universität
wien

DISSERTATION

Titel der Dissertation

"Frege and Metatheory"

Verfasser

Mag. phil. Günther Eder

angestrebter akademischer Grad

Doktor der Philosophie (Dr. phil.)

Wien, Oktober 2012

Studienkennzahl lt. Studienblatt: 092 296

Studienrichtung lt. Studienblatt: Philosophie

Betreuer: Ao. Univ. Prof. Dr. Richard Heinrich

Für meine Eltern

Contents

1 Introduction

> Logic is an old subject, and since 1879 it has been a great one. (Quine, *Methods of Logic*[1].)

1879 is, of course, the year in which Frege's *Begriffsschrift* ([15]) first appeared. Although some have questioned the relative significance of the year 1879 for the development of modern logic, the *Begriffsschrift* is clearly a landmark in the history of logic. Frege was the first to introduce a *formal language* of truth-functional propositional logic and higher-order quantification theory, which (*modulo* notational differences) is still in use today. Moreover, he devises a *calculus* for this formal language, i.e. a system of syntactically specified axioms and rules of inference, in which proofs can be represented and assessed as to their validity. Again, axiomatic systems of formal logic such as Frege's are standard in modern logic. The propositional, as well as the first-order part of his *Begriffsschrift* calculus are *sound* and *complete* with respect to standard-semantics. Although Frege never proved the latter (and, as we shall see shortly, according to some *could not prove* for principled reasons), this shows Frege's exceptionally good logical sense.

But Frege did not just devise a formal language and deductive system of higher-order logic up to modern standards but also *used* it to prove things that hitherto were believed to rest on substantial assumptions drawn from intuition. In particular he defined – within his formal language – the *ancestral* of a given relation and proves its basic properties. These results, together with Frege's conviction that there is a major difference between the truths of arithmetic and the truths of geometry, led him to pursue a program now called "logicism".[2]

The remaining task after the *Begriffsschrift* – to define in purely logical terms the number 0 and the successor-relation S – was subsequently met in his celebrated *Grundlagen der Arithmetik 1884* ([20]). By means of the *ancestral* of a given relation R, which Frege had already defined in his *Begriffsschrift*, the natural numbers could then be defined as

---

[1] To be found in [48], p. vii. The quoted remark has been dropped in later editions.

[2] See Frege's dissertation for the *venia docendi* in [21], p. 56, for Frege's first articulation of the logicist thesis.

those objects that follow 0 in the S-sequence.[3] In his *Grundlagen der Arithmetik*, which, according to Dummett is "Frege's masterpiece, [...] his most powerful and most pregnant piece of philosophical writing, composed when he was at the very height of his powers"[4], Frege reduced the problem of defining the number Zero and the successor-relation to the problem of defining the number operator "the numbers of F's".[5] The last step in his reduction of arithmetical concepts to (purportedly) purely logical concepts was taken in his *Grundgesetze der Arithmetik* (1893, [18]), where he introduced *extensions of concepts* in order to define the number operator and proved *Hume's Principle*, the basic principle governing the number operator. In his *Grundlagen* Frege found extensions to be dispensible[6], but at the time of *Grundgesetze* he came to believe that they were necessary for his logicist program of reducing arithmetic to logic.[7]

Although the system of his *Grundgesetze* turned out to be inconsistent, essentially due to the introduction of *extensions* and *Basic Law V*, the only axiom governing extensions (or "value-ranges" more general), the fact that he was the first to devise explicitly axioms governing sets (or extensions), shouldn't be underestimated.

Firstly, it provided the first step towards an increase of "truthfulness" with regards to set-theoretic principles. It seems to me that Frege was right in claiming in the preface to his *Grundgesetze*:

> Ein Streit kann hierbei, soviel ich sehe, nur um mein Grundgesetz der Werthverläufe (V) entbrennen, das von den Logikern vielleicht noch nicht eigens ausgesprochen ist, obwohl man danach denkt, z.B. wenn man von Begriffsumfängen redet.[8]

and, after receiving the letter from Russell, he was, for all intents and purposes, also

---

[3]The *strong ancestral* $R^<$ with respect to a given 2-place relation R can be defined in second-order logic by $R^<xy :\leftrightarrow \forall F((\forall u(Rxu \rightarrow Fu) \wedge \forall u \forall v(Fu \wedge Ruv \rightarrow Fv)) \rightarrow Fy)$. The *weak ancestral* $R^\leq$ is then defined by $R^\leq xy :\leftrightarrow R^<xy \vee x = y$. For a discussion of the formal achievements in Frege's *Begriffschrift* see for instance [6].

[4][13], p. 1

[5]The number 0 can be defined by $Nu\,[x : x \neq x]$ and the successor-relation by $Sxy \leftrightarrow \exists F \exists z(Fz \wedge y = [u : Fu] \wedge x = Nu\,[u : Fu \wedge u \neq z]$. For a thorough discussion (including formal details) of Frege's treatment of the natural numbers see for instance [6] and [7]. An exposition including philosophical discussion can also be found in [13], in particular chapters 9, 10, 11, 13 and 14.

[6]See [20], p. 80.

[7][18] p. 449 (S. 253). The number in the brackets indicates the original page-number.

[8][18] p. 4 (S. VII)

right in his diagnosis in the appendix:

> *Solatium miseris, socios habuisse malorum.* Dieser Trost, wenn es einer ist, steht auch mir zur Seite; denn Alle, die von Begriffsumfängen, Klassen, Mengen (footnote: Auch die Systeme des Herrn R. Dedekind gehören hierher.) in ihren Beweisen Gebrauch machen, sind in derselben Lage.[9]

He was just the only one to explicitly state the assumptions leading to inconsistency.[10]

Secondly, it mediately prompted the revision of informal set-theoretical reasoning ultimately leading to consistent axiomatizations of set-theory.

But Frege was not just the first to devise workable formal systems of higher-order logic, but he was also interested in the *semantics* of this newly created systems. To this end, Frege famously introduced – in his celebrated *Über Sinn und Bedeutung* (1892) – the distinction between *sense* and *reference*.[11] Although the original purpose of introducing this distinction was the rather narrow one of Frege's felt need to revise certain stipulations he made in his *Begriffsschrift* with regards to the relation of identity, it seems fair to say that it also served the broader purpose of making explicit the built-in semantics of his newly created formal language. And although in his *Über Sinn und Bedeutung*, Frege was for the most part interested in *natural* language, this interest seems to have been motivated by Frege's conviction that natural language has major shortcomings which had to be remedied in order to devise a scientifically workable *formal* language. So getting a clear picture of the intended semantics of his *Begriffsschrift* was necessary in order to get a clear grasp of its intended applications.[12]

Another motive for his interest in semantics starting at the end of the 1880s might be due to his awareness of its *explanatory potential*, as it might be called. The adoption of a rule permitting the inference from $\phi(a)$ to $\exists x\phi(x)$ ("a" being some name) for instance, might be explained by appeal to the stipulation according to which every name that occurs within a language like his *Begriffsschrift* has to have a *unique referent*. Still more basic

---

[9] [18] p. 549 (S. 253)

[10] Note that Dedekind, as well as Hilbert explicitly concede that they had fallen prey to the same mistake. See Hilbert's letter to Frege in [19], p.80. For Dedekind and Zermelo see the discussion in [46], pp. 46-58.

[11] See [21], pp. 157-177.

[12] For instance, by Frege's stipulations, *non-referring terms* or *vague predicates* are to be excluded from a language that should be adequate for scientific purposes.

are principles like the *law of excluded middle*, the adoption of which can be explained by appeal to the two-valued semantics Frege had in mind in setting up the rules and axioms of his formal language.

