an arithmetical sentence is true, or what it means to say things like this.

The relationship between truth and proof came to the fore in philosophy in the aftermath of several great achievements in the field of mathematical logic. Specifically, these were Frege's invention and calculization of the concept of logical consequence, and Gödel's metaproofs of logic's completeness and incompleteness. (By logic, I mean logic in its primordial shape, i.e. before the first- and higher-order partitioning which, to some extent, was a subconscious response to Gödel's discovery.) According to Gödel, logic and arithmetic are not only incomplete, but incompletable in the sense of strong (essential) ineffectivity. As a consequence, one can differentiate between strong *effective syntax* on the one hand and more *liberal semantics* on the other. The corresponding distinction, however, between (abstract) truth and (concrete) proof that approximates only partially to truth, seems less straightforward. We need go no further here to appreciate the range of potential problems: What does it mean for a sentence to be essentially unprovable, and yet true? Can we 'prove' this state of affairs or not? Does the 'Gödelian' true/proof distinction entail a Platonistic standpoint (as Gödel himself believed)? How are we acquainted with (abstract) semantics? Is it still possible to maintain the constructivist and pragmatist concept of truth? Etc.

Our point here is that, no matter how impressive some of Frege's, Cantor's or Gödel's and Tarski's achievements may appear, conundrums such as 'What is true?' and 'What does it mean to prove it?', or 'What is meaning?' and 'How do we understand it?', are not thereby provided with factual answers in the sense that, say, scientific facts may be disclosed. The explanations they give us are, of course, facts of their own nature, but in a very modest, artificial sense, which fails to reflect directly all the wide range of our experience of using the words "truth", "proof", "meaning", etc. To illustrate the point, let us take a closer look at Bolzano's famous proof of the intermediate value theorem:

> Let a and b be reals such that $a < b$, let f be a function continuous on a closed interval $[a, b]$ and such that $f(a)$ and $f(b)$ are of opposite signs. Then there exists a number $x \in [a, b]$ such that $f(x) = 0$.

According to the official legend, Bolzano proved something self-evident on an analytic (purely conceptual or verbal) basis merely to avoid a usage of the spatial intuition. But this does not make sense for two fundamental reasons: (1) the new explicit (not the intuitive) definitions of conceptualized analysis (such as those of continuity, convergence and derivative) are by no means self-evident; (2) the theorem does not hold in general. In accordance with the well-known 'epsilon-delta-type' definition of continuity, the function

$$f(x) = \begin{cases} 1 & \text{if} \quad x^2 < 2 \vee x < 0, \\ -1 & \text{if} \quad x^2 \geq 2 \wedge x \geq 0, \end{cases}$$

is continuous on the rational line and yet it fails to have the intermediate value there. A similar argument holds for other non-Cantorian continua like the Euclidian one (consisting of numbers constructed by means of rule and compass), the Cartesian one (consisting of real algebraic numbers) or that built on the basis of *lawlike* sequences of rational numbers. Of course, we have not yet said what a continuum is, but then neither did Bolzano (nor Leibniz or Cauchy), so we might as well use the ancient standards according to which 'being continuous' amounts to the simple opposite of 'being finitely divisible' and is therefore met already by the totality of rational numbers. Hence what I am saying is that since Bolzano did not possess a clear concept of the real number, one cannot conclude that he tried to prove — or even proved — a self-evident truth by analytical means. He could not provide a definite 'choice' between the truth and falsity of a conjecture since he had no clear concept of truth for arithmetical sentences at all. What we are authorized to claim is merely that Bolzano *indicated* that in order to maintain the intermediate theorem the continuum must be complete in a very specific (order-complete) sense. So, instead of a proof we are facing a decision to define real numbers in a certain holistic way.

Note a subtle point here. Due to Tarski, the theory of real numbers, i.e. of a complete ordered field, is complete and decidable as soon as the first-order schema replaces the second-order completeness axiom. However, the very same theory is obtained from the axioms of an ordered field together with an intermediate value scheme for polynomials

$$\exists x(p(x) \leq 0) \wedge \exists x(p(x) \geq 0) \rightarrow \exists x(p(x) = 0).$$

By the Löwenheim-Skolem theorem, this theory has a countable model and the smallest of such models is, in fact, the field of real algebraic numbers, i.e. the Cartesian continuum. Hence, we can now see that in order to define the continuum we must make one more decision, namely the decision of what is (admissible) function. The question is: Should we go along with the most liberal concept of Fourier's and Dirichlet's, as Cantor did, or should we accept Euler's and Lagrange's additional requirements for which the polynomial is the prototype?

It is precisely because of this complicated relational structure between (1) the theories of meaning, proof and truth, or of logic in general, and (2) their applications to language and reality, that the possibilities, profits (advantages) and limits (disadvantages) of this projection are repeatedly subjected to comment and re-assessment. Such comments go back to Bolzano's critique of Kant's analytic/synthetic distinction, his concept of direct proof and his definition of truth, and to Frege's creation of the logical calculus and his foundations of mathematics. They were continued by Wittgenstein in his critique of logicism and set theory, as well as in his later remarks on rule-following; by Carnap in his logical syntax project, in the constructivism of Brouwer and Weyl and their alternative (effective) conceptions of truth; by Hilbert within his finitist program, in Tarski's model theory and Gentzen's consistency proofs.

In view of this, the aim of our anthology is to collect comments and insights concerning not only the truth/proof distinction itself, but also the long and respected tradition of its interpretation, with the modest hope that it will contribute to a better understanding of the tradition's 'twists and turns', i.e. its contradictory statements, weak arguments and further philosophical trivia.

Prague, October 2006

Vojtěch Kolman

One

Consequence & Inference

Jaroslav Peregrin

1 Inference as an explication and as a counterpart of consequence

Logic is usually considered to be the study of logical consequence — of the most basic laws governing how a statement's truth depends on the truth of other statements. Some of the pioneers of modern formal logic, notably Hilbert and Carnap, assumed that the only way to get hold of the relation of consequence was to reconstruct it as a relation of inference within a formal system built upon explicit inferential rules. Even Alfred Tarski in 1930 seemed to foresee no kind of consequence other than that induced by a set of inference rules:

> Let S be an arbitrary set of sentences of a particular discipline. With the help of certain operations, the so-called *rules of inference*, new sentences are derived from the set S called the *consequences of the set S*. To establish these rules of inference, and with their help exactly define the concept of consequence, is again a task of special metadisciplines; in the usual terminology of set theory the schema of such definition can be formulated as follows: The set of all *consequences of the set S* is the intersection of all sets which contain the set S and are closed under the given rules of inference. [Tarski, 1930, p. 63]

The work on this paper was supported by the grant No. 401/06/0387 of the Grant Agency of the Czech Republic.

Thereby also the concept of truth came to be reconstructed as inferability from the empty set of premises. (More precisely, this holds only for non-empirical, necessary truth; but of course logic never set itself the task of studying empirical truth.) From this viewpoint, logic came to look as the enterprise of *explication of consequence in terms of inference.*

This view was soon shattered by the incompleteness proof of Kurt Gödel and by the arguments of Tarski himself which seemed to indicate that inference can never be a fully satisfying explication of consequence and that hence we must find a more direct way of dealing with consequence. And the model theory, established by Tarski, appeared to be the requisite tool. Consequence and inference thus started to appear as two things to be confronted and one of the main issues of logic became the investigation of the results of various instances of the confrontations — various kinds of soundness and completeness theorems. However, as Bencivenga duly points out, "completeness theorems are a dime a dozen, but there is hardly any discussion of what it means to prove such a theorem."[1)]

2 Consequence

What is consequence? When is a statement a consequence of other statements? As we have already noted, consequence is generally understood as truth-preservation:

> A is a consequence of $A_1, \ldots, A_n$ if and only if A is true whenever $A_1, \ldots, A_n$ are true.

However, what does the "whenever" amount to here? It appears to be a matter of a universal quantification over some universe of cases — so what are these cases supposed to be?

On first sight, it seems that the "whenever" has to mean simply "in all possible circumstances". This construal is plausible for empirical sentences, but it is important to realize that for non-empirical, especially mathematical statements it would identify consequence with material implication: thus, any true mathematical statement would be a consequence of any mathematical statements whatsoever and any mathematical statement whatsoever would be a consequence of every false mathematical statement. This sounds quite implausible; and hence it is worth paying attention to other explications of the relevant "whenever".

1) [Bencivenga, 1999, p. 16].

An alternative idea can be traced back to Bolzano [1837][2)] who proposed, in effect, that the "whenever" should be construed as "under every interchange of certain parts of expressions in question." Bolzano's direct target was *analyticity* which, however, is interdefinable with consequence, at least within 'usual' languages:[3)]

> A is a consequence of $A_1, \ldots, A_n$ if and only if the sentence 'if $A_1, \ldots, A_n$, then A' is analytic.

Hence the generality alluded to here is not a factual one, consisting in considering possible states of affairs, but rather a linguistic one, consisting in considering possible substitutions of expressions for other expressions (and hence substitutional variants of the relevant sentences). Note that this latter generality partly emulates the former one. Consider the sentence:

> Dumbo is an elephant.

This sentence is true *wrt* some circumstances (those in which Dumbo is an elephant) and is false *wrt* others (those in which Dumbo is not an elephant). But these kinds of circumstances can be seen as 'emulated' by kinds of interpretations — namely the first by interpretations interpreting "Dumbo" as the name of an (actual) elephant and the second by interpretations interpreting "Dumbo" as the name of something else.

The problem is how to draw a line between the part of vocabulary that we should hold fixed and which we are to vary. Take the following obvious instance of consequence:

> Dumbo is an elephant,
> Every elephant is gray,
> ―――――――――――――
> There is something that is gray.

Intuitively, this is an instance of consequence for it is inconceivable that Dumbo is an elephant, every elephant is gray and yet there exists nothing that is gray. To this there corresponds the fact that substituting names of other entities for "Dumbo" and names of other sets of entities for "elephant" and "gray" we cannot make the sentences in the antecedent true while that in the consequent at the same time false. Hence it might seem

2) Or perhaps it can be traced even much farther back — see King [2001].

3) Languages having a logical connective of the kind of "if ..., then ...". Obviously, this can be taken for granted for any natural language, and it is also fulfilled by common logical languages. But, of course, there may be artificial languages (or 'languages'?) lacking any such connective.

that what we should vary is simply empirical vocabulary — however, we have already noted that this would make it impossible to make any nontrivial sense of consequence within mathematics. Moreover, take the following similar case of consequence:

Dumbo is an elephant,
Every elephant is an animal,
―――――――――――――
There is something that is an animal.

Could there be a situation in which an elephant is not an animal? Hardly — if something were not an animal, we could not reasonably call it an elephant. (Note that it would be wrong to reason: "if 'elephant' represented not elephanthood, but, say, trainhood then elephants would *not* be animals" — as Abraham Lincoln observed, a dog would keep to have four legs even if we decided to call his tail a leg.) Hence it seems that this instance of consequence is tantamount to:

Dumbo is an elephant,
―――――――――――――
There is something that is an animal.

And it is clear that the only part which can safely be varied here is the name "Dumbo" — the identity of the predicates "elephant" and "animal" becomes substantial. (In other words, it seems we must acknowledge what Sellars [1948] called *material* instances of consequence.) Hence precisely *which* kinds of expressions should we require to be varied?

Bolzano seemed to imply that analyticity emerges wherever there is a *salva veritate* variation of *any* part of a sentence. However, this cannot be correct. Take the sentence:

Dumbo is sleeping and nothing that George Bush thinks of it can change it.

It would seem that if it is indeed the case that Dumbo is sleeping then the truth of the sentence cannot be affected by substituting in it a name other than George Bush — but should we see the sentence therefore as analytic?

A possible way of avoiding this problem is to narrow our focus and concentrate on merely *logical* consequence — consequence 'in force of' logical vocabulary alone. In a sense, this does *not* narrow down the scope of instances we can consider for, as Frege [1879] already noticed, B's being a consequence of A can also be understood as its being a *logical* consequence of A and 'if A, then B'.[4)] This, admittedly, institutes no *sharp* boundary for there is no strict criterion for distinguishing between logical and extralogical words; however, loose criteria, such as topic-neutrality, are at hand.

But there is also a more serious shortcoming to the Bolzanian method, brought forth by Bolzano himself: it makes consequence depend on the contingent fact of the richness of the language in question (something might cease to be an instance of consequence by the introduction of a new expression enabling us to articulate a counterexample). We can call this the problem of the (possible) 'poverty of language'. Bolzano avoided it by basing his definition on an 'ideal' language, language *per se*, which as such cannot lack anything.

Bolzano's modern successors, in particular Alfred Tarski,[5)] avoided this recourse to an ideal language by offering an alternative solution. The point of interchanging expressions within Bolzano's approach is in gaining other sentences with the same structure as the original one; and clearly the same thing could be effected by varying the meanings of the original expressions. (Replacing the meaning of "Dumbo" by that of "Batman" has clearly the same effect as replacing the word "Dumbo" itself by the word "Batman".) Hence we could replace Bolzanian substitutions by *interpretations*: assignments of appropriate kinds of objects to expressions as their denotations. This has the advantage that we solve the problem of the 'poverty of language' without having to presuppose some such entity as an ideal language *per se*.

3 Inference

The term "inference" is in itself problematic because of its multiple ambiguities. First, there is inference in the sense of the acts of inferencing carried out by concrete people in concrete circumstances. This is not a topic we are usually interested in when doing logic — it is rather a matter of cognitive psychology.[6)] Let us use the term "inference$_1$" for the inference conceived in this sense (perhaps we could also use the term

[4)] This led many logicians to the conclusion that the former case of consequence is only a disguised version of the latter — that the step from the former to the latter amounts to disclosing a covert presupposition. However, as Sellars [1953] pointed out, it may be more plausible to see the situation completely in reverse — to see logical consequences as explicative of the material ones.

[5)] See Tarski [1936].

[6)] Criticism of logic based on the very assumption that this is what logic is after (see, e.g., Perkins [2002]) is, however, perennial.

"reasoning"; and then we would have to conclude, together with Harman [2002], that logic is not a theory of reasoning).

Second, there is inference in the sense of the relationship of *correct inferability* — as the result of the fact that we often hold the acts of inferencing (reasoning) people undertake for right or wrong. Inference in this sense can be seen as a relation between (sets of) sentences and sentences, which is, however, in the sense just mentioned intimately connected to people's doings. We will refer to it as "inference$_2$".

Third, there is inference in the sense of an arbitrarily defined abstract relation, usually generated by a system of inferential rules related to an artificial language. This last relation can be used as an explication of the previous one; but it might also be brought into being by purely mathematical interests. We will refer to inference in this sense as "inference$_3$".

Confusing these three senses of "inference" often leads to fatal perplexities within philosophy of logic; and, unfortunately, it is rather common. Let me mention briefly two very common kinds of confusion which I have discussed in greater detail elsewhere.

Firstly, there is the confusion of inference$_1$ with inference$_2$. As an example we can take the recent discussion of Brandom's[7)] inferentialism (analyzed in detail in Peregrin [2006*c*]). Fodor and LePore [1993], along with other philosophers, argue that inference cannot be a basis for meaning — however, the trouble is that while Brandom construes the term "inference" as inference$_2$, Fodor and LePore base their critique on a construal amounting to inference$_1$. Needless to say, given this, both sides cannot but talk past each other.

Secondly, in Peregrin [2006*a*] I pointed out, in effect, that there is also a common confusion between inference$_2$ and inference$_3$ which leads to the claim that inference is a purely syntactic matter. This surely holds about inference$_3$ but not about inference$_2$. While inference$_3$ is based on a wholly arbitrary set of rules; the rules of inference$_2$ are, by definition, truth-preserving. I argued that it is this kind of confusion which leads to the arguments that computers cannot think because they can have only syntax and not semantics (see Searle [1984]): computers can clearly have *rules* (and hence inference$_2$) which, if truth-preserving, are more than syntax (inference$_3$).

Sellars [1953] duly points out that the concept of inference as put forward by Carnap [1934*b*] unsuccessfully tries to ride all the three horses at the same time. Carnap's inference (as well as the inference of many

7) Brandom [1994].

contemporary logicians) is in fact inference$_3$, and Carnap makes a deep point of the fact that the relation is fully arbitrary. On the other hand, he refers to it as the relation of *derivability*, which, as Sellars points out, alludes to the fact that it expresses what *is permitted to be derived* (which would be appropriate for inference$_2$). Moreover, as what is or is not permitted are human actions, see those of inferring, it further alludes to the classification of human inferential performances, which are a matter of inference$_1$. But Carnap pays no attention to any constraints which would be implied for his definition of inference by any normative considerations or empirical studies of human inferential activities. As a result, Sellars concludes that "Carnap's claim that he is giving a definition of 'directly derivable in *S*' is a snare and a delusion."[8)]

Keeping these distinctions in mind, what can we say about the relation between consequence and inference? Consequence has very little to do with inference$_1$. Consequence is an objective matter (what follows from what does not depend on whether I or you believe it to follow),[9)] whereas inference$_1$ is purely individual. The fact that somebody announces that he will soon cease seeing me, from which I infer that he is about to kill me has very little to do with consequence; and the fact that somebody is disposed to infer "This is a fish" from "This is a cat" does not undermine the fact that (in English) "This is a fish" is *not* a consequence of "This is a cat".

On the other hand, consequence and inference$_2$ can be seen as simply two sides of the same coin (at least when we restrict ourselves to a finite number of premises). The reason is that "to be correctly inferable from" is nothing else than "to be a consequence of" — on the one hand we can infer B from A if the truth of A guarantees the truth of B; on the other hand, we can do so only if there is such a guaranty. Hence insofar as inference$_3$ is a suitable tool for the explication of inference$_2$, it is a suitable tool for the explication of consequence.

Elsewhere[10)] I called this kind of explication *criterial reconstruction*; and I still find this term very instructive. What is going on here is that a concept the extension of which is not delimited by an explicit rule (but only in terms of practical know-how) is associated with a *criterion*. The explicit criterion cannot exactly replicate the boundaries of the implicitly

8) [Sellars, 1953, p. 329].

9) Of course it depends on the existence of the shared language and thereby on certain 'beliefs' of members of the relevant linguistic community. However, this does not make consequence not objective — at least it is surely no less objective than chess or NATO or money, which all also depend on certain 'beliefs' of people.

10) Peregrin [1995].

delimited extension — for the extension does not have exact boundaries, it is more or less fuzzy.

4 The cleft

Let us review the usual reasons for asserting the existence of an unbridgeable cleft between consequence and inference. I cite the most explicit argument according to Tarski in full:[11)]

> Some years ago I gave a quite elementary example of a theory which shows the following peculiarity: among its theorems there occur such sentences as:
>
> A_0 0 possesses the given property P,
>
> A_1 1 possesses the given property P,
>
> and, in general, all particular sentences of the form:
>
> A_n n possesses the given property P,
>
> where 'n' represents any symbol which denotes a natural number in a given (e.g., decimal) number system. On the other hand the universal sentence:
>
> (A) Every natural number possesses the given property P
>
> cannot be proved on the basis of the theory in question by means of the normal rules of inference. This fact seems to me to speak for itself. It shows that the formalized concept of consequence as it is generally used by mathematical logicians by no means coincides with the common concept. Yet intuitively it seems certain that the universal sentence A follows in the usual sense from the totality of particular sentences $A_0, A_1, \ldots, A_n, \ldots$. Provided all these sentences are true, the sentence A must also be true. [Tarski, 1936, p. 411]

Can we say that this argument shows that inference and consequence can never coincide? Well, there is a rather shallow sense in which this is obvious: while it does not appear to make sense to talk about inference with an infinite number of premises (insofar as the inference relation

---

11) Cf. also my discussion of this passage in Peregrin [1997].

amounts to the correctness of human inferrings and no human could actually handle an infinite number of sentences), there seems no reason not to consider consequences with infinite numbers of premises (this follows among other things from the fact that it seems reasonable to admit that if something follows from some premises then it also follows from any *more* premises). So the *nontrivial* difference between consequence and inference would obtain only if there existed a statement which followed from an infinite set of premises without following from any of its finite subsets — in other words if consequence were not *compact*. And it seems that Tarski's example shows that this is indeed the case.

A parallel case against the identifiability of consequence with inference is entailed by Gödel's incompleteness proof. One of the direct consequences of the proof is usually taken to be the fact that for any axiom system of arithmetic, there is an arithmetical sentence which is true, but not provable within the system (intuitively, it is the sentence which 'codifies' the claim that it itself is unprovable). Moreover, as the truth of mathematical sentences does not depend on states of the world, this sentence must be true necessarily, i.e., it must be entailed by the empty set. However, it is not inferable from the empty set — hence again, inference would seem to lag behind consequence.

Both Tarski's and Gödel's cases concern, at least *prima facie*, arithmetic, hence not directly *logical* consequence. However, while in the former case this seems to be essential (though Edwards [2003] argues that what Tarski had in mind was a specific, disguised case of logical consequence), the latter is easily convertible to the domain of pure logic. It would imply that the undecidable sentence is a logical consequence of the axioms of arithmetic (and within second-order logic where arithmetic is finitely axiomatizable it thus follows logically from a finite number of sentences) without being inferable from them.[12)]

These facts made the majority of modern logicians accept Tarski's proposal to investigate consequence (and truth) not via inference (proof theory), but via *model theory* — Tarski's[13)] widely accepted explication of consequence is:

A is a consequence of $A_1, \ldots, A_n$ if and only if every model

[12)] It would be argued that the truth of Gödel's sentence does not follow from the axioms of arithmetic, for it is not true in all models of the axioms. However, this objection turns on the first-order regimentation of arithmetic (which allows for non-standard models), which cannot be equated with arithmetic as such. Within second-order arithmetic, there are no models of the axioms in which Gödel's sentence would be false.

[13)] Tarski [1936].

(i.e. verifying interpretation) of $A_1, \ldots, A_n$ is also a model of A.[14)]

This is a reconstruction of the concept of consequence which is also ‘criterial’, but in a weaker sense: it not only does not allow us to always decide whether a given sentence is a consequence of some other given sentences, but it even disallows us to generate all instances of consequence. This also opened up the problem of the relationship between consequence and inference for individual calculi as an important research topic — to what extent can we turn the ‘loosely criterial’, model-theoretic delimitation of consequence into the ‘strictly criterial’, proof-theoretic one?

5 The bridge of meaning

The cleft thus opened between model theory and proof theory paved the way into vast new realms of interesting mathematics; however, should we read it as showing that consequence is wholly independent of inference, being only — in a better or worse way — mimicked by inference? There is a substantial argument against this conclusion, an argument turning on the concept of meaning.

Why is a sentence, say “There is something that is gray”, a consequence of other sentences, say “Every elephant is gray” and “Dumbo is an elephant”? This must clearly be in virtue of the meaning of all the sentences involved. Hence how come that the sentences do have the meaning they have? Again, this must be a matter of the way we, as the members of the relevant community, treat them — sounds and inscriptions do not mean anything without our endeavour. Hence, what have we done to the sentences that they mean what they do?

There are two popular kinds of answer to this question, both backed by a host of advocates. The first is that our words mean something because we let them *stand for* this something; the second is that they mean something because we use them in a certain way.[15)] In this general form, these two answers do not seem to be intrinsically incompatible — perhaps we use words in such a way that they come to stand for something. But the usual elaborations of these two kinds of answer do lead to incompatible approaches.

The first standpoint standardly leads to the view that we make words meaningful in that we *conventionally establish* that they become *symbols* for various kinds of extralinguistic entities which thus become

14) See also Priest [1999].
15) See Peregrin [2004].

their meanings. The second one leads to the view that meanings of words should be construed not as something represented by them, but rather as the very ways in which they are used, as their functions or roles within our linguistic transactions.

How would these two answers fare when answering the subsequent question about the origin and nature of consequence? It seems that the first must claim that any semantic relation between expressions, such as that of consequence between statements, cannot but be a mimic of a relationship between the entities they represent. One may, for example, want to claim that sentences entail other sentences because they stand for facts and some facts contain other facts — hence that the relation of consequence is a linguistic reflection of the nonlinguistic relation of containment.

Is this answer viable? Hardly. It would mean that "Dumbo is an elephant" entailing "Dumbo is an elephant or is a rhino" is the results of the following three facts:

(1) that we have introduced the former sentence as a name of a fact;

(2) that we have introduced the latter as a name of *another* fact (*entirely independently* of the first naming); and

(3) that the first of these facts happens to contain the second.

Needless to say, even if we disregard all problems connected to the concept of fact and admit that a sentence like "Dumbo is an elephant" can be reasonably seen as a name of a fact, it is hard to lend any credibility to a theory which assumes that the fact named by "Dumbo is an elephant" can be only empirically discovered to be part of the fact named by "Dumbo is an elephant or is a rhino".

It seems much more plausible to assume that whatever the meaning of "Dumbo is an elephant" is, if it gets combined with another sentence by means of "or", the meaning of the result derives from the meanings of the parts in a way determined by "or", and the determination consists especially in (though it is perhaps not reducible to) the consequential links between the complex sentence and the parts. Hence we move to the second paradigm of meaning: meaning consists in the way an expression is used, and the meaning of at least some sentences comes purely from the *inferential rules* which govern them. From this viewpoint, meanings are to be identified by the roles conferred on words by those very rules (just as the roles of chess pieces consists exclusively in the fact that we treat them according to the rules of chess). And the rules which are basically relevant from the viewpoint of meaning are inferential rules — logical (and perhaps some other) words are governed exclusively by this kind of rules, whereas empirical words are governed by these rules together with different types of rules maintaining 'links between language to the world', i.e., stating in what circumstances it is correct to utter some sentences and perhaps what kinds of action are appropriate in response to an utterance.

If we agree with this, then conferring a meaning on a logical word is accepting a basic inferential pattern governing the usage of some sentences containing the word. Conferring the usual meaning on *and* is accepting that it is correct to infer 'A and B' from A and B, and that it is correct to infer both A and B from 'A and B'. Now, however, it seems that consequence must be entirely brought into being by inference: for consequence is a product of meaning and meaning is a product of inferential rules. Is this viable?

6 From an inferential pattern to the relation of inference

Note that the inferential pattern associated with a statement is not a list of everything inferable from that statement and everything from which the statement is inferable. Rather, it is a collection of the most basic inferences which are supposed somehow to 'establish' many others. However, in what sense do they 'establish' them? Take conjunction. The pattern governing it is:

$$A \wedge B \vdash A$$
$$A \wedge B \vdash B$$
$$A, B \vdash A \wedge B$$

and we take it to 'establish' a number of other inferences, including:

$$A \wedge B, C \vdash A$$
$$(A \wedge B) \wedge C \vdash A$$
$$A \wedge B \vdash B \wedge A$$
$$\text{etc.}$$

What kind of inferences are these? Let us first sharpen our terminology. Where S is a class of statements, an *inference* over S will be an ordered pair of a finite set of elements of S and an element of S. An *inferential relation over* S will be a set of inferences over S. What we are now looking for is a characterization of how a basic inferential relation ('inferential pattern') R naturally leads to ('establishes') a wider inferential relation R^*. There are several ways of approaching this problem. The most common characterization is that R can be taken to underlie the wider relationship of 'provability in terms of R'. Hence, we may say,

> $\langle M, s \rangle \in R^*$ if and only if there is a sequence $s_1, \ldots, s_n$ of statements such that $s_n = s$ and each s_i is either an element of M or there is an $\langle X, s_i \rangle \in R$ such that $X \subseteq \{s_1, \ldots, s_{i-1}\}$.

We may reach a variant of this answer by characterizing the general relationship between R and R^* in terms of closure conditions: if we find some operations on $\mathrm{Pow}\,(S) \times S$ such that R^* is always the smallest superset of R closed to these operations. We may call these operations *metainferences*, for they infer inferences from inferences. And in fact we do not need to seek them, for they were disclosed long ago by Gentzen [1934] — they are nothing else than the structural rules of his sequent calculus. In other words, R^* is the smallest set of inferences which contains R and contains[16)]

(a) $\langle \{s\}, s \rangle$ for every s,

(b) $\langle X', s \rangle$ whenever it contains $\langle X, s \rangle$ and $X \subseteq X'$,[17)]

(c) $\langle X \cup (X' \setminus s), s' \rangle$ whenever it contains $\langle X, s \rangle$ and $\langle X', s' \rangle$ and $s \in X'$.

It is very plausible to assume that it is via these closure conditions that an inferential pattern establishes a wider class of valid inferences than those constituting the pattern. (Though there are also well known objections — like the relevantists' rejection of (b) etc.) However, might there not be *an additional*, less obvious, way in which an inferential pattern can establish *a still wider*, 'quasi-inferential' relation?

7 Consequence is quasi-inference . . .

In fact, this idea is not without precedent. In his *Logische Syntax der Sprache* Rudolf Carnap distinguished between *Beweisbarkeit* (*provabili-*

16) As we treat inferences as pairs each consisting of a *set* of statements and a statement, we do not need the contraction and permutation.

17) This rule may, or may not, be restricted to finite sets of premises — in the latter case it leads us to the abandonment of the realm where we can speak about inference straightforwardly; however, as it cannot escape compactness, this abandonment is merely trivial.

ty) and *Folgerung* (*consequence*), both being established by the axioms and rules of the relevant system. The difference between them was precisely the one hinted at above — the way leading from the system to consequence is a generalization of the one leading from it to provability. In the first of the two model languages presented in Carnap's book, the distinction consists merely in the fact that a general number-theoretic sentence is a consequence of all its number-theoretic instances, though it is not provable from them. (In the second one he replaced his 'quasi-inferentialist' approach to consequence with a wholly different one, basing his definition of consequence on an almost model-theoretical definition of analyticity — see [Coffa, 1991, chap. 16] for a thorough discussion.) This amounts to the infinitary rule of inference (where N is the predicate delimiting natural numbers):

$$P(0),\ P(1),\ P(2),\ \ldots\ \vdash (\forall x)(N(x) \to P(x))$$

or, equivalently:

$$N(x),\ P(0),\ P(1),\ P(2),\ \ldots\ \vdash P(x). \qquad (\omega_1)$$

Thus, it seems that Carnap simply took the very same case which troubled Tarski and made it into a (quasi-)inferential rule. In fact Tarski did consider the very same possibility. (Tarski [1933] realized that the addition of this rule would align inference with consequence, but rejected considering the rule as a rule *of inference*, for it is "infinitistic". However, it seems to follow that at least for arithmetic he would see the 'quasi-inference' constituted by the normal inference plus the ω-rule equivalent to his model-theoretically defined consequence.)

As I pointed out elsewhere,[18] there is also another available explanation for how an inferential pattern can be envisioned to establish a relation of consequence. Arguably, at least some inferential patterns can be seen as sorts of exhaustive lists, as exhaustive lists of either the principal grounds of a given sentence or its principal corollaries. Thus, putting forward:

$$A \vdash A \vee B$$
$$B \vdash A \vee B$$

as a pattern might be read as giving a list of sentences from which $A \vee B$ follows, which is exhaustive in the sense that any C which also entails $A \vee B$ must entail either A or B.

[18] Peregrin [2006*b*].

This particular case is not interesting from the current viewpoint, as it establishes no new inferences or quasi-inferences (though it does manage to pin down the meaning of $\vee$ to classical disjunction which is impossible otherwise). However, in a similar way, putting forward the induction scheme *as a self-contained inferential pattern* may be read as not only stating that 0 is a number and a successor of any number is a number, but also that nothing else is a number. This is to say that stipulating the pattern:

$$\begin{aligned} &\vdash N(0) \\ N(x) &\vdash N(x'), \end{aligned}$$

we implicitly stipulate that nothing is a number unless it follows from these inferences, i.e., unless it either is 0, or is accessible from 0 by iteration of the successor operation. It might seem that this requirement can be expressed in terms of the stipulation that if $N(x)$ then either $N(0)$ or there is a y such that $x = y'$, but this is not the case as this would obviously be valid even in non-standard models of arithmetic. What is needed to express the requirement in its full strength is the infinitary sequent (the notation is the usual Gentzenian one — 'if every item before $\vdash$, then at least one item thereafter'):

$$N(x) \vdash x = 0,\ x = 1,\ x = 2,\ \dots\ . \tag{ω_2}$$

And it is clear that this is precisely what is needed to exclude the non-standard models and hence pin down the model of the theory to the intuitive natural numbers.