To sum up: Frege was the first to devise a workable formal language meeting modern standards – in fact a formal language that in a way *defined* modern standards. Note though, that Frege was not interested in developing a formal language just for the sake of *calculation*, a purpose he often attributes to Boole and other logicians in the "algebraic tradition". Rather, he was interested in a language in which *content* could be expressed, although with more presicion than can be done in natural language. As he puts it:

> Boole wollte [...] eine Technik ausbilden, durch welche logische Aufgaben in systematischer Weise gelöst werden könnten, ähnlich wie die Algebra eine Technik der Elimination und der Berechnung von Unbekannten lehrt. [...] Ich hatte von vornherein den Ausdruck eines Inhaltes im Auge. Der Zielpunkt meiner Bestrebungen ist eine *lingua characterica* zunächst für die Mathematik, nicht ein auf reine Logik beschränkter *calculus*. Der Inhalt aber soll genauer als durch die Wortsprache wiedergegeben werden.[13]

The fact that Frege considered his *Begriffsschrift* as a *lingua characterica*, built upon the model of *natural* language, has led many scholars to think that Frege was barred *in principle* from certain investigations that have come to the fore in 20th century logic.

1.1 Metatheory

As we saw, the purpose of inventing a *Begriffsschrift* was to devise a language in which content could be expressed more precisely than is possible in colloquial languages. Clearly, the principle goal was to show rigorously that the truths of arithmetic are provable from logical principles (together with definitions). But the invention of the *Begriffsschrift* should not only serve the narrow purpose of helping to justify the logicist thesis, but it was also meant to provide a framework in which, eventually, *every* part of scientific discourse should have its place.[14]

---

[13] [22], p. 13

[14] In [15], p. XII for instance he writes: "Ich verspreche mir überall da eine erfolgreiche Anwendung meiner Begriffsschrift, wo ein besonderer Werth auf die Bündigkeit der Beweisführung gelegt werden muss, wie bei

Now, suppose we were given some particular part of scientific discourse, and we have properly axiomatized it within the *Begriffsschrift*. That is, we have delineated some *basic concepts* and we have chosen some truths about this part of discourse as *axioms*. Given such a set of axioms and the laws of logic, we can now start to derive theorems from the given set of axioms according to the rules specified by the *Begriffsschrift*. It then seems entirely natural to ask questions like the following:

1. Are *only* truths of the given part of discourse derivable from the axioms? (In particular: Is it impossible to derive a *contradiction* from the axioms?)
2. Do the axioms suffice to derive *every* truth of the particular field in question?
3. Is the set of axioms *minimal* in the sense that no axiom can be derived from the others?
4. Is the set of basic concepts *minimal* in the sense that no basic concept is definable in terms of others?

Natural as these questions might seem, they are often not easy to answer. How can it be shown, for instance, that *every* truth of a given part of science can be derived from a given set of axioms for this part? Of course, as mortal beings that we are, it is not possible to actually "look at every truth" of the field in question. And even if we could: how can we come to know that a given truth is or is not derivable, if we haven't found a derivation *yet*? Maybe there *is* a proof but we just haven't been able to *find* it yet.

Or take the even simpler question 1: how can we know that *only* truths are derivable from a given set of axioms? One might of course argue that axioms are *per definitionem* basic truths. Taking a quick look at the logical axioms and the rules of inference of, say, those codified in the *Begriffschrift* will make it apparent that the axioms are *true* (regardless of the specific content of the basic concepts of the given science) and that the rules of inference are *truth-preserving*. That is, they always lead from true premises to

---

der Grundlegung der Differential- und Integralrechnung. Noch leichter scheint es mir zu sein, das Gebiet dieser Formelsprache auf Geometrie auszudehnen. Es müssten nur für die hier vorkommenden anschaulichen Verhältnisse noch einige Zeichen eingefügt werden. [...] Der Uebergang zu der reinen Bewegungslehre und weiter zur Mechanik und Physik möchte sich hier anschließen."

true conclusions. Hence, it might be argued, it is obvious that we can only prove truths, for the axioms stated at the outset are true and the logical laws lead from truths to truths.

Now it is clear that the argument just given was not an argument *within* the science under consideration. We took a step back and considered *from the outside* what can be proved by means of the axioms of the given science and the logical principles codified in the rules of the *Begriffsschrift*. Now, one of the questions that is at stake here is simply this: was it possible for Frege to take this "step back" and argue just like we did? If so, is this kind of argument to be counted as a genuine *proof*? If not, can it be supplemented or modified so as to count as a genuine proof? In short: Can questions of the sort 1 – 4 be adressed *scientifically*, i.e. within the bounds of a *lingua characterica* like Frege's *Begriffsschrift*?

According to an influential scholarly tradition (initiated by Van Heijenoorts seminal paper *Logic as Language vs. Logic as Calculus*[15]), Frege was not just not interested in such questions, but was in fact barred from even *asking* them in a way that permits an answer that is subject to proof. This is said to follow from Frege's *universalist conception* of logic. In the indroductory note to Gödel's dissertation by Van Heijenoort and Dreben we can read for instance this:

> For Frege, and then for Russell and Whitehead, logic was universal: within each explicit formulation of logic all deductive reasoning [...] was to be formalised. Hence, not only was pure quantification theory never at the center of their attention, but metasystematic questions as such, for example the question of completeness, could not be meaningfully raised. [...] we have no vantage point from which we can survey a given formalism as a whole, let alone look at logic as a whole.[16]

Warren Goldfarb writes:

> If the system [of logicism] constitutes the universal logical language, then there can be no external standpoint from which one may view and discuss the system. Metasystematic considerations are illegitimate rather than simply undesirable.[17]

---

[15] See [31].
[16] [23] p. 44.
[17] [25], p. 353

In Joan Weiner's *Frege in Perspective* we can read this:

> Frege's view of the nature of logical laws precludes the existence of a substantive metaperspective for logic [...] he would refuse to regard any metatheoretic reasoning about primitive logical laws as expressing an objective inference.[18]

And Stewart Shapiro writes:

> More important, perhaps, is the fact that metatheory and model-theoretic semantics are foreign to logicism.[19]

This should suffice to give an impression of the view under consideration.[20] It is hard to pin it down more exactly, for it is notoriously vague, and there are of course variations of the main theme. But the core of this point of view seems to lie in the following line of reasoning, which might be called the *Universality-Argument*:

1. According to Frege the *Begriffsschrift* should be a universal language, that is, a language in which every part of scientific reasoning should take place.
2. To step outside the *Begriffsschrift* is therefore leaving the realm of scientific justification.
3. But metatheoretic questions *force* us to step outside the *Begriffsschrift*.
4. Hence, metatheoretic questions cannot be treated scientifically.

I admit that the matter is complex, and clearly more complex than this little argument suggests, but I take it for granted that the view under consideration appeals to something along the lines of this argument.

Now, as it stands, the argument does no seem to be sound in its (apparently) intended generality. The reason is that the third premise is taken in far too much generality. It is clearly false that "metatheoretic questions" *in general* necessitate a "step outside" the

---

[18] [65] p. 227.
[19] [59], p. 178
[20] Another prominent proponent of this view is Ricketts. See his [53]. For a more balanced account of Fregean metatheory see [3], [60] and [55].

logic of *Begriffsschrift* (or a language that would suffice as an adequate *lingua characterica*). In fact, quite the *opposite* seems to be the case *in general*. Metatheoretic questions like *completeness* for instance arose by *restricting* attention to particular *fragments* of what counted as "logic", say the *propositional* or *first-order fragment*. Recall that, at least until the 40's, what belongs to "logic" was conceived of very broadly and was taken to consist in something like *type theory*, even including an axiom of *infinity* and a substantial theory of *classes*.[21] And this theory was of course taken by most logicians as a fully interpreted "universal language". So it would seem that *most of the logicians* of the 20's or 30's, say, could not have raised any "metatheoretical questions", for *most of them* held the view that this kind of logic is "universal" in the sense that it is needed as a "background-theory" for *any* kind of scientific investigation.