Now it can be shown that (ω_1) and (ω_2) are equivalent (and hence that the assumption of the induction rule is the same assumption as the assumption of the validity of the ω-rule). To do so, suppose first that (ω_1) holds and take $P(y)$ to be $x \neq y$:

$$N(x),\ x \neq 0, x \neq 1,\ x \neq 2,\ \dots \vdash x \neq x.$$

Using the Gentzenian rules of negation elimination, this yields

$$N(x),\ x = x \vdash x = 0,\ x = 1,\ x = 2,\ \dots$$

and hence, as $x = x$ is valid, (ω_2). Now suppose, conversely, that (ω_2) holds. As it is the case that

$$P(0),\ x = 0 \vdash P(x),$$

we can use cut on this sequent and (ω_2) to obtain

$$N(x),\ P(0) \vdash P(x),\ x = 1,\ x = 2,\ \dots\ .$$

In the same way we can use

$$P(1),\ x = 1 \vdash P(x)$$

to further obtain

$$N(x),\ P(0),\ P(1) \vdash P(x),\ x = 2,\ \dots$$

and ultimately

$$N(x),\ P(0),\ P(1),\ P(2),\ \dots\ \vdash P(x),$$

i.e. (ω_1). Note that as we consider the list on the sides of $\vdash$ as representing *sets*, contraction is automatic.

8 ...but quasi-inference is not really inference!

Consider (ω_1) once more. It is clearly not a case of logical consequence — it says that "Every natural number is P" follows from $\{n \text{ is } P\}_{n=1}^{\infty}$. But it is clear that "Every natural number is P" does *not* follow from $\{n \text{ is } P\}_{n=1}^{\infty}$ logically; for the logical form of the argument would be

$$P(T_1),\ P(T_2),\ \dots\ \vdash (\forall x)(Q(x) \rightarrow P(x))$$

or

$$Q(x),\ P(T_1),\ P(T_2),\ \dots\ \vdash P(x),$$

which is obviously invalid. If it had a finite number of premises, based on the terms T_1, T_2, ..., T_n, then it could be logically valid with one more premise added, namely

$$(\forall x)(Q(x) \rightarrow (x = T_1 \vee \dots \vee x = T_n)),$$

guaranteeing that T_1,..., T_n are all the Qs there are. However, if the premises and hence the terms are infinite in number, such a premise cannot be formulated and the argument can not even be made valid in this way.

Hence it seems that (ω_1) cannot pretend to *logical* validity, but merely to some more general kind of validity, assuming the intended content of the term *natural number* and all the numerals; and this is also how Etchemendy [1990] interpreted the original proposal of Tarski [1936]. In fact Tarski confessed that he did not see a sharp boundary between logical and extralogical words — and concluded that perhaps the line, and hence the concept of logical consequence is wholly arbitrary. So imagine we draw the boundary so that the arithmetical vocabulary falls on the logical side. Can we say that Tarski's case shows that this ('logico-arithmetic') kind of consequence cannot be captured by inferential rules?

However, even this is implausible. For can we really say that (ω_1) is an intuitively valid instance of consequence? Well, if intuitive, then only for those who understand what the three dots are supposed to be a shorthand for. But what are they a shorthand for? Shorthands save us labor — allow us to write something short instead of something long. But the three dots of the above inference schema are *indispensable* — we cannot expand them to a full wording. This was observed by Wittgenstein:

> We should distinguish between the "and so on" which is, and the "and so on" which is not, an abbreviated notation. "And so on ad inf." is not such an abbreviation. The fact that we cannot write down all the digits of — is not a human shortcoming, as mathematicians sometimes think. [Wittgenstein, 1953, §208][19)]

But do we not *understand* the inferential pattern above and do we not *have* the intuition that it is valid? Granted; but how do we manage to understand what is beyond the three dots? We have learned how to write the natural number sequence starting from 0 and going on as far as requested; and analogously we know how to continue the beginning of the list of statements in the antecedent of (ω_1). But how do we know where to stop? This is the point: we are *never* to stop — and hence there is no way to expand the three dots except with the help of a universal quantification:

> The collection of premises consists of '$P(x)$' for every natural number x.

Informally expressed, the rule amounts to

$$\frac{(\forall x)\{\ulcorner P(x)\urcorner\}}{(\forall n)P(n)}$$

[19)] Cf. also [Peregrin, 1995, chapter 9].

where, of course, the quasi-formula above the line is not an object language statement, but rather a metalinguistic description of the set of premises.[20)]

Hence, one is tempted to say, the instance of consequence which is allegedly not capturable as an instance of inference, *is*, after all, an inference — but one transgressing the boundary of languages, namely that between metalanguage and the object language. It can be seen as the explication of the universal quantification of the object language only if we take for granted that we can make a free use of universal quantification of the metalanguage. In this sense, consequence can be said to be 'a kind of' inference (or perhaps better quasi-inference), which, nevertheless, does not fall under what is normally understood as inference proper.

9 Conclusion

It is misguided to see the relationship between consequence and inference as that between the subject matter of logical investigations and their tool; to see consequence as something that exists 'out there' and inference as something that we devise to approximate it. It is, I have argued, inference that should be seen as the basic concept, underlying even the concept of consequence. Therefore, it is more adequate to see consequence as a 'quasi-inference', as what becomes of inference if we relax the concept of inferential rule. This also vindicates inferentialism: the view that it is inferential patterns that furnish our words with their semantics and that are, consequently, responsible for the entire logical structure of our language.

[20)] If metalanguage becomes incorporated into the object language (via gödelization), then such rules can underlie Feferman's *reflection principles*, cf. Feferman [1962].

Two

Inferential Semantics in the Pragmatic Theory of Truth and Reference

Pirmin Stekeler-Weithofer

1 Task and method of inferential semantics

We can read Carnap's *Aufbau* as a counter-model to metaphysical correspondence theories of meaning and truth. The formal backbone of the model is elaborated in Carnap's *Logical Syntax of Language*. The basic concept is that of Hilbert's implicit definitions. The meaning of sentences and subsentential words is given in terms of their deductive roles or use in a holistic system — an axiomatic theory. Language itself — or rather its semantics — is seen as such a system of *inference rules*. On the ground of Modus Ponens, the detachment rule, an axiomatic system can be seen as a generator of formulas or sentences. Its theorems can be directly read either as premises or as inference rules with or without premises that we are entitled to use by the system.

According to *Aufbau*, the empirical point of touch where language or theory meets the real world is a *Humean* kind of memory of resemblances of sensations. Judgments of resemblance are presupposed in our concept of perception. Perceptions are sensations that can be reported or described by ('non-inferential') observation sentences or 'Konstatierungen'. They can be predicted or expected if they are consequences of (logically complex) statements in the framework of a theory.

In what follows, I try to sketch the philosophical place of an inferentialist account of the concepts 'meaning', 'intention', 'representation', and ‘truth’ and the reason for some major changes in the overall picture. I do this in view of the following questions: Which problems lead us to an inferentialist approach to semantics? Which problems lead to the behavioural, operational, cognitive, social, and normative turns given to it by Quine and Sellars, Dummett and Davidson, Rorty and Brandom?

The basic *opponents* of inferential semantics are the two main versions of classical metaphysics. The first is *Platonism* with its different versions of correspondence theories of truth and representation theories of meaning and reference. The second is *Cartesianism* with its subjectivist approach to thought and experience. As I see things, modern physicalism is a version of Platonism. It turns mathematical entities into ‘real’ objects. Empiricism in its Lockean and Humean form is a version of Cartesianism. Cartesianism *presupposes* the concept of objectivity and truth, reality and existence. It only asks *what we can know* — if we accept our actual human situation as given. It does not ask what knowing, reality, and truth *are*. Even Kant is still too much concerned with the *epistemological* questions: ‘What can we know?’ and ‘What are sufficient *reasons to believe* something?’ A more profound question would be to ask what it is to believe and what we do when we distinguish belief from knowledge, appearance from nature as such.

Rorty’s critique of epistemological representationalism and of its idea of mirroring nature attacks both, Platonism and Cartesianism, empiricism as well as what a mere cognitivistic reading of Kant says. For many readers, Rorty’s debunking of epistemology may be hard to accept. But his deep insight is the following: Precisely the brave intention of limiting the scope of human knowledge leaves a back door open for metaphysical faith. The seemingly humble acknowledgement that human knowledge is fallible hides a not at all modest belief in a world behind the scene of human experience, practice and life.

It may be difficult, indeed, to see how critical philosophy has to turn from mere epistemology and faithlike theories about real and possible worlds to a *conceptual analysis* of *onto-logical* commitments. The problem is to make ontological presuppositions explicit, to find *hidden metaphysical* beliefs and to show why they are misleading. Here, however, I do not intend to explain or defend these claims about an implicit metaphysics in modern epistemologies further. They only show where the whole consideration belongs to and what is at stake. The basic *contention* of inferential semantics is this: Semantics altogether is *not* concerned with a *metaphysical* relation between

(1) signs or symbols or words or sentences on the side of semiotic

and

(2) *linguistic practice*, thoughts or ideas on the side of *mental representations*, things and

(3) properties, states of affairs and events on the side of the *world*.

To presuppose such metaphysical relations always means not to go deep enough in a reflective analysis of what we do when we *talk about* meaning(s) and truth(s) and what we do when we *use* signs and words in different modes of speech and contexts. An important further question is what we do when we *judge in retrospect* that a given use (a token) or usage (a form or type of possible use) is meaningful, correct, or 'true'.

The basic *goal* of inferential semantics is to *replace* the assumption of a metaphysical relation of representation between words and ideas (or thoughts) and between words (or ideas) and their objects by an analysis of the things we do with words and by an investigation of how we make up *worlds as realms of (possible) objects, properties and states of affairs.* The question is how to distinguish between the real and the unreal, between the actual and the possible, between the concrete and the ideal. We should not presuppose these distinctions as given and clear. They have to be placed into the context of our usage of words in a practice of individual and joint experience(s), individual and cooperative action(s) and life(s) — and not mere solipsistic, animal, behaviour.

2 Problems of empiricism and behaviourism

Descartes' *subjective turn* had been aiming for a better understanding of the difference between the presupposed performative stance of authentic human thinking and any claim about the 'outer world'. *Hume* wants to describe human actions and cooperation just as we describe animal *behaviour*. But by this very move, the progress of the Cartesian insight into the primacy of the perspectival stance of the subject, of the person as a speaker, hearer, actor and juror, is taken back. Hume's critique of a mystification of thinking has lead us to a kind of *disengaged attitude* with respect to the behaviour of others and of ourselves — as we find it until today in anthropological descriptions and explanations from a third person's point of view. The danger of this third person stance is this: we forget about the fact that any such point of view *implicitly* presupposes an *engaged* understanding of *what we* do when we *describe and explain behaviour*. We as describers, composers of natural history and science and we as readers, already are situated in the world. Therefore, we should defend the following Cartesian and Leibnizian (and, for that matter, Heideggerian) idea. In the *one world we live in*, there are two worlds to be addressed in different ways:

(1) the world of objects and events we talk *about* and

(2) the world of persons we talk *with*, the world of Us.

The first world is usually called 'nature', the second does not have a suitable name because all the grand titles like Kant's 'Reason' with capital "R" or his or Fichte's '(transcendental) I or Self', or Hegel's '(subjective and objective) Spirit' name only some of its particular aspects. For my consideration I will be content with using the title-world 'Culture' as it was proposed by the Neo-Kantians to serve the purpose at stake. In fact, any *philosophy of mind and language* is parochial and unreflective if we do not embed it into a *philosophy of culture*.

The basic problem addressed by the talk of two worlds in the one world, nature and culture, is to see that the *personal stance* is a very peculiar one. It cannot be understood outside culture and practice. This holds in particular for the personal relation of a speaker to a hearer. In the end, there is no content to be understood and no claim possibly true if we do not already presuppose that we take part as persons in a public and cooperative enterprise of speaking and acting, controlling assertions, proposals and claims, of describing and explaining experience. In a sense, this is an insight of Descartes, expressed in a non-Cartesian way.

Moreover, the very concept of *experience* is misunderstood, as Kant already knows, if we see experience as a set of individual *sensations* or even as real or possible or expected *perceptions*. Experience is always already 'formed' by a system of conceptually ordered and culturally learned common 'knowledge' (or common 'sense'), by 'theory', as the more modern metaphor has it. Such theories are needed to form expectations and make intentional actions. Their meaning (significance, importance) lies in their function, the orientation of judgements and actions.

The problem of philosophy now seems to be this: The basic difference between behaviour and action, between animal and personal life, between the 'mere sentience' of 'Them' and possible sapience of 'Us' lies outside the scope of the peculiar form of mere description of behaviour and theoretical explanation. We cannot give an account of it in a mere descriptive form. This may sound dogmatic. But it is not difficult to see the reasons why we should be unsatisfied with any 'disengaged account' that only describes behaviour and tells stories about the genesis of the cultural world of Us: We first have to understand the peculiarity of this 'world of Us' before we can explain its genesis. This, of course, is a Kantian argument. In fact, a structural account of present affairs, especially of understanding and thinking and dealing with the difference between belief and knowledge methodologically precedes any cosmological story and explanation of natural evolution and cultural development. On what grounds should we believe the stories of cosmology? What consequences are we supposed to draw? And why should we be committed to do so? Just because we do not have a 'better' story? It is not only the difference between sentience and sapience we have to account for, it is the peculiar form of accounting for sapience we need to learn. What we have to learn is why it is not enough to remind us of the peculiar social behaviour which is presupposed in any epistemological game of making knowledge claims or believing something and giving and asking for reasons.

On the other hand, the *empiricist* or *behavioural* or *Humean* turn in philosophy was and is an important first step away from *mentalism*. But empiricism was (and still is) in need of quite some further *turns* in order to put things straight. The first turn is most famously called the linguistic turn. It starts with Frege's anti-mentalist and, at least in my assessment, anti-Platonist logical analysis of abstract objects and mathematical truths. (I admit, however, that there are important differences between what Frege really accomplished and what he was after resp., what I would attribute as an intention to him.) Russell's, Wittgenstein's and Carnap's (three different) versions of *logical atomism* have turned Frege's logicistic analysis of mathematical language and truth into *logical empiricism*. Logical empiricism is a movement into the direction of an *anti-metaphysical* and *anti-representationalist* reflection on meaning, reference, and truth — not only for mathematical statements. The idea is that *logical form* of truth- or verification- or at least falsification-conditions on the ground of *perceptions* and *basic sentences* ('Konstatierungen') constitute meaning, reference, and truth. This logical form is learned implicitly in language use. It can be made explicit by logical analysis or reconstructions of semantical deep structure.

Quine's well known critique of the 'two dogmas of (logical) empiricism' has shown, however, that *logical atomism* has serious problems with its two basic concepts, *logical form* and *empirical basis*, or, if we apply the differentiation to natural language, with the distinction between *conceptual schemes* and *perceptual experience*. Applied to sentences, the question is how to differentiate between analytical statements that are true for conceptual reasons and synthetic statements that are true or false by empirical or *a posteriori* reasons. The result of this criticism was and is a behavioural, holistic, procedural and cognitive turn in the analysis of linguistic meaning and competence.

The third and more recent turn could be placed in the context of a critique of *social atomism* in Quine's behaviouristic picture of language use and in Davidson's cognitive picture of language learning. To label Davidson's interpretationism as 'cognitive' and 'atomistic' might seem too crude. It abstracts from Davidson's later developments with their open or hidden retreats. But it helps to remain oriented at large. For we can say now that the semantical theory of pragmatism gets a peculiar '*social*' twist if we follow Rorty and say that the basic function of the word "true" lies in the *appraisal* of what is taken to be true by the speaker. This twist is, in a certain sense, not behaviouristic (non-Quinean) and no interpretationism (non-Davidsonian). The idea is this: The speaker is *committed* to what he appraises and he *entitles* the hearer to rely on it. This is, in a way, an assimilation of *true (enough)* to *good (enough)* and stands as such in the tradition of William James. In a sense, it is similar to Stevenson's appraisal theory of the basic use of the word "good": To say that some action is good means to say something like: "I approve it — and you should do so too."

It is to be expected that such a *pragmatist* analysis of the basic meaning of "true" and "good" will immediately be attacked — because it does not seem to distinguish between the True and the Good, nor between knowledge and mere persuasion or between justice and mere propaganda. But the appraisal theory of the True and the Good is a *core insight* of the American Pragmatism. It is the first and most important step in a reflection of the real constitution of the concept of truth and the idea of the good.

The social approach to a pragmatistic semantics has, however, openly to face the problems raised by the attacks mentioned. The performatory function of appraisal certainly does not exhaust the use of the word "true". The appraisal theory of truth does not solve at all the problem of *meaning* and *reference*, of what there *really* is or exists independently of our subjective thoughts and beliefs. Moreover, there is a danger of *wishful thinking* or *instrumentalism*, of *relativism* and of *linguistic idealism.* Some errors and lies can be beneficial — an argument Nietzsche is famous for. Russell uses it in his attack of William James. Even if a large society takes something as true, this does not make it true. "Saying does not make things so." Therefore there is a challenge to avoid or overcome relativism and arbitrary beliefs without leaving the insights of the previous turns, the linguistic, behavioural and social turns behind. We should resist, in fact, the tendency to return to some open or hidden metaphysical faith, to any sort of objectivism, be it theological, historical, or expertocratic. Instead, it is necessary to develop a fourth, a *normative*, spin to the social twist of pragmatism. According to it, the use of phrases like "is correct", "is true", "does exist" or "does really refer" belong to a normative and, if you wish, meta-level, discourse in which we *assess* claims or assertions in a kind of control game of giving and asking for reasons. On this ground we can account for the difference between appraisal and correctness, 'subjective' commitment and 'objective' entitlement. This and more is shown in Brandom's work *Making It Explicit*. After this book, the pragmatic twist of the linguistic turn in our analysis of meaning and truth, of belief, knowledge, and intention cannot be turned back, especially not by arguments of the type Bertrand Russell could still use against William James.

Belief and knowledge, understanding and acknowledging take place in a kind of already pre-formed or pre-existing deontic language game. As individuals, we learn to take part in undertaking commitments, granting entitlements to theoretical (verbal) and practical (action-orienting) inferences and assessing the commitments of a speaker in a kind of control-game. These control games of 'scorekeeping' may lead me as a hearer to an actual acknowledgement of the commitment. The idea is to define the difference between belief and knowledge, appearance and truth along these lines. But this is possible only if we add appropriate concepts of counterfactual *idealizations*. Idealization is a peculiar verbal and logical operation which I would like to call a *de-limitation*. In fact, I miss the analysis of the logical form of idealization in Brandom's book. Without it, the idea and ideal of absolute, situation-independent, timeless, truth cannot be understood in pragmatic terms. We need, and we use, ideals in order to make generic, infinite, delimited, goals of a cooperative enterprise explicit. Or rather, by talking about such an infinite goal we talk about directions in orientations, just as we talk about such directions in mathematics by talking about infinite points in space. The cooperative enterprise in question is the development of knowledge, science, technology and all related human institutions like language and morals, education and the legal system. We describe the ideal goals by ideal statements. But we always need to deal with such descriptions in a proper way. We have to limit their use in real orientations. We do so by judging what is relevant in the singular case. It is these judgments that explain what it means to say that something is ideally true or true in principle but not in reality. Therefore, an analysis of the logical form of relevant judgement is necessary for a full understanding of our (ideal and real) talk about truth and meaning, the good and the just.

In practice, we give and ask for reasons when we control the truth of a claim. This practice of assessment defines, so to speak, the real difference between '*Schein*' and '*Sein*', or rather, the difference between applying norms of correctness in judgment and action and mere attempts to do so. In retrospect we talk about following rules of criteria. But the very concept of a rule is defined in this framework. I.e., a rule is an expression that we can use properly or correctly. Only in such a normative framework defining correctness we can distinguish between (merely subjective) beliefs and (jointly accepted and corroborated) knowledge, between merely purported reference and sufficiently successful reference.

3 Developments of pragmatism

In a sense, what I have said until now amounts to a rough differentiation between three main types of pragmatisms in modern philosophy after the linguistic turn. It can be outlined by relation to earlier stages of philosophy. The first type stands in the tradition of David Hume and comprises the 'behavioural' approach of Quine, the 'procedural' of Dummett and the 'instrumental and interpretational' of Davidson. Rorty's social and educational pragmatism stands in the tradition of John Dewey. Brandom's pragmatic normativism stands in the tradition of Kant and Hegel. But it was Hegel, not Kant, who turned from epistemological questions to a logical analysis of the real, and social, and normative meaning of our claims about what there is, and of our assessments of such claims. The following remarks try to give an outline of how we arrive from Carnap's formal inferentialism to Brandom's social normativism.

In order to overcome the subjectivist myth of the given (Sellars) and the two dogmas of sensation-based empiricism (Quine) we have to accept that observation is already theory-laden (Neurath, Popper) and that in a system of verbalized 'knowledge' like in an axiomatic system there is no clear distinction between conceptual (analytical or linguistic) truth and factual material truth. On the other hand, any theory is itself a kind of linguistic behaviour that has to be understood pragmatically, namely in terms of disposition of accepting sentences and inferences in appropriate situations. This is, very roughly, Quine's 'pragmatistic' turn in philosophical semantics with its thesis of indeterminacy of translation.

According to Davidson, radical translation begins at home. Thus it turns into radical interpretation. Like Quine, Davidson wants to give an account for how we learn a common language. His interpretationism turns, in the end, into a cognitive theory of language-acquisition — in a similar way as Chomsky had turned structuralist linguistics into a part of cognitive science.

In fact, Davidson wants to present a formal framework by which we can imagine how we as singular learners can form a theory of syntax and a theory of truth if we are exposed to a pregiven social practice of language use. To do this, Davidson uses Tarski's axiomatic definition of truth for a formal, i.e. axiomatic, system 'upside down'. He assumes that the sentences we hear are (mostly) syntactically correct and semantically true. On this assumption he wants to give an account of how we may, perhaps, identify, for example, the logical connectives like "and" or "or" or "not" or "for all" and how we may, perhaps, identify semantical deep structure or structural truth conditions (down to subsentential terms) and inferential postulates (e.g., for concrete one- and more-place predicates). In a sense, Davidson presupposes a kind of semantical language-learning machinery or universal semantics just as Chomsky does for syntax. With Chomsky and Dummett, Davidson wants to explain how we can learn to produce and understand infinite many sentences and utterances correctly by being exposed to finite examples. This seems to be possible only if we learn operational schemes of producing and analysing well-formed syntactic structures and of drawing and controlling correct inferences in such a way that these schemes can be applied recursively. I for my part doubt that there can be an explanation of the acquisition of linguistic competence in this sense. This is so because we do not generate and control 'correct' utterances by applying prefixed rules. Linguistic norms of correctness do not exist internally as rules in the mind or brain as computer programs do. They exist in an external way, just as other *social norms* do.

This is the deep insight of Brandom's normative turn. In it, he really follows Wittgenstein who had said that we should not look for abstract meanings nor for processes in the brain, but for real use. Hence we see that getting rid of Platonistic interpretations in our talk about meaning does not suffice. We have to put the logical and methodological order of rule following and the use of symbols for rules straight. A third thing is the differentiation between the (individual) faculty to follow schematic rules and a more general practice in which social norms are 'applied' in non-schematic ways.

The basic questions of Brandom's account is indeed this: What are rules? What are social norms? What do we do when we (learn to) use them? And what do we do when we make them explicit? The answer is this: We learn to take part in an already established social game with its 'implicit' norms of correct play and incorrect moves. These norms appear as norms in a social behaviour that can be seen as 'sanctioning' incorrect moves in one way or other. It is not the expression R of a rule or norm or criterion but the way we distinguish between correct and incorrect moves (perhaps following the expressive intention to follow rule R) that defines what it is to follow the rule. Hence, implicit rule following can be made explicit by use of 'names' or 'symbols' for the rules and a practice of controlling the relation between symbol and performance in an appropriate practice. If we follow explicit rules we say that we do this consciously. This means that a possibility or faculty of (common) control is actualized. Accordingly, Brandom says that intentions (acting intentionally, acting with an intention and acting for an intention) are *normative* and presuppose normative control of entitlement and commitment. If there is no 'rule' or 'policy' or 'scheme' for my behaviour that I can follow correctly (or miss), then there is no (intentional) action at all. We see here also the importance of Kant's insight into the distinction between *saying* that one does X (or should do X) and *doing* X (or not). More precisely, we have to account for the difference between the following three cases:

(1) In the case, in which I praise a maxim of form of action, I *explicitly* acknowledge it as a rule for orientation — but this does not mean by itself that I really act according to it.

(2) Only in what I do I show practical acknowledgment of a certain action scheme. In a sense, this acknowledgment is implicit.

(3) Acting in accordance with what I verbally acknowledge is, therefore, a peculiar form of consequential or inferential coherence.

Saying something and acting according to some generic action, lying and being sincere, making an error or being correct or true: all these things are possible only if we presuppose a common realm of (at first dialogical) assessment of correctness. We presuppose a normative practice in which there are entitlement and commitment and inferences of entitlements and commitments that cross the limits of a person not only insofar as they can be passed from person to person but insofar as the assessment of correctness must always be thought of as interpersonal — if only as a possibility, when I talk to myself or play a multi-person-game alone (as I can play chess with myself).

Brandom's normative turn makes it even clearer than Wittgenstein why there is no pure descriptive perspective by which we can give an account of the concept of correct rule following. We do not only need the perspective of an actor, but a second perspective of a person who can assess the correctness of my doing — if only by not objecting to what I do. But we need even more. Giving and asking for reasons presupposes a whole culture of human cooperation. But despite this progress of Brandom's normativism, there is still a problem. For it seems as if Brandom, like Quine, Davidson, and Rorty, gives physics the last word in all explanations and all assessments of reality and objectivity. By this move, physics becomes the ultimate place for judgement about what things really are and what objective truth is. Physics becomes *the* ontological science: In physics we decide what there really is and what not, what is concrete and what is abstract. In fact Rorty wants to explain culture as a continuation of natural evolution. Brandom wants to explain normativity on the ground of the one and only physical world, mediated by social behaviour of positive and negative sanctions as reactions to the behaviour of others. There is a danger, indeed, that now the whole enterprise of giving a non-Platonist account of the human world collapses.

We arrive at a different perspective if we would put our philosophical reflections into the context of a therapeutic and educational enterprise. Such point of view which is more modest starts in the middle of life and practice and takes part in improving a certain culture of language use and language design for the purpose of a better, more reflective understanding of science, culture, and ourselves. If we do this, we avoid mixturing and muddling logical and cosmological reasoning. The problem is the metaphilosophical one. We can see it if we really draw the consequences of contextual, holistic and inferential theory of meaning and truth to the sentences of philosophical 'theories'. If such a theory entails 'cosmological' statements, the radical questions should be: What do they entitle us to? What do they commit us to believe and to do — if we accept them? Why should we accept them? How are the corresponding commitments assessed? Is acknowledgment of corresponding theories a question of science, of aesthetics, or a new form of religion? I.e., is the resulting physicalism and scientism just a form of arbitrary dogmatism?

4 Pragmatism in logical empiricism

In the following passage, I want to give a more detailed story that adds some flesh to the skeleton presented above. Carnap already had banned any external question of how words relate to 'the' world or to 'a' world (of naive set theory for example) out of the scope of a logical syntax. In logical syntax, the meaning of the words and the 'truth' of sentences is defined implicitly by a holistic system of deductive or inferential rules. Truth is replaced by deducibility in an axiomatic system T. On the metalevel, Tarski considered the possibility of an implicit definition of a kind of truth-predicate

$N(s)$ is true

together with a prosentential operator $N(s)$ such that for any closed formula s of $\mathscr{L}(T)$ the formula

$N(s)$ is true if and only if s

can be derived in a larger axiomatic system $M(T)$. It is important to see that a Tarskian truth theory, properly understood, does not leave the framework of Hilbert's concept of implicit or axiomatic definitions. This idea of implicit definitions goes back to Hilbert's *Foundations of Geometry*.

We can look at axioms in formal deductive systems as holistic semantical rules expressed by sentences. In fact, sentences derived in an axiomatic system always can be seen as possible inference rules themselves, expressed in the form of sentences. This is clear for sentences of the forms 'if A then B' (in symbols: '$A \to B$') or 'for all x if $A(x)$ then $B(x)$' (in symbols: '$(\forall x)(A(x) \to B(x))$') and for conjunction '$A \wedge B$'. The only problem is how to read negated sentences and disjunctive and existentially quantified sentences as rules. This is not too difficult to explain. But we do not need the details here. It suffices to see them as rules without premises.

Any rule with n premises $A_1, \ldots, A_n$ and one conclusion A can, so it seems, *in principle* be *made explicit* by a 'sentence' of the form '$A_1 \wedge \ldots \wedge A_n \to A$' and by one and only deduction rule, the Modus Ponens. This deductions rule has two premises: A and $A \to B$. It has one conclusion: B. The use of this inference rule cannot be explained by a sentence of the form '$A \wedge (A \to B) \to B$' without presupposing the competence to use this rule already. In other words, the practical competence to use the inference scheme Modus Ponens correctly is a crucial part of the meaning of an arrow or of a conditional phrase of the form 'if A then B' by which we make an inference that allows us or commits us to go from A to B explicit.

A deductive game T would lack any meaning if we were only interested in the question which sentences of $\mathscr{L}(T)$ are theorems of T, and which are not. Hence, at least some sentences should have a relevant use outside the deductive game of T. One possibility of such a use is this: Some sentences derivable in T 'report', or 'predict', possible perceptions — such that the theory can *contradict* perceptible 'reality'.

If this is so, then we have to distinguish two components of 'meaning' and 'truth', the internal one, defined by the deductive system, and the external one, given by some projective relation between (basic) sentences or propositions and perceptions. Hence, if we still want to say that the meaning of the sentences and subsentential terms consists in inferential roles, then we have to say that language entry 'rules' that lead from perceptions to reports are also inferences. And the same has to be said of language exit rules that lead from sentences to actions.

An axiomatic theory T can be seen now as a deductive system that produces or generates sentences that express, in turn, inference rules we are entitled by the theory to use. In a similar way we can look at a *belief* as a kind of implicit theory. It is implicit in the sense that it does not have to be made explicit yet. If it were made explicit, the belief would correspond to acknowledged 'axioms'. They would tell what consequences I would be committed to accept as being inferred by my belief — even if I never actually had thought of them yet. Like a theory, a belief can be wrong.

The question which axioms are to be chosen turns into a mere *pragmatic* question. They *define* the inferential roles of the words in the formulas or sentences. We would wish, of course, that the axiomatic system as a system of inference rules is *consistent* or *coherent*. Hence, questions of deductive consistency and coherence with experience become central in the meta-theoretical investigation of formal theories.

The position of such a *pragmaticist* inferential semantics is therefore this: If consistency and coherence is secured, the question which theory or semantic or inferential framework is to be chosen, is a matter of free decision. We are liberal here — if only formal exactitude is achieved, i.e., if the inferential rules and axioms are made explicit. All other external questions turn into questions of *pragmatics* in the sense of investigations of *usefulness* of the semantic system or theory in playing its role for structural representations of empirical knowledge, for example. Thus, the quest for Truth and Certainty is replaced by the pragmaticist evaluation of *empirical relevance*, *fruitfulness*, simplicity, and *beauty*. Truth with capital "T" is expelled out of the realm of semantics. The same holds for Reality with capital "R" or Objectivity with capital "O". This fact appears as the anti-metaphysical success of inferential semantics and formalist pragmaticism.

Let us now assume that T expresses a kind of common knowledge of speakers together with a common form of (formal, conceptual, and material) inferential competence. Let us assume moreover that one speaker a 'tells' other persons about his particular beliefs $T(a)$. By this he is 'committed' to the reliability of $T(a)$ with respect to the inferences that can be drawn out of $T + T(a)$. We as hearers are 'entitled' now to draw these inferences. This is the language game of assertion.