Hence, even if we admit that the argument given above would be cogent if properly restricted, the range of its applicability would be severly limited: it would *only* pertain to metatheoretic questions relating to the universal language *as a whole*. But again, the most fruitful "metatheoretical questions" were raised concerning *restricted parts* of what has been counted as "logic" and *theories* that were formulated within such restricted parts. Examples for the former include the questions of soundness and completeness of propositional logic or quantification theory (first-order logic). Examples of questions of the latter kind are provided by axiomatizations of *parts* of mathematics like Euclidean geometry or elementary arithmetic. It seems that there is no *a priori* reason (that is, a reason based on the *Universality-argument*) that someone who believes in the universality of logic is committed to a the view according to which metatheoretical investigations *tout court* would be impossible.[22]

---

[21] It is for this reason that *Hempel*, even in the 40's, could still claim that arithmetic could be reduced to "logic". See his [32].

[22] It should be mentioned that there is a least one passage in Frege's writings where he seems to be concerned with a decidedly metatheoretical question concerning his *Begriffsschrift as a whole*, namely in the famous §§ 29-31 of his *Basic Laws*. In these paragraphs, Frege apparently tries to prove that every expression that can be formed by the primitives of his system has a *unique reference*. So if the proof were correct, Frege would have established the consistency of his system, for – by *uniqueness* of reference – no sentence could be both true and false. (See for instance [44], [30] or [43], 125 - 130.) Adherents of the view that, according to Frege, *logic as a whole* would not be subject to scientific investigation, typically reject such passages as belonging to the realm of unscientific "elucidations". Although I will not pursue this line of thought here, it seems to me that the issue of metatheory

This seems to suffice as a refutation of the *global* argument accordung to which Frege's "universalist view of logic" implies that *no metatheory whatsoever* would be possible. Still, there might be *specific* reasons why a *particular* metatheoretic question might be ill-posed from Frege's perspective. This is in fact the approach I would recommend in investigating Frege's stance towards metatheory: to look *case by case* which kind of metatheoretic questions were open for investigation – *in principle* – to Frege. In doing so one has to be careful to avoid anachronism. The fact that Frege simply wasn't interested in certain questions, or that he might have done things differently than is done nowadays should not lead us to draw significant conclusions about Frege's conception of logic. Furthermore, it is important to bear in mind that it is one thing to be able to *pose* certain questions but still another one to assess their *significance*. Take for instance the completeness theorem for first-order logic. That some given notion of *derivability* is provably *coextensive* with the set-theoretically defined notion of *semantic consequence* is not *in itself* significant. Rather the completeness theorem derives its significance from the felt *priority* of the notion of semantic consequence. But it is not at all clear why this should be the case. A major tradition in modern logic has it that it is quite the *other way round*. That is, it takes it that the notion of logical consequence should be based on the notion of *correct inference* rather than the set-theoretically defined notion of semantic consequence.[23] It has been doubted if the set-theoretically defined concept of consequence should even be taken as explicating the notion of *semantic* consequence correctly.[24]

My point here is not to take sides in the discussion as to which notion of logical consequence is "more basic". The point is rather that sometimes it is simply a matter of controversy as to the relative *significance* we ascribe to a particular metatheoretic result. And that the importance we attribute to certain questions might be due to philosophical considerations that do not prevent us from *posing* such questions.

---

pertaining even to *logic as a whole* is not *that* clear. It seems to me that attributing a view to Frege, according to which such passages were *unscientific*, does no justice to Frege's careful argumentation, for it clearly *seems* as if he was trying to give something like an informal "proof".

[23]See for instance [47].

[24]See [14] for an influential critique of the concept of semantic consequence.

1.2 Semantics vs. Model-theory

There is another point which must be borne in mind in assessing Frege's stance towards metatheory. There is a salient tendency among adherents of the no-metatheory view to conflate metatheory with *semantic* metatheory and semantic metatheory with *modeltheoretic* semantics.[25] Approximatetely, metatheory (in a wide sense) can be said to be the investigation of formal languages or theories by means of logico-mathematical methods.[26] Obviously, nothing is thereby said about which *aspect* of the language or system in question is to be studied. One might be interested in purely *syntactical* aspects of the language in question, for instance if a particular string of signs is "producible" from a given set of *basic sequences* and syntactical rules that allow us to transform these basic sequences. In asking this kind of questions we are regarding the given formal language temporarily as a mere *game*. But in doing so we are *not* committed to a view according to which the formal language in question *is* nothing but a game, we just restrict our attention to certain *aspects* of it.[27]

Metatheoretic questions of quite another kind are concerned with the intended *semantics* of a given formal language. Say, one adopts a classical, two-valued semantics for a given formal language. In adopting such a semantics, every sentence that can be formed by means of the logical particles and the non-logical vocabulary of this language is assumed to be either *true* or *false*. We can then ask, for instance, if *every* sentence of a particular syntactic shape (say, every sentence of the form $\alpha \rightarrow \alpha$) is *true*. Now, there is no *prima facie* reason to believe that this sort of question has anything to do with *model-theoretic semantics*. In particular, in posing such a question, one does not have to invoke any notion of "truth in an interpretation" or something alike. The only devices to formulate such a question is a means to specify certain subclasses of the wellformed

---

[25] To distinguish between the latter two is particularly important, for there is a clear sense of "semantic metatheory", which Frege would have endorsed (to which he was in fact *committed*, as I will argue in the second paper of this dissertation), but which is quite different from *modeltheoretic* semantics.

[26] A narrower notion of "metatheory" is related to the distinction (bound up with Tarski) between *object-* and *metalanguage*, which will concern us later on.

[27] Natural languages like German are clearly *not* "meaningless formalisms". Still, there is no problem in restricting attention to purely syntactical features of the german language. For Frege's discussion of formalism and the "game-metapher" see his *Basic Laws Pt. II*, §§ 90, 91 ([18], pp. 407-408).

sentences and a predicate for truth *simpliciter*. One might argue that *truth simpliciter* is just *truth in the intended interpretation*, but this seems to turn things upside down. Of course, someone who thinks that model-theoretic semantics is the *only* way to do semantics might identify *truth simpliciter* with *truth in the intended interpretation*.[28] But nothing *forces* us to do so! So the possibility of studying semantical relations without being engaged in model-theory is not ruled out thereby.[29]

The discussion here is, again, not intended as to take sides in such a debate, but merely to point to the fact that metatheoretical questions concerning *semantics* cannot simply be equated with questions concerning *modeltheoretic* semantics. So even if something in Frege's conception of logic inherently forces him to reject *model theory*, there is no reason to believe that he was precluded from enganging in semantic investigations, understood as the study of the relationship between a language (formal or otherwise) and its intended intepretation (or to use a more neutral terminology: between a language and what it is *about*). And in particular, there is no reason to think that this cannot be done *scientifically*, that is, within the bounds, set out by, say, the *Begriffsschrift*.

Still, it is true that model-theory occupies a central place in 20th century logic. Somewhat quote-mining Vann McGee, one might even say: we *need* model-theory not just semantics![30]

Technically, the study of various models of axiomatic theories has turned out to be extremely fruitful in various areas of mathematics. To name just some of the most prominent examples: the investigation of models of axiomatic systems for *first order-arithmetic*

---

[28]From this perspective, "semantics" would be something like a subdiscipline of modeltheoretic semantics: whereas *modeltheoretic* semantics is concerned with *all* interpretations of a given language, semantics would be concerned only with questions relating to *one particular* interpretation, viz. the *intended* one. Of course, this leaves the model-theorist with the problem of determining what is meant by the "intended interpretation" of a theory.

[29]One might be inclined to doubt that the notion of *truth simpliciter* is in fact *intelligible*. This has been doubted for instance by *intuitionists* (and other "anti-realists") like Dummett, who believe that the semantic value of a statement is bound to certain *epistemic* features. But the claim here is not that semantics can "in fact" be done, but rather that there is no reason to believe that semantics can *not* be done *without doing model-theoretic semantics*.

[30][45] p. 569

has become a major field of interest[31] as well as the study of models of first-order theories for the *real numbers*, even leading to a certain "rehabilitation" of *infinitesimals*.[32] Model-theory also turned out to have a fruitful influence on abstract algebra.[33] Hence, if it could be shown that Frege didn't grasp what model-theory is all about, it would be shown thereby that he would be cut off from a major tradition of 20th century logico-mathematical research.

It is often suggested that, in order to address certain metatheoretic questions concerning axiomatic systems, a certain conception of axioms as *reinterpretable schemes* is prerequisite. Surely there is more than a grain of truth in this. But one has to be cautious here: in the context of the question at hand one has to bear in mind that only because Frege (or any other pre-50's researcher for that matter) might have done things differently, this does not imply that he had no grasp of what modeltheoretic *reasoning* amounts to. Also, one has to be careful not to make too great a deal of terminological issues (in particular regarding the word "axiom") in asking what kind of questions Frege "could have" adressed.

As an example, consider the question of *categoricity* of a system of axioms. A system of axioms is said to be *categorical* if any two of its *models* are identical from a structural point of view, i.e.

> **(C)** An axiomatic theory T is *categorical* iff there is an isomorphism between any two of its models

It seems that if *any* question necessitates a genuine "model-theoretic point of view", the question of categoricity is one of the first candidates to consider, for *models* and structure-preserving mappings between models are built into the very concept of categoricity. So if any problem should be foreign to Frege, it should be this.

Now it is clear that Frege did not address the question of categoricity of systems of axioms in this form. Frege clearly dismisses the notion of *truth in a model/interpretation*. On Frege's conception, a sentence of a language properly so called is *inherently interpreted*:

---

[31] See [38] for an extensive discussion of models of first-order Peano arithmetic.

[32] See [54] for the classical exposition of "Non-Standard Analysis".

[33] See [37] for a modern (but not quite up-to date) presentation of model-theory and its main fields of application.

it expresses a definite sense and is definitely either true or false.[34] In particular, an *axiom* expresses a definite thought and is true *by fiat*, for this is exactly what an axiom is supposed to be: an unprovable, basic truth. So according to Frege, in considering *proper axioms* (axioms "in the Euclidean sense", as he sometimes calls them[35]) we are not dealing with uninterpreted schemes, which only afterwards are being supplemented with an interpretation (nor do we consider sentences as being capable of being re-interpreted), but with meaningful propositions. So it seems that any question of categoricity of axiom systems should be out of reach of Frege's "conceptual framework" as it were.

To be more specific let's consider as a concrete example the (conjunction of the) following four sentences $\mathcal{A}$:

A_1: $\forall x \exists y Sxy$

A_2: $\forall x \forall y \forall z (Sxy \wedge Sxz \rightarrow y = z)$

A_3: $\neg \exists x S^{<} xx$

A_4: $\forall x (Nx \leftrightarrow S^{\leq} 0x)$

Here 0 stands for the number *Zero*, Sxy for the two-place relation *y is a successor of* x and Nx for the concept *x is a natural number*.[36] Note that we are *not* considering A_1 - A_4 as "schemes" or something alike. In A_1 - A_4 the concepts Nx, Sxy and 0 are assumed to have a *definite meaning*, they are *basic concepts* as Frege would have conceived of basic concepts. The first axiom for instance expresses the basic arithmetical truth that every natural number has a successor and the second that this successor is *unique*. Were Frege not a logicist (hence believing that Nx, Sxy and 0 could be *defined* and A_1 - A_4 *proved*), these sentences could have well served as a basis for arithmetic. That is, they could have been accepted by him as *proper axioms*.

As I have stated earlier, it would have made no sense for Frege to ask about "models"

---

[34]For more on Frege's *fixed meaning conception* see [1].

[35]See Frege's [16] and [17]. More on this topic will be said in the second article of this dissertation.

[36]Note that, on a modern formulation the mentioning of the predicate N is suppressed A_4 would thus be represented by $\forall x S^{\leq} 0x$. The restriction of the quantifiers to some *domain* is effected *metatheoretically* by letting the quantifiers range over some *domain*. More on this topic will be said later. It is well known that the axioms A_1 - A_4 are just a variant of the famous (second-order) *Peano-axioms* for arithmetic, i.e. the Peano-axioms can be derived from A_1 - A_4 and *vice versa*. See [28].

of this axiom system. On Frege's account a proper axiom expresses a particular proposition and as such "leaves no room for different interpretations".[37] But now consider the following move. Instead of asking if the system of proper axioms $\mathcal{A}$ is categorical, we consider the set of *conditions* $\mathcal{A}_C(X, Y, z)$ corresponding to $\mathcal{A}$. That is, in every axiom of $\mathcal{A}$, the basic concepts are replaced by *variables* of the appropriate type.

A_1^C: $\forall x \exists y Y x y$

A_2^C: $\forall x \forall y \forall z (Y x y \wedge Y x z \rightarrow y = z)$

A_3^C: $\neg \exists x Y^{<} x x$

A_4^C: $\forall x (X x \leftrightarrow Y^{\leq} x z)$

It seems then that nothing has changed, except that letters are replaced by other letters. But from the Fregean perspective it is important to notice that N, S and 0 are supposed to have a particular meaning, whereas X, Y and z are *variables*, designating nothing at all. A_1 for instance does no longer express the proposition that every natural number has a successor, but now expresses a *condition* which a particular relation can have or not. That is, $\mathcal{A}_C$ as a whole is no longer an *axiom system* in the Fregean sense but a set of *higher-order conditions* and we can ask if a given sequence of meaningful concepts has the property defined by these conditions or not. The triple $\langle N, S, 0 \rangle$ for instance has the property defined by $\mathcal{A}_C$, for if these concepts are substituted for the variables in $\mathcal{A}_C$ we just get back to the axioms $\mathcal{A}$. Furthermore we can ask if any two triples consisting of a *concept*, a *relation* and a distinguished *object*, satisfying these conditions, are *isomorphic*.

Now, as a matter of fact, Frege was not only not precluded from asking this question, but he *answered* it in the course of proving his Theorem 263 of his *Basic Laws of Arithmetic*, which essentially states that the number of a "simply infinite system" is $\aleph_0$. He did so by proving first a general theorem justifying definition by recursion and then shows that for any triple $\langle D, R, a \rangle$ satisfying the condition $\mathcal{A}_C$ we can define by recursion an isomorphism between this triple and $\langle N, S, 0 \rangle$ (which trivially implies that D must have the cardinality of the natural numbers). It is worth quoting Frege himself:[38]

---

[37] [40] p. 79.

[38] For a detailed reconstruction of Frege's proof of theorem 263 see [29].