Some of the sentences produced by the system $T + T(a)$ could (and should) be basic in the sense that they tell us to expect some peculiar facts that can be perceived or controlled as 'true in experience'. To express this relative independence of basic sentences, let us call sentences 'elementary' if they have a kind of *stimulus meaning* attached in the following sense: We are already trained to *expect certain perceptions* if they are uttered with reference to a present situation (in space and time). And we are trained to be disposed to utter them or to agree to them or to evaluate them as true if the appropriate situation is presently perceived. (I do not care for the question here "How to perceive a situation?" and I do not go into the question if Quine's term "stimulus meaning" can lead us into a behaviouristic perspective; I rather find it helpful here.) The elementary sentences express what Wittgenstein calls 'states of affairs'. They are called 'facts' if they are *actual*, i.e., if the *observation sentence* that accompanies the corresponding perception is (to be) judged as *true*. The truth- or rather *assent-conditions* are, of course, a matter of *social practice*. Mastering this practice is taking part in *a social art*.

Observation sentences are, in a sense, no sentences at all. Or rather, their use is a very special one, as Wittgenstein had observed already. If we want to be more precise, we should not speak of observation *sentences*, but of a *predicative use of sentences as observation sentences*. We do not inform anybody about anything by their performances — if not about the use of language in the given situation. We do not express *claims* by such predicative speech acts. We rather *establish* and *reconfirm* basic agreements in our projections of words or strings of words or sentences to the perceptible world. Or we *control* possible disagreements. In a sense, observation sentences are rather elementary predications by which we articulate more or less general and vague distinctions of qualities in our perceptions and experience. We distinguish by them situations — if we keep in mind that such situations are, as such, not yet identified or identifiable as possible objects of speech. They are present — or they are nothing at all. Hence, elementary predications do not yet presuppose a differentiation between objects that can be named and qualities that can be expressed by predicates. Therefore I have coined the special expression "predicative use of (observation) sentences" for them. But I continue to use the shorter expression "observation sentence" when it is sufficiently clear what is meant.

What I call "basic sentences" here are (expressions for) *speech acts* that already come with an implicit or explicit limitation of space and time to which a possible predicative use of observation sentence refers.

In the explicit case, such speech acts might contain deicticals like “here” and “there”, “now” and “yesterday” and “tomorrow”, and, later, even names of place and time like “in Paris” or “1960”. (Remember that the sentences provided by the speaker may implicitly refer to the situation of utterance.) By such basic sentences we can say that the qualities expressed obtain at the place and time mentioned, quite similar as in Otto Neurath’s *protocol sentences*. Thay say, in a sense, that here and now (or then and there) a certain observation is really possible. Basic sentences, as they are understood here, already are spatially and chronologically determined. They can be used to express the content of certain expectations.

The formal evaluation of basic sentences can now be twofold. One is dependent on perception, the other is dependent on theory. The first is, so to speak, truth evaluation from below, *a posteriori*, the second from above, *a priori*. The *a priori* evaluation is inferential, the a posteriori one is, in a sense, non-inferential (even if we made it inferential above). Basic sentences form a ‘Humean’ kind of empirical foundation of theories. They do so not in the sense of verification or justification of theory, but rather in the sense of Popper’s falsificationism: Basic sentences together with their corresponding realm of reference (in the world of possible experience) express empirical truth-claims. Basic sentences of this sort can be generated by a theory. If a theory generates enough of them, it can make a large number of ‘reports’ and ‘expectations’ explicit. In fact, theories can be seen as a kind of shorthand for expressing a large amount of knowledge.

Knowledge always rests on a certain logical language design and on the method of inferring sentences according to a certain social technique. We have learned to perform corresponding inferences and to control their correctness. This technique is the reason why only an ‘externalist’ stance in a theory of meaning can be adequate: Nobody knows all licences of inferences that come with an assertion or information. But everybody appeals to some public domain in which we can figure out if the inference is valid or not. This is the reason why the meanings of my words are not in my head and not in yours. They are nothing that can be pointed to in actuality because they are a public form we appeal to when we judge about *what was said* and what can be inferred from what was said.

I can, however, mean some other things with my utterance than what follows from it in a standard reading. But this does not mean that my intention determines the meaning of my utterance. It only means that in any singular and specific case there are possible filters of relevance and reasonability. Only relevant and reasonable forms of inference are to be used by the hearer if he is charitable and cooperative enough in his understanding of what the speaker has said. I.e., the hearer must find out, partly on his own, which of the generic norms of inferences are relevant in the specific case. I.e., the hearer is not just entitled to use schematic rules in order to infer conclusions from what the speaker has said or when ascribing certain commitments to him. Sometimes he even has to ask the speaker if the speaker acknowledges certain normal forms of inference or not. The answer of the speaker is important, often decisive. But it is frequently not the last word: The first person authority regarding the meaning of what is said is even more limited than the generic norms of normal inferences. All in all, intentional semantics is a latecomer riding on the broad back of externalist and intersubjective (cooperative) inferential semantics.

Any user of a theory is committed to its inferences and predictions. If we ask for the reasons that speak in favour of the theory, he should at least be able to tell us something about *the absence* of too gross falsehoods in the sense of already known disorientations or non-fulfilment of expectations or truth conditions of observation sentences. Hence, it can be worth while to *test* a given or proposed theory, to control its predictions by artificial attempts of refutations — more or less in the way Popper describes them. This practice of control is a trans-personal enterprise.

5 Analytical and empirical parts of a theory

It is the very concept of implicit definition in whole theories that does not allow for a sharp line of demarcation between synthetic and analytic sentences. Just think of the question which axioms or which theorems of our joint background system T should be called analytic, and which synthetic. What is a part of general meaning, what is a particular 'empirical hypothesis'? Could we say, for example, that the empirical hypotheses are a kind of *special* information presented in a system of sentences U that is *added* to an already accepted system T of common knowledge that defines meanings? Then T would be a system of 'formal' or 'meaning-related' sentences that can be used to express 'semantically valid' inferences. U would be a system of 'material' or world-related sentences expressing empirical inference. T and U would be parts of a complex theory $T + U$ that produces some basic sentences. The question would remain how to distinguish between T and U if not in terms of general belief, i.e. joint commitment and entitlement, and particular claims and commitments that still are to be controlled empirically.

Let us assume for a moment that there were certain analytical sentences. Some of them may 'say', for example, that two expressions t and t^* are synonymous in the sense that the one can replace the other in certain or even all contexts *salva veritate*. This can happen if $t = t^*$ is derivable in T. But then the question arises again: How do we know if $t = t^*$ or if the replacements we feel entitled to by the (formal) 'truth' of equalities really hold for conceptual or semantical and not rather for empirical reasons? Replacement holds relative to the theorems of the belief-system $T + U$ and the corresponding basic sentences. They may be derivable in $T + U$ or not. The question is once again this: Are all axioms and theorems of T 'analytically true', if they are part of a 'definition' of the meanings of the terms in $\mathscr{L}(T+U)$, the merely syntactically defined 'language' of T and U? Or does only a part T_1 of T determine the meanings, such that another part T_2 articulates general, i.e. generally accepted 'empirical' hypotheses or 'truths' that we teach and learn verbally, whereas a third set of beliefs, $U(i)$ and $U(j)$, respectively, remains subjective in the sense that it differs also from person i to person j? But how would we have to figure out these parts? How could we decide which axioms in a holistic theory should count as meaning postulates, which should count as empirically true or practically fruitful, and which are personal beliefs? We cannot decide this arbitrarily — if the distinction between analytic and non-analytic statements, and between general (even if non-ideal, hence, non-absolute) 'knowledge' or 'truths' and mere subjective (even though frequently shared beliefs) should have real content at all.

This is already the master argument of Quine's attack against the dogma of an allegedly clear distinction between analytic and synthetic sentences. It puts into question the idea of a clear distinction between knowledge and belief. And it questions an allegedly well-defined notion of synonymy by which we could define what we call the meaning of an expression, namely by use of the well-known method of abstraction with respect to an equivalence relation. Quine's result is that analytic sentences can only be defined by an explicit and arbitrary stipulation of verbal definitions and rules of formal inference.

In empirical theories, most axioms or axiom-schemes have a Janus-head position. They are neither empirically verified nor mere verbal stipulations or formal consequences. This shows that they are of a third kind. They *constitute* the theory as a means for re-presenting past or present or future or possible experience. Therefore the sentences express rules that play an *a priori* role in the theory just like the mathematical sentences do in Wittgenstein's *Tractatus*. They are presupposed when we *use* the theory. We can call them 'trancendental' and, at the same time normative (pre)conditions of objective, i.e. articulated, experience. The sentences expressing these (pre)conditions or rules, however, do not *represent* any reality or object as such, neither transcendent nor empirical. They just help to re-present possible experience in a whole inferential system. This, I take it, was the holistic position of Kant with respect to synthetic *a priori* sentences. But Kant's explanation of the double-head status of synthetic and *a priori* truths might be too short. He did not discuss the possibility of a change. If preconceptions that counted as norms of allowed inferences until now have consequences that contradict empirical observation, we must make changes. And we do make changes. This is the reason why the norms of 'conceptual' inferences that 'define' the *form* of objective experience have a *history*, as Hegel teaches us and Thomas Kuhn reminds us again. The historical status of general rules that form the framework or paradigm of a theory shows that these rules or the corresponding sentences are not *analytic* in the strong sense of mere verbal truth: Synthetic *a priori* sentences have *some* kind of empirical status.

The second part of Quine's critique of the dogmas of empiricism is concerned with our assumption that there are basic propositions that can directly be judged as true or false on the ground of sensory inputs. We may think of the perceptions of an eye-witness, for example, who already masters our language and even might be controlled by others. Even if we have stressed the relative independence of observation sentences and basic sentences from theory T (or $T + U$), we have to admit that there is no absolute independence of the empirical basis from theory. Our attention and focus of perception depends already on theoretic preconceptions or expectations. If a basic statement derived from a theory turns out to contradict some observation-based truth, it is often not yet clear *what* has to be revised, the theory or the basic propositions. This is so because we can always err perceptually, or rather, that we have to render our observations in proper words. But it is not at all clear which part of the theory, which 'axioms' or 'maxims' or 'basic contentions' or 'beliefs' should be revised then.

Moreover, even what counts as basic sentences and the way we have to evaluate their truth by observation often depends heavily on a theory, on pre-knowledge or prejudice. It depends, moreover, on relevance considerations concerning core interests we connect with the theory in question. This fact makes things complicated as the practice of experimental physics shows: Here, like everywhere, we first learn the theories (by heart, if you wish). Only later we learn to test the theories and to make proposals to improve the theory if we see that it does not give us as good enough orientations for what is to be expected or done.

One problem of Quine's behaviouristic account, however, is a 'flattening' effect in his 'naturalism'. Quine defends the idea of the unity of science even more radically than Carnap did. For Quine, there is not even a clear difference between formal and empirical parts of science. But theories used for reporting and predicting the behaviour of things, of animals, and of men are not distinguished from theories that help us act and cooperate in a reasonable, and responsible, way. I.e., reporting and predicting the behaviour of humans is not distinguished from giving orientations. But giving orientations is an essential part of our individual and joint action. If we do not see the principal difference between natural science and philosophical or cultural reflection, we also do not see the difference between merely technical orientations and orientations in cooperative actions.

It is a wholly general fact that in the real world there are no exact and almost no absolute, differences at all. But this fact proves much less than Quine thinks — like any overcritical sceptic of our always fairly rough differentiations does. The relative differentiations between conceptual and empirical statements are as such fairly helpful — if we use them in each singular case in a proper, pragmatical, way. Moreover, there are many relative levels of conceptual presuppositions that can be made explicit. This helps us to see a fine structure of logical, pragmatical and methodological order. To say this in a simile: Our house of language is no bungalow with centre and peripheries as in Quine's flat picture. It is rather a building in which many higher regions rest on lower foundations, at least in the sense that more complicated faculties, techniques, or theories presuppose the mastery of less complicated ones.

Moreover, empiricism and methodological individualism overestimate individual perception. Therefore, Cartesian and Empiricist epistemology runs counter to the social reality of science. It misses the practice of our tradition of language, rules, and knowledge. This insight, which was made explicit by Hegel, already stands behind Kant's arguments against Hume. As individuals, we first have to learn to take part in a joint effort consisting in a correct, i.e. common, use of language, its rules of inferences and of verbally expressed knowledge. Then we learn to take part in the theory control and theory revision, i.e. in *developing* human knowledge and practice. Finally we learn to teach forms and norms of practice and to take part in judgements about the correctness of attempts to fulfil such normative conditions by others or by ourselves. Such judgements often are difficult: We do not just follow rules or try to fulfil or control schematic criteria. We rather take part in a free jury. As persons, we are all such jurors.

If a rather schematic control is possible, such a jury does only do some monitoring. 'Scorekeeping' in the sense of Brandom is controlling the fulfilment or non-fulfilment of already acknowledged rules or norms of correctness. David Lewis has proposed the paradigm of the role of an umpire in cricket or baseball. The problem I have with Lewis's and Brandom's approach is this: Scorekeeping is just a kind of notation. It makes the history of a language game between two persons in a certain way explicit. It is a kind of evaluation which works similar as the evaluation of truth values for logically complex sentences: In both cases the criteria are presupposed that tell us how to evaluate a 'further step' — given the 'earlier' or 'shorter' ones. A clearer example is dialogical logic: There we give the *tableaux method* of Beth a dialogical interpretation: Two persons play a game of asking and giving reasons. In such a game, they control if a proponent can defend his commitments against real or possible questions of an opponent. If the dialogue is played according to certain rules, the notation of it is a scorekeeping device by which we can look in a systematic way for 'winning strategies' of the proponent. If he has one, he is entitled for his claim because he can defend it.

Real argumentation, however, cannot be reduced to this kind of scorekeeping or dialogical control of rules of inferential entitlement and commitment. Such games and scorekeeping, it is true, can help make explicit how the rules are used or applied. And they show how any concept of rules (which appeared at first as monological) implicitly refers to some norms for leading reasonable dialogues about correct rule following. This only seems to lead to an infinite regress: Now we seem to have to make the norms of the dialogue explicit. And we do this, indeed, by turning them into rules. But in all these cases, there is an appeal to already accepted (and mastered) norms that tell us what are correct moves in the play.

Discussions *about* the norms or rules are of a different kind. In fact, mastering the meta-game of joint reflective judgement about rules and maxims in their relation to proper actions is to do something else than to make correct moves according to some norms and avoiding the sanctions that enforce correctness.

Brandom leaves the resulting question open where the acceptance of the norms comes from and how the norms themselves and the corresponding practice of sanctions change and develop, or rather, how we change and develop them in argumentations. The problem concerns the *place* of norms in a free cooperative practice. Any norm is founded in the autonomy of *us* as jurors. To take part in this autonomy is not just to make a subjective decision, or play the role of a scorekeeper in the right way, but to jointly develop and save intersubjective forms of human practice.

6 Indeterminacy of reference and translation

In Quine's picture, it is as clear as in Popper's that any theory is hopelessly underdetermined if we only look at elementary predication and basic sentences. We can put the things around and assume that we have learned some basic words in common situations by establishing joint assent and dissent with respect to observation sentences like: "(this is a) lion", "(lo, a) bird", "(here is a) rabbit". Even if we always agree in *these* uses, we do not yet master the inferential role of a sentence like "The lion in the zoo has caught an ostrich". The meaning or inferential role of names or name-like expressions in sentences is quite different to simple observation sentences. More precisely, the conditions of 'truth' and the entitlements for further inferences of sentences containing such names are logically much more complex than the stimulus meaning and truth of observation sentences. Naming (empirical) things is already a rather complex theoretical enterprise. In a sense it involves a whole theory about the normal, usually expected, properties of the named thing (over time). This is the reason why we can say that an elementary predication such as "this is Rabbit" behaves to "Rabbit is at rest now" in a similar way as basic sentences to whole theories.

Indeed, this analogical comparison leads to the master argument of Quine in favour of his famous *indeterminacy of reference*: We do not know just by observations of utterances like "lo, Rabbit" in appropriate situations what is named by "Rabbit" in sentences of the form "Rabbit has the property P". There always is a certain set of inferential expectations or deductive theory behind the use of "Rabbit" as a name in such sentences. We may even not be sure if Rabbit is a rabbit, or rather a person (as in Updike's sequel of novels).

Kant analysed some of the problems of objective reference in his *Critique of Pure Reason*, even though his picture does not explicitly refer to linguistic rules of inference. In any case, Kant's concept of rationality ('Verstand' or Understanding) is nothing else than a competence to *master rules of (verbal and practical) inferences properly.* And his distinctions between sensation, perception and intuition ('Anschauung') correspond to the different truth conditions of observation sentences and basic sentences, i.e. to "I see or perceive P now", "I and we can perceive P now", "I and we could perceive P then and there". The famous principle of apperception just says that it must be possible that any (re)presentation (of some determined object) can be accompanied by thinking, i.e. by a corresponding *sentence*. Moreover, language was already for Kant the paradigm case for making thought explicit. Also, thinking *is* in almost all cases using language — for Kant as well as for Hobbes or already for Plato.

Indeterminacy of translation is an even more straightforward result of our inability to make out clear and distinct meaning postulates and distinguish them from empirical truths. Quine's example is an anthropologist who lives, let us say, the first few weeks with a native tribe. She hears what people say, sees where they point to, and learns after some time what are signs of agreement and disagreement. For a long time she will not be able to construct an *uniquely determined* theory which 'predicts' not only the observation sentences to which the native in the jungle would agree but also translates more complicated sentences into her mother tongue. But assume that she does well after some time. Then her theory would be a kind of translation manual that is supported by the agreement and disagreement to complex and simple sentences in relation to possible observation. Such a manual would have to identify many elementary predications (observation sentences) and basic sentences, the logical connectives and the differences between names, predicative contexts (complex predicates) and other operators (predicates of higher order).

Quine shows, then, that for any such manual there can be others that are mutually inconsistent and still agree with the linguistic data hitherto experienced. It seems obvious that Quine sees a parallel between establishing a translation manual and constructing a theory. In both cases we construct theories that hopefully are 'true to the facts' (of possible observation) in the sense described above. In both cases we explain things by predicting possible observations under certain conditions in the theory. In both cases there are possible indeterminacies or inferential incommensurabilities between different theories — even if both were equally good with respect to the empirical basis or observation. (This is Quine's version of Popper's scepticism against the possibility of inductive logic.) Theories are constructions. As such, they rest on a kind of 'fantasy', not on calculation. Therefore they are in wide respects 'arbitrary'. Their construction and change take place in a process of cultural development. There are no schematic methods of rational theory formation starting with observation and statistics — as Carnap still seems to believe despite Popper's criticism. But perhaps we should be more charitable and read Carnap's 'inductive logics' just as a filter. It filters out unreasonable theory constructions, especially in the realm of stochastic theories. In fact, Bayesianism is a filter-theory for 'irrational' assumptions of probabilities — in the contexts of decision- and game-theory.

7 Radical interpretation as a theory of language acquisition

When we ask how we learn a language or a theory, the question of verificationism or inductivism takes a new form. In acquiring the social art of language we have to know what to say and what to do when. We have to learn to draw inferences. But how do we do that? How do we learn to assess commitments and entitlements? How do we learn the language structure, what the words mean, the material inferences, the meaning postulates and inferential forms? We seem to have to explain how we do these things when we are exposed to a linguistic practice. How do we learn, for example, what "and" means and what "not" means and what "not both" means in the language use of the native speaker?

In Quine's picture, at least the logical connectives are relatively easy to learn. We can learn them in connection with observation sentences. If p, q are observation sentences that can be learnt by normatively controlled joint assent and dissent, it is not too difficult to learn the inferential roles of "p and q" and "p or q" and "not p" with respect to entitlements and commitments. Knowing the scheme seems to be enough to apply it recursively. The problem is how to learn the difference between syntactical surface structure and logical deep structure, and how to avoid mystical entities like intensions and possible worlds by an 'extensional' analysis of different rules for substitution and inference, of anaphorical relations and by the logical machinery of cross-references, and quantifying-in. With respect to singular terms or anaphora, we must, moreover, distinguish between contexts of *de-re-* and contexts of *de-dicto-*types of use.

Quine and Davidson agree that a non-metaphysical concept of linguistics and semantics should start with linguistic behaviour and learning. But Davidson wants to develop a kind of generative and inferential semantics. He assumes that the general form of language is fixed: It is some kind of an axiomatic deductive system that comes together with a kind of social and joint evaluation of observation sentences. Therefore, Davidson does not seem to lose too much when he even includes into his 'definition' of a language the postulate that any language has deduction rules and axioms that correspond somehow to first order predicate logics. By this, Davidson turns Fregean semantics into a universal feature of any human language.

In a sense, Brandom gives an answer to an open question in Davidson's account when he shows why we need names and predicates. This shows, in a sense, why there is no language without predicative structure and without logical connectives, by which we can make implicit norms of correct inferences explicit in the form of rules and logically complex sentences.

A part of the question how we learn a language then turns into the following question: How do we identify the names as names? How do we identify the predicative sentence frames as such and the logical connectives and operators? If one reads Davidson, it seems as if all these features are presupposed to exist in a universal grammar of human languages. We learn them by learning to deal with certain substitutions and corresponding inference rules, just as we learn syntactical rules by substitutions and transformation (*pace* Chomsky). Later we have to learn and master particular concepts with their peculiar inferential roles.

According to Davidson's picture, we never can be sure that we understand the other correctly — if "correct" does not mean that we both are happy with our cooperation right now but that we shall 'understand' us further. The real matter of understanding consists, in the end, in not much more than an absence of misunderstanding or frustration. If we see things in this light it seems to say that there are no prefixed meanings or conceptual schemes.

On the other hand, if we did not 'believe' what the other says, we never would be able to learn 'what he means', and what his words refer to. This is a rather trivial fact if we look at observation sentences. I cannot learn what "nádraží" means if my host always utters the word when I sneeze and not when I am standing in front of the train station. When learning a language I must assume that most sentences I hear are "true". Or rather, I must take them for correct. This shows the priority of sincerity over insincerity, truth over falsehood, common sense over individual disagreement. Without an overwhelming amount of agreement and cooperation there is no understanding at all.

A Davidsonian ideal learner of a language (an axiomatic system) starts by listening to the verbal and looking to nonverbal behaviour of the speakers. He assumes (as a kind of basic maxim) that all or almost all sentence he hears are true and all or almost all deductions he observes are correct. His task now is to understand or 'interpret' the system of language. This is not to be assumed as a conscious enterprise. We do this even when we do not know what we do in the sense that we do not give an account yet of how we proceed.

It seems natural to attack Davidson by saying that original language acquisition or radical translation cannot be interpretation. Interpretation is done ‘consciously’ and therefore presupposes a language. Therefore, Wittgenstein distinguished later (conscious) interpretation from implicit understanding. But radical interpretation in Davidson’s sense is different from both. He only seems to use the conscious or explicit case in order to elucidate the unconscious or implicit process. The overall picture is this. The learner tries to figure out a syntactic structure that can be a basis or framework of logical semantics. Then he tries to find out the observation sentences, to identify the logical form and logical connectives, corresponding relevant observations, material meaning postulates (that cannot be distinguished from complex statements having empirical consequences) and other inferential rules for the sentences and subsentential terms. We may start, perhaps, with basic words of logic like “and”, “not” and “for all”. Like Quine’s linguist in the jungle of natives, Davidson’s ideal learner or hearer does not really take part in joint judgments about correctness and incorrectness. He only works with sense-related inputs and inferential outputs. This makes Davidson’s picture mentalistic and at the same time materialistic. Davidson seems to think that a kind of axiomatic system of inferential semantics is in some way or other wired into the brain, but that the personal history of the learner plays a crucial role in how this wiring comes about. This idea is his basic reason for his thesis of anomalous monism.

Three

Gödel's Theorems and the Synthetic-Analytic Distinction

Vojtěch Kolman

When posing the old question 'What are arithmetical truths about?' ('What is their epistemic status?' or 'How are they possible?') we find ourselves standing in the shadow of Gödel, just as our predecessors stood in the shadow of Kant. Of course, this observation may be a bit misleading if only for the reason that Gödel's famous incompleteness theorems are not of a philosophical nature, at least not in the first place. There are plenty of texts, however, explaining them as philosophically relevant, i.e. as having some philosophical implications.

In this article I am not aiming to add a new interpretation to the old ones. Rather, I am proposing to see the incompleteness as a link in the chain of certain great (positive or negative) foundational results such as Frege's calculization of logic, Russell's paradox, Gödel's completeness theorem, Gentzen's proof of consistency etc. The foundational line described in this way can then be critically examined as relatively successful with respect to some of its leading ideas and as unsuccessful with respect to others. What I have particularly in mind here is the idea of *reducing*

I would like to thank Pirmin Stekeler-Weithofer, Jaroslav Peregrin and Karel Procházka for helpful comments and conversations. Work on this paper has been supported in part by the grant No. 401/06/0387 of the Grant Agency of the Czech Republic and in part by the research project MSM 0021620839 of the Ministry of Education of the Czech Republic.

arithmetic to logic, with its decisive influence on the rebirth and subsequent development of modern (mathematical) logic. Hence, the key issue of this article may be formulated as follows: 'What do Gödel's theorems tell us about the alleged analyticity or syntheticity of arithmetic?'.

1 Frege's thesis

In the 19^{th} century there were many programs announcing the need of reducing arithmetic to logic (Jevons, Schröder, Dedekind, Peano), yet ironically there was no logic capable of competing with arithmetic in rigor and self-sufficiency. Frege overcame this difficulty by simply inventing it, but his new logic remained virtually unknown until his death, so there must be another explanation of this sudden enthusiasm for logical methods, and, in fact, there is: a general disappointment with the Kantian intuitive conception of mathematics, i.e. with Kant's attempt to ground both arithmetic and geometry in the spatio-temporal structures of reality.

As much as Frege's contemporaries and Frege himself disagreed with Kant on the nature and sources of mathematical knowledge, they were unable to put this disagreement in other than Kantian terms, discarding one side (the left one) of his fundamental distinction between

constructive	vs.	discursive,
intuition	vs.	concept,
mathematic	vs.	logic,
synthetic	vs.	analytic

while endorsing the other. In the light of this observation we can rephrase the main task of Frege's logicism as follows: 'How can one present arithmetic in a non-constructive way?' or 'How can one show that arithmetic is not synthetic, but analytic (not *world*-dependent, but *word*-dependent)?' Or more specifically: 'How can one avoid intuition in the process of justifying basic arithmetical concepts (the intuitive number construction 0, $0+1$, $0+1+1$, etc.) and propositions ($2+2=4$, $34\times 2=68$)?'

The general answer: 'by conceptual means' was, for the first time, successfully implemented in Frege's *Begriffsschrift*, in particular in his ancestral definition and his axiomatization of logic. Proceeding from the (parental) relation R to the explicit second-order definition of its arbitrary finite iteration $RR\dots R$ (ancestral), Frege was able to rephrase the predicate "x is a number" as "x is a successor of the number 0" or "$(\forall X)[X(0)\wedge(\forall y)(X(y)\rightarrow X(y+1))\rightarrow X(x)]$", with "$x+1=y$" as the basic (parental) relation. This initial success encouraged Frege to declare logicism a feasible *hypothesis*, according to which, prospectively,

(1) numbers are to be *conceptually* separated by means of the aforementioned predicate as a certain *species* of the more general *genus* 'logical object' or of the most general concept 'object' (the same holds for functions whose 'intuitive' recursive — formation was proved to be logistically admissible in Dedekind's famous recursive theorem),[1)]

(2) arithmetical proposition are to be deduced from logical axioms, the conceptual truths of Frege's new logic, by logical rules alone. Among these axioms, the so-called *Grundgesetz V* — Frege's Axiom of Extensionality — is in charge of the ontological basis from which numbers as logical objects are to be separated.

Stages (1) and (2) mirror the expressive and deductive parts, respectively, of the logicist project.

It seems to be clear that Russell's paradox decimated in the first place the second, *proof-theoretical* part of the project. That is why set theorists like Georg Cantor did not regard the antinomy as a serious problem and why some modern logicians, like Crispin Wright or George Boolos, still hope to resurrect logicism in a *model-theoretical*, structuralistic way. The second-order logic they are using ('the set theory in sheep's clothing', as Quine put it) is deductively incomplete (thus proof-theoretically unacceptable), but semantically very strong. This seems to be in accord with the expressive part of the original project.

Although I agree with the neologicists that Frege's system is not affected by the paradox as badly as we thought it was, I claim that this does not warrant it as successful according to Frege's own standards. Let me indicate why.

In a sense, both Frege and Cantor proposed a set-theoretical (i.e. a kind of a semantical) solution to the foundational problems of arithmetic: their numbers are the so-called pure sets or what Frege called 'logical objects'. Frege *unlike* Cantor, however, realized that now we have to face up to a new problem. Instead of 'What are numbers and how are they given to us?' we have 'What are sets and how are they given to us?'. I cannot go into details here,[2)] so I will merely claim that in this respect neither Frege nor anyone else could succeed in a desirable way and that

---

[1)] Analyzed in detail in Kolman [2007].

[2)] Cf. Kolman [2005].

it is Russell's paradox that indicates why. To put it simply, there are no objects justifiable by logic alone; there are no pure sets, but only sets of concrete objects like cats, atoms or numbers. And make no mistake! Axiomatic set theory does not justify pure sets, but presupposes them in order to be consistent.

2 Wittgenstein's antithesis

Does all of this imply that the idea of analytical (language-dependent) arithmetic is definitely dead? As a matter of fact, yes, but I do not want to commit myself to such a strong claim yet. Contrary to Poincaré, Brouwer or Wittgenstein, neither do I want to criticize the logicist thesis as such. I pointed out earlier that Frege (unlike Poincaré, Brouwer and Wittgenstein) at least gave his reasons for treating logicism as a bold, but promising *hypo*thesis and that, accordingly, he did not stick dogmatically to it after it was found wrong. However, in spite of this self-critical approach, Frege was clearly too absorbed in his methods to recognize where and why they failed. In this respect, Wittgenstein was more successful.

Like Poincaré and Brouwer in their destructive approach to logicism, Wittgenstein was not overly specific either, but the main idea of his critique is clear: in order to avoid the spatio-temporal intuition Frege strives to proceed as abstractly as possible, pushing the practical arithmetic aside as scientifically irrelevant. But his abstract second-order definitions such as

$$(\forall X)[X(0) \wedge (\forall y)(X(y) \rightarrow X(y+1)) \rightarrow X(x)]$$

work correctly only under the condition that their second-order variables ('X') range over sets specified in pre-scientific, spatio-temporal or practical fashion. In our example it is

> 'the Set of all and only those objects obtained by iterating the operation +1 a finite number of times, with 0 at the beginning'.

Since the soundness of these definitions ought to be established before they are employed, the desirable elimination of the practical — presumably Kantian — element is only imaginary, as Wittgenstein clearly recognized in the context of his later reflections on 'following a rule'. His critique of the logicist foundations, however, was already anticipated in his *Tractatus*-theory of internal properties, relations and operations.

According to Wittgenstein, concepts such as number are not explicitly definable, i.e. graspable by means of an explicit formula. Rather, they are categorical notions, and as such they are describable only by means of a rule, i.e. implicitly as a potentially infinite process generating their instances. (This is Wittgenstein's early doctrine of inexpressibility and at the same time Kant's doctrine of concepts as rules.) Wittgenstein represents these internal notions with the help of a complex variable

$$[a, x, O(x)]$$

which can be transcribed in Lorenzen's operativist style as

$$\Rightarrow a, \qquad \text{(starting rule)}$$
$$x \Rightarrow O(x). \qquad \text{(inductive rule)}$$

In the case of number we get the rules '$\Rightarrow |$', '$x \Rightarrow x|$'. Although without any direct *factual* difference (on the arithmetical object level), these transcriptions are very important *philosophically* (on the metalevel). They provide the missing link between the abstract notions of arithmetic and their practical applicability in everyday life. Grasping the concept of number now amounts to mastering the 'counting' rules mentioned above. Wittgenstein, however, did not systematically develop these suggestions into a full-fledged theory since he simply did not believe in such a thing as the 'foundations of arithmetic'. In a sense he may have been right but generally he committed the same error as Frege did, only the other way around: instead of the practice he undervalued the theory.

3 Hilbert's synthesis

The need for the mutual support between practice and theory was already partially recognized by the later Hilbert in his metamathematics. As a study of derivability in certain calculi it was not only a new *mathematical* discipline but also a new *philosophy* of mathematics with the ambition to supersede the old versions of formalism and possibly Brouwer's mentalistically misconceived constructivism. In Lorenzen's later 'operativist' elaboration of Hilbert's ideas[3)] we find certain calculi (collections of rules) as the basic form of mathematical practice (ability to operate according to the rules) and a theory of this operating as 'arithmetic' itself. So, for instance, to say that the arithmetical formula "$2 + 2 = 4$" is true, according to Lorenzen, is to assert the derivability of the figure

3) See Lorenzen [1955].