> Wir beweisen nun die Umkehrung des Satzes (207), dass nämlich Endlos [the number of natural numbers, i.e. $\aleph_0$; author] die Anzahl ist, die einem Begriffe zukommt, wenn sich die unter diesen Begriff fallenden Gegenstände in eine Reihe ordnen lassen, die mit einem gewissen Gegenstande anfängt und endlos fortläuft, ohne in sich zurückzulaufen und ohne sich zu verzweigen. Es kommt darauf an zu zeigen, dass Endlos die Anzahl ist, die einem Begriffe *Glied einer solchen Reihe zukommt.* [...] Wir benutzen hierzu den Satz (32) und haben eine Beziehung nachzuweisen, welche die Anzahlenreihe in die mit x [z in our formulation $\mathcal{A}_C$; author] anfangende q-Reihe [Y-series in our formulation; author] und deren Umkehrung diese in jene abbildet. Es liegt nahe, die 0 dem x, die 1 dem auf x nächsfolgenden Gliede der q-Reihe und so immer die nächstfolgende Anzahl dem nächstfolgenden Gliede der q-Reihe zuzuordnen. Wir fassen immer ein Glied der Anzahlenreihe und ein Glied der q-Reihe zu einem Paare zusammen und bilden aus diesen Paaren eine Reihe. Die reihenbildende Beziehung ist dadurch bestimmt, dass ein Paar zu einem zweiten Paare dann in ihr steht, wenn das erste Glied des ersten Paares zum ersten Gliede des zweiten Paares in der nf-Beziehung [the successor-relation; author] und das zweite Glied des ersten Paares zum zweiten Gliede des zweiten Paares in der q-Beziehung steht.[39]

Frege's sketch here is straighforward and can be illustrated like this:

$$\begin{array}{ccccccccccc} 0 & \xrightarrow{nf} & 1 & \xrightarrow{nf} & 2 & \xrightarrow{nf} & 3 & \xrightarrow{nf} & 4 & \xrightarrow{nf} & \cdots \\ \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow & & \\ x & \xrightarrow[q]{} & x_1 & \xrightarrow[q]{} & x_2 & \xrightarrow[q]{} & x_3 & \xrightarrow[q]{} & x_4 & \xrightarrow[q]{} & \cdots \end{array}$$

Bearing in mind that the axioms A_1 - A_4 are equivalent to more standard formulations of arithmetic (the *Dedekind-Peano axioms*), it should be clear that what Frege is talking about in this passage is just a formalized version of a standard proof of categoricity for second-order arithmetic.

What are we to do with this? If it is true that model-theoretic reasoning was utterly foreign to Frege, we should be able to make out some point that disqualifies the reasoning

---

[39] [18], pp. 210-211 (S. 179).

behind Theorem 263 as "genuinely" model-theoretic. To see what this point might be let us look at a little bit closer at modern model-theory.

1.3 Truth in a Model, Validity and Semantic consequence

The notion of *truth in a model* is essential for modern logical theory, for it seems clear that certain metatheoretical questions cannot even be stated without it. A case in point is the already mentioned question of *categoricity* of a system of axioms (or "conditions"). But there are others, like the question of *completeness* or *soundness* of some given deductive system, concepts that relate *deductive consequence/theoremhood* to *logical consequence/logical truth*, notions which are defined in terms of *being true in a model*. A precise formulation of these concepts is not just needed for theoretical reasons, but it is necessitated in order to spell out what is involved in "applications" of these concepts, for instance in independence proofs.

Now, it is hard to make out a core-set of features that qualifies some part of logico-mathematical reasoning as genuinely *model-theoretic*. The problem of delineating what kind of arguments should count as "model-theoretic" becomes even fuzzier if *historical figures* are assessed. Again, care has to be taken to avoid anachronism here. Historically, the development of 20th century model-theory can be traced back on different developments within logical theory and mathematics such as

1. The emergence of Non-Euclidean Geometry in the 19th century (Gauss, Bolyai, Lobachevski, Klein, etc.)
2. The adoption of algebraic methods for investigations in logic (Schröder, Peirce, Boole, etc.)
3. The isolation of restricted parts of what earlier counted as "logic"; in particular the emergence of *first-order logic*[40]

[40]Note that 20th century model-theory is deeply connected with first-order logic. One of the reasons for this seems to be the simple fact that the model-theory of higher-order languages/theories is just not that *interesting*. Many interesting mathematical structures can be captured up to isomporphism in higher-order languages, so there are not much "different" models to investigate. First-order logic on the other hand has the *Löwenheim-Skolem-property* and is *compact*, both properties implying that each first-order theory has a wide range of structurally

4. The development of *set-theory* as a discipline in its own right; in particular the emergence of the notion of an *arbitrary set* as opposed to sets as *extensions of concepts* (Cantor, Zermelo, etc.)

5. The clarification of the concepts of *truth in a model* and *logical consequence* in terms of *satisfaction* (Tarski)

Here I will not try to tell a convincing story about the origins and conceptual structure of modern model theory, taking into account all the developments listed above.[41] What I am interested in here is the "spirit" behind model-theoretic reasoning and the question if Frege was able to grasp this spirit (or at least *some* of it).

A general feature that is often said to be basic for the "model-theoretic point of view" is a largely *structuralist* conception of axiomatic systems. What is meant by a structuralist conception is that axiom systems are characterized, not by their *deductive* consequences, but by the *structures* that satisfy these axioms. Prerequisite for such an understanding of axiom systems is that *axioms* be conceived of as uninterpreted (or re-interpretable) *schemes*. So on this account, there is no longer a difference between a system of axioms, say, for *geometry* and "axioms" for *groups* or *vector spaces*. It is clear that Frege would have dismissed such a blurring, given his old-fashioned conception of axioms as basic truths. According to Frege this would amount simply to a *confusion* between *proper axioms* and *conditions* which can be satisfied or not. But suppose now we would substitute the word "axiom" by the word "condition" whenever the "modern" model-theorist would use the word "axiom", just like we did with $\mathcal{A}$ and $\mathcal{A}_C$ some paragraphs earlier. It seems then that "model-theoretic" talk about *axioms* and the structures that satisfy them could be translated into talk about *conditions* and the structures that satisfy them. In particular, there would be no presuppostions of a *conceptual* nature whose non-appreciation would prevent Frege from engaging in "model-theoretic" questions.

In order to get a better grip on what is the matter here, let us be a little bit more

---

different models. Another reason might be a *general* tendency to focus investigations on first-order logic as the "core" of logic, due to philosophical worries concerning higher-order logic. To this end see Quine's classical *Philosophy of Logic* ([49]), chapter 5.

[41] See [11] or [35] for attempts to do so. Compare also [56] for a discussion of model-theory in the 30's, a formative period of modern model-theory.

careful and restrict – for the moment – attention to *first-order logic* and its modern modeltheoretic semantics and see how the notion of *truth in a model* is defined for such a standard-language. A first-order language $\mathbf{L}$ is specified by the following items: 1. The (so-called) *logical constants* $\forall, \wedge, \neg, =$ 2. A denumerable set of *individual variables* $Var(\mathbf{L}) := \{x, y, z, ...\}$ and 3. a denumerable set of (so-called) *non-logical constants* σ (the "signature"). The set of well-formed formulas is then recursively defined as usual. A *model* (or more precisely, a σ-*model*) $\mathfrak{M}$ for such a language is then defined as an ordered pair $\langle D, I\rangle$, consisting of some *set* D (the *domain*) and a function I (the *interpretation-function*), which assigns

1. an element $I(a) = a^* \in D$ to each individual constant $a \in \sigma$
2. a set $I(R) = R^* \subseteq D^n$ to each n-ary relation sign $R \in \sigma$ and
3. a function $I(f) = f^* : D^n \longrightarrow D$ to each n-ary function sign $f \in \sigma$

In order to define the notion of *truth in a model* then, one has to define the notion of *satisfaction* first. To do so, define (for any given model $\mathfrak{M}$) an $\mathfrak{M}$-*assignment* to be a function $s : Var(\mathbf{L}) \longrightarrow D$, and for any given assignment s, define an x-*variant* of s to be an assignment which is just like s, exept (possibly) for the variable x.

The satisfaction relation $\vDash_s$ (relative to some assignment s) between a σ-model $\mathfrak{M}$ and a σ-formula ϕ is then defined recursively as follows:

1. $\mathfrak{M} \vDash_s Rt_1, ...t_n$ iff. $\langle I^s(t_1), ...I^s(t_n)\rangle \in I(R)$ (for an n-ary relation R and terms $t_1, ...t_n$[42])
2. $\mathfrak{M} \vDash_s \neg\phi$ iff. $\mathfrak{M} \nvDash_s \phi$
3. $\mathfrak{M} \vDash_s (\phi \wedge \psi)$ iff. $\mathfrak{M} \vDash_s \phi$ and $\mathfrak{M} \vDash_s \psi$
4. $\mathfrak{M} \vDash_s \forall x\phi$ iff. for all x-variants s': $\mathfrak{M} \vDash_{s'} \phi$

The notion of a *sentence* ϕ (a formula containing no free individual variables) *being true*

---

[42]The class of σ-*terms* is defined recursively, just like the interpretation $I^s(t)$ of a term t (in $\mathfrak{M}$) relative to an assignment s. That is $I^s(t) = s(t)$ if t is a variable x; $I^s(t) = I(a) = a^*$ if t is a constant a and $I^s(t) = f^*(I(t_1), ...I(t_n))$ if $t = f(t_1, ...t_n)$ for some function sign f and terms $t_1, ...t_n$.

in a model $\mathfrak{M}$, in symbols $\mathfrak{M} \vDash \phi$, is then defined by "quantifying away" the assignment-parameter s:

Definition 1. $\mathfrak{M} \vDash \phi$ *iff. for all assignments* s*:* $\mathfrak{M} \vDash_s \phi$

If Φ is a *set* of σ-sentences, we define

Definition 2. $\mathfrak{M} \vDash \Phi$ *iff. for all* $\phi \in \Phi$*:* $\mathfrak{M} \vDash \phi$

Based on this notion of truth in a model, the relation of (first-order) *logical consequence* between a σ-theory Φ and a σ-sentence ϕ (in symbols, $\Phi \vDash^1 \phi$) and the notion of *satisfiability*, are defined thusly:

Definition 3. $\Phi \vDash^1 \phi$ *iff. for all models* $\mathfrak{M}$*: If* $\mathfrak{M} \vDash \Phi$*, then* $\mathfrak{M} \vDash \phi$

Definition 4. T *is satisfiable iff. there is a model* $\mathfrak{M}$*, such that* $\mathfrak{M} \vDash T$

In particular, a σ-sentence ϕ is said to be *valid* (i.e. a *logical truth*), if it follows logically from the "empty theory", i.e. if it is *true in all models*.

The extension of these definitions to *higher-order* languages is straightforward (assuming a standard account of higher-order quantification). A model for a second-order language for instance, is the same thing as a model for a first-order language, i.e. a domain together with an interpretation function. The only thing that has to be done additionally, is to redefine an *assignment* to be a function that assigns to every first-order variable some element $d \in D$ *and* every n-ary second-order variable some subset S^* of D^n.[43]

In the recursive definition of satisfaction one adds the following clause for atomic formulas

$\mathfrak{M} \vDash_s Xt_1...t_n$ iff. $\langle I(t_1), ...I(t_n)\rangle \in s(X)$ for each n-ary relation variable X

as well as a clause for second-order quantifiers

$\mathfrak{M} \vDash_s \forall X\phi$ iff. for all X-variants s': $\mathfrak{M} \vDash_{s'} \phi$

---

[43]If one adopts a *non-standard account* of second-order quantification, then not every $S \in \mathcal{P}(D)$ might be a possible value of the assignment function s. The domain of the second-order variables might be restricted to some subset $\mathcal{S}$ of $\mathcal{P}(D)$.

The notions of (second-order) *truth in* $\mathfrak{M}$, *logical consequence* (in symbols, $\Phi \vDash^2 \phi$)[44], *validity* and *satisfiability* are defined completely analogous to their first-order counterparts.

Note first that the notion of first-order *validity* can be restated in *pure* second-order logic. Say ϕ is some sentence of a first-order language $\mathbf{L}$ containing as only non-logical constant the predicate P. Then ϕ is first-order valid if and only if the *pure* second-order sentence $\forall X\phi(X)$ is second-order valid (where $\phi(X)$ is the result of replacing each occurence of the predicate P with the variable X). Similarly, ϕ is first-order *satisfiable* if and only if $\exists X\phi(X)$ is second-order satisfiable.

Similar remarks apply to the notion of *logical consequence* for finitely axiomatized theories Φ as well. In particular: For every *finitely axiomatized* first- *or* second-order theory Φ that includes only "first-order non-logical constants" (like $\mathcal{A}$ from above) we have: $\Phi \vDash \phi$ iff. $\Phi \rightarrow \phi$ is *valid* iff. $\forall \vec{X}(\Phi(\vec{X}) \rightarrow \phi(\vec{X}))$ is *valid*.[45] So in discussing finitely axiomatized theories we can, without loss of generality, restrict attention to the concept of *validity* of sentences of pure second-order logic. Note in particular that a *model* for a *pure* second-order sentence is just a *set* D.

Now, if one wants to be careful, and makes explicit the *metatheory* $\mathcal{M}$, in which these definitions are given and which is taken to be some standard set-theory (like ZFC), what one gets is essentially this:

Definition 5. *A first-order sentence ϕ is valid iff. $\forall DSat(D, \ulcorner \forall \vec{X}\phi(\vec{X}) \urcorner)$ is a theorem of $\mathcal{M}$*

and

Definition 6. *A first-order sentence ϕ is satisfiable iff. $\exists DSat(D, \ulcorner \exists \vec{X}\phi(\vec{X}) \urcorner)$ is a theorem of $\mathcal{M}$*

Here $\ulcorner \psi \urcorner$ stands for the "$\mathcal{M}$-code" of the pure second-order formula ψ and $Sat(x, y)$ is simply the "$\mathcal{M}$-coded" version of the relation $\vDash$.

Similar "definitions" of *validity* and *satisfiability* can be given *mutatis mutandis* for

---

[44] If no confusion is to be expected, superscripts, indicating the order of the consequence relation, are dropped.

[45] Here Φ is to be understood as the conjunction of the finitely many sentences in Φ and $\forall \vec{X}$ as a string of quantifiers binding the variables in $\phi(\vec{X})$.

languages of still higher order as well. For an n-th order language one defines *validity* by $\forall DSat(D, \ulcorner \forall \vec{X}_1 ... \vec{X}_n \phi(\vec{X}_1 ... \vec{X}_n) \urcorner)$, where the pure $(n+1)$st-order formula $\phi(\vec{X}_1 ... \vec{X}_n)$ is the result of replacing all the "non-logical constants" of order $\leq n$ by variables of the appropriate type.[46]

Now, keeping in mind the points just made, what "model-variation" essentially comes down to can be seen to be metatheoretic *domain*-variation. That is, the metatheoretic "model-quantifiers" in the definitions of *validity*, *satisfiability* and *logical consequence* are essentially just quantifiers ranging over the sets whose existence is implied by the metatheory $\mathcal{M}$. So in fixing a domain D, the possible values of the "interpretation function" are thereby fixed as well, for the "interpretations" with respect to a given domain D are just the possible *assignments* over that domain.