"$|| + || = ||||$" in the following calculus $(+)$:

$$\Rightarrow x + | = x|, \tag{+1}$$

$$x + y = z \Rightarrow x + y| = z|. \tag{+2}$$

Here the variables x, y, z range over the figures manufactured by this calculus $(|)$:

$$\Rightarrow |, \tag{|1}$$

$$x \Rightarrow x|. \tag{|2}$$

This time, however, the variable x is the so-called 'eigenvariable' ranging over the figures so far manufactured by the *same* calculus (the so-called numerals $|$, $||$, $|||$, etc.). As a result, to justify the truth of a sentence such as "$2 + 2 = 4$", one simply has to write down the respective derivation in $(+)$.

So far so good, but the apparent simplicity of this radical syntactic account ends with the next step, namely with the question: 'What is one supposed to write down to justify the falsity of some formula?' Obviously, there is no 'negative' derivation (Russell's negative fact) available. Lorenzen solved this problem by supplementing his operative arithmetic with operative logic, i.e. with something he would have to add later anyway.[4)] His solution is nevertheless very elegant.

According to it (and also according to the approach of the late Wittgenstein), to justify the truth of a sentence does not mean in general just to say the sentence or to write it down, but to justify it to somebody. This somebody needn't be a passive listener, but may potentially disagree, i.e. become an opponent. Constructively interpreted, this implies that the opponent of $\neg A$ commits himself to justifying A. An elementary arithmetical sentence A is true if its proponent can justify it, while its negation $\neg A$ is true if an opponent cannot justify A. Now we can apply the same dialogical approach to the sentential connectives and afterwards build an alternative operational semantics of complex sentences. In this connection Lorenzen uses the name 'dialogical logic'. Instead of giving a systematic account, let me briefly demonstrate the principles it uses by justifying some complex arithmetical sentences. These sentences turn out to be the so-called Peano axioms of arithmetic.

Let us first take the formula $m + 1 = n + 1 \rightarrow m = n$. Translated into the unary notation $m| = n| \rightarrow m = n$, this formula turns out not

4) Lorenzen and Lorenz [1978].

to depend on the calculus (+), but on another one, which can briefly be described as

$$\Rightarrow | = |, \tag{=1}$$

$$x = y \Rightarrow x| = y|. \tag{=2}$$

"Briefly" means that we shall avoid details concerning the general problem of interpreting identity. — By means of semantic tableaux we can now unfold a proponent's justification of the sentence $m| = n| \rightarrow m = n$ as a justification of its consequent $m = n$ in the situation of an opponent's simultaneous committing to its antecedent $m| = n|$, where m and n stand for some specific numerals. The proponent's commitment, hence, is conditioned, i.e., he can demand the opponent's reasons before giving his own justification. But this immediately leads to the proponent's victory since every derivation which justifies the antecedent can be converted into a derivation justifying its consequent, simply by deleting the last row. Moreover, since the indicated winning strategy is completely general, i.e. independent of the choice of m, n, we can take our example as a way of justifying the sentence

$$(\forall x, y)(x = y \Rightarrow x| = y|). \tag{P1}$$

The expression "winning strategy" indicates that we are not interested in a victory achieved with the help of good luck but in a victory achieved according to the rules allowing us to win not only against this or that opponent, but against every opponent possible. Otherwise it wouldn't make much sense to call any sentence unambiguously true or false. In the case of a quantified sentence $(\forall x)A(x)$, the relevant rule amounts to a general strategy telling us how to win (a game associated with) the sentence $A(n)$ for a numeral n suggested by a random opponent. Since the sentence

$$(\forall x)(x + 1 \neq x) \tag{P2}$$

doesn't need any additional explanation, we can proceed directly to the induction schema

$$A(|) \wedge (\forall x)(A(x) \rightarrow A(x|)) \rightarrow (\forall x)A(x). \tag{PI}$$

In order to justify a sentence of the form '$A \wedge B$' one needs to know the winning strategy for both A and B. The beginning of the dialog can be described accordingly as follows:

(O)	(P)	
$A(\vert)$	$(\forall x)A(x)$	(1)
$(\forall x)(A(x) \to A(x\vert))$		(2)

In the next step the opponent chooses some numeral m, thereby attacking the proponent's sentence $(\forall x)A(x)$. The proponent's defence consists simply in writing down the sentence $A(m)$ as a new claim he commits himself to. If we fix some concrete value of m, say $m := |||$, we can present the further steps schematically in the following tableau, with attacks (?) noted on the right:

	$A(\vert\vert\vert)$	(1)	$\vert\vert\vert$?	(3)
$A(\vert) \to A(\vert\vert)$		(2)	$\vert$?	(4)
$A(\vert\vert)$	$A(\vert)$	(4)	?	(5)
$A(\vert\vert) \to A(\vert\vert\vert)$		(2)	$\vert\vert$?	(6)
$A(\vert\vert\vert)$	$A(\vert\vert)$	(6)	?	(7)

This strategy is completely general; i.e., the choice of m does not affect the final result of the dialog, but only its length. By iterating steps (4) and (5), the proponent can always force his opponent to claim an elementary sentence $A(m)$ which the proponent was forced to claim before. In this way the tableau 'closes' and we are done.

4 Syntheticity and incompleteness

The relevancy of Lorenzen's operativism to our opening problem is twofold. Firstly, here we have an account of arithmetic which does not start with some undefined or principally indefinable first principles, such as, e.g., Peano axioms, as we have been used to since Hilbert, but tries to justify all the arithmetical propositions (including those 'axioms') by more basic, pragmatic means instead. Being totally in accord with the 'original' characterization of arithmetic as a science dealing with calculations (Kant would add: calculations in time), this account provides us with a *positive* support of the syntheticity thesis. Secondly, we can interpret Gödel's incompleteness result along Lorenzen's line, i.e., as saying that arithmetic does not belong to sciences employing the axiomatic method. This, under appropriate circumstances, can be regarded as support for the syntheticity thesis, too, but in a more *negative* way, which depends partly on the last update we will make to our analyticity/syntheticity distinction and partly on our interpretation of Gödel's theorems. Let us elaborate these two points.

Lorenzen's approach, i.e. his operative arithmetic supplemented with dialogical logic, provides us with a convenient theoretical device for bringing the old distinction between 'analytic' and 'synthetic' or between logic and arithmetic closer to modern standards. First of all, for a given language there is always a certain class of sentences that are justifiable only on the grounds of dialogical rules alone, i.e. their truth (the proponent's winning strategy) doesn't depend on the truth of the elementary sentences in question (that's why it doesn't matter which language is being used). Every sentence of the form '$A \to \neg\neg A$', for instance, belongs to this class of the so-called logical truths.[5] The arithmetical truth, on the contrary, is defined by means of dialogical rules *plus* arithmetical calculi (|), (=), (+), etc. The winning strategy for (PI), for instance, presupposes the proponent's familiarity with the number-construction via (|). In fact, that's why Lorenzen calls arithmetic 'synthetic' — it is based on construction (*synthesis*). Logical truth, by contrast, is based only on the argumentation-governing rules, i.e. on linguistic norms. As such it may be called 'analytical'.

Instructive as it is, this differentiation doesn't seem to cut deep. Its conventionality points rather in the opposite direction, namely that arithmetic and logic are closely related. They are both dealing with symbols and accordingly may be called 'formal'. In fact, this is Lorenzen's own proposal. His alternative differentiation takes advantage of Gödel's incompleteness result, thereby putting both the technical and the ideological part of modern logic in perspective. The basic idea goes as follows: Just as we have calculized basic arithmetical concepts (such as number, addition, etc.), we can attempt to calculize arithmetical and logical truth. In other words, one may try to describe true arithmetical and logical sentences *only* by means of some mechanical device, manufacturing them as mere syntactic figures according to the respective schematic rules. Knowing that this is *possible* for logic and *impossible* for arithmetic (both due

[5] Lorenzen's original idea, however, was to show that only the truths of Brouwer's or Heyting's intuitionistic logics are justifiable by the dialogical, i.e. pragmatic means. In the course of the development it turned out that one can dialogically justify both classical and intuitionistic (and in fact many other) concepts, depending on the additional rules which specify (not only *how*, but this time also) *when* it is allowed to attack the opponent and *when* it is allowed to defend oneself against his attack. Hence, speaking of logical truth we need to specify, in advance, what additional dialogical rules we are using.

to Gödel), the completeness/incompleteness distinction seems to be finally the required solution to our introductory problem. In the rest of the article, we will discuss this promising possibility, and we will soon discover that things are less easy-going than we might wish. A straightforward presentation of Gödel's incompleteness results turns out to be a necessary part of this enterprise.

If we interpret 'analytical' as something like 'free of intuition', thereby indicating that we can possibly relinquish the non-schematic, 'material' controllability of the relevant concepts, then arithmetic post-Gödel is certainly a non-analytical discipline. As we shall emphasize, arithmetical methods of proof *provably* transcend any attempted schematization. However, as we have already pointed out, one cannot conclude from this *negative* evidence that arithmetic *unlike* logic is synthetic, if only for the reason that the original distinction between the constructive and the conceptual is rather blurred in our Hilbertian operative update: they are both dealing with symbols.

My point is that we should first rather examine closer the alleged independence of arithmetic of pure schemata. After all, there are at least two extreme epistemological doctrines of arithmetical truth which consider themselves to be vindicated by the bare fact of incompleteness. The first of them is mathematical mentalism (instantiated in the intuitionism of Brouwer) basing arithmetical truth on mental constructions as opposed to linguistic conventions which, according to Brouwer, are totally heterogeneous with respect to mathematics. The second one is mathematical Platonism with its stress on the independence of arithmetic not only of the individual human subject (which has a point), but of the whole of mankind as well. According to both of them arithmetic is non-analytical. But neither intuitionism nor Platonism, provides us with a satisfactory analysis of how *we* can know that an arithmetical sentence is true and what such a proclamation should mean, not to mention their respective treatments of Gödel's theorems.[6] Let us begin with this.

5 What is an arithmetical rule?

First of all, Gödel's theorems apply only to the theories that are *axiomatized* effectively. This doesn't imply any kind of strong finitism, because we don't want, for example, to rule out axiomatized theories with in-

[6] By this I mean especially the most popular versions of their interpretation such as: 'we know that we cannot know everything' (Socratian modesty) or reference to the so-called intuition as a kind of mysterious power available only to the chosen few (the mathematician's 'sixth sense').

finitely many axioms or rules. These axioms and/or rules,[7] however, should be mechanically testable (recognizable), which already implies that they have to be finite sequences of symbols and, as a consequence, that formal deductions (arithmetical proofs) have to be mechanically testable, too. Speaking rather more technically, a theory is effectively axiomatized if and only if it has a *decidable* set of axioms and uses a proof system in which it is *decidable* whether a sequence of well-formed formulas is a proof.

On the other hand, there seems to be no theoretical reason why some infinite rules couldn't under appropriate circumstances be regarded as 'effectively' manageable, too. Let us take, for example, the so-called ω-rule

$$A(|), A(||), A(|||), \text{ etc.} \qquad \Rightarrow \qquad (\forall x)A(x). \tag{ω}$$

As an arithmetical rule it is transparent and sound enough, as long as one interprets the "etc." correctly. In fact, the Gödel-Tarski idea of semantics employs this kind of infinite rules systematically, with the ω-rule as a special case of the more general

$$A(N) \text{ for all substituents } N \qquad \Rightarrow \qquad (\forall x)A(x). \tag{$\forall$}$$

The $\forall$-rule is then nothing but the well-known part of the so-called semantical definition of truth. But let us be careful! Contrary to the ω-rule, the $\forall$-rule doesn't partake in any concrete definition of truth but represents only a truth-schema. The semantical definition ignores the evaluation of elementary sentences.

The whole point of the last paragraph is to make us think about semantic definitions such as ($\forall$) as special (more generously conceived) systems of rules (proof systems) which — starting with some elementary sentences — evaluate the complex ones by *exactly one* of two truth values: true or false. Constructive (effective) or intuitionistic logic denies this very possibility, arguing that from the mere *non-existence* of a winning strategy for $A(x)$ one cannot validly conclude that there is a concrete strategy for some $\neg A(N)$ or, in particular, that the existence of concrete strategies for winning or refuting every $A(N)$ doesn't entail the existence of a general strategy for $A(x)$.

To illustrate my point by a more familiar example, let us consider this: There is no problem in demonstrating whether, for any given even number M, it is the sum of two primes. However, the truth value of

[7] Axioms can easily be interpreted as rules and *vice versa*.

the general judgment that every even number is the sum of two primes (the so-called Goldbach conjecture), is still unknown, 250 years after the problem was first posed. Hence, although we could potentially access individual strategies for every single number, we still do not know the general strategy of how to win a proposition concerning them all. Consequently, a decision must be made whether the infinite vehicles of truth and judgment such as (ω) or ($\forall$) should be referred to as rules

(1) only in the case when we positively *know* that all their premises are true, i.e., when we have at our disposal some general strategy for winning all of them at once, or,

(2) more liberally, if we *know* somehow that all their premises are positively true *or* false.

Since the concept of elementary arithmetical truth — as defined by the calculi (=), (+), etc., plus the dialogical rule for negation — is strongly effective, i.e. mechanically testable, we can choose it as our basis. Starting from it we can subsequently arrive at the concept of constructive or classical arithmetical truth, depending on how we interpreted the infinite rule (ω).

Moreover, the classical concept of truth obtained by a more liberal reading of (ω) allows us to articulate the important distinction between the truth we already know and the truth which has not been recognized yet, but is recognizable in principle. (Incidentally, Frege's semantical pair of sense and reference — of truth-conditions and truth-value — aims at the same thing.) What seems to beg the question now is

(1) the compatibility of this mild 'semantical Platonism' (as Stekeler once called it) with the possibility of arithmetical truths which are not only unknown at the moment, but unknowable in principle, on the one hand, and

(2) the widespread opinion that such a strong 'ontological Platonism' is validated by Gödel's theorems, on the other.

In what follows, we can forget about the first part of this question as long as we remember that the problem with Platonism doesn't lie in its compatibility or incompatibility with our experience, but in the lack of better arguments in its favour. The second part, however, is relevant here and we want to refute it in the next section.

6 What is arithmetical truth?

Although the more liberal reading of (ω) gives us a better idea about the so-called standard model of arithmetic, which as we are usually told can be described only in an intuitive way, the constructivists do have a point when saying that the words “rule” or “inference pattern” refer ordinarily to something one can actually follow, hence that a rule which one can follow only *in principle* is not in fact a rule. Imposing the condition of their effective controllability on the premises of (ω) we actually obtain the above mentioned constructive meaning of the quantified sentence $A(x)$: it is true (justified) if and only if there is some general winning strategy for every substitutable name N, i.e. for every sentence $A(N)$.

Since in the constructivist reading the concept of winning strategy remains to a large extent deliberately open, there is always room for an effective, yet liberal enough semantics and a strong effective finite or ‘mechanical’ syntax or axiomatics. These axiomatics or strong finite rule-systems (i.e. systems with finite rules) are called *full-formalisms*, and those more liberal ones (i.e. systems with infinite rules, no matter if classically or constructively interpreted) are referred to as *semi-formalisms*; both distinctions are due to Schütte [1960]. The important thing is that Gödel's theorems affect only the full-formal systems.

Gödel came up with a general metastrategy of how to construct, for every full-formal (hence schematically given) system of winning arithmetical strategies (i.e. axioms and/or rules), a justifiable arithmetical sentence not winnable by them. This (meta)strategy rests on the so-called diagonal construction and on a presupposition that the starting system is sufficiently strong, since the weak systems are incomplete by definition.

The basic idea behind Gödel's proof looks like this: After devising an appropriate coding scheme we can express many sentences about a certain arithmetical full-formalism in its language and even deduce them in it in accord with the truth. Firstly, we associate arithmetical expressions with numbers (codes) in such a way that a particular number fulfils a concrete arithmetical condition if and only if the encoded expression fulfils a certain syntactic condition, e.g., ‘to be an axiom’, ‘to be a proof’ etc. Secondly, we take into consideration the syntactic relation holding between two expressions if and only if the first of them is the proof of the second one and we name the corresponding arithmetical condition

$$\text{Proof}\,(x, y).$$

It holds for two numbers m, n just in case m codes the proof of the formula coded by n. And finally, we need an operation

$$\text{subst}\,(x, y)$$

which yields, for two numbers m, n as arguments, the code p of the result of substituting the numeral for n (i.e. the arithmetical expression 'n') for any occurrence of the sole free variable "x" in the formula $F(x)$ coded by m. Therewith we have the formula

$$(\forall y)\neg\,\text{Proof}\,(y, \text{subst}\,(x, x))$$

at our disposal. Let us abbreviate it as $G(x)$. — So far so good. The formula $G(x)$ has "x" as its sole free variable, it is associated with some code g and it holds of m if the formula coded by subst (m, m) is not provable in the relevant full-formalism, i.e., if there is no number n coding the proof of the formula in question. Substituting the numeral for g in $G(x)$ yields the formula

$$G(\mathrm{g}) \rightleftharpoons (\forall x)\neg\,\text{Proof}\,(x, \text{subst}\,(\mathrm{g}, \mathrm{g})).$$

This is, in fact, the critical formula we are looking for, because $G(\mathrm{g})$ is true (justifiable) if and only if there is no number m such that $\text{Proof}\,(\mathrm{m}, \text{subst}\,(\mathrm{g}, \mathrm{g}))$ holds, in other words: if and only if the formula coded by subst (g, g) is unprovable in the given full-formalism. But this formula is $G(\mathrm{g})$ itself! Since the full-formalism is constructed in such a way that it deduces only true (justifiable) arithmetical sentences, the formula $G(\mathrm{g})$ cannot be provable, otherwise we would have a false theorem. Hence, $G(\mathrm{g})$ is not provable, hence $G(\mathrm{g})$ is true!

The reason why we retell the whole story in terms of the semi- and full-formalisms and the winning strategies lies in the observation that the unprovable yet true arithmetical sentence of Gödel's theorem is an unprovable sentence of the full-formalism but a provable (i.e. justifiable or true) sentence of the semi-formalism: there is a strategy of how to win $G(\mathrm{g})$, but also of how to construct a new unprovable yet true formula in the case when $G(\mathrm{g})$ is added *ad hoc* to the original full-formalism as a new axiom. Now it is quite clear how the concept of essential unprovability is *not* to be understood, namely Platonistically, no matter if in a strong or a mild sense. The reason is that Gödel's proof does not even cross the border of the constructive semi-formalism. — Hereafter, therefore, we will carefully differentiate between the deducibility in a full-formalism and in a semi-formalism, using the well-established symbols "$\vdash$" and "$\models$", respectively. To sum up, the theses we have established so far are the following:

(1) The essential incompleteness or incompletability of arithmetic affects only the arithmetical full-formalism, which means that there is always a true arithmetical formula which is not a theorem.

(2) 'Incomplete' thereby always means 'incomplete with respect to some semi-formalism' that defines which sentences are to be evaluated as true and which not.

(3) For this reason the semi-formalism itself cannot be incomplete. However, it can be incomplete with respect to some other semi-formalism, which is actually the case of the constructive semi-formalism in relation to the classical one.

In this particular case, however, we are not able to prove this incompleteness to be essential, so we can keep on believing with Hilbert that every mathematical problem is solvable and we can prospectively announce this general solvability as a kind of regulative (optimistic) hypothesis ("wir müssen wissen, wir werden wissen").[8)]

(4) Even if we interpret arithmetical truth classically, i.e. not effectively, the unprovable yet true formula of Gödel's theorem remains constructively true, i.e. provable in the constructive semi-formalism.

Therefore we should formally differentiate between $\vDash_{\mathcal{C}}$ and $\vDash_{\mathcal{L}}$ ($\mathcal{C}$ standing for "classical", $\mathcal{L}$ for "Lorenzen").

7 Is arithmetic consistent?

Along the same lines we have explained Gödel's theorem we can now handle its corollary, better known under the name of the Second Incompleteness Theorem. This is in fact the famous slogan that the consistency of arithmetic cannot be proved by appealing to arithmetical means alone, or even that it cannot be established at all. Of course, we ought to be cautious here again. A sober reading of the corollary entails only something like this:

> There is an actual strategy of how for every full-formalization T of arithmetic (which is, again, strong enough) to construct a sentence S_T of which the following conditions hold:

8) [Hilbert, 1935, p. 387].

(1) S_T is justifiable ($\models$) if and only if T is consistent,

(2) S_T is unprovable ($\vdash$) in T if and only if T is consistent.

Then, on closer look, it is clear that the condition (2) alone amounts to a triviality: we can simply take S_T to be any contradiction we like. Adding the condition (1) we want the unprovable sentence of (2) to be true. But there is a sentence like this already due to the (First) Theorem. So what is the point of the corollary?

To understand its significance we have to look at it with Hilbert's eyes, i.e. from the point of view of a proponent of strong effective — full-formal — systems. Then there is no place for a semi-formal justification and for a general, non-effective concept of arithmetical truth in particular. According to Hilbert, semantical concepts such as this one could lead us only badly astray (as Russell's paradox had shown) and just for this reason they should eventually be replaced with something more secure, e.g., with syntactical consistency. With this in mind we can first rephrase the concept of incompleteness in syntactic terms and say that an axiomatic (full-formal) theory T is incomplete if there is a sentence S of its language such that neither S nor $\neg S$ are provable ($\vdash$) in T. More importantly, Gödel's theorem can be generalized in this way, i.e. without assuming the truth of the full-formal system in question. The argument goes roughly like this.

Firstly, one must show that the elementary syntactic properties (formulahood, axiomhood, proofhood) are not only adequately expressible in arithmetical language, which means that the arithmetical formulas assigned to these properties are provable or refutable in the arithmetical semi-formalism in accordance with truth, but that these formulas are provable or refutable in the relevant full-formal system T as well. This second characteristic is known as the case-by-case capturing in T and holds of the aforementioned syntactic properties already in the Robinson Arithmetic (Q), i.e. in the (first-order) Peano Arithmetic (PA) without the induction.

Now consider the sentence $G(\mathrm{g}) = (\forall y)\neg\,\mathrm{Proof}\,(y, \mathrm{subst}\,(\mathrm{g}, \mathrm{g}))$ again and suppose it is provable in T. Then there is some number m which codes its proof and, by definition, the sentence $\mathrm{Proof}\,(\mathrm{m}, \mathrm{subst}\,(\mathrm{g}, \mathrm{g}))$ is provable in T, too. But the provability of $(\forall y)\neg\,\mathrm{Proof}\,(y, \mathrm{subst}\,(\mathrm{g}, \mathrm{g}))$ entails the provability of $\neg\,\mathrm{Proof}\,(\mathrm{n}, \mathrm{subst}\,(\mathrm{g}, \mathrm{g}))$ for every n, particularly for $n := m$. Hence

$$T \vdash \mathrm{Proof}\,(\mathrm{m}, \mathrm{subst}\,(\mathrm{g}, \mathrm{g})) \text{ and } T \vdash \neg\,\mathrm{Proof}\,(\mathrm{m}, \mathrm{subst}\,(\mathrm{g}, \mathrm{g})),$$

which means that T is (syntactically) inconsistent. Hence, if T is consistent, which is the formalist's substitute for truth, there can be no proof of $G(\mathrm{g})$ in T. In order to show that $\neg G(\mathrm{g})$ is unprovable too, we need a stronger assumption than a mere consistency of T. — Gödel introduced the concept of ω-consistency, defining a theory to be ω-consistent if the fact that $T \vdash \neg A(n)$ for each n excludes the possibility of $T \vdash (\exists x)A(x)$. Note that ω-consistency implies plain consistency, because T is inconsistent if and only if all formulas are provable in it, so in the case of an inconsistent theory the forbidden combination of provable formulas mentioned in the definition of ω-consistency obtains automatically. — Suppose that T is ω-consistent and $\neg G(\mathrm{g})$ is provable in T. Then $G(\mathrm{g})$ is unprovable, because of plain consistency, which means that no number m can code a proof of $G(\mathrm{g})$. Hence, by definition, $\neg\,\mathrm{Proof}\,(\mathrm{n}, \mathrm{subst}\,(\mathrm{g}, \mathrm{g}))$ holds ($\vDash$) for each n and, by the requirement that T captures the relation 'to be a proof of', even

$$T \vdash \neg\,\mathrm{Proof}\,(\mathrm{n}, \mathrm{subst}\,(\mathrm{g}, \mathrm{g})) \text{ for each } n.$$

But we are assuming that $\neg G(\mathrm{g})$ is provable in T, which is equivalent to $T \vdash (\exists y)\,\mathrm{Proof}\,(y, \mathrm{subst}\,(\mathrm{g}, \mathrm{g}))$, and that makes T ω-inconsistent, contrary to the hypothesis.

Now for the Second Theorem. According to Hilbert, syntactic consistency is a sufficient and in fact the only requirement one can impose on an axiomatic theory. This requirement can be coded in a familiar manner as the unprovability of a contradictory formula in T, i.e. as $(\forall x)\neg\,\mathrm{Proof}\,(x, \mathrm{m})$, where m is, e.g., the code of "$0 = 1$". (This is, of course, because we assume that the negation of "$0 = 1$" is provable in T.) Let us abbreviate $(\forall x)\neg\,\mathrm{Proof}\,(x, \mathrm{m})$ as $\mathrm{Con}\,T$. — The content of the First Theorem in its syntactic version can now be formalized as

$$\mathrm{Con}\,T \rightarrow G(\mathrm{g})$$

and eventually proved in T, if T is at least as strong as PA, hence stronger than Q. Assuming T is consistent, this already yields that $\mathrm{Con}\,T$ must be unprovable, since $T \vdash \mathrm{Con}\,T$ and $T \vdash \mathrm{Con}\,T \rightarrow G(\mathrm{g})$ entail $T \vdash G(\mathrm{g})$, which, as we know from the First Theorem, is not true.

No matter how interesting the unprovability of $\mathrm{Con}\,T$ in T can be, it should be clear that it has nothing to do with the consistency or inconsistency of arithmetic. It makes no sense to construct a formula provable in T if and only if T is consistent since an inconsistent theory entails every formula. In his controversy with Hilbert about the nature of axioms, i.e. of axiomatic theories in general, Frege was right to point out that the consistency of axioms is secondary to their truth: When we are devising an axiomatization of arithmetic, we are obliged, of course, to pick up only true sentences as axioms. Their mutual consistency does not suffice (for them to be picked out), and, as a matter of fact, follows from the fact they are true.

With the proof-theoretical jargon of semi- and full-formalisms at our disposal and with hindsight on the Second Theorem we can now articulate the problem of arithmetical consistency as follows: the already established full-formal systems such as PA or Q are consistent simply because their axioms are provable in the arithmetical semi-formalism. This is, in fact, the usual model-theoretical argument: a theory is consistent because there is a model for it.

The main use of our proof-theoretical diction lies in its relativity. If PA is inconsistent, then arithmetic full-formalism is inconsistent. In the model-theoretical jargon, where semi-formalism is replaced by the so-called standard model, we are usually told that this possibility is precluded simply by definition. Eventually an appeal is made to some kind of intuition. In the proof-theoretical case we do not confine ourselves to such vague justifications, because we can actually prove that the rules of the semi-formalism do not evaluate arithmetical sentences inconsistently. Our method is an easy metainduction:

(1) Elementary arithmetical sentences ($m + n = p$, $m \times n = p$) are evaluated unambiguously as true or false only on the basis of the arithmetical calculi we have set down before,

(2) Tarski's evaluation of complex sentences is correct, too, though we can still argue about whether they assign one and only one value to each sentence (as Stekeler's semantical Platonism maintains)[9] or at most one value to each sentence (as Brouwer and his followers believe).

In fact, this is the difference between the classical and constructivist conception of truth, logic and arithmetic.

8 Conclusion

The import of the Second Theorem seems to lie in its support for the evidence that arguments for the consistency of arithmetic cannot avoid appeal to the infinite, semantical methods. Consistency proofs by Gentzen

[9] See Stekeler-Weithofer [1986].

build on this very idea. In fact, this inevitability is possibly already given in the infinite construction of $|$, $||$, $|||$, $\dots$, and hence one can, on the one hand, repeat Kant's doubts about arithmetical sentences being deducible from a single formal definition and, on the other hand, expect with Poincaré that this fact has something to do with the complete induction.

But we ought to be cautious not to carry this observation too far, as Poincaré [1908] did, having declared complete induction a dividing line between logical and mathematical methods. Of course, complete or mathematical induction is indispensable in mathematics, but this is because numbers (arithmetical operations etc.) are defined inductively, not because it is somehow essential for arithmetic *per se*. Realizing this we can see that induction is indispensable in logic too as long as the basic concepts like formula, theorem or proof are employed. In his *Tractatus* Wittgenstein availed himself of this observation by interpreting natural numbers to be indexes of a sentence-forming operation. Shall we conclude from this evidence that there is no difference between logic and mathematics?

Wittgenstein's later answer would certainly have been negative: Arithmetic and logic are, of course, different, simply by definition: the first one makes computations, the second one inferences. The completeness and incompleteness phenomenon has little to do with this. In fact, instead of interpreting the variable X in the formula

$$(\forall X)[X(0) \wedge (\forall y)(X(y) \rightarrow X(y+1)) \rightarrow X(x)]$$

schematically, i.e. as ranging over expressions of a formal language, as the first-order version of the induction axiom (schema) does, we can make use of the original, indefinite way, which amounts to the second-order axiom. This is the part of the full-formal system of second-order Peano Arithmetic (PA_2) which is, of course, incomplete due to Gödel's Theorem. We can, however, argue that in contrast with PA the axioms of PA_2 describe arithmetical semi-formalism so well (or technically: up to isomorphism), that every semi-formalism which entails them (under some structure-preserving translation) already entails all theorems of the original semi-formalism (under the same structure-preserving translation). This is nothing but the well-known categoricity theorem for the second-order Peano Arithmetic. As a consequence, to justify an arithmetical sentence we do not need elementary arithmetical calculi any more, we need only Peano's axioms. The point is that for any true arithmetical sentence S the conditional

$$\mathsf{PA}_2 \rightarrow S$$

becomes justifiable by logic alone, which means that it is a tautology! Assuming the underlying logic is complete in the sense that all and only the tautologies are deducible, the arithmetic becomes complete, too, which is impossible. Hence, by contraposition, second-order logic is also incomplete. So, Gödel's Theorem, in fact, proves the incompleteness not only for arithmetic, but for logic as well.

But then, we need to ask, what is the moral of Gödel's Incompleteness Theorem with respect to the question of the epistemic status of arithmetic or logic? Are there any essential differences between them? Of course there are, but they are not easily to be found, given the condition that there are remarkable similarities between counting and judging and these undoubtedly are built into the foundations of modern logic. The problem of our question and of the postlogicist philosophy of mathematics in general lies precisely in the fact that they do not take this into account.

So Poincaré was actually right. Modern logic was successful because of availing itself of methods peculiar to mathematics, especially of complete induction. But this did not turn logic into mathematics, as Poincaré suggested; neither did it turn mathematics into logic as the logicists thought. The relations between them were nevertheless changed or distorted, if we wish, and it made them, or certain parts of them, simultaneously more powerful with respect to some problems and objections and more vulnerable with respect to others. But this seems to be a necessary epiphenomenon of any scientific development. — Nevertheless, under the influence of logic, mathematics has become more sensitive to the syntactic design of its theories, making them available for intersubjective checking by devising a transparent, uniform concept of (deductive) proof. The negative side of this move was, of course, the above mentioned identification of arithmetical truth with deductive consistency. Gödel's theorems, as we have interpreted them above, are only the symptoms of some consequences of this decision.

In this article I have tried to argue that they tell us nothing fatal about the nature of our reason, nor anything about logic and arithmetic as its prominent offspring, as long as we are aware that they are (by definition) disciplines of their own, as Wittgenstein used to stress. Simultaneously I want to point out, *pace* the radical scepticism of Poincaré and Wittgenstein, that the story of modern logic shows us how fruitful the possible crossovers of these two disciplines of pure reason can be if they are interpreted in a modest, dialectical way, i.e. not as the reduction of the whole of logic to arithmetic nor *vice versa*, but as the projection of a part of the former onto part of the latter, leading eventually to a discipline of a new, somewhat mixed kind, as already displayed by subjects such as metamathematics, proof- or model-theory, computational complexity and many others.

Four

The Discrete Charm of Non-standard Real Numbers

Kateřina Trlifajová

The absence of the theory of real numbers before the late nineteenth century was one of the main barriers in the development of modern mathematics. Although studies in analysis, differential geometry and algebra, all of which utilized real numbers, were quite advanced, mathematicians often operated on the basis of their intuitive understanding. There were at least three main requirements stimulating the creation of a rigorous foundation of the reals:

(1) The arithmetization of real numbers would give an understanding of the continuum. Reals represent a metaphorical bridge between discreteness and continuity, numbers and space, arithmetic and geometry. They express in a numerical way what points, lines and space are.

(2) A theory of reals was needed in order to serve as an exact foundation of calculus. For over two hundred and fifty years this powerful instrument had been used, and new branches had stemmed from it: the mathematical analysis, the theory of differential equations, differential geometry, calculus on variation, functions of complex variables

The work on this paper is a part of the research project MSM 0021620845 financed by the Ministry of Education of the Czech Republic.

and some others. Mathematicians did not have a correct mathematical basis for their investigation, but were guided by an intuitive and physical insight. Their insight was based on the original Leibniz and Newton calculus as a computation with ideal infinitely small entities, *infinitesimals*. It was necessary to compute carefully, as there was no exact theory of infinitesimals, nor a set of exhaustive rules on how to deal with them. Their investigation was thus accompanied by errors and confusions of the creative process.

(3) The creation of non-Euclidean geometry had shaken geometry's status of absolute truth. It still seemed that mathematics based on the ordinary arithmetic should preserve this position. For instance, Gauss distinguished arithmetic from geometry in that only the former was purely *a priori*, only laws of arithmetic were necessary and true. The foundation of a general number system would disperse doubts about the truth of arithmetic.

The fulfilment of this third requirement appeared to have failed in the twentieth century when Gödel published his major and dismaying discovery in 1931. Gödel's famous Incompleteness Theorem demonstrated that in any system rich enough to contain formal elementary arithmetic there were always theorems that could never be proven or disproven. Arithmetic was relegated to the position of geometry. It is relatively true but it cannot claim to absolute truth.

We can speak about truth in mathematics in two ways: as the consistency of a given theory and as the correspondence of a theory with our insight into reality. The former must naturally hold for any mathematical theory, while the latter is considered to be less important. In this paper, we will investigate how later constructions of real numbers satisfy the first two requirements and how they correspond to our intuition of reality.

1 Bolzano's theory

The question of the structure of real numbers came up in Bolzano's paper of 1813. He tried to prove by purely analytical ways (that is arithmetic here) that if a function $f(x)$ is continuous in the interval $[a,b]$, $f(a) > 0$, $f(b) < 0$ then there is always a c such that $a < c < b$ and $f(c) = 0$. This means: if a continuous function represented by a line is positive on one side of the interval and negative on the other side then it must cross the x coordinate necessarily. This assertion is not too surprising, and mathematicians have used it quite often, but their reasoning was only geometrical. Bolzano formulated here the so-called Bolzano-Cauchy criterion of an infinite sequence $\{a_n\}_{n=1}^{\infty}$ as follows:

$$(\forall k > 0)(\exists n > 0)(\forall m > 0)\left(|a_n - a_{n+m}| < \frac{1}{k}\right).$$

But he could not prove that Bolzano-Cauchy sequences are convergent because there was no arithmetic construction of real numbers available. He did not have a name for the point — real number — to which such sequence converges.

Bolzano was aware of this gap in his proof. It was one of the reasons for his attempt around 1836, to build his own theory of real numbers. Although Bolzano's conception of infinity is unusual, his theory is correct, and can be interpreted both in a standard and in a non-standard way. However, his study remained only in a manuscript, as much of Bolzano's work was banned, and thus had no influence for more than a century.

2 Cantor's conception of completion

Cantor is very famous as the inventor of set theory, a mathematical theory of actual infinity. The first step toward this theory was the foundation of the theory of real numbers. His construction — the so-called completion of rational numbers — is well-known: he added to all Bolzano-Cauchy sequences (he called them "fundamental") their limits. If two fundamental sequences $\{a_n\}_{n=1}^{\infty}$, $\{b_n\}_{n=1}^{\infty}$ had the property that for any $\varepsilon > 0$ there was an n such that for all $m > n$ one could say that $|a_m - b_m| < \varepsilon$, Cantor defined that these sequences had the same limit. The limits represent real numbers. They keep all characteristic properties:

(1) linear ordering,

(2) density: $(\forall x, y \in \mathbb{R})(\exists z \in \mathbb{R})(x < z < y)$,

(3) completeness, i.e., all Bolzano-Cauchy sequences of real numbers have their limits in $\mathbb{R}$, and moreover

(4) they are Archimedean: $(\forall x, y \in \mathbb{R})(\exists n \in \mathbb{N})\left(\frac{1}{n}|x| < |y| < |nx|\right)$.

Their geometric interpretation is that of points in a line. If we denote rational numbers by $\mathbb{Q}$, their completion by $\hat{\mathbb{Q}}$ and real numbers by $\mathbb{R}$ we can symbolically describe Cantor's construction as:

$$\mathbb{Q} \quad \longrightarrow \quad \hat{\mathbb{Q}} \cong \mathbb{R}.$$

The investigation of the structure of sets of real numbers led Cantor to the discovery of ordinal and cardinal numbers, and to the creation of set theory. It was accepted as the right theory of mathematical infinity by the end of the nineteenth century.

3 Cantor's rejection of infinitesimals

A hope had risen for some mathematicians. Mathematical analysis could indeed be based on infinitesimals because the theory of infinite numbers intrinsically justifies infinitely small numbers as their inverse values. But Cantor denied this idea very firmly. Ordinal or cardinal numbers cannot serve as inverse values of infinitesimals as Cantor well knew. They can be added and multiplied and it would be difficult to find consistent arithmetical laws for their inverse values. Infinitesimals contradict the Archimedean Axiom, for this axiom follows from the completeness of real numbers. If they were admitted as a new sort of numbers sandwiched between rational and irrational numbers, they would only further complicate the already complicated problem of the continuum hypothesis.

Cantor was sure that his characterization of the infinite, and of real numbers, was the only characterization possible. He claimed that the theory of infinitesimals was on a par with the attempts to square a circle, i.e. impossible. His arguments were so persuasive that Bertrand Russell argued in his *Principles of Mathematics* that mathematicians, fully understanding the nature of real numbers, could safely conclude that the non-existence of infinitesimals was firmly established. He was wise enough to add, however, that if it were ever possible to speak of infinitesimal numbers, it would have to be in a radically new sense.

4 Foundations of the calculus

The classical model of the continuum — either by Cantor or by Dedekind — is correct and consistent. The continuum is composed of infinitely many (symbolic) real numbers. Reals are arithmetic expressions of points of a line. They satisfy all we expect from real numbers in terms of linear ordering, density and completeness. Calculus is based on the so-called ε-δ approach that is now generally accepted:

> "For every $\varepsilon > 0$, there is a $\delta > 0$ such that …"

is a typical phrase by which definitions of limits, continuity, differentiation and convergence of functions begin. Surprisingly, these formulations evoke a potential infinity, and make no use of Cantor's infinite numbers representing the actual infinity. But we can query the transparency of this approach. The idea of infinitely small seems to appeal more naturally to our intuition. The use of infinitesimals was widespread during the rise of the differential and integral analysis for nearly three centuries. It is simpler and clearer to compute with infinitesimals than to describe the computation in terms of limits.

The question now is whether it is possible to find a theory of real numbers which is a consistent model of the continuum, a consistent base for the calculus based on infinitesimals and which is appropriate to our intuition. This is not only a question of truth as a consistency, but also a question of truth as a comprehensibility and an insight.

5 Non-standard constructions

Non-standard real numbers are extensions of standard reals that contain infinitesimals. Our examples are Robinson's Non-standard Analysis, Vopěnka's Alternative Set Theory and Nelson's Internal Set Theory. Their constructions are based on a similar principle.

Russell was right that if infinitesimals could be accepted it must be done in an entirely new sense not only as quantities sandwiched between standard reals. Thus, we introduce a class of new entities, non-standard numbers, a *substratum*, and we denote it by S. The structure of real numbers is defined on it. — Rather than being just symbols denoting limits or cuts, real numbers are represented by the so-called *monads*,[1)] which are collections of non-standard numbers from the *substratum* S. The natural numbers $\mathbb{N}$ and the rational numbers $\mathbb{Q}$, or their isomorphic representatives, are subsets of S. We can define a non-standard number $a \in S$ as:

(1) infinitely small (an infinitesimal) $\rightleftharpoons (\forall n \in \mathbb{N}) \left(|a| < \frac{1}{n}\right)$,

(2) finite $\rightleftharpoons (\exists n \in \mathbb{N})(|a| < n)$,

(3) infinite $\rightleftharpoons (\forall n \in \mathbb{N})(|a| > n)$,

(4) two non-standard numbers $a, b \in S$ are infinitely close, denoted by $a \doteq b$, if and only if their difference is infinitely small,

(5) $\mathrm{Mon}\,(a) \overset{\mathrm{def}}{=} \{b \in S \mid b \doteq a\}$.

$\mathrm{Mon}\,(a)$ is called the *monad* of a. It contains all non-standard numbers that are infinitely close to a. The monad of zero $\mathrm{Mon}\,(0)$ contains all

[1)] Abraham Robinson used this poetic name, having borrowed it from Leibniz.

infinitely small numbers, we denote it also by S_i. The relation $\doteq$ is an equivalence. The class of finite non-standard numbers is denoted by S_f. We factorize the class S_f by the equivalence $\doteq$. The classes of this factorization are monads that represent real numbers

$$S_f|_{\doteq} \cong \mathbb{R}.$$

We can also describe this construction from an algebraic point of view. Non-standard numbers S form a commutative ring. Finite non-standard numbers S_f form a subring in S. Infinitely small non-standard numbers S_i form a maximal ideal in S, the monad of zero.

A well-known algebraic theorem says that the factorization of a commutative ring modulo its maximal ideal yields a field. Thus, factorizing S_f modulo S_i, we obtain a field, namely the field of real numbers $\mathbb{R}$:

$$S_f|_{S_i} \cong S_f|_{\doteq} \cong \mathbb{R}.$$

If we define an ordering on $S_f|_{S_i} = \mathbb{R}$ we can prove all necessary properties of real numbers: (1) linearity, (2) density and (3) completeness. The following is the symbolic scheme of the construction:

$$\mathbb{Q} \quad \longrightarrow \quad S \supseteq S_f \supseteq S_i \quad \longrightarrow \quad S_f|_{S_i} \cong \mathbb{R}.$$

Comparing the standard and the non-standard construction above, we receive the following commutative diagram:

$$\begin{array}{ccc} \mathbb{Q} & \longrightarrow & S \supseteq S_f \supseteq S_i \\ & \searrow \quad \swarrow & \\ & \hat{\mathbb{Q}} \cong \mathbb{R} \cong S_f|_{S_i} & \end{array}$$

This is a good philosophical model of a continuum. The basic example of a continuum is a line. Points of a line are expressed as monads. All points (monads) together fill a line. Any of its part is infinitely divisible, in accordance with Aristotle's characteristic of a continuum. Also according with the conceptions of Democritus and Zeno, the line is assembled from infinitely many infinitely small points. These points are not only symbolic, they have their own 'body' composed from non-standard entities.

The last but not least task remains: the construction of the universe of non-standard numbers S. There are several ways to achieve this, differing in their perspectives: a mathematical construction, a philosophical reasoning, a simplicity.

6 Robinson's non-standard analysis

The first mathematical construction of a non-standard theory of real numbers was created in 1963 by Abraham Robinson. In fact, he did not give an arithmetization of a continuum, a construction of reals from rationals. But it is a construction of non-standard reals from standard reals. Robinson employs the Zermelo-Fraenkel Set Theory with the Axiom of Choice (ZFC). He uses a free (nontrivial) ultrafilter $\mathscr{U}$ on natural numbers $\mathbb{N}$, the existence of which is guaranteed by the Axiom of Choice. Here is Robinson's construction in brief:

We proceed from real numbers $\mathbb{R}$. $\mathbb{R}^{\mathbb{N}}$ is the set of all the real sequences, or equivalently, the set of all functions from $\mathbb{N}$ to $\mathbb{R}$,

$$a = \{a(i)\}_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}.$$

The equivalence relation $\sim$ on sequences a, b in $\mathbb{R}^{\mathbb{N}}$ is introduced via the ultrafilter:

$$a \sim_U b \rightleftharpoons \{i \in \mathbb{N} \mid a(i) = b(i)\} \in \mathscr{U}.$$

$\mathbb{R}^{\mathbb{N}}$ modulo the equivalence relation $\sim$ yields the ultraproduct

$$\mathbb{R}^{\mathbb{N}}|_{\mathscr{U}} = \mathbb{R}^*,$$

called the non-standard reals. — If $a \in \mathbb{R}^{\mathbb{N}}$, its equivalence class in $\mathbb{R}^*$ is denoted by $a_{\mathscr{U}}$, every element of $\mathbb{R}^*$ is of the form $a_{\mathscr{U}}$ for any $a \in \mathbb{R}^{\mathbb{N}}$. The role of *substratum* S plays the class of all non-standard real numbers $\mathbb{R}^*$. We define naturally an embedding of $\mathbb{R}$ into $\mathbb{R}^*$ that assigns every r from $\mathbb{R}$ the element $r^* = r_{\mathscr{U}}$ that is the equivalence class of the constant function $r(i) = r$ for all $i \in \mathbb{N}$. We extend the ordering $<$ from $\mathbb{R}$ on $\mathbb{R}^*$ by

$$a_{\mathscr{U}} < b_{\mathscr{U}} \rightleftharpoons \{i \in \mathbb{N} \mid a(i) < b(i)\} \in \mathscr{U}.$$

We can easily prove that the extended domain $\mathbb{R}^*$ is linearly ordered. As an example we verify the transitivity of $<$ in $\mathbb{R}^*$:

> If $a_{\mathscr{U}}, b_{\mathscr{U}}, c_{\mathscr{U}} \in \mathbb{R}^*$ are such that $a_{\mathscr{U}} < b_{\mathscr{U}}$ and $b_{\mathscr{U}} < c_{\mathscr{U}}$ then $A = \{i \in \mathbb{N} \mid a(i) < b(i)\} \in \mathscr{U}$ and $B = \{i \in \mathbb{N} \mid a(i) < b(i)\} \in \mathscr{U}$, hence $A \cap B \in \mathscr{U}$ and $A \cap B \subseteq \{i \in \mathbb{N} \mid a(i) < c(i)\} \in \mathscr{U}$. Thus $a_{\mathscr{U}} < c_{\mathscr{U}}$.

Likewise, we can verify the trichotomy of $<$, i.e., if $a_{\mathscr{U}}, b_{\mathscr{U}} \in \mathbb{R}^*$ then either $a_{\mathscr{U}} < b_{\mathscr{U}}$, or $a_{\mathscr{U}} = b_{\mathscr{U}}$, or $b_{\mathscr{U}} < a_{\mathscr{U}}$. By the similar way, we extend the arithmetical operations $+$, $\times$, from $\mathbb{R}$ on $\mathbb{R}^*$:

$$a_{\mathscr{U}} + b_{\mathscr{U}} = c_{\mathscr{U}} \rightleftharpoons \{i \in \mathbb{N} \mid a(i) + b(i) = c(i)\} \in \mathscr{U}\,,$$
$$a_{\mathscr{U}} \times b_{\mathscr{U}} = c_{\mathscr{U}} \rightleftharpoons \{i \in \mathbb{N} \mid a(i) \times b(i) = c(i)\} \in \mathscr{U}\,.$$

Thanks to the properties of the ultrafilter, $\mathbb{R}^*$ is a linearly ordered field, an extension of $\mathbb{R}$. As described above, we define (1) infinitesimal, (2) finite and (3) infinite non-standard real numbers, (4) the relation of infinite closeness $\doteq$ and (5) monads. For instance, the function $a(i) = \frac{1}{i}$ defines a class of the ultraproduct that represents an infinitesimal, the function $b(i) = \frac{1}{i^2}$ represents another infinitesimal. The functions $c(i) = i, d(i) = i^2$ represent infinite numbers.

Let $x \in \mathbb{R}^*$ be finite. Thanks to the linear ordering of $\mathbb{R}^*$, we can define $D_1 = \{r \in \mathbb{R} \mid x > r^*\}$ and $D_2 = \{r \in \mathbb{R} \mid x < r^*\}$. The pair (D_1, D_2) forms a Dedekind cut in $\mathbb{R}$, hence determines a unique $r_0 \in \mathbb{R}$. We can easily prove that $r_0^* \doteq x$. Consequently, for every finite non-standard number x from $\mathbb{R}^*$ there is exactly one standard real number $r_0 \in \mathbb{R}$ such that the difference $x - r_0^*$ is infinitely small. We call this unique element r_0 the *standard* part of x. Of course, two infinitely close finite elements of $\mathbb{R}^*$ have the same standard parts.

Let $\mathbb{R}_f$ denote the set of all finite elements of $\mathbb{R}^*$, $\mathbb{R}_i$ the set of all infinitesimals. The relation of $\doteq$ is an equivalence. Monads are its equivalence classes. By factorizing $\mathbb{R}_f$ modulo $\doteq$ or equivalently modulo $\mathbb{R}_i$ we obtain the structure isomorphic to real numbers $\mathbb{R}$. The isomorphism is the mapping that assigns every monad the standard part of their elements:

$$\mathbb{R}_f|_{\doteq} = \mathbb{R}_f|_{\mathbb{R}_i} \cong \mathbb{R}.$$

The whole construction can be described symbolically as follows:

$$\mathbb{N} \subseteq \mathbb{R} \quad \longrightarrow \quad \mathbb{R}^{\mathbb{N}} \quad \longrightarrow$$
$$\mathbb{R}^{\mathbb{N}}|_{\mathscr{U}} = \mathbb{R}^* \supseteq \mathbb{R}_f \supseteq \mathbb{R}_i \quad \longrightarrow \quad \mathbb{R}_f|_{\mathbb{R}_i} \cong \mathbb{R}.$$

The main result about ultrafilter extensions is the theorem of Łoś. Its immediate corollary is the general transfer principle. It says that the same properties of the first-order logic hold for elements of $\mathbb{R}$ and $\mathbb{R}^*$:

> Transfer Principle. Let $\phi(X_1, \ldots, X_m, x_1, \ldots, x_n)$ be a formula of the first-order logic. Then for any $A_1, \ldots, A_m \subseteq \mathbb{R}$ and $r_1, \ldots, r_n \in \mathbb{R}$, $\phi(A_1, \ldots, A_m, r_1, \ldots, r_n)$ is true in $\mathbb{R}$ if and only if $\phi(A_1^*, \ldots, A_m^*, r_1^*, \ldots, r_n^*)$ is true in $\mathbb{R}^*$.

The extension of any $A \subseteq \mathbb{R}$ to $A^* \subseteq \mathbb{R}^*$ is defined as:

$$a \in A^* \rightleftharpoons \{i \in \mathbb{N} \mid a(i) \in A\} \in \mathscr{U}\,.$$

This principle enables us to define the same structures in $\mathbb{R}$ and $\mathbb{R}^*$, and guarantees their correspondence.

7 The non-standard constructions of reals from rationals

Robinson's construction is not a construction of reals from rationals. Nevertheless, the same way can be used for this case. The role of the *substratum* S is now played by the class of non-standard rational numbers. We proceed from the rational $\mathbb{Q}$ and the natural numbers $\mathbb{N}$. $\mathbb{Q}^{\mathbb{N}}$ is the set of all sequences of rational numbers. The ultraproduct $\mathbb{Q}^{\mathbb{N}}|_{\mathscr{U}}$ represents the class of non-standard rational numbers $\mathbb{Q}^*$:

$$\mathbb{Q}^{\mathbb{N}}|_{\mathscr{U}} = \mathbb{Q}^*.$$

As in the case of $\mathbb{R}^*$, we can define (1) a natural embedding of $\mathbb{Q}$ into $\mathbb{Q}^*$, (2) a linear ordering $<$ of $\mathbb{Q}^*$, (3) infinitesimals $\mathbb{Q}_i$, (4) infinite numbers, (5) finite rational numbers $\mathbb{Q}_f$, (6) infinite closeness $\doteq$ and (7) monads. But of course because $\mathbb{Q}$ is not complete not all elements of $\mathbb{Q}_f$ have their standard part in $\mathbb{Q}$.

By the factorization of finite rational numbers modulo infinitesimals or equivalently by the equivalence of infinite closeness, we receive the structure isomorphic to real numbers. If $x \in \mathbb{Q}_f$ we can define a Dedekind cut in $\mathbb{Q}$, $D_1 = \{q \in \mathbb{Q} \mid x > q^*\}$ and $D_2 = \{q \in \mathbb{Q} \mid x < q^*\}$. The pair (D_1, D_2) determines a unique $r_0 \in \mathbb{R}$ such that $r_0^* \doteq x$. Two infinitely close elements have the same standard part. The isomorphism is the mapping that assigns every monad this unique element:

$$\mathbb{Q}_f|_{\doteq} = \mathbb{Q}_f|_{\mathbb{Q}_i} \cong \mathbb{R}.$$

We can describe the whole construction symbolically as follows:

$$\mathbb{N} \subseteq \mathbb{Q} \quad \longrightarrow \quad \mathbb{Q}^{\mathbb{N}} \quad \longrightarrow \quad \mathbb{Q}^{\mathbb{N}}|_{\mathscr{U}} = \mathbb{Q}^* \supseteq \mathbb{Q}_f \quad \longrightarrow \quad \mathbb{Q}_f|_{\doteq} \cong \mathbb{R}.$$

This approach can be compared to the standard construction of Cantor. He proceeds also from the sequences of rational numbers from $\mathbb{Q}^{\mathbb{N}}$. He deals only with the set of all Bolzano-Cauchy sequences, which is a commutative ring, denoted by B. Two sequences have the same limit if their difference is a sequence converging to zero, and the set of all these sequences is a maximal ideal, denoted by C. We can say that Cantor makes a factorization of B modulo C immediately. He excludes infinitely small entities at the beginning and obtains a commutative field $B|_C$, i.e. real numbers $\mathbb{R}$. Cantor did not need any ultrafilters, the Bolzano-Cauchy condition sufficed:

$$\mathbb{N} \subseteq \mathbb{Q} \quad \longrightarrow \quad \mathbb{Q}^{\mathbb{N}} \supseteq B \supseteq C \quad \longrightarrow \quad B|_C \cong \mathbb{R}.$$

Cantor's completion can be interpreted in terms of non-standardness.

8 Infinitesimal calculus

The transfer principle provides a very good foundation for the infinitesimal calculus. Functions defined on $\mathbb{R}$ can be extended to $\mathbb{R}^*$. All their first-order properties are preserved. Thus we can define limits, continuity, derivative and integral of functions in a simple and natural way. For instance, a function f has a derivative in $a \in \mathbb{R}$ if there is $f'(a) \in \mathbb{R}$ such that for all non-zero infinitesimals $i \in \mathbb{R}_f$ the following condition is satisfied

$$f'(a) \doteq \frac{f(a+i) - f(a)}{i}.$$

Now you can enjoy proving the standard results on derivative of functions, for instance:

$$(x^2)' \doteq \frac{(x+i)^2 - x^2}{i} = \frac{x^2 - 2xi + i^2 - x^2}{i} = 2x + i \doteq 2x.$$

Nevertheless, the situation in the case of non-standard rational numbers is different. The transfer principle does not hold between $\mathbb{R}$ and $\mathbb{Q}^*$. We cannot easily extend functions defined on $\mathbb{R}$ to $\mathbb{Q}^*$ and consequently the infinitesimal calculus cannot be introduced directly. Although $\mathbb{Q}^*$ is a good arithmetical model of continuum, it cannot serve as a direct basis for the calculus. Of course, we can use the standard ϵ-δ approach.

9 Vopěnka's philosophical approach

A similar construction of real numbers appears in Vopěnka's Alternative Set Theory. Vopěnka does not work in ZFC or in any other modification of Cantorian Set Theory and he does not use ultrafilters. His concept of infinity is based on the phenomenological notion of the horizon encompassed by so-called *semisets*. He employs two types of natural numbers, the finite (standard) natural numbers $\mathbb{N}$ and all natural numbers $\mathbb{N}^*$ that involve also infinite numbers $\mathbb{N}^* \setminus \mathbb{N}$. Both kinds of numbers, $\mathbb{N}$ and $\mathbb{N}^*$, are models of the Peano Arithmetic. Finite numbers are 'in front of the horizon', they are accessible in a way. The class of all finite numbers $\mathbb{N}$ is a typical example of a semiset. Infinite numbers are 'beyond the horizon'.

The finite rational numbers $\mathbb{Q}$ form the quotient field of $\mathbb{N}$, the rational numbers $\mathbb{Q}^*$ form the quotient field of all natural numbers $\mathbb{N}^*$. Because elements of $\mathbb{N}^* \setminus \mathbb{N}$ are infinite, their inverse values in $\mathbb{Q}^*$ are infinitely small. While infinite numbers are 'beyond the horizon of the distance', infinitely small numbers are 'beyond the horizon of the depth'. Two elements of $\mathbb{Q}^*$ are infinitely close if and only if their distance is infinitely small. The relation of infinite closeness is an equivalence relation. Classes of this equivalence are semisets, and are called monads.

Again, we shall deal only with rational numbers that are smaller than any finite number $n \in \mathbb{N}$, and denote them by $\mathbb{Q}_f$. It is a commutative ring. Infinitely small rational numbers $\mathbb{Q}_i$ form a maximal ideal. By the factorization of $\mathbb{Q}_f$ modulo $\mathbb{Q}_i$ or equivalently by the equivalence of infinite closeness, we obtain the commutative field of real numbers. Their elements are represented by monads of infinitely close rational numbers:

$$\mathbb{N} \subseteq \mathbb{N}^* \quad \longrightarrow \quad \mathbb{Q}^* \supseteq \mathbb{Q}_f \supseteq \mathbb{Q}_i \quad \longrightarrow \quad \mathbb{Q}_f|_{\mathbb{Q}_i} = \mathbb{Q}_f|_{\doteq} \cong \mathbb{R}.$$

Although Vopěnka's construction is supported by a good philosophical reasoning, it is formally similar to the non-standard construction of reals from rationals described above. He uses semisets rather than ultrafilters, so his construction is simpler and clearer. Nevertheless, it cannot serve as a direct basis for the calculus using infinitesimals for the same reason. The transfer principle cannot be used here either.

10 Axiomatic non-standard theory

The properties of the ultraproducts are fascinating, but not straightforward. It is nearly impossible to imagine free ultrafilters. Thus if our preference is a simplicity we can use an axiomatic approach. An example of a non-standard theory based axiomatically is provided by the Internal Set Theory (IST) described in 1977 by Edward Nelson. This axiomatizes a basis of Abraham Robinson's non-standard analysis. It works in ZFC and constructs a theory extending ordinary mathematics that can serve as a good base for calculus.

IST adds only one predicate "standard" and three new axioms: the transfer principle, the idealization principle and the standardization principle. Using this new predicate, all the necessary notions are introduced. The axiomatic approach was also used by Vopěnka in his book *Calculus infinitesimalis*. He introduced four principles that are similar to Nelson's axioms. Then, he built a calculus based on infinitesimals.

11 Conclusion

There are many ways to introduce real numbers. The mathematical means applied for this purpose are quite different: ultrafilters, semisets, and fundamental sequences. A rare consensus prevails among mathematicians concerning the shape of the continuum. Although their conceptions of infinity may differ, the formal mathematical properties of real numbers remain the same: linearity, density and completeness.

However, the *substratum* on which real numbers are defined varies. In fact, Cantor' theory has no *substratum* and consequently no space for infinitesimals. Robinson's theory is a perfect base for non-standard analysis, but it is not a construction of reals from rationals, it is an extension of real numbers with help of ultrafilters. Vopěnka's construction is well-founded philosophically, though unsuitable for non-standard analysis. Axiomatic systems are the simplest, they formally describe properties of non-standard continuum, but avoid its construction.

■ Five

Consequence and Semantics in Carnap's Syntax

Karel Procházka

Given the title of Carnap's 'opus magnum', *The Logical Syntax of Language*, it is not surprising that one expects that what the text is about is (logical) syntax. These expectations are further strengthened by Carnap's own words:

> As soon as logic is formulated in an exact manner, it turns out to be nothing other than the syntax either of a particular language or of languages in general. [Carnap, 1937, p. 233]

And further:

> [T]he logic of science is nothing other than the logical syntax of the language of science. [Carnap, 1937, p. xiii]

It appears that Carnap's message is clear and simple: logic is syntax. And since philosophy is just the logic of science, philosophy is syntax, too. In a word, everything seems to be syntax.

Today we are wiser; we know that Carnap was wrong and that syntax is just a part of the story. We need semantics, and we need pragmatics, too. However, we have to be broadminded and not to reprobate Carnap too strongly. He himself realized the mistake and a few years later, influenced by Tarski, joined the semantic camp, leaving his 'syntactic period' far behind. *The Logical Syntax of Language* thus stands as a fine monument of philosophical error, and philosophers mention it from time to time to show how, by learning from the past mistakes, our knowledge advances.

This is the received picture. Alberto Coffa strove to show how ill-founded it is,[1] but I am afraid that despite his effort the received picture is still prevailing. In the book on logical consequence Etchemendy, for instance, says:

> According to Carnap's picture, a deductive system for a language, its logical syntax, was essentially independent of the language's semantics. The question of whether one sentence followed logically from another came down to the question of whether a derivation of the one from the other could be constructed by means of the conventionally adopted logical rules. [...] [I]t makes no sense to ask whether a deductive system, a language's logical syntax, is sound and complete. Since the deductive system is what gives rise to the consequence relation for the language, it automatically gets that relation exactly right. [Etchemendy, 1990, pp. 156–157]

However, there are a number of signs suggesting that perhaps something else than just investigations in pure syntax is going on in Carnap's *chef d'œuvre*. The mere fact that Carnap refers to Chwistek's semantics as to a system developed with the same aim as his syntax,[2] the fact that he originally entitled the second part of his manuscript of *The Logical Syntax of Language* "Semantics", the fact that he would use the term "semantic" instead of "syntactical" while discussing the issues related to his work in progress in his correspondence with Gödel,[3] as well as the fact that Quine opens his review of *Logische Syntax der Sprache* by these words: "By *logical syntax* Carnap means semantic[s]",[4] should make us cautious as to what exactly "syntax" stands for in the Carnapian project.

What I want to suggest is the rather obvious point that it is not reasonable to make conclusions only on the evidence as weak as the occurrence of a word in the title of the book, all the more so if the exact meaning of this word is not very clear. The thesis I want to defend is that Carnap's treatment of the consequence relation, together with his

---

1) Cf. Coffa [1976], Coffa [1987] and [Coffa, 1991, pp. 285–305].

2) [Carnap, 1937, p. 249].

3) Cf. [Gödel, 2003, pp. 346–347, 350–351]. One example for all: Gödel tells Carnap that "one can *not* view the higher functional calculus semantically." [Gödel, 2003, p. 347; Gödel's italics]

4) [Quine, 1935, p. 394; Quine's italics].

treatment of quantification are fundamentally semantic in the contemporary sense of the word, and therefore that there is no fundamental gap between Carnap the syntacticist and Carnap the semanticist.

This paper has two aims, both modest. The wider aim is to contribute to the clarification of several aspects of Carnap's logical syntax and to show in what sense and to what degree it is syntax as this term is usually understood. The other, narrower aim is to elucidate Carnap's treatment of logical consequence, and to exhibit in what sense it does not coincide with the syntactical relation of derivation. It should be stressed at the outset that what I am primarily after is not an exact historical reconstruction of Carnap's ideas, but rather a systematic exemplification of some of the valuable lines of thought which, as I believe, *The Logical Syntax of Language* contains.