The possiblity of this kind of metatheoretic "domain-variation" is in fact by many regarded as some key ingredient of the "model-theoretic viewpoint".[47] So if it could be shown that domain variation is in no way intelligible to Frege, it would be shown thereby that anything even close to model-theory in spirit would be foreign to Frege as well. Although Frege never conceives of *metatheoretic* "domain variability" in the sense just explicated, he seems to have been aware of something very close to it in spirit. This should come as no surprise in the light of Frege's Theorem 263, where *something* like "domain-variation" *has to be* involved. So it might be instructive to look more closely on what this surrogate might be.

The basic idea here ist this: instead of using a *metatheoretic* quantifier, quantifying *in the metatheory* over the possible domains of the *objectlinguistic* quantifiers, something similar to domain-variation can be achieved by restricting the object-linguistic quantifiers *in the object-language* to some *domain predicate*. For this, one defines recursively the *relativization* of a formula ϕ to some predicate P. The crucial clauses are those for the quantifiers:

---

[46] As an example, consider the second-order sentence $\forall X \forall Y(Nu(X) = Nu(Y) \leftrightarrow X \approx Y)$. Here Nu is the non-logical *number-operator*, and $\approx$ the (second-order definable) relation of *equinumerosity*. The satisfiability of this sentence (*Hume's Principle*, as it is called nowadays) can be expressed as the satisfiability of the pure *third*-order sentence $\exists f \forall X \forall Y(f(X) = f(Y) \leftrightarrow X \approx Y)$, i.e. Hume's Principle is satisfiable if and only if $\exists DSat(D, \ulcorner \exists f \forall X \forall Y(f(X) = f(Y) \leftrightarrow X \approx Y) \urcorner)$ is a theorem of $\mathcal{M}$.

[47] See for instance [35]. For further discussion compare [56].

$(\forall x\phi)^P := \forall x(Px \to \phi^P)$ and $(\exists x\phi)^P := \exists x(Px \wedge \phi^P)$[48]

The "domain-variability" required in the definitions of validity, satisfiability and logical consequence is then effected by object-linguistic quantification over the predicate P. Moreover, analogues of the definitions of *validity*, *satisfiability* and *logical consequence* (for finitely axiomatized theories) given earlier that are compatible with this broadly Fregean view can then be given thusly:

Definition 7. *ϕ is valid iff. $\forall P(\forall \vec{X}\phi(\vec{X}))^P$ is a theorem of logic*

Similarly

Definition 8. *ϕ is satisfiable iff. $\exists P(\exists \vec{X}\phi(\vec{X}))^P$ is a theorem of logic*

and finally

Definition 9. *ϕ is a logical consequence of Φ iff. $\forall P(\forall \vec{X}(\Phi(\vec{X}) \to \phi(\vec{X})))^P$ is a theorem of logic*

The "domain-variability" needed in Frege's theorem 263 can then be seen to be effected as follows: The key here is axiom 4., $\forall x(Nx \leftrightarrow S^{\leq}0x)$, which states that x is a natural number if and only if x can be reached from zero in a finite number of successor-steps. Note that in a modern standard-formulation, axiom 4. would be simply $\forall x S^{\leq}0x$, leaving the domain implicit, whereas on the Fregean formulation a "domain-predicate" N is used. This carries over to the *condition* $\mathcal{A}_C$ corresponding to $\mathcal{A}$, where the concept of natural number N is replaced by a *variable* X. Speaking anachronistically, a "model" for A_C is the same thing as a model for the theory formulated *without* the use of a domain-predicate. The difference is that, on a modern view, a domain is provided by "interpreting" the "quantifiers", whereas what is "interpreted" *here* is the variable X. "Domain-variability" is then achieved simply by *object-linguistic* quantification over the "variabilized" domain predicate.

---

[48]The relativization to a predicate P for Higher-order quantifiers is defined similarily. The clauses for second-order quantifiers for instance are: $(\forall X\phi)^P := \forall x(\forall x(Xx \to Px) \to \phi^P)$ and $(\exists X\phi)^P := \exists x(\forall x(Xx \to Px) \wedge \phi^P)$. Relativization to some predicate is of course a standard tool in modern axiomatic set-theory and is treated informally as being of one kind with the notion of being true in a *model*. See [42], p. 112.

It is quite safe to say that this reconstruction of *logical consequence* (*basic concepts* are represented by *variables*, *domain-variability* is achieved by *relativization*) is essentially the way Frege reconstructs the informal notion of *logical consequence* and *validity* as used for instance by Hilbert in his independence- and consistency proofs concerning his axiomatization of Euklidean geometry.[49] This can be seen most clearly in the second part of Frege's 1906-paper *On the foundations of geometry*. There we can read for instance this:

> If, as we have assumed, the words "point", "straight line", etc. do not designate but merely are to lend generality, like the letters in arithmetic, then it will be conducive to our insight into the true state of affairs to actually use letters for this purpose. Let us therefore stipulate the following: Instead of "the point A lies in the plane α", let us say "A stands in the p-relation to α". Instead of "the point A lies on the straight line a", let us say "A stands in the q-relation to a". Instead of "A is a point" let us say "A is a Π".
>
> Hilbert's axiom I.1 can now be expressed like this:
>
> If A is a Π and B is a Π, then there is something to which both A and B stand in the q-relation.[50]

What Frege suggests here is of course the strategy explained so far, i.e. to view "basic concepts" as variables ("letters"). This is in particular so for the "domain-predicate", i.e. the "points". Frege goes on to review Hilbert's methodology thusly reconstructed and explores what it means to speak of *theorems* of a thusly understood "axiom system" and comes essentially to Definition 7.[51]

It is to be noted that variants of this reconstruction of logical consequence were extremely common at least until the 30s of the 20th century.[52] Even Hilbert himself seems

---

[49] In retrospect, Bernays confirmed this in his [2].

[50] [40] pp. 83-84.

[51] For further discussion of Frege's reconstruction see [52] or [39].

[52] See [56] for an exposition of the situation concerning logical consequence and validity in the 20s and 30s. Particularly interesting in this context is Carnap's [9], which is something like a border-stone between *traditional axiomatics* on the one hand and *formal axiomatics* on the other hand. For a discussion of Carnap see [57] and [58].

to have adopted a view quite similar to Frege's. This can be seen for instance in Hilbert's and Bernays' *Grundlagen der Mathematik* from 1934. After having axiomatized a part of plane geometry by means of (contentually understood) 3-place relations Zw ("betweenness") and Gr ("lie on"), the axiom system is represented by $\mathfrak{A}(Gr, Zw)$. We can then read this:

> [...] if in axiomatic geometry the respective names for relations in intuitive geometry like "lie on" or "between" are used this is only a concession to custom and a means of simplyfying the connection of the theory with intuitive facts. In fact, however, in formal axiomatics the fundamental relations play the role of *variable* predicates. [...]
>
> The axiom system consists of a demand on two such predicates expressed in the logical formula $\mathfrak{A}(R, S)$, that we get from $\mathfrak{A}(Gr, Zw)$ when we replace $Gr(x, y, z)$ with $R(x, y, z)$, $Zw(x, y, z)$ with $S(x, y, z)$.[53]

Further,

> From this point of view a sentence of the form $\mathfrak{S}(Gr, Zw)$ corresponds to the *logical statement* [emphasis by the author] that for any predicates $R(x, y, z)$, $S(x, y, z)$ satisfying the demand $\mathfrak{A}(R, S)$, the relation $\mathfrak{S}(R, S)$ also holds; in other words, for any two predicates $R(x, y, z)$, $S(x, y, z)$ the formula
>
> $$\mathfrak{A}(R, S) \to \mathfrak{S}(R, S)$$
>
> represents a true statement. In this way a geometrical sentence is transformed into a sentence of pure predicate logic.[54]