1 Language II

Carnap first works in *The Logical Syntax* with two concrete languages, imaginatively entitled "Language I" and "Language II",[5] and then he develops the results obtained from their investigation into the theory of general syntax. Before going any further it is better to say a few words about Language II. It is construed as a simple theory of types, inspired by Ramsey, provided with a special notation for different types. The types of complex expressions are determined by the types of simple expressions in such a way that if t_1 is (a variable for) the type of the argument $\mathfrak{Arg}_1$, the sentence $\mathfrak{Pr}_1(\mathfrak{Arg}_1)$ has the type (t_1); if t_1 and t_2 are (variables for) the types of the arguments $\mathfrak{Arg}_1$ and $\mathfrak{Arg}_2$, respectively, the type of $\mathfrak{Pr}_1\,(\mathfrak{Arg}_1, \mathfrak{Arg}_2)$ is (t_1, t_2) and the type of $\mathfrak{Fu}_1\,(\mathfrak{Arg}_1) = \mathfrak{Arg}_2$ is $(t_1 : t_2)$.[6] Another prominent characteristic of Language II is that it is a so-called 'coordinate language'. This is to say that its constants of type 0 — the type of individuals — are all numerals, or, to be more precise, the so-called 'accented expressions', $\mathfrak{St}$. These are construed as "0" followed by a number of strokes "′", and they form the series "0", "0′", "0″", etc. The variables of type 0 range over the fixed domain of positions in this series of accented expressions. Obviously, the members of this domain can be correlated in the 1–1 manner with natural numbers.

[5] An observation borrowed from [Potter, 2000, p. 261].

[6] The Gothic symbolism is Carnap's device for referring, in the metalanguage, to symbols and expressions of the object-language. It is not, therefore, a part of the symbolism of Language II itself. Throughout the present paper I will keep using more or less Carnap's symbolism with the exception that I will utilize "$(\forall \mathfrak{v})$" instead of Carnap's "$(\mathfrak{v})$" for the universal quantifier.

In Language II we do not speak of any kind of extralinguistic entities directly, but only via designating the positions in what is, in effect, a one-dimensional coordinate system, which can easily be extended to an n-dimensional coordinate system by using ordered n-tuples of accented expressions. It is the positions in the coordinate system that predicates and functors apply to, e.g., "Blue($0'''$)" means "the position $0'''$ is blue" and "te($0'''$) $= 0'''''$" means "the temperature at the position $0'''$ is $0'''''$".

The primitive vocabulary of Language II consists of the propositional connectives, "=", the stroke "$'$" (successor), the universal and existential quantifiers — both restricted and unrestricted — for variables of any types and the minimum operator "K" — also both restricted and unrestricted ("$(Kx)(\mathfrak{S}_1)$" reads as "the least x of which $\mathfrak{S}_1$ is true"). The constants of type 0 are the constant numerals $\mathfrak{St}$, often abbreviated as "0", "1", "2", etc., to which numerical variables $\mathfrak{z}$ correspond. The constants of type (0) are predicates $\mathfrak{pr}$ to which predicate variables $\mathfrak{p}$ correspond; and the constants of type $(0:0)$ are functors $\mathfrak{fu}$ to which functor variables $\mathfrak{f}$ correspond. We can also form expressions of any number of other types, e.g., of types $(0,0)$, $((0))$, etc.; there will always be constants and variables associated with each type.

The rules of formation of Language II, which determine what counts as a well formed formula, are standard, and it is not necessary to mention them explicitly. The first part of the rules of transformation, the aim of which is to provide Language II with a deductive system, consists of the axioms and the rules of inference. The axioms, or, in Carnap's terminology, 'primitive sentences', include the axioms of propositional logic and predicate logic with identity, plus the axioms governing the use of the K-operator. In addition, they comprise the Peano axioms, including the complete induction schema, and versions of the Axioms of Choice and Extensionality. The rules of inference are Modus Ponens and the 'rule of the universal operator', permitting the transition from an open formula $\mathfrak{S}_1$ to $(\forall \mathfrak{v})(\mathfrak{S}_1)$.[7] The remaining part of the transformation rules — the rules of consequence — requires a rather more complex discussion which I will provide in the subsequent sections of this paper.

What I have described so far forms the logical basis of Language II. The primitive vocabulary as well as the aforementioned transformation rules count as logical. An expression is considered to be logical if,

[7] Here it should be noted that Carnap distinguishes between an open formula such as "$\varphi(x)$", which would be in his Gothic symbolism simply $\mathfrak{S}$, and a sentence schema which represents all the substitution instances $\mathfrak{S}(\mathfrak{v}\,/\,\mathfrak{A})$. The free variable in $\mathfrak{S}$ does not prevent Carnap from asserting $\mathfrak{S}$ as a proper sentence; hence $\mathfrak{S}_1$ and $(\forall \mathfrak{v})(\mathfrak{S}_1)$ are seen as demonstrably equivalent.

roughly speaking, it belongs to the largest class of expressions for which it is characteristic that every sentence formed exclusively out of these expressions is determinate;[8] a sentence is called determinate if it is either analytic or contradictory.[9] If a language contains only logical vocabulary, it is called logical, otherwise descriptive. A descriptive language can have a part that can be isolated which makes use only of logical expressions; such a part is a logical sublanguage (of a descriptive language). If a language contains only logical transformation rules, i.e., those that determine just the concept of logical consequence, it is called L-language. If it involves also physical transformation rules, it is called P-language. Carnap constructs his Language II as a logical L-language which can be, however, very easily extended by the introduction of descriptive vocabulary as well as other, extralogical, rules of transformation into a descriptive P-language. In what follows, however, I will only be dealing with its unextended version.

It is important to realize that Language II is actually much more than just a language, as this term is ordinarily understood, and even more than just a system of logic. It comprises a language, a deductive apparatus, the theory of Peano Arithmetic and several axioms of the theory that can be interpreted both as theory of properties or relations and set theory. Hence, to be precise, from the standard point of view, Language II is *de facto* a formalized theory. Since it contains Peano Arithmetic, it is capable of defining the sequence of natural numbers as well as the basic operations of addition and multiplication. In Carnap's words, Language II "contains an arithmetic."[10]

This position is, indeed, rather peculiar. Not only Carnap does not explicitly distinguish between a language, a deductive system and a theory, but he makes the latter two inextricably interconnected. They are both part of the definition of the consequence relation; logical conse-

---

8) [Carnap, 1937, pp. 177–178].

9) Potter in [Potter, 2000, p. 267] criticises this definition as blatantly circular. The vocabulary is said to be logical if every sentence constructed only by its means is determinate. However, being determinate is defined with the help of the sum of the logical transformation rules, which already make use of the logical vocabulary. Potter is, indeed, right that Carnap's definition is circular in the sense that it makes us no wiser as to what the nature of being logical is. However, such an explanation does not seem to have been Carnap's concern. He merely attempts to construct, in an informal metalanguage, a formalized language, and he needs to sort out some vocabulary as logical, set out the rules of transformation and check out whether the requirements of determinateness are met. If they are not met, he can change the original selection and start again.

10) [Carnap, 1937, pp. 205–256].

quence and 'arithmetical' consequence are two sides of the same coin. Arithmetic is a part of logic and, as Potter puts it, everything is "loaded onto the consequence relation."[11] From this point of view, it is therefore quite misleading to say that Language II is a theory, i.e., that it contains a set of sentences that are put forward as fundamental assumptions from which other sentences can be obtained. Carnap's position represents a special kind of logicism. Although he no longer attempts to show that arithmetical concepts can be defined in terms of logic, he joins the two together in order to show that what they give rise to is nothing else than the property of being logically determinate by the consequence relation alone.

2 Definiteness

One of the key concepts of *The Logical Syntax of Language* is definiteness. A property is said to be definite if its possession or non-possession by any object can be determined "in a finite number of steps by means of a strictly established method."[12] In other words, a property is definite if we have an algorithm for which it is guaranteed that it can be determined in a finite number of steps whether the property holds or does not hold of an object. Carnap's 'determinateness' thus corresponds to 'decidability'. Formalized theories are usually constructed so that some of the basic properties, such as 'being a well-formed formula', 'being an axiom', 'being a proof', were decidable. This is also true of Language II.

A chief task of *The Logical Syntax of Language* is to provide the logical foundations of mathematics, and

> [o]ne of the chief tasks of the logical foundations of mathematics is to set up a formal criterion of validity, that is, to state the necessary and sufficient conditions which a sentence must fulfil in order to be valid (correct, true) in the sense understood in classical mathematics. [Carnap, 1937, p. 98]

Now if we were able to develop this formal criterion of validity as a definite criterion, "we should then possess a method of solution for mathematical problems."[13] For any sentence which is logical in the above described sense of being composed only of logical symbols it would then be calculable whether it is true or not. However, it is well known that this

---

11) [Potter, 2000, p. 266].

12) [Carnap, 1937, pp. 98–99].

13) [Carnap, 1937, p. 99].

can be done only for very weak deductive systems, e.g., for propositional logic where the concept of the logically valid sentence coincides with the concept of tautology, which is clearly definite. In propositional logic the property of being (even indirectly) derivable is definite.

If we are to capture the criterion of validity for systems as rich as arithmetic, we have to give up definiteness of the relation of derivation. We will not be in possession of an effective solution for an arbitrary logical sentence, but we will be able to determine whether a given sequence of logical sentences is or is not a valid derivation chain, i.e., we will have a definite concept of *direct* derivability. This gives rise to the definite concept of proof. A "proof" is defined as a finite derivation chain without premises;[14] the word "premises", however, is understood here as referring to extralogical assumptions, and not to axioms or primitive sentences which are part of the definition of the definite part of the transformation rules. To be provable thus means to be determinate by means of the axioms and the rules of inference alone. Now to say that being a Language II proof is a definite property is to say that when we are actually presented with a concrete sample of a derivation chain, we are effectively able to decide whether it is or it is not a proof, i.e., whether it is or it is not a valid derivation chain. This is something quite different from determining, when we are given a well-formed formula, whether this formula can be proved from the axioms of Language II. Carnap expresses this difference by saying that the property of being directly derivable — i.e. derivable in a single step — is definite, while the property of being derivable without qualification is indefinite in Language II.

The concept of proof together with other concepts based on the definite term "directly derivable" are called "*d*-concepts"; and the method of dealing with symbols that makes use only of the *d*-concepts is called a "method of derivation or the *d*-method". We have to face the following question concerning the deductive strength of the *d*-concept of demonstrability: Does the *d*-method make it possible to develop the concept of demonstrable sentence in Language II so that this concept coincided with that of logically valid sentence? In other words, can it be shown that a logical sentence is true if and only if it is demonstrable? This question is closely connected to yet another question. Since the logical sentences of Language II are all assumed to be either true or false, i.e., either the logical sentence itself or its negation is assumed to be true, we can ask whether Language II is strong enough to prove every logical sentence or its negation. In other words, is Language II a (negation) complete theory

[14] [Carnap, 1937, p. 29].

in the sense that every well-formed logical sentence of Language II or its negation is demonstrable? And if not, can Language II be modified so that it became (negation) complete?

Of course, it is well known that the answer to both questions raised in the previous paragraph is negative. Carnap is not only aware of that, but he also presents an informal proof of the incompleteness of Language II with respect to the relation of derivation, as well as the proof that neither Language II nor any other theory with certain properties can ever be made complete by any additions of new axioms.

3 The incompleteness argument

The argument is, as it might be expected, founded upon Gödel's proof, but it takes a slightly different path via the Diagonalization Lemma (sometimes also called the "Fixed Point Theorem"), whose first formulation can actually be attributed to Carnap.[15] The method of arithmetization of syntax is the standard one. We assign different term-numbers to different primitive symbols of the language in such a way that various kinds of variables are assigned prime numbers greater than 2, or their second, third and higher powers, according to their respective kinds, and the undefined logical constants are assigned, in quite an arbitrary manner, numbers that are not primes. Sequences of symbols are assigned so-called series-numbers which are construed as products of powers of primes in the following way. If "p_1", "p_2", …, "p_n" refer to prime numbers in the order of magnitude, and "k_1", "k_2", …, "k_n" refer to the term-numbers of the individual symbols in the sequence, then the following number is the series-number of the whole sequence:

$$p_1^{k_1} \cdot p_2^{k_2} \cdot \ldots \cdot p_n^{k_n}.$$

Since the factorization into the prime factors is unique, we are always able, when given a series-number, to decode it and get the wording of the original sequence of symbols. The same procedure which was applied to term-numbers in order to obtain the series-numbers can be repeated once more to obtain the series-series-numbers, i.e. the numbers associated with the sequences of sequences of symbols, e.g., with proofs. Now since the arithmetization of syntax, as outlined above, does not involve anything other than natural numbers, their properties and relations and operations on them, we can say that Language II, which 'contains' the arithmetic

[15] Cf. [Goldfarb, 2003, p. 338].

of natural numbers, has the means to arithmetize its own syntax, i.e. to express, in the syntactical interpretation, its own logical properties.

Carnap starts his construction of the so-called Gödel sentence ($\mathfrak{G}$) by making a general remark that "if the syntax of a language is formulated in that language itself, then a syntactical sentence may sometimes speak about itself, or more exactly, it may speak about its own design."[16] In other words, for any property whatsoever[17] that is expressible in the given language it is possible to construct a sentence which attributes this property to itself. To put it differently, if the given system satisfies certain conditions (among which is the Carnapian condition that it can express its own syntax), then for any given open formula $\varphi(x)$ with just the variable x free, there is such a closed formula ψ that it can be proved in the given system that $\psi \leftrightarrow \varphi(\ulcorner\psi\urcorner)$, $\ulcorner\psi\urcorner$ signifying the numeral referring to the series-number of the formula ψ. This is known as the Diagonalization Lemma.

Carnap's way of presenting the lemma is slightly more complex. He starts by constructing the functor $\mathrm{subst}\,[x, 3, \mathrm{str}\,(x)]$ in which x is a numerical variable, 3 is the numeral standing for the term-number associated with the very same numerical variable x and $\mathrm{str}\,(x)$ expresses the function whose value is the series-number of the accented expression which has the value x. Syntactically interpreted, the value of the whole function is the series-number of the expression that we will obtain when we substitute, in the expression with the series-number x, the accented expression associated with the expression with the series-number x for the (occurrences of) the free variable x. Let us say now that the diagonalization of an open formula with one free variable is a closure of that formula by means of the substitution of that formula's series-number for the free variable. We can then reformulate the explanation of the aforementioned functor in the following way: the functor $\mathrm{subst}\,[x, 3, \mathrm{str}\,(x)]$ expresses the function which — when applied to the series-number x of

[16] [Carnap, 1937, p. 129]. It should be remarked that Carnap's incompleteness argument was left out from the original German edition of *The Logical Syntax of Language*, Carnap [1934*b*]. It was published separately in the form of two articles as Carnap [1934*a*] and Carnap [1935], and it is contained in the enlarged English edition Carnap [1937].

[17] Carnap speaks of syntactical properties only. Nevertheless, it is clearly an unnecessary limitation since, given the syntactical properties are the properties that are true of linguistic expressions rather than of extra-linguistic objects, the worst thing that can happen to a non-syntactical property in this context is that the sentence in which it is attributed to a linguistic expression will not be true. However, we are interested here in the equivalence of two sentences, not in their respective truth or falsity.

some open formula — yields the series-number of that formula's diagonalization.

The second step in getting at the lemma is the introduction of a formula ($\mathfrak{S}_1$), which is an arbitrary open formula with the free variable x. To make the reasoning clearer I will designate this formula by $\varphi(x)$. Thus we have:

$$\varphi(x). \qquad (\mathfrak{S}_1)$$

This formula has a certain series-number, which is among the values of the variable x. By means of the functor $\mathrm{subst}\,[x, 3, \mathrm{str}\,(x)]$ it is possible to take this particular series-number and construct a formula which expresses the fact that φ holds of this series-number. To do this, we simply substitute, in ($\mathfrak{S}_1$), our functor $\mathrm{subst}\,[x, 3, \mathrm{str}\,(x)]$ for (the occurrences of) x:

$$\varphi(\mathrm{subst}\,[x, 3, \mathrm{str}\,(x)]). \qquad (\mathfrak{S}_2)$$

Syntactically interpreted, this open formula means that if we have an open formula with the series-number x, φ holds of its diagonalization. Now ($\mathfrak{S}_2$), too, will have a certain series-number. Let us designate it by b, and let us diagonalize it. The diagonalization of b is $\mathrm{subst}\,[b, 3, \mathrm{str}\,(b)]$, which is a series-number of a certain closed formula, which can, in turn, be decoded from this number. Thus we obtain the closed formula:

$$\varphi(\mathrm{subst}\,[b, 3, \mathrm{str}\,(b)]). \qquad (\mathfrak{S}_3)$$

($\mathfrak{S}_3$) is a closure of ($\mathfrak{S}_2$), yet the trick is that the closure is performed by the series-number of ($\mathfrak{S}_3$) itself. In other words, ($\mathfrak{S}_3$) is true if and only if φ holds of its series-number, which is a form of the Diagonalization Lemma:

$$\mathfrak{S}_3 \leftrightarrow \varphi(\ulcorner \mathfrak{S}_3 \urcorner).^{18)} \qquad \text{(DL)}$$

Having obtained this result, we can construct the Gödel sentence in just a few simple steps. Take the relation of being a (Language II) proof, expressed by the two-place predicate $\mathrm{BewSatzII}\,(r, x)$. Carnap's exact definition of this predicate is too complex to be presented here, but we

18) However, the standard way of putting the Diagonalization Lemma would be $T \vdash \mathfrak{S}_3 \leftrightarrow \varphi(\ulcorner \mathfrak{S}_3 \urcorner)$, T designating the theory in question. We have said, though, that Carnap does not properly distinguish between a language and a theory, and includes the primitive sentences of Peano Arithmetic into the definition of the relation of derivation itself. In Language II, then, it suffices to prove $\mathfrak{S}_3 \leftrightarrow \varphi(\ulcorner \mathfrak{S}_3 \urcorner)$.

can paraphrase it informally in the following way: BewSatzII (r, x) says that r is a (Language II) proof of x, which is to say that r is the series-series-number of the sequence of formulas which is a derivation without premises and whose last sentence is the sentence with the series-number x.

What we need, though, for the construction of the Gödel sentence is the negation of this predicate, namely $\neg$ BewSatzII (r, x), which is to be read "r is not a (Language II) proof of x". In order to express that x is not demonstrable, i.e., that there is no number which is the series-series-number of the proof of x, we need to assert this open formula of all values for the free variable r. The standard way of achieving this is prefixing the universal quantifier to obtain the open formula with one free variable, i.e. $(\forall r)\neg$ BewSatzII (r, x). However, Carnap's slightly idiosyncratic approach we mentioned above permits considering the assertion of an open formula for all values of the free variable(s) as a sentence with a definite truth-value; $(\forall x)\varphi(x)$ and $\varphi(x)$ are, when asserted, considered equal. Therefore, in order to express that x is not demonstrable, it suffices merely to assert that $\neg$ BewSatzII (r, x).

The method of arithmetization of syntax guarantees the association of a unique series-number with any expression composed of the primitive symbols according to the given rules. Now since the formula $\neg$ BewSatzII (r, x) is a well-formed expression of Language II, it will have a unique series-number, too. With this in mind it is not difficult to construct the Gödel sentence $\mathfrak{G}$ of which we will be able to prove that $\mathfrak{G} \leftrightarrow \neg$ BewSatzII $(r, \ulcorner\mathfrak{G}\urcorner)$. We simply diagonalize the formula by substituting the functor subst $[x, 3, \mathrm{str}\,(x)]$ for the free variable x, which will give us the open formula $\neg$ BewSatzII $(r, \mathrm{subst}\,[x, 3, \mathrm{str}\,(x)])$ whose series-number we will designate by b. Now the series-number of the formula $\mathfrak{G}$ we are looking for is subst $[b, 3, \mathrm{str}\,(b)]$, which is the series-number of the following formula:

$$\neg\,\mathrm{BewSatzII}\,(r, \mathrm{subst}\,[b, 3, \mathrm{str}\,(b)]). \qquad (\mathfrak{G})$$

It is clearly visible that the formula with the series-number subst$[b, 3, \mathrm{str}\,(b)]$ is true if and only if the formula $\neg$ BewSatzII $(r, \mathrm{subst}\,[b, 3, \mathrm{str}\,(b)])$ is true. If we read it as asserted for all values of the free variable r, $\mathfrak{G}$ is indeed the Gödel sentence we strove to construct. It is true if and only if there is no number r which is the series-series-number of the proof of the formula which is the diagonalization of the formula with the series-number b. However, the formula with the series-number b is $\neg$ BewSatzII $(r, \mathrm{subst}\,[x, 3, \mathrm{str}\,(x)])$, and its diagonalization is $\mathfrak{G}$. In other words, $\mathfrak{G}$ is true if and only if there is no proof of $\mathfrak{G}$. "Thus $\mathfrak{G}$ means that $\mathfrak{G}$ is not demonstrable."[19)]

Having constructed $\mathfrak{G}$, the syntactical argument for its non-demonstrability is straightforward. If $\mathfrak{G}$ were demonstrable, there would be a series-series-number of its proof, say c, and the sentence

$$\mathrm{BewSatzII}\,(c, \mathrm{subst}\,[b, 3, \mathrm{str}\,(b)])$$

stating the demonstrability of $\mathfrak{G}$ would be true. However, from the universal assertion of the Gödel sentence $\neg\,\mathrm{BewSatzII}\,(r, \mathrm{subst}\,[b, 3, \mathrm{str}\,(b)])$ it is derivable that $\neg\,\mathrm{BewSatzII}\,(c, \mathrm{subst}\,[b, 3, \mathrm{str}\,(b)])$, which is a negation of the aforementioned sentence $\mathrm{BewSatzII}\,(c, \mathrm{subst}\,[b, 3, \mathrm{str}\,(b)])$. Therefore, if Language II is non-contradictory, $\mathfrak{G}$ is non-demonstrable (in Language II).[20)]

Yet, this is not enough to establish the incompleteness result, since were the negation of $\mathfrak{G}$ demonstrable, $\mathfrak{G}$ would be false and the system in question, in our case Language II (in which the principle of bivalence holds), would indeed be complete. It also needs to be shown that $\neg\,\mathfrak{G}$ is non-demonstrable. In order to see that this cannot be proved either, let us assume that $\neg\,\mathfrak{G}$ is demonstrable. From this assumption it follows that $\neg(\forall r)\neg\,\mathrm{BewSatzII}\,(r, \mathrm{subst}\,[b, 3, \mathrm{str}\,(b)])$ is demonstrable, which is equivalent to holding that $(\exists r)\,\mathrm{BewSatzII}\,(r, \mathrm{subst}\,[b, 3, \mathrm{str}\,(b)])$ is demonstrable. However, and this is the key point, $\mathfrak{G}(\mathfrak{z}_1\,/\,\mathfrak{St})$ — i.e. any sentence which results from the substitution of a particular numeral for the free variable occurring in $\mathfrak{G}$ — is demonstrable. This is to say that we can always prove for any particular number that this number is not a series-series-number of the proof of $\mathfrak{G}$. If "0", "1" and "n" are defined numerals in Carnap's sense, we can prove all of the following sentences:

$$\begin{aligned}
&\neg\,\mathrm{BewSatzII}\,(0, \mathrm{subst}\,[b, 3, \mathrm{str}\,(b)]),\\
&\neg\,\mathrm{BewSatzII}\,(1, \mathrm{subst}\,[b, 3, \mathrm{str}\,(b)]),\\
&\qquad\qquad\vdots\\
&\neg\,\mathrm{BewSatzII}\,(n, \mathrm{subst}\,[b, 3, \mathrm{str}\,(b)]).
\end{aligned}$$

Every such particular instance obtained from $\mathfrak{G}$ is demonstrable. Hence, on the assumption that $\neg\,\mathfrak{G}$ is demonstrable, we are bound to hold, on the one hand, that $\mathfrak{G}(\mathfrak{z}_1\,/\,\mathfrak{St})$ is true for any particular number whatsoever while, on the other hand, $\neg\,\mathfrak{G}$ is true when asserted as a universal sentence. But we have said that the latter is equal to holding that

19) [Carnap, 1937, p. 131].

20) This is Carnap's Theorem 36.1, [Carnap, 1937, p. 132].

$(\exists r)$ BewSatzII $(r, \text{subst}\,[b, 3, \text{str}\,(b)])$, which contradicts our finding that $\mathfrak{G}(\mathfrak{z}_1 \,/\, \mathfrak{St})$ is true for any number. To put it generally, we would have proved, for a particular formula $\varphi(x)$, that $\varphi(n)$ is true of any number n whatsoever but, at the same time, that $\neg(\forall x)\varphi(x)$ holds. In other words, we would have proved that Language II is in a rather specific sense inconsistent. This special kind of inconsistency is called ω-inconsistency. Therefore, on the assumption that Language II is not ω-inconsistent, $\neg\,\mathfrak{G}$ is non-demonstrable.[21)]

4 The ω-rule

Around the time Carnap was working on his logical syntax, a debate was going on about the so-called ω-rule, introduced by Tarski and Hilbert.[22)] In the article "On the Concept of Logical Consequence" Tarski gives the following example of a failure of the formalized concept of consequence (which can be seen as corresponding to Carnap's concept of derivation) to replace the intuitive concept of consequence based on "the common usage of the language of everyday life."[23)] Take, e.g., a theory with standard rules of inference which contains among its theorems sentences such as "0 possesses the property P", "1 possesses the property P" and, in general, all sentences of the form "n possesses the property P", where "n" is any numeral designating a natural number. Although such a theory is strong enough to prove that any particular natural number whatsoever possesses the property P, it is incapable to prove the universal sentence "Every natural number possesses the property P". In order to prove this universal sentence, we would need an essentially stronger rule which Tarski calls "the rule of infinite induction".[24)] In what follows I

21) This is stated by Theorem 36.3. However, Carnap proves his theorem by what would be considered semantic considerations based on the truth of $\mathfrak{G}$ resulting from the truth of Theorem 36.1 established earlier. The notion of ω-consistency is dealt with in connection with the proof of Theorem 36.5.

22) Tarski introduced the ω-rule in his lecture "Remarks on some notions of the methodology of the deductive sciences" given at the Second Conference of the Polish Philosophical Society in Warsaw in 1927. Hilbert stated it in his Hilbert [1931*a*] and Hilbert [1931*b*].

23) [Tarski, 1956*b*, p. 409].

24) [Tarski, 1956*b*, p. 411]. The ω-rule might be reminiscent of Peano's fifth axiom, usually referred to as complete or mathematical induction, which is in Carnap's Language II given the shape of the following transformation schema:

$$F(0) \wedge (\forall x)(F(x) \rightarrow F(x')) \rightarrow (\forall x)F(x). \qquad \text{(PSII 20)}$$

The basic difference between this (first-order) induction schema and the ω-rule is that the former is justified by the structure of the sequence of natural numbers and will, though, use the term "ω-rule". Carnap formulates the ω-rule in the following way:

> If $\mathfrak{S}_1$ contains exactly one free variable $\mathfrak{z}_1$, and if every sentence of the form $\mathfrak{S}_1(\mathfrak{z}_1 \,/\, \mathfrak{St})$ is demonstrable, then $(\forall\mathfrak{z}_1)(\mathfrak{S}_1)$ may be laid down as a primitive sentence. [Carnap, 1937, p. 173]

This rule permits a transition from the infinite number of particular premises to a universal conclusion, and it is, therefore, apparently infinitary. We are thus, properly speaking, never in the position to actually reach the desired conclusion. Tarski[25)] suggests ways of getting around this difficulty. It may be possible to prove a sentence $\mathfrak{S}_2$ of a finitary metatheory stating that all the particular instances $\mathfrak{S}_1(\mathfrak{z}_1 \,/\, \mathfrak{St})$ are demonstrable, without having to carry out their actual proofs in the theory in question. The ω-rule would then consist in the transition from $\mathfrak{S}_2$ to $(\forall\mathfrak{z}_1)(\mathfrak{S}_1)$. Such a special version of the ω-rule is sometimes called "formalized ω-rule",[26)] and it apparently loses its infinitary character. An expected objection to the first suggestion, namely that the reformulation of the ω-rule involves not only Language II but also an appropriate metatheory, and thus the rule ceases to be applicable within Language II, can be met on the condition that Language II is rich enough to include the arithmetic of the natural numbers, which it is. On this condition the metatheory can be interpreted within Language II, and the ω-rule regains its applicability within its limits. Another way of getting round the difficulty might be based on considering the ω-rule to be applicable only if there is a function assigning to every accented expression $\mathfrak{St}$ the series-series-number of the proof of the corresponding instance $\mathfrak{S}_1(\mathfrak{z}_1 \,/\, \mathfrak{St})$. It may be said against this suggestion that it might not be clear whether there is the function required.

A further problem with the ω-rule is that its correct application requires that we be able to name every individual in the domain over which we quantify. From this it follows that, since the set of the names we are able to form in our language is always at most denumerable — i.e., its members can be put into 1–1 correspondence with natural numbers —, the domain of individuals to which the ω-rule applies has to be at most denumerable, too. It is, however, well known that there are non-denumerable sets. It is true that we have said that there is a 1–1

it already contains the universal condition $(\forall x)(F(x) \rightarrow F(x'))$ in the antecedent. Mathematical induction is thus not, strictly speaking, induction in the sense of judging from the individual to the universal, but a purely deductive principle.

25) [Tarski, 1956*b*, p. 411].

26) Hazen [1998].

correspondence between the natural numbers and the accented expressions $\mathfrak{St}$ which play the role of the primitive numerals of Language II and which exhaust its domain of individuals. Hence, in our particular case of Language II, it is guaranteed that we possess sufficient means for naming every individual whatsoever in the domain, and a key condition for the application of the ω-rule thus has been met. Yet, this cannot be taken for granted in general.

Nevertheless, neither the question whether the ω-rule can be justified on the finite grounds nor the difficulty with its limited applicability has any direct bearing upon the points I want to make in what follows. Thus I can leave the discussion open. The important thing is to ask whether the introduction of some acceptable version of the ω-rule, whatever shape it would take, could in principle result in the completeness of Language II. Would our non-demonstrable sentence $\mathfrak{G}$ become demonstrable if the ω-rule were introduced in Language II? It is clear that if we applied the ω-rule to the succession of the particular instances of the Gödel sentence $\mathfrak{G}$ listed above, we would clearly be able to reach the conclusion in the form of the universally asserted formula $\neg\, \mathrm{BewSatzII}\,(r, \mathrm{subst}\,[b, 3, \mathrm{str}\,(b)])$. $\mathfrak{G}$ would then have been manifestly proved. Does this mean, though, that Language II, enriched by the addition of the ω-rule, would no longer allow for an incompleteness proof? Does it mean that Language II enriched by the ω-rule would be complete?

In the article "Some Observations on the Concepts of ω-consistency and ω-completeness", Tarski stresses[27)] that when "we add to the rules of inference the so-called rule of infinite induction, [...] [e]very deductive system will then be ω-complete by definition." However, he rushes to assure us that "this would have no fundamental advantage." In the article "On the Concept of Logical Consequence" Tarski generalizes:

> In every deductive theory (apart from certain theories of a particularly elementary nature), however much we supplement the ordinary rules of inference by new purely structural rules, it is possible to construct sentences which follow, in the usual sense, from the theorems of this theory, but which nevertheless cannot be proved in this theory on the basis of the accepted rules of inference. [Tarski, 1956*b*, pp. 412–413]

Why is this so? Let us start with the obvious. Consider the most trivial way of attempting to make the system in question, in our case Language

27) [Tarski, 1956*b*, pp. 295–296].

II, complete, namely by adding the Gödel sentence itself as a new axiom. The resulting system of Language II $+\mathfrak{G}$ will be strong enough to prove, albeit trivially, the original Gödel sentence $\mathfrak{G}$. However, will not our making $\mathfrak{G}$ demonstrable immediately give rise to a contradiction? The answer to this question is obviously "No". The concept "demonstrable" is a very precise one, and it has been given an accurate formal definition. The whole system of primitive sentences and rules of inference is to be seen as constitutive of this concept. The relation referred to as "BewSatzII" is defined (in and) for Language II, i.e. for the system without the Gödel sentence as an additional axiom. When we add $\mathfrak{G}$ as a new axiom, our system will become Language II $+\mathfrak{G}$, and we will no longer be dealing with the relation "BewSatzII", but with a new, augmented relation BewSatzII $+\mathfrak{G}$. Yet, the new system, though it will be able to prove $\mathfrak{G}$, will be incomplete again, since there is nothing to prevent us from constructing a new Gödel sentence $\mathfrak{G}'$. Therefore, the addition of the Gödel sentence $\mathfrak{G}$ does not get us very far.

Now back to the ω-rule. An addition of this rule to the transformation rules of Language II is permissible in the sense that we do not need to worry about the resulting system becoming, in the aforementioned fashion, immediately inconsistent. But does it get us anywhere? Without going into detail, it can be said that it is generally accepted that first-order arithmetic with a standard deductive apparatus enriched by (some appropriate formulation of) the ω-rule is complete.[28)]

How is this possible? Why cannot we just repeat the procedure of applying the method of construction of a new Gödel sentence to the new, augmented system? I will refer to the system enlarged by the ω-rule by "Language II$^\omega$". Assume that Language II$^\omega$ is complete, i.e., that for the series-number n of every true logical sentence expressible in Language II$^\omega$ there is a series-series-number such that the relation designated by "BewSatzII$^\omega$" holds between it and n. Now what precludes us from constructing, in quite a straightforward manner, the following sentence $\mathfrak{G}'$, analogous to the Gödel sentence Carnap constructed for Language II:

$$\neg\,\mathrm{BewSatzII}^{\omega}\,(r, \mathrm{subst}\,[b, 3, \mathrm{str}\,(b)])? \qquad (\mathfrak{G}')$$

It is, indeed, obvious that in Language II$^\omega$ this sentence would be easily demonstrable. Thus there would have to be a series-series-number of its proof. However, as we know, this is precisely what the sentence, when interpreted, denies. Therefore, if we are to exclude the possibility that

28) Cf. Hazen [1998].

Language II^ω is complete in the most trivial way, namely by being inconsistent, it has to be the case that $\mathfrak{G}'$ is not constructible in it. But why not?

Since the expressive power of Language II^ω is definitely not smaller than that of the former system, there remains just one possibility: Language II^ω is no longer definite in the sense Language II was. We have said that the Language II properties of being a sentence, an axiom, a direct derivation or a proof are all definite. This is to say that the conclusion we have reached is that the introduction of the ω-rule into Language II would make the property of being a proof indefinite. In particular, valid proofs of Language II^ω will no longer be effectively enumerable. We lose the ability to take the numbers, one by one, and determine, by an effective procedure, whether the particular number we pick up is a series-series-number of a valid proof. Hence what the completeness result for our system shows "is of course that the ω-rule cannot be formal in Gödel's sense,"[29] i.e., that the whole system which has it as a component does not fulfil the conditions Gödel sets for formal theories.

5 Analyticity

We have seen that if arithmetic which includes the ω-rule is both consistent and complete, then it can no longer be, as Tarski puts it, "purely structural", or, as Carnap puts it, the relation of direct derivation involved can no longer be definite. However, why should we need to enrich the theory we are working with, so that it were made complete and able to prove the Gödel sentence $\mathfrak{G}$? It is important to realize that what is, after all, exciting about the proof we sketched above is that in a certain sense we have already 'proved' $\mathfrak{G}$. When proving that $\mathfrak{G}$ is non-demonstrable we showed that, given that Language II fulfils the explicitly stated conditions, $\mathfrak{G}$ is bound to be true. This, of course, could not have been proved by the definite proof methods available in Language II, since precisely for these methods the sentence $\mathfrak{G}$ is demonstrably out of reach. So how did we show the truth of $\mathfrak{G}$? The key notion was that of ω-consistency; had we denied $\mathfrak{G}$, the system of Language II would have become ω-inconsistent. Therefore, it seems intuitively perfectly justified to conclude that the truth of $\mathfrak{G}$ has been firmly established.

The truth of $\mathfrak{G}$ has thus been established with the help of the same principle that was behind the development of the ω-rule. In his article

29) [Potter, 2000, p. 247].

"On the Concept of Logical Consequence", Tarski formulates the principle in the following way:

> Yet intuitively it seems certain that the universal sentence A follows in the usual sense from the totality of particular sentences $A_0, A_1, \ldots, A_n, \ldots$. Provided all these sentences are true, the sentence A must also be true. [Tarski, 1936, p. 411]

I will call this intuitive, pre-theoretical principle "ω-principle". Now it is crucial to realize that exploiting this principle in terms of the addition of the ω-rule to the deductive apparatus of the theory in question, which results in enlarging the theory's deductive power at the cost of casting away its specifically formal character, is just one way of making use of the ω-principle. It has one particular deficiency, namely that it does not sufficiently appreciate the fact that we have already established the truth of $\mathfrak{G}$ without actually proving it within the original theory itself.

Now what if there were also another way of exploiting the ω-principle? Instead of trying to capture the strong intuitive plausibility that the transition from the sum of the particular instances $\mathfrak{S}_1(\mathfrak{z}_1 \,/\, \mathfrak{St})$ to the universal conclusion $(\forall \mathfrak{z}_1)(\mathfrak{S}_1)$ seems to have by means of creating a new theory with a larger deductive strength, we could attempt to leave the notions of provability and truth apart. In other words, instead of empowering the deductive strength of the theory by means of creating a new system, such as Language II$^{\omega}$, with a reinforced relation of derivation, we could try to keep the system in question intact, and we could set about to establish, for the very same system, the property of being a true sentence. And this is precisely what Carnap identifies as one of his principal goals in the already quoted passage where he says that his chief task is to "set up a formal criterion of validity" or truth as it is understood in classical mathematics.[30] As suspected, the concept Carnap constructs that is to serve as an explicatum for the intuitive notions of truth, correctness or validity of sentences of Language II is the concept of analyticity.

The definition of analyticity in its entirety is far too complex to be presented here.[31] I will present just those points that are relevant for our purposes. Carnap's strategy involves three preliminary stages. In the first we carry out, by the successive application of the so-called rules of

30) [Carnap, 1937, p. 98].

31) Cf. [Carnap, 1937, pp. 102–114]. In the aforementioned article [Tarski, 1936, p. 414] Tarski makes the following comment on the other half of the concept of analyticity, i.e. on contradictoriness: "Carnap's definition of this concept is too complicated and special to be reproduced here without long and troublesome explanations."

reduction, the reduction of the sentence in question. This is to say that we transform the sentence to what is, in effect, the prenex normal form. When the sentence receives the desired form it is called "reduced", which is designated by $^{\mathfrak{R}}\mathfrak{S}$. The second stage consists in providing, by means of the so-called rules of valuation, a valuation for every free individual variable $\mathfrak{z}$, predicate-variable $\mathfrak{p}$ and functor-variable $\mathfrak{f}$ which occurs in the given $^{\mathfrak{R}}\mathfrak{S}$. In the third stage we perform, applying the so-called rules of evaluation, the evaluation of the $^{\mathfrak{R}}\mathfrak{S}$ with respect to the valuations we established in the previous step. The evaluation is a process allegedly leading in every case to the ultimate replacement of the sentence in question, relative to the given valuations, by either $0 = 0$ or $0 \neq 0$, which are to stand for the conspicuous truth or falsehood, respectively.[32] Eventually, on the basis of these three steps Carnap defines a successive procedure for determining whether the property of being analytic, as well as its contrary, the property of being contradictory, holds for different types of sentences and sentential classes.

The pivotal role in Carnap's conception of analyticity is played by the ω-principle. Now since Language II is a coordinate language, we are always, by definition, able to assign names to all individuals, i.e. to all possible values of its individual variables, and therefore first-order quantification can be understood in terms of infinite conjunction or disjunction. However, this is not the case in general. When we allow for quantification over properties, which is possible in Language II, we have to face the problem that there are numerical properties that are not definable within the system. Therefore, we cannot primarily and generally understand quantification substitutionally, i.e. in terms of infinite conjunction or disjunction. Where we cannot generally guarantee the 1–1 correspondence between names and individuals of the domain of quantification — which, as we have just said, no longer obtains already in the case of Language II when we are dealing with the second-order quantification over properties — we have to define the truth of a universally asserted formula $M(F)$ so that this formula were true if and only "if M holds for every numerical property irrespective of the limited domain of definitions which are possible in [Language] II."[33]

How can this be achieved? The key notion is the notion of the val-

[32] The ultimate sentences $0 = 0$ and $0 \neq 0$ (in Carnap's symbolism $\mathfrak{N}$ and $\neg\,\mathfrak{N}$, respectively), as Coffa notes in [Coffa, 1991, p. 291], are Carnap's way of syntactically representing Frege's Truth and Falsehood. However, as Coffa puts it, "[t]he choice of logically determined sentences as representatives is a further confusing aspect of Carnap's strategy."

[33] [Carnap, 1937, p. 107].

uation for the quantifier-free formulas. Take, for instance, the (second-order) quantified sentence

$$(\exists M)(\forall F)M(F).$$

We will obtain the quantifier-free matrix simply by leaving out the quantifiers, hence we get $M(F)$. The concept of valuation is defined for all the different logical types of expressions; to make the matter as simple as possible, let us just stick to predicate-expressions, where predicate-variables belong. These are expressions of type (t_1), and their valuation is defined as the class of valuations of type t_1. If we are dealing with the predicates whose argument-expressions are numerical expressions, i.e. with the properties of individuals, the type t_1 of the argument-expression is equal to the type 0, whose valuation is defined as an $\mathfrak{St}$. Hence the valuation of the first-level predicates is a class of accented expressions $\mathfrak{St}$. In a similar manner, the valuation of the second-order predicates is defined as a class of classes of accented expressions $\mathfrak{St}$, etc. To generalize the notion of valuation to n-ary predicates, we can say that if the predicate in question has n arguments of types t_1, $t_2, \ldots, t_n$, the valuation of its arguments is an ordered n-ad of the valuations of types $t_1, t_2, \ldots, t_n$. For illustration, let us return to the example $M(F)$. The valuation $\mathfrak{B}_1$ of the argument-expression F is of the type t_1 — which is, if F is a first-order property, a class of accented expressions $\mathfrak{St}$ — and the valuation $\mathfrak{B}_2$ of the predicate-expression M is of the type (t_1) — which is, under the same assumption, the class of classes of accented expressions $\mathfrak{St}$.

When we have provided for the valuations for a given quantifier-free formula we can perform its evaluation. This is done, for predicates, by replacing the original formula by $0 = 0$ when the valuation $\mathfrak{B}_1$ of the argument-expression is an element of the valuation $\mathfrak{B}_2$ of the predicate-expression, and by $0 \neq 0$ when otherwise. Since the evaluation is defined for reduced sentences without quantifiers, it is always a finite process. In other words, it "leads in every case, in a finite number of steps, to the final result; this is either $\mathfrak{N}$ or $\neg\mathfrak{N}$."[34] To put it differently once again, given a set of valuations of a quantifier-free reduced formula, the evaluation of this formula with respect to these valuations always leads to either Truth or Falsehood. Let us return to our sample formula $M(F)$. We were already able to get the valuations $\mathfrak{B}_1$ and $\mathfrak{B}_2$ for the components of this formula, and now we have learnt how to assign either Truth or Falsehood (i.e. $0 = 0$ or $0 \neq 0$) to $M(F)$ with respect to the chosen

[34] Theorem 34c.1, [Carnap, 1937, p. 110].

valuations. This is to say that we are able to deal with quantifier-free formulas.

We are now ready to get hold of the full specification of analyticity. The natural starting point is to ascribe the properties of being analytic and contradictory to the two minimal sentences $0 = 0$ and $0 \neq 0$, respectively. As we have at our disposal the method of the evaluation of formulas with respect to given valuations, which always results in either $0 = 0$ or $0 \neq 0$, we have already almost reached the desired end. The general strategy is, of course, to consider a given formula to be analytic with respect to the given valuation if the result of its evaluation is $0 = 0$, and contradictory, if the result is $0 \neq 0$.

Now, since analyticity is primarily a property of sentences, and since sentences are usually identified with closed formulas, the subject matter that remains to be dealt with is quantification. For a reduced sentence $\mathfrak{S}_1$ in which merely logical constituents appear and which contains only bound variables analyticity is defined in the following way. If $\mathfrak{S}_1$ has the form of the universally quantified sentence $(\forall \mathfrak{v}_1)\,\mathfrak{S}_2$, the sentence $\mathfrak{S}_1$ is considered analytic if the open formula $\mathfrak{S}_2$ is analytic in respect of every valuation of the variable $\mathfrak{v}_1$ which occurs as free in $\mathfrak{S}_2$. This is to say that $\mathfrak{S}_1$ is analytic if the evaluation of the open formula $\mathfrak{S}_2$ results always, i.e. with respect to every valuation of $\mathfrak{v}_1$, in the minimal (analytic, true) sentence $0 = 0$. And if $\mathfrak{S}_1$ has the form of the existentially quantified sentence $(\forall \mathfrak{v}_1)\,\mathfrak{S}_2$, the sentence $\mathfrak{S}_1$ is considered analytic if the open formula $\mathfrak{S}_2$ is analytic in respect of at least one valuation of the variable $\mathfrak{v}_1$. Once again, $\mathfrak{S}_1$ is analytic if the evaluation of $\mathfrak{S}_2$ results at least in a single case, i.e. with respect to at least one valuation of $\mathfrak{v}_1$, in the minimal sentence $0 = 0$. Let us return, for illustration, to our sample sentence $(\exists M)(\forall F)M(F)$ mentioned above. In order to establish whether it is analytic, we need to carry out the evaluation of the open formula $M(F)$ and see whether there is at least one valuation of M such that every valuation of F is its member. If this is the case, then our sentence $(\exists M)(\forall F)M(F)$ is analytic. This terminates the presentation of the definition of analyticity.

It is clearly visible that Carnap's conception of analyticity has one very significant feature, namely that it no longer belongs to Language II. As it is defined, the concept of analyticity is metalinguistic in the sense that we have defined a concept which applies to the expressions of Language II, but we have done this in a different language. We have achieved this in an appropriate metalanguage. Now the following question needs to be asked. As we are in possession of the method of arithmetization of syntax via the system of Gödel numbering, cannot this method be used to translate the whole definition of analyticity back to Language II? Cannot the whole concept of analyticity be systematically rid of its metalinguistic dimension?

The answer to this question is negative, or, to be more precise, it is negative unless we want our language to become contradictory. The reason is quite simple. Assume that we are able to define the property of being analytic as well as its reverse, the property of being contradictory, within Language II. Then we would have a predicate such as ContII at our disposal, and we could use it to form formulas such as $\mathrm{ContII}\,(x)$ where x is a numerical variable. This formula would mean that the sentence with the series-number x is contradictory. Now we can apply again the diagonalization functor $\mathrm{subst}\,[x, 3, \mathrm{str}\,(x)]$ as a value for x. In this way we will obtain $\mathrm{ContII}\,(\mathrm{subst}\,[x, 3, \mathrm{str}\,(x)])$. This open formula will have a certain series-number, say b, which will be among the possible values of the free variable x. We can substitute the numeral corresponding to the number b for the occurrences of x, and we will get the following sentence:

$$\mathrm{ContII}\,(\mathrm{subst}\,[b, 3, \mathrm{str}\,(b)]).$$

This sentence would be logical, i.e. either analytic or contradictory. However, syntactically interpreted, it says that the sentence with the series-number b is contradictory, and the sentence with the series-number b is this sentence itself. Therefore, on the assumption that this sentence is analytic, it would have to be true, hence contradictory. On the opposite assumption, namely that it is contradictory, it would have to be true, hence analytic. The sentence $\mathrm{ContII}\,(\mathrm{subst}\,[b, 3, \mathrm{str}\,(b)])$ thus turns out to be a version of the antinomy of the liar. Furthermore, by constructing also Grelling's antinomy[35)] and Richard's antinomy[36)] Carnap shows that the antinomy of the liar is not the only antinomy that arises if we permit the definitions of analyticity to be given within the same language to which it applies. This fact is formulated in the form of a general conclusion:

> Theorem 60c.1. If S is consistent, or, at least, non-contradictory, then "*analytic (in S)*" *is indefinable in S*. The same thing holds for the remaining *c*-terms [...] in so far as they do not coincide with *d*-terms [...]. [Carnap, 1937, p. 219; Carnap's italics]

This is Carnap's version of Tarski's Theorem which is, however, stated for truth, not for analyticity. "Analytic (in Language II)" thus cannot

35) See [Carnap, 1937, p. 218].
36) See [Carnap, 1937, pp. 219–220].

even be a predicate of Language II. It cannot be part of the vocabulary of the language for which it is defined, but only of a different, syntactically more powerful language.

6 Logical consequence

We have seen that if we are to preserve the strict formality of the deductive system, we cannot perform a derivation corresponding to the intuitively apparently justified transition from the truth of all the particular instances $\mathfrak{S}_1(\mathfrak{z}_1 \,/\, \mathfrak{St})$ to the truth of the universal conclusion $(\forall\mathfrak{z}_1)(\mathfrak{S}_1)$. Yet, it has been suggested that we can define a relation which is wider in extension than the relation of derivation but which does not, at the same time, contribute to the deductive power of the system in question understood in the strictly formal manner associated with Carnap's *d*-concepts. The starting point was the key insight that, when proving the non-demonstrability of $\mathfrak{G}$, we have already established its truth or validity in the usual sense. We have shown that the Gödel sentence $\mathfrak{G}$ does, indeed, follow somehow from the theory in question. What remains to be shown is how this new, wider relation, which Carnap calls, in opposition to the relation of derivation, the relation of logical consequence, is to be defined.

As we already have the definitions of analyticity and contradictoriness in Language II, the relation of logical consequence can be defined quite easily. The definition involves the notion of a contradictory class of sentences; a class of *logical* sentences $\mathfrak{K}_1$ is considered contradictory if among the sentences of $\mathfrak{K}_1$ there is at least one which is contradictory. The size of the class $\mathfrak{K}_1$ is not limited; it can be either finite, or infinite. Now a Language II sentence $\mathfrak{S}_1$ is a consequence of the class of sentences $\mathfrak{K}_1$ if $\mathfrak{K}_1$ plus the negation of $\mathfrak{S}_1$ is contradictory. If the sentence $\mathfrak{S}_1$ is open, we add to $\mathfrak{K}_1$ its universal closure.[37)]

Let us illustrate the definition on the ω-principle. Is the transition from the infinite class $\mathfrak{K}_1$ of all the particular sentences $\mathfrak{S}_1(\mathfrak{z}_1 \,/\, \mathfrak{St})$ to the universal conclusion $(\forall\mathfrak{z}_1)(\mathfrak{S}_1)$ an instance of the relation of logical consequence? It is, provided the class $\mathfrak{K}_2$ obtained by the addition of the sentence $\neg(\forall\mathfrak{z}_1)(\mathfrak{S}_1)$ to $\mathfrak{K}_1$ contains at least one sentence that is contradictory. The interesting case is, indeed, just the one in which the class $\mathfrak{K}_1$ is not itself contradictory. On the assumption that we are only dealing with logical sentences, we need to establish that, if the transition is to be an interesting instance of the relation of logical consequence, the

37) [Carnap, 1937, p. 117].

sentence $\neg(\forall \mathfrak{z}_1)(\mathfrak{S}_1)$ is itself contradictory, or, which is equal, that the conclusion $(\forall \mathfrak{z}_1)(\mathfrak{S}_1)$ is analytic. And we already know how to do this; $\mathfrak{S}_1$ is analytic if it holds for all valuations of the free variable $\mathfrak{z}_1$.

From this it is easily perceived that an analytic sentence is a consequence of any class of sentences whatsoever. The reason is that any non-contradictory class of sentences becomes contradictory when we add to it the negation of an analytic sentence. Therefore, for logical sentences, if the transition from the non-contradictory infinite class $\mathfrak{K}_1$ of all the particular sentences $\mathfrak{S}_1(\mathfrak{z}_1 \,/\, \mathfrak{St})$ to the universal conclusion $(\forall \mathfrak{z}_1)(\mathfrak{S}_1)$ is an instance of logical consequence, then it is so regardless of the contents of the class $\mathfrak{K}_1$. In other words, when we are dealing only with logical sentences, the relation of logical consequence turns out to be a rather degenerate one in the sense that there is no nontrivial transition from a set of sentences to a conclusion involved. The conclusion is strong enough to stand on its own. Therefore, an analytic sentence can also be alternatively defined as a logical consequence of the null class of sentences because the class that results from the addition of the negation of an analytic sentence to the null class of sentences is contradictory. Similarly, a contradictory sentence can be defined as a consequence of every class of sentences whatsoever, for every class containing a contradictory sentence is contradictory.

Now how are we to show that Language II is complete with respect to the relation of consequence? The important requirement of Carnap's strategy that needs to be stressed is that in order to establish that the relation of consequence obtains, we actually have to *prove* that it obtains. And this is a matter of derivation, not of consequence. To put it differently, consequence must be proved formally, and since, as we know, such a proof cannot be conducted in the object-language, it remains that it has to be performed in the metalanguage.[38)] By the definition of analyticity for Language II we obtain a new predicate which we will designate by AnII$\,(x)$, reading "x is a series-number of an analytic sentence of Language II." Of course, the language containing this new predicate can no longer be Language II, but an appropriate syntax-language; let us designate it by S. The syntax-language S thus corresponds to Language II enlarged, on the level of vocabulary, by the analyticity predicate and, on the level of transformation rules, by the definition of analyticity. To

[38)] Carnap emphasizes in [Carnap, 1937, p. 39]: "[E]very demonstration of the applicability of any term is ultimately based upon a derivation. Even the demonstration of the existence of a consequence-relation — that is to say, the construction of a consequence-series in the object-language — can only be achieved by means of a derivation (a proof) in the syntax-language."

put it briefly, S corresponds to Language II enriched by the analyticity theory.

The system S has to satisfy the Tarskian material adequacy condition. When discussing the concept of truth, Carnap explicitly states the Convention T for the truth-predicate $\mathfrak{W}(`A\text{'})$ where A is, in effect, an arbitrary sentence of the object-language and 'A' is its syntactical designation in the metalanguage:[39)]

$$\mathfrak{W}(`A\text{'}) \leftrightarrow A. \tag{T}$$

Since analyticity coincides, within the logical part of Language II, with truth, i.e., every true logical sentence is analytic and vice versa, we can, in quite a straightforward manner, adapt the Convention T for logical sentences for our analyticity predicate $\mathrm{AnII}\,(x)$:

$$\mathrm{AnII}\,(\ulcorner \mathfrak{S}_1 \urcorner) \leftrightarrow \mathfrak{S}_1\,. \tag{$\mathrm{T_A}$}$$

This reads: the (logical) sentence with the series-number $\ulcorner \mathfrak{S}_1 \urcorner$ is analytic (in Language II) if and only if $\mathfrak{S}_1$. The purpose of the definition of the predicate $\mathrm{AnII}\,(x)$ is to furnish the language S with a theory which would make it possible to derive, i.e. to prove formally, all the sentences of the form ($\mathrm{T_A}$). This is to say that all the analytic sentences of Language II are to be demonstrable in S. This requirement entails that we have to be able to express the definition of analyticity in S in a formal manner, which involves, among other things, the requirement that the property of being an axiom of the analyticity-theory should be effectively decidable, and that it should be possible to arithmetize the theory itself via the system of Gödel-numbering.

In order to show that Language II is complete in respect of the relation of consequence, we need to be able to prove $\mathfrak{G}$ of Language II in S. Carnap does not present such an argument in the shape of a formal proof, and we will not do so either. Nevertheless, we need to emphasize that what the theory of analyticity is to achieve is precisely to enable us to prove such a result formally.[40)] Therefore, we can summarize that the property of being analytic (or contradictory) in Language II holds if and only if the translation of the Language II sentence in question into the metalanguage S is demonstrable (or refutable) in S, S being a

39) [Carnap, 1937, p. 214].

40) There is no longer an immediate danger of contradiction, which was to mar any such attempt to do so in Language II itself, since we are now using a different language. A proof that $\mathfrak{G}$ is demonstrable in S is, of course, perfectly compatible with its being non-demonstrable in Language II.

system containing the analyticity theory for Language II. Accordingly, the relation of logical consequence in Language II obtains if and only if the relation of derivation obtains in S between the translations of the Language II sentences in S.

7 Truth

Carnap's theory of analyticity can rightly be seen as a theory of truth for logical languages. The theory can easily be extended to capture the logical truth in languages containing also descriptive constants. A sentence $\mathfrak{S}_1$ will be said analytic, i.e. (logically) true, if the universal closure of the sentence obtained by the substitution of variables of appropriate types for the descriptive constants contained in $\mathfrak{S}_1$ will be analytic. Moreover, the theory could be extended even further to capture not only the concept of logical truth, but also the concept of truth in general. However, in *The Logical Syntax of Language* Carnap did not develop such a theory of truth without qualification, i.e., a theory applicable also to sentences that were merely factually true. This might look a bit surprising because he had everything that was needed to do so. One more step, and Carnap would have reached the definition of the general concept of truth.[41)]

And it was not that such an idea merely did not occur to Carnap. He actively rejected a mere consideration of it when Gödel made a suggestion along these lines, informing Carnap in a letter from September 1932 that he "will give a definition for 'truth'" in the planned sequel to his incompleteness article, which was never published. For Carnap this clearly represented a misuse of terminology:

> The term "true" seems to me very unsuitable; in any case, its usage would not be in accord with general linguistic usage. For according to the latter, the sentence "Vienna has so and so many inhabitants" is of course true, whereas the definition proposed by you surely does not apply to it. [Gödel, 2003, p. 353]

[41)] As Coffa puts it: "Truth *can* be defined in what he [Carnap] called the syntax of language. This is obvious as soon as we realize that Carnap allowed his syntax languages to include translations of their object languages (see, e.g., p. 228). As noted earlier, Carnap thought that this portion of the language was simply excess baggage, a repetition of what one already had at a different level. But because they included this excess baggage, Carnap's languages were indistinguishable from Tarski's semantic metalanguages, and truth is consequently definable in them." [Coffa, 1991, p. 303; Coffa's italics]

Carnap then corrects Gödel that what he must have meant was not the general concept of truth, but rather the concept of logical truth or analyticity.

Why did Carnap repudiate the very possibility, which was clearly within his reach, of constructing a definition of the general concept of truth along Tarski's lines? What kept the "scales on his eyes" with respect to the generally applicable notion of truth? An influential answer to this question given by Coffa[42] is that the chief reason was his verificationist prejudice. The definition of truth is to determine the class of true sentences, which may be understood in terms of deciding which particular sentences are true and which are not. It is obvious that no definition of the concept of truth can decide *a priori* which factual sentences are true. This can be done at most for the *a priori* sentences, i.e., for those that are analytic in Carnap's sense. However, there are, in general, two ways of determining classes of objects that share a certain property. One consists in fixing the members of such a class, while the other consists in defining a property which is coextensive with the original one, i.e., which selects exactly the same objects. If we were to proceed according to the first method, we would indeed have to decide upon each sentence whether it is true or not before putting it into the class of true sentences. Yet if we followed the other procedure, we could merely identify a precisely defined property which would — by its structural, or, according to Carnap's terminology, "syntactical" design — guarantee that only true sentences will get into the desired class. This could, nevertheless, leave us entirely in the dark concerning what the elements of this class are.[43]

However, in *The Logical Syntax of Language* there are clear marks of yet another tendency, which is definitely no less important and which I have already touched upon in this paper. Analytic and contradictory sentences were said to be instances of a certain degenerate use of language. In order to see precisely in what sense they are degenerate, we need to introduce Carnap's explicatum for the concept of meaning, namely the concept of content. The content of a sentence $\mathfrak{S}_1$ or of a sentential class $\mathfrak{K}_1$ is defined as the class of their respective non-analytic consequences.[44] From this definition it follows that the content of an analytic sentence is null, while the content of a contradictory sentence is the total class of all sentences whatsoever. To paraphrase it rather bluntly, we can say that analytic sentences do not mean anything, while contradictory sentences

42) [Coffa, 1991, p. 304].
43) Cf. [Coffa, 1976, p. 229].
44) [Carnap, 1937, p. 120].

mean everything. In neither case we are any wiser, and in both cases something has clearly gone awry. To see what will help us answer the aforementioned question.

Carnap attempts to construct Language II so that it were possible to obtain all the logically true sentences merely by exploitation of the relation of logical consequence alone, i.e. without making use of any special axioms. The goal of the rules of inference is to capture the narrower relation of derivation, while the totality of the transformation rules is to capture the wider relation of logical consequence. It is within this outline that a significant shift from a certain tradition takes place. The tradition I have in mind is the axiomatic tradition characteristic of the early twentieth-century logic.[45] Dummett puts the blame on Frege and Russell, saying that they "formalized logical systems on the quite misleading analogy of an axiomatized theory."[46] This approach involves a stipulation of a certain set of logical truths as axioms, from which are then derived, with the help of a few simple rules of inference, other truths as theorems. The focus is placed on the concept of logical truth and its justification. In contrast with this perspective, the significant shift in emphasis is that when Carnap lays down the axioms, it is not with the view to formulate the most fundamental universal truths, but merely as a technically advantageous means to capture the relations of derivation and consequence. The importance of the whole system of primitive sentences as well as of the inference of other truths from them now becomes merely technical.[47] It is consequence which is in the true heart of logic.[48]

Hence, within this project, not only truth and falsehood are of no fundamental interest for logic, but they do not, strictly speaking, apply to logical sentences at all — since, as we have seen, these, being meaningless, can be said to be true or false only in a rather degenerate sense of the word. So long as we get the relations of derivation and consequence right, it is not substantial what the set of logical truths will look like. For Carnap, logic is not primarily a science of the properties of analyticity and

[45] Cf. [Hacking, 1979, p. 290].

[46] [Dummett, 1981, p. 432].

[47] Carnap explains: "For reasons of technical simplicity, it is customary not to formulate the entire system of rules of inference, but only a few of these, and in place of the rest to set up certain sentences which are demonstrable (on the basis of the total system of rules), the so-called *primitive sentences*. The choice of rules and primitive sentences — even when a definite material interpretation of the calculus is assumed beforehand — is, to a large extent, arbitrary." [Carnap, 1937, p. 29; Carnap's emphasis]

[48] [Carnap, 1937, p. 168].

contradictoriness, but the science of the relation of consequence. In the Carnapian project of logical syntax, truth and falsehood are captured by the auxiliary properties of analyticity and contradictoriness, respectively, and this is where their importance terminates. To sum it up, I believe that one of the reasons why Carnap did not develop a Tarski-style general theory of truth in his logical syntax, which he clearly could have done, was also the fact that, for him, the philosophical interest of the property of truth simply did not extend beyond the auxiliary notion of analyticity.

8 Quantification

We have seen that the relation of consequence is explicated via the relation of derivation that obtains on the metalinguistic level of a syntax-language which is irreducibly "richer in modes of expression."[49] I have also mentioned a key requirement concerning the concepts of analyticity and contradictoriness, namely that the evaluation should be performed in respect of the appropriate valuations for the so-called convaluable symbols (free variables and descriptive predicates and functors) of the given theory, irrespective of the possibly limited modes of expression available in that theory. This is to say that in order to perform the evaluation of a certain formula of the object-language, we need to run through all *possible* valuations, and not just through those we are actually able to define in the syntax-language.[50]

Yet, how are we to achieve this? Have we not also seen that no single formal theory is capable of capturing arithmetic, i.e., that there will always be arithmetical terms which are indefinable and sentences which are irresoluble in the theory?[51] Have we not established that in order to define the terms that were indefinable in our original language, we always need a richer language, and in order to prove the sentences that were non-demonstrable in the original theory, we always need a

[49] See [Carnap, 1937, p. 219].

[50] This dispels the relatively widespread criticism of Carnap's definition of logical consequence, voiced, e.g., by Tarski who says in [Tarski, 1956*b*, p. 416] that Carnap's definitions of logical consequence and a series of derivative concepts seem to be "materially inadequate, just because the defined concepts depend essentially, in their extension, on the richness of the language investigated."

[51] Carnap stresses: "[T]here exists neither a language in which all arithmetical terms can be defined nor one in which all arithmetical sentences are resoluble. [...] In other words, *everything mathematical can be formalized, but mathematics cannot be exhausted by one system*; it requires an infinite series of ever richer languages." [Carnap, 1937, p. 222; Carnap's italics]

stronger theory?[52] How are we to run through all properties whatsoever if we have to do that in a specific theory and if it is certain that there are properties that are, so to speak, out of our reach in the given theory, no matter how rich it is? Carnap words this concern in the following way:

> But do we not by this means arrive at a Platonic absolutism of ideas, that is, at the conception that the totality of all properties, which is non-denumerable and therefore can never be exhausted by definitions, is something which subsists in itself, independent of all construction and definition? [Carnap, 1937, p. 114]

Have we not reached the ultimate ground where we can no any longer restrict ourselves to the methods available in the logical syntax of language, and where we have to start talking about objects, properties and relations as about something irreducibly extra-linguistic, and investigate the relations between them and the modes of expressing them in language from a different, external, extra-linguistic perspective? It is essential for Carnap to show that these worries are misplaced. This is the answer he offers:

> We have here absolutely nothing to do with the metaphysical question as to whether properties exist in themselves or whether they are created by definition. The question must rather be put as follows: can the phrase "for all properties ..." (interpreted as "for all properties whatsoever" and not "for all properties which are definable in S") be formulated in the symbolic syntax-language S? This question may be answered in the affirmative. The formulation is effected by the help of a universal operator with a variable $\mathfrak{p}$, i.e. by means of "$(F)(\dots)$", for example. [Carnap, 1937, p. 114]

In other words, we are told that the heart of the matter lies in the use of unrestricted quantification, interpreted objectually. We have seen that what the quantifiers run through depends on the type of the variable involved. The quantifier binding the individual variables runs through the infinite sequence of accented expressions $\mathfrak{St}$; the quantifier binding the predicate-variables runs through classes of accented expressions. We have also seen that, in the case of Language II, if we restrict ourselves to the first-order quantification, i.e., if we permit only individual variables

[52] This is what Carnap's Theorem 60d.2 states. See [Carnap, 1937, pp. 221–222].

at the positions of arguments, there is no substantial difference between the substitutional treatment of quantification and the treatment based on the method of evaluation of a formula in respect of a valuation. It is not until the second-order quantification is introduced that we come to face the universal realm that is insurmountable in a single language. And it is precisely the impossibility to give the second-order language a substitutional treatment which was Carnap's chief motivation for developing the theory of valuations and evaluations.

As Language II contains second-order quantification, we have to face the problem of clarification of what it is that we quantify over when we bind predicates by quantifiers. To explicate this, Carnap uses what is, in effect, an informal set-theoretical language.[53] Now, when we have such a set-theoretical language, the solution most at hand seems to be to take simply some set of individuals as the domain of the first-order quantifiers and the power set of this domain as the domain of the second-order quantifiers. Yet, this approach is somewhat problematic. The basic reason is the well-known fact that the naive conception of arbitrary sets leads to contradictions. Hence in order to make sure that our theory is not blatantly inconsistent, we would have to put it on a firm basis, presumably by means of an axiomatization, which would restrict the far too extensive notion of an arbitrary set. In other words, we would have to develop a full-blooded set theory.

Carnap does not formalize the analyticity theory for Language II, and the expressions such as "is an element of" or "class" are left without explanation. However, he develops a specific formal theory of properties and relations.[54] This is a theory aiming at formalizing the basic properties and relations of properties and relations. Its development is clearly inspired by Schröder's algebraic tradition in the sense that it is to be interpreted both as a calculus of classes and as a calculus of properties or relations.[55] The theory of properties and relations is not, though, explicitly used in the metatheory to provide a formal treatment of quan-

[53] Recall the definition of the concept of valuation of the type (t_1) as "a class of valuations of the type t_1", cf. [Carnap, 1937, p. 108], and that the expression "$\mathfrak{B}_1$ is an element of $\mathfrak{B}_2$" occurs in the rules of evaluation, cf. [Carnap, 1937, p. 110].

[54] The theory of properties and relations is defined as the syntax of one- and many-termed predicates. A theory of single-term predicates is developed as a part of Carnap's treatment of Language II, [Carnap, 1937, pp. 134–136], but it does not bear any name; the title of the section is just "Predicates as Class-Symbols". A theory of relations is developed as a part of general syntax, [Carnap, 1937, pp. 260–267]. To make the terminology as simple as possible, I use the term "theory of properties and relations".

[55] [Carnap, 1937, p. 134].

tification in the object-language. Furthermore, it is not even clear from Carnap's explanations what the actual status of this theory is in terms of the level of the predicates. What are seemingly second-order properties and relations, such as Trans (F) which means "a (two-place) relation F is transitive", can be unpacked, by going all the way down along the definition chain, as involving only first-order quantification. On the other hand, Carnap maintains that his definitions can be framed for arbitrary types of expressions. To conclude, since Carnap does not develop his theory of properties and relations with the same questions in mind that I have asked, it should not be seen as representing a direct answer to the problems concerning a formalization of the analyticity theory in the metalanguage.

It is important to make more precise the nature of the problem Carnap is facing. To make the matter as simple as possible, in what follows I will stick to Zermelo-Fraenkel set theory. It is well known that in Zermelo-Fraenkel set theory we can prove the existence of non-denumerable sets. The classical arguments for their existence are the diagonal arguments due to Cantor. Non-denumerable are the sets whose members cannot be paired off with natural numbers. Such sets are said to have a greater cardinality than the set of natural numbers, the best known examples of them being the power set of the set of natural numbers or the set of real numbers. Now if Carnap's syntacticism is to succeed, it needs to meet this challenge and represent the non-denumerable sets only with the help of the means available in syntax. He needs to show that despite their being at most denumerable, the syntactical objects suffice to represent the whole hierarchy of ever greater infinite sets, of which the denumerable sets form just a portion. The same difficulty can also be seen from a different point of view as the problem of providing for the possibility of conclusively proving something of all objects when some of the objects or the properties of objects over which we quantify are apparently not representable within any single theory.

The answer to this problem presented by Carnap is founded on the definitions, formulated within the theory of properties and relations, of the concepts of L-isomorphism and syntactical isomorphism. It is not necessary to go into the technical details.[56] They are both syntactical concepts in the sense that they are defined in the metatheory for predicates of the object-theory. Roughly speaking, two predicates of the object-theory are L-isomorphic if it is possible to define in the object-theory a special relation called L-correlator, which conveys an effective

[56] For the details see [Carnap, 1937, pp. 264–265].

1–1 correlation of the arguments of which the two given predicates hold. To be sure, the role of the correlator can be viewed analogically to that of the set which carries out the pairing off elements of two different sets in set theory. Now, in contrast to L-isomorphism, two predicates are said to be syntactically isomorphic if there is a syntactical correlator which conveys the 1–1 correlation but, and this is an important point, the syntactical correlator does not have to be definable in the object-theory — it can also be definable in a richer theory formulated in the metalanguage. Hence syntactical isomorphism is a more extensive notion than L-isomorphism.

And this is the key to the solution Carnap puts forward. The set-theoretical property of being a non-denumerable set is transformed into the denial that the given predicate and the predicate "is a natural number" are L-isomorphic. However, this does not mean that they cannot be syntactically isomorphic. The correlator which does not exist in the object-theory can be found in the metatheory. Denumerability within a certain theory and syntactical denumerability understood in terms of an arbitrarily powerful metatheory are two distinct things. And it is also this distinction that represents Carnap's solution to the famous Skolem's paradox. Let us take, as an example, quantification over the set of real numbers. In every theory of real numbers the set of real numbers is non-denumerable, i.e., it is not possible to construct, within the theory in question, the correlator of real numbers and natural numbers. This deficiency applies also to Language II. However, if we shift onto the level of an appropriate metatheory S which is stronger in terms of expressive power, we can perform the pairing off by means of a syntactical correlator constructed in S. The set of real numbers as defined within the metatheory S will still be non-denumerable, but the axioms and theorems of Language II will be, within the metatheory S, satisfiable only by a denumerable domain. In other words, any real number definable in Language II — even those that do not fit in the numbered list of Language II — will fit in the numbered list of S. And this is all we need for the moment.[57)]

To sum it up, the set of real numbers expands with every new level reached in the object-language/metalanguage hierarchy, and within every level it remains non-denumerable. However, from the vantage point of an appropriate metalanguage the sentences of the object-language can be in-

[57)] Crossley puts it quite nicely: "[W]henever we describe a real number as the unique number having some property defined by a formula in the formal language, this number is already in A [a countable domain of the model of the formal language]. The remaining reals are thus in a sense redundant." [Crossley *et al.*, 1972, p. 28] terpreted only with the help of sets that are denumerable within the metalanguage. Thence our quantifiers can run through a non-denumerable domain, but they still can be interpretable in a metatheory as running through the domain which is only denumerable.[58] So Carnap feels safe to declare that "[i]n syntax it is always possible to effect a denumeration of expressions of any kind."[59] Nothing seems to be out of reach for what Carnap calls the logical syntax of language.

However, there remains a fundamental difficulty with this theory of Carnap's the solution of which, unfortunately, exceeds the scope of this paper. The problem is that Löwenheim-Skolem theorem is valid only for those (consistent) theories that do not presuppose that their quantifiers run through *all* of the subsets of an infinite set. This is to say that 'full-blooded' higher-order theories are not affected by the theorem. Yet, when Skolem's paradox is discussed in *The Logical Syntax of Language*, there is not a single mention of this condition on the applicability of the theorem. Carnap simply does not seem to be aware of it. Therefore, in order to guarantee that Löwenheim-Skolem theorem really applies, it would be necessary to clarify the question of what exactly the higher-order quantifiers of Language II run through. We know that Language II is a higher-order theory of types; but should we see it — or construe it — as a 'full'-higher-order theory?

The debate, still ongoing, between the supporters of higher-order logics and those who claim that only first-order logic is the right one is very complex, and I neither want to enter it, nor I want to search for the indices that would suggest what Carnap's position might have been if he had felt pressed to decide for one of the paths to follow. The suggestion I wish to make is merely that Carnap's substitute for set theory — his theory of properties and relations — is almost entirely formulated as first-order. The sentences containing properties of second-order are in

[58] It is important to understand that this does not involve making the concepts involved less applicable, precise or 'real'. A.W. Moore, for instance, describes in his otherwise remarkable book the solution along the lines proposed by Carnap in the following way: "It is *as if*, within M, there is something which *looks like* ω and there is something which *looks like* its power-set and it *looks as though* their members cannot be paired off. When we step outside M, or perhaps we should say, from our vantage point already outside M, we can see that what *looks like* ω down there *really* is ω, whereas what *looks like* its power-set is a Set that is *in fact* only countable, and the reason they *look* to be of different sizes is that there is no Set of pairs to establish this countability within M." [Moore, 2001, p. 164; my italics] A.W. Moore's explication clearly entails the need to distinguish between sets that are what they seem to be and sets that are not. Since this cannot be distinguished within any fixed system of set theory, it implies that no proof within this theory can establish what a set really is.

[59] [Carnap, 1937, p. 269].

most cases just abbreviations of first-order sentences. Now as second-order languages are usually interpreted in terms of set theory which is in turn, in the classical shape of Zermelo-Fraenkel set theory, a first-order theory, it might be possible to construct the analyticity theory for Language II also as only a first-order theory. However, since I do not judge it feasible to decide upon the workability of such a suggestion on the means that Carnap actually provides in his logical syntax, and since a thorough systematic investigation of the matter would take us too far from our original subject of syntax and the relation of logical consequence, I will leave the problem open.

9 Conclusion

Carnap's project of logical syntax is an intersection of several different philosophical ambitions. One of the chief ambitions is to challenge Wittgenstein's Tractarian doctrine that logic is a system of tautologies and contradictions. According to Carnap, logic is an elaborate system whose aim is to represent the relation of logical consequence, which is seen as constitutive of meaning of descriptive expressions. By-products of this effort to capture logical consequence are the concepts of analyticity and contradictoriness, degenerate cases of truth and falsehood. Contrary to Wittgenstein's position in the *Tractatus*, the relation of logical consequence as well as the concepts of analyticity and contradictoriness are construed as indefinite. We can, indeed, attempt to construct a formal theory and hope that, by means of some explicitly and precisely stated rules, we will succeed in making the relation of consequence definite. Yet, what we can make definite, if our system is sufficiently rich and not contradictory, is at most the relation of direct derivation and the property of being a proof, and not the concept of consequence itself. It is not within our finite powers to establish in general what the consequences of a certain sentence are, whether a given sentence is a logical consequence of another sentence, or whether a sentence is analytic or contradictory. Carnap nicely formulates this point when he discusses Wittgenstein's and Schlick's error consisting in seeing logic as a definite discipline:

> "In the case of an analytic judgment, to understand its meaning and to see its *a priori* validity are one and the same process." [Schlick] tries to justify this opinion by quite rightly pointing out that the analytic character of a sentence depends solely upon the rules of application of the words concerned, and that a sentence is only understood when the rules of application are

> clear. But the crux of the matter is that it is one thing to be clear about the rules of application without at the same time being able to envisage all their consequences and connections. [Carnap, 1937, pp. 101–102; Carnap's italics]

As we have seen, the reaction to the impossibility of formalization of the relation of logical consequence is a transition onto the metatheoretical level. Here the analyticity predicate is defined for the object-language together with the rules governing its application. Carnap's undertaking can be seen, in analogy to Tarski, as an attempt to formulate the theory of truth for logical languages. The core notion of this theory, namely that of being analytic in respect of certain valuations, is without any debate semantic, and not syntactical, as these terms are ordinarily understood.[60] Hence we seem to be justified in claiming that the theory of logical truth in *The Logical Syntax of Language* is in fact a semantic theory of truth, and that the relation of logical consequence is a semantic relation. Tarski praises this theory of Carnap's as "[t]he first attempt to formulate a precise definition of the proper concept of consequence"; and by "the proper concept of consequence" is meant the semantic concept.[61]

However, it is of essential importance for Carnap's semantics in his syntactical period that semantic concepts should be treated only by the means that in fact belong to syntax. Recall that the objective is to explicate logical consequence as derivation in the metatheory and analyticity as provability in the metatheory. Thus Carnap's strategy is to reduce the semantic concepts involved in the object-language to the syntactical concepts available in the metalanguage. This tendency, too, goes hand in hand with the approach developed by Tarski. Church's description of this attitude is quite pertinent:

> [T]here is a sense in which semantics can be reduced to syntax. Tarski has emphasized especially the possibility of finding, for a given formalized language, a purely syntactical property of the well-formed formulas which coincides in extension with the semantical property of being a true sentence. And in Tarski's *Wahrheitsbegriff* the problem of finding such a syntactical property is solved for various particular formalized languages. But

[60] Cf. Tarski's definition of "semantics" in "The Establishment of Scientific Semantics" published in 1936: "We shall understand by semantics the totality of considerations concerning those concepts which, roughly speaking, express certain connexions between the expressions of a language and the objects and states of affairs referred to by these expressions." [Tarski, 1956*b*, p. 401]

[61] [Tarski, 1956*b*, p. 413].

> like methods apply to the semantical concepts of denoting and having values, so that syntactical concepts may be found which coincide with them in extension. Therefore, if names expressing these two concepts are the only specifically semantical (non-syntactical) primitive symbols of semantical metalanguage, it is possible to transform the semantical metalanguage into a syntax language by a change of interpretation which consists only in altering the sense of those names without changing their denotations. [Church, 1956, pp. 65–66]

This can be seen as the ultimate goal also of Carnap's philosophical project of *The Logical Syntax of Language*. Carnap strives to explicate the semantic concepts by constructing syntactical concepts belonging to appropriate syntactical metalanguages. Semantics is not to involve any undefined primitives. Semantic concepts that would not be dissoluble by syntactical definitions would, for Carnap as well as for Tarski, represent an attempt to adopt an external, extra-linguistic perspective, which is epistemologically very problematic in the least. Semantics based on primitive semantic concepts would have to face strong philosophical objections concerning the nature of these concepts and challenging our justification for them. The metaphysical spectre of the synthetic *a priori*, repudiated by the ideal of the unity of science, would be feared to be creeping in through the back door.

Linguistic meaning can only be grasped in language and by the means available in language. There is nothing transcendent or ineffable about it. We do not have to *say* things in order to see what the sentences *show* us. It is true that meaning cannot be exhausted by one system. Nevertheless, the potentially infinite hierarchy of languages makes it possible to express everything that there is about meaning in any language whatsoever, and it is possible to do this syntactically.

Bibliography

ALBEVERIO, Sergio; *et al.* (1986): *Nonstandard Methods in Stochastic Analysis and Mathematical Physics.* Academic Press, New York.

BENCIVENGA, Ermanno (1999): “What is logic about?” In *The Nature of Logic* (Achille C. VARZI, ed.), *European Review of Philosophy*, vol. 4, pp. 5–19. CSLI Publications, Stanford, CA.

BENTHEM, Johan van; *et al.* (eds.) (2006): *The Age of Alternative Logics. Logic, Epistemology, and the Unity of Science*, vol. 3. Kluwer, Dordrecht.

BERG, Jan (1992): *Ontology without Ultrafilters and Possible Worlds. An Examination of Bolzano’s Ontology.* Academia Verlag, Sankt Augustin.

BOLZANO, Bernard (1817): *Rein analytischer Beweis des Lehrsatzes, daß zwischen je zwey Werthen, die ein entgegengesetztes Resultat gewähren, wenigstens eine reelle Wurzel der Gleichung liege*. G. Haase, Prag.

— (1837): *Wissenschaftslehre. Versuch einer ausführlichen und größtenteils neuen Darstellung der Logik mit steter Rücksicht auf deren bisherigen Bearbeiter.* Seidel, Sulzbach.

— (2004): *The Mathematical Works of Bernard Bolzano* (Steve RUSS, transl. & ed.). Oxford University Press, Oxford.

BRANDOM, Robert (1994): *Making It Explicit. Reasoning, Representing and Discursive Commitment.* Harvard University Press, Cambridge, MA.

— (2000): *Articulating Reasons. An Introduction to Inferentialism.* Harvard University Press, Cambridge, MA.

Cantor, Georg (1883): *Grundlagen einer allgemeinen Mannigfaltigkeitslehre. Ein mathematisch-philosophischer Versuch in der Lehre des Unendlichen.* Teubner, Leipzig.

Carnap, Rudolf (1934*a*): "Die Antinomien und die Unvollständigkeit der Mathematik". *Monatshefte für Mathematik und Physik*, vol. 41(2): pp. 263–284.

— (1934*b*): *Logische Syntax der Sprache.* Springer, Wien.

— (1935): "Ein Gültigkeitskriterium für die Sätze der klassischen Mathematik". *Monatshefte für Mathematik und Physik*, vol. 42(1): pp. 163–190.

— (1937): *The Logical Syntax of Language.* Clarendon Press, London. Amethe Smeaton, transl.

— (1942): *Introduction to Semantics.* Harvard University Press, Cambridge, MA.

— (1956): *Meaning and Necessity.* University of Chicago Press, Chicago, second edition.

Church, Alonzo (1956): *Introduction to Mathematical Logic.* Princeton University Press, Princeton, NJ, revised and enlarged edition.

Coffa, J. Alberto (1976): "Carnap's *Sprachanschauung* circa 1932". In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* (Frederick Suppe; Peter Asquith, eds.), vol. 2, pp. 205–241. East Lansing, MI.

— (1987): "Carnap, Tarski, and the search for truth". *Noûs*, vol. 21(4): pp. 547–572.

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

Crossley, John N.; *et al.* (1972): *What is Mathematical Logic?* Oxford University Press, Oxford.

Dauben, Joseph Warren (1979): *Georg Cantor. His Mathematics and Philosophy of Infinite.* Princeton University Press, Princeton, NJ.

Davidson, Donald (1970): "Semantics for natural languages". In *Linguaggi nella società e nella tecnica* (Bruno Visentini, ed.). Edizioni di Comunità, Milan.

— (1984): *Inquiries into Truth and Interpretation.* Oxford University Press, Oxford.

Davidson, Donald; Harman, Gilbert (eds.) (1972): *Semantics of Natural Language.* Reidel, Dordrecht.

Donnellan, Keith (1966): "Reference and definite descriptions". *Philosophical Review*, vol. 75: pp. 281–304.

Dummett, Michael (1981): *Frege: Philosophy of Language.* Harvard University Press, Cambridge, MA, second edition.

— (1993): *The Seas of Language.* Harvard University Press, Cambridge, MA.

Edwards, Jim (2003): "Reduction and Tarski's definition of logical consequence". *Notre Dame Journal of Formal Logic*, vol. 44: pp. 49–62.

Etchemendy, John (1990): *The Concept of Logical Consequence.* Harvard University Press, Cambridge, MA.

Evans, Gareth (1973): "The causal theory of names". *Aristotelian Society Supplementary*, vol. 47: pp. 187–208.

— (1982): *The Varieties of Reference.* Oxford University Press, Oxford.

Evans, Gareth; McDowell, John (1976): *Truth and Meaning. Essays in Semantics.* Clarendon Press, Oxford.

Feferman, Solomon (1962): "Transfinite recursive progressions of axiomatic theories". *Journal of Symbolic Logic*, vol. 27: pp. 259–316.

Fodor, Jerry; LePore, Ernest (1993): "Why meaning (probably) isn't conceptual role". In *Science and Knowledge* (Enrique Villaneuva, ed.), pp. 15–35. Ridgeview, Atascadero.

Frege, Gottlob (1879): *Begriffsschrift.* Nebert, Halle. English translation in Heijenoort [1971].

— (1918/9): "Der Gedanke". *Beiträge zur Philosophie des deutschen Idealismus*, vol. 2: pp. 58–77.

French, Peter A.; Uehling, Theodore E.; Wettstein, Howard K. (eds.) (1979): *Contemporary Perspectives in the Philosophy of Language.* University of Minnesota Press, Minneapolis.

Gabbay, Dov M.; *et al.* (eds.) (2002): *Handbook of the Logic of Argument and Inference.* Elsevier, Amsterdam.

Geach, Peter T. (1975): "Names and identity". In Guttenplan [1975], pp. 1939–1958.

Gentzen, Gerhard (1934): "Untersuchungen über das logische Schliessen I–II". *Mathematische Zeitschrift*, vol. 39: pp. 176–210.

Gödel, Kurt (2003): *Collected Works. Vol. IV: Correspondence A–G* (Solomon Feferman; John W. Dawson, Jr., eds.). Clarendon Press, Oxford.

Goldfarb, Warren (2003): "Rudolf Carnap". In Gödel [2003], pp. 337–341.

Guttenplan, Samuel (ed.) (1975): *Mind and Language.* Clarendon Press, Oxford.

Hacking, Ian (1979): "What is logic?" *Journal of Philosophy*, vol. 76(6): pp. 285–319.

Harman, Gilbert (2002): "Internal critique: A logic is not a theory of reasoning and a theory of reasoning is not a logic". In Gabbay *et al.* [2002], pp. 171–186.

Hazen, Allen P. (1998): "Non-constructive rules of inference". In *Routledge Encyclopedia of Philosophy.* Routledge, London.

Heijenoort, Jean Van (1971): *From Frege to Gödel. A Source Book in Mathematical Logic.* Harvard University Press, Cambridge, MA.

Hilbert, David (1931*a*): "Beweis des *Tertium non datur*". In *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse*, pp. 120–125.

— (1931*b*): "Die Grundlegung der elementaren Zahlentheorie". *Mathematische Annalen*, vol. 104: pp. 485–494.

— (1935): *Gesammelte Abhandlungen*, vol. 3. Springer, Berlin.

King, Peter (2001): "Consequence as inference. Mediæval proof theory 1300–1350)". In *Mediæval Formal Logic. Consequences, Obligations and Insolubles* (M. Yrjönsuuri, ed.), pp. 117–145. Kluwer, Dordrecht.

Kline, Morris (1972): *Mathematical Thought from Ancient to Modern Times.* Oxford University Press, Oxford.

Kolman, Vojtěch (2005): "Lässt sich der Logizismus retten?" *Allgemeine Zeitschrift für Philosophie*, vol. 30(2): pp. 159–174.

— (2007): "Logicism and the recursion theorem". In *The Logica Yearbook 2006* (Radek Honzík; Ondřej Tomala, eds.). FILOSOFIA, Praha. To appear.

Kripke, Saul (1972): *Naming and Necessity.* Harvard University Press, Cambridge.

— (1979): "Speaker's reference and semantic reference". In French *et al.* [1979], pp. 6–27.

— (1982): *Wittgenstein on Rules and Private Languages. An Elementary Exposition.* Harvard University Press, Cambridge, MA.

Lance, Mark; Wolf, Michael P. (eds.) (2006): *The Self-Correcting Enterprise. Essays on Wilfrid Sellars.* Rodopi, Amsterdam.

Laugwitz, Detlef (1982): "Bolzano's infinitesimal numbers". *Czechoslovak Mathematical Journal*, vol. 32: pp. 667–670.

Lorenzen, Paul (1955): *Einführung in die operative Logik und Mathematik.* Berlin, Springer.

Lorenzen, Paul; Lorenz, Kuno (1978): *Dialogische Logik.* Wissenschaftliche Buchgesellschaft, Darmstadt.

McDowell, John (1994): *Mind and World.* Harvard University Press, Cambridge, MA.

— (1998): *Meaning, Knowledge and Reality.* Harvard University Press, Cambridge, MA.

Moore, A.W. (2001): *The Infinite.* Routledge, London, second edition.

Nelson, Edward (2006): "Internal set theory". Available from `www.math.princeton.edu/~nelson/books/1.pdf`.

Penrose, Roger (1989): *The Emperor's New Mind. Concerning Computers, Minds, and the Laws of Physics.* Oxford University Press, New York.

Peregrin, Jaroslav (1995): *Doing Worlds with Words. Formal Semantics without Formal Metaphysics.* Kluwer, Dordrecht.

— (1997): “Language and its models”. *Nordic Journal of Philosophical Logic*, vol. 2: pp. 1–23.

— (2001): *Meaning and Structure. Structuralism of Postanalytic Philosophers.* Aldershot, Ashgate.

— (2004): “Pragmatism und Semantik”. In *Pragmatisch denken* (A. Fuhrmann; E.J. Olsson, eds.), pp. 89–108. Ontos, Frankfurt am Main. Available in English from `jarda.peregrin.cz/mybibl.php`.

— (2006*a*): “Developing Sellars' semantic legacy. Meaning as a role”. In Lance and Wolf [2006].

— (2006*b*): “Semantics as based on inference”. In Benthem *et al.* [2006], pp. 25–36.

— (2006*c*): “Viable inferentialism”. Available in English from `jarda.peregrin.cz/mybibl.php`.

Perkins, David N. (2002): “Standard logic as a model of reasoning: the empirical critique”. In Gabbay *et al.* [2002], pp. 186–223.

Poincaré, Henri (1908): *Science et méthode.* Flammarion, Paris.

Potter, M. (2000): *Reason's Nearest Kin. Philosophies of Arithmetic from Kant to Carnap.* Oxford University Press, Oxford.

Priest, Graham (1999): “Validity”. In *The Nature of Logic* (Achille C. Varzi, ed.), *European Review of Philosophy*, vol. 4, pp. 183–206. CSLI, Stanford.

Putnam, Hilary (1975): “The meaning of ‘meaning’ ”. In *Language, Mind and Knowledge* (Keith Gunderson, ed.), pp. 131–193. Minneapolis University Press, Minneapolis.

Quine, Willard Van Orman (1935): “Logische Syntax der Sprache”. *The Philosophical Review*, vol. 44(4): pp. 394–397.

— (1960): *World and Object.* MIT Press, Cambridge, MA.

— (1961): *From a Logical Point of View. Logico-philosophical Essays.* Harvard University Press, Cambridge, MA, second, revised edition.

— (1969): *Ontological Relativity and Other Essays.* Columbia University Press, New York.

— (1974): *The Roots of Reference.* La Salle, Open Court.

Robinson, Abraham (1974): *Non-standard Analysis.* North-Holland Publishing Company, Amsterdam.

Rootselaar, Bob van (1962–1966): "Bolzano's theory of real numbers". *Archive for History of Exact Sciences*, vol. 2: pp. 168–180.

Rorty, Richard (1991): *Objectivity, Relativism and Truth. Philosophical Papers I.* Cambridge University Press, Cambridge.

Russell, Bertrand (1903): *The Principles of Mathematics.* Cambridge.

Salanskis, Jean-Michel (1999): *Le constructivisme non-standard.* Presses Universitaires du Septentrion, Septentrion.

Savigny, Eike von (1988): *The Social Foundations of Meaning.* Springer, Berlin.

Schneider, Hans Julius (1975): *Pragmatik als Basis von Semantik und Syntax.* Suhrkamp, Frankfurt am Main.

— (1992): *Phantasie und Kalkül. Über die Polarität von Handlung und Struktur in der Sprache.* Suhrkamp, Frankfurt am Main.

Schütte, Kurt (1960): *Beweistheorie.* Springer, Berlin.

Searle, John (1984): "Can computers think?" In *Minds, Brains and Science*, pp. 28–41. Harvard University Press, Cambridge, MA.

Šebestík, Jan (1992): *Logique et mathématique chez Bernard Bolzano.* J. Vrin, Paris.

Sellars, Willfrid (1948): "Concepts as involving laws and inconceivable without them". *Journal of Philosophy*, vol. 15: pp. 287–315.

— (1953): "Meaning and inference". *Mind*, vol. 62: pp. 313–338.

Stalnaker, Robert (1979): "Pragmatics". *Synthèse*, vol. 22: pp. 272–289.

Stekeler-Weithofer, Pirmin (1986): *Grundprobleme der Logik. Elemente einer Kritik der formalen Vernunft.* Walter de Gruyter, Berlin.

Tarski, Alfred (1930): "Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften I". *Monatshefte für Mathematik und Physik*, vol. 37: pp. 361–404. Quoted from the English translation "Fundamental concepts of the methodology of the deductive sciences" in [Tarski, 1956*b*, pp. 60–109].

— (1933): "Einige Betrachtungen über die Begriffe ω-Widerspruchsfreiheit und der ω-Vollständigkeit". *Monatshefte für Mathematik und Physik*, vol. 40: pp. 97–112. Quoted from the English translation "Some observations on the concept of ω-consistency and ω-completeness" in [Tarski, 1956*b*, pp. 279–295].

— (1936): "Über den Begriff der logischen Folgerung". *Actes du Congrès International de Philosophique Scientifique*, vol. 7: pp. 1–11. Quoted from the English translation "On the concept of logical consequence" in [Tarski, 1956*b*, pp. 409–420].

— (1956*a*): "The concept of truth in formalized languages". In Tarski [1956*b*], pp. 152–278. J.H. Woodger, transl.; John Corcoran, ed.

— (1956*b*): *Logic, Semantics, Metamathematics: Papers from 1923 to 1938.* Clarendon Press, Oxford. J.H. Woodger, transl.; John Corcoran, ed.

Trlifajová, Kateřina (2005): "Teologické zdůvodnění Cantorovy teorie množin". *Filosofický časopis*, vol. 2: pp. 195–218.

Trlifajová, Kateřina; Vopěnka, Petr (2001): "Alternative set theory". In *Encyclopedia of Optimization*, pp. 43–47. Kluwer Academic Publishers, Amsterdam.

Vopěnka, Petr (1989): *Mathematics in the Alternative Set Theory.* Teubner, Leipzig.

— (1996): *Calculus infinitesimalis.* Práh, Praha.

Wittgenstein, Ludwig (1922): *Tractatus logico-philosophicus* (Frank P. Ramsey; C.K. Ogden, transl., eds.). Routledge, London.

— (1953): *Philosophische Untersuchungen* (G.A. Anscombe; G.H. von Wright; R. Rhees, eds.). Blackwell, Oxford.

— (1974): *Bemerkungen ber die Grundlagen der Mathematik* (G.A. Anscombe; G.H. von Wright; R. Rhees, eds.). Suhrkamp, Frankfurt am Main.

PhDr. Vojtěch Kolman, Ph.D. (ed.)
FROM TRUTH TO PROOF

Vydala Univerzita Karlova v Praze,
Filozofická fakulta
v roce 2006 v nákladu 200 výtisků.

Tisk Reprostředisko UK MFF,
Sokolovská 83, 186 75 Praha 8-Karlín.
Objednávky vyřizuje edice@ff.cuni.cz.