Just like Frege had suggested nearly 30 years earlier, Hilbert here speaks of "variables" instead of "basic concepts". So Hilbert seems to have come in agreement with Frege over his own methodology (at least to a certain extent) after all.[55]

---

[53] [34], p. 7

[54] [34], p. 7

[55] I say "to a certain extent" because there are still differences: One point of divergence relates to the problem of domain-variability. Hilbert (at least in [34]) conciously formulates geometry *without* the invocation of an explicit

Now, there are still clear differences to make out between the modern definitions and the tentative suggestions made by Frege (and between the definitions given by any other pre-50s logician for that matter). For one thing, the modern definitions are framed for a particular *object theory* $\mathcal{O}$, (formulated in some *object-language* L_O) in a particular *metatheory* $\mathcal{M}$ (formulated in some *meta-language* L_M), and which is usually taken to be some standard *set theory* like ZFC, if made explicit. On the modern account object-and metatheory are kept strictly seperated. By contrast, the Frege-style definitions are framed within a *single* higher-order framework. This leaves us with some important issues.

1.4 Object- and Metalanguage

It is well known that the distinction between object- and metalanguage has been introduced in a rigorous way by Tarski in his *The concept of Truth in formalized languages.*[56] Now, it might have become entirely natural for modern logicians to precisely delineate what is to be counted as *object-language* and what is to be counted as *meta-language* when it comes to metatheoretic issues. But the reason why Tarski put so much emphasis on the distinction in the first place was rather specific. As Tarski has shown, an adequate definition of truth for an arbitrary (*interpreted*) *object-language* could always be given in a (sufficiently rich) *meta-language*, but not in the object-language *itself*. This is of course just Tarski's famous *undefinability theorem*. Moreover, it can be seen that it is not only not possible to *define* truth for a given object-language in that very same language, but it is not even possible for a language to *contain* its own truth predicate, i.e. as an undefined *primitive*. Given these limitative results concerning the possibility of developing the semantics of an interpreted language *within this very language*, it is understandable that Tarski was dwelling on the distinction between object- and metalanguage (or object- and meta*theory* as we would say nowadays).

Now, usually, modern logicians are not that interested in defining truth (*simpliciter*)

"domain-predicate" for points. Instead, Hilbert speaks of a "hidden variable". "It is to be observed that along with the determination of the predicates the *domain of individuals* over which the variables $x, y, ...$ range has to be fixed. This enters into a logical formula as a kind of *hidden variable*." ([34], p.13). Although this is still not the modern account, it is clearly closer to it in spirit. For further discussion compare [56].

[56] [63], pp. 152-278.

for interpreteted languages, but they are interested in defining *truth in a model* $\mathfrak{M}$, for *variable* $\mathfrak{M}$. For the most part, they are not interested in truth in the "intended model". In fact, the notion of "intendend model" seems to be eschewed altogether. The interest in the notion of *being true in a model* is not due to an interest in "scientific semantics", but due to the interest in metatheoretic notions like *validity*, *satisfiability* and *logical consequence*. As we have seen, it is not that clear that for *this* purpose the distinction between object- and metalanguage is *that* important, for a lot of what seems to be intended with speaking about various models can be simulated to a certain extent in Higher-Order logic.

To be sure, there are limits of – or *problems* at least – with this reconstruction of metatheoretical concepts within higher-order logic. As we saw, one of the main reasons for this reconstruction of metatheoretical concepts was to make more precise the notions of validity and logical consequence as they were used informally in independence- or consistency proofs. But in order to serve this purpose the "logic" employed therein has to be rather strong. To give a simple, yet still instructive example, consider the first-order schemes

$$\alpha := \exists x Rxx$$

$$\sigma := \forall x \exists y Rxy$$

$$\tau := \forall x \forall y \forall z (Rxy \wedge Ryz \rightarrow Rxz)$$

We then ask if $\tau \wedge \sigma \vDash \alpha$ or if $\tau \wedge \sigma \nvDash \alpha$, which, on the suggested reconstruction, become the question if the higher-order sentence

$$(*)\colon\ \forall X(\forall R(\tau(R) \wedge \sigma(R) \rightarrow \alpha(R)))^X$$

is a theorem of logic or not. Now informally, $(*)$ is clearly *not* valid, for there are (necessarily *infinite*) domains in which τ, σ and $\neg\alpha$ are satisfied. Hence $\tau \wedge \sigma \wedge \neg\alpha$ should be *satisfiable*, that is

$$(**)\colon\ \exists X(\exists R(\tau(R) \wedge \sigma(R) \wedge \neg\alpha(R)))^X$$

should be a theorem of logic.

From a modern point of view, two questions arise immediatly: What are the quantifiers in this formula supposed to range over (in particular the initial "domain-quantifier")? And

what is meant by the locution "theorem of *logic*"? The Fregean answer to the first question seems to be more or less straightforward (at least for the pre-May-1902-Frege): the intial quantifier ranges over (absolutely) *every* possible referent of a predicate-expression (Or, to put it in un-Fregean terms, over (absolutely) every *set*.) And a "theorem of logic" is understood as a sentence provable from the general laws of logic, i.e. higher-order logic *including* a theory of *extensions*. That higher-logic, including a theory of extensions, is Frege's "background-theory" for investigations of independence of conditions, fits nicely with a passage of his *Grundgesetze*, where he comments on his definition of a "Positivalklasse":

> With the installation of this definition, I have taken the trouble to fix only the necessary conditions, and only those that are independent from each other. That this has succeeded can not admittedly be proven, but it becomes likely however, if attempts to derive one of these conditions from others fail many times.[57]

Apparently thinking that this could be misinterpreted, he includes the following remark:

> It should not necessarily have been stated that the independence of the stated conditions from one another could not be proven. It is of course conceivable that one could find classes of relations, to which every condition would apply but one, and that every condition would fail in one of the examples. But it should be questioned whether at this stage of the investigation it is possible to give such examples without presupposing geometry, or fractional, negative and irrational numbers, or facts of experience.[58]

As Tappenden remarks, the only things that Frege in this passage explicitly *excludes* for a proof of independence of the stated conditions, are things that he back then didn't regard to have established rigorously as belonging to "logic".[59] So the "counter-models"

---

[57] [18] pp. 467-468 (§ 175, S. 172)
[58] [18], p. 534. (Anmerkung zu Ende § 175, S. 172)
[59] See [60], p. 216

that would instantiate an independence claim are required to be drawn from *logic alone* (or what Frege *thinks* as belonging to "logic" at this stage).

After receiving Russells letter, things become more fuzzy. Although this is not entirely clear, Frege seems to have abandoned the view that there are *logical objects* (like *extensions*), i.e. objects whose existence is implied by the basic laws of logic.[60] This implies that the ressources to provide "models" (for instance to instantiate the existence claim $(**)$ from above) by *logical reasoning alone* might be no longer available.

To state the obvious: the role of providing the needed models is, on the modern account, played by some sufficiently strong *set theory* $\mathcal{M}$ as metatheory.[61] Although this is clearly an important difference between modern model-theoretic metatheory and earlier accounts, it does not seem to be *that* important from a *conceptual* point of view. It is more a question of *what can be done* by adopting a particular background-theory, that is: does it, for instance, provide enough models to instantiate independence claims for *arbitrary* axiom systems (or conditions). As I have already mentioned, it was a widely held view until the early 40s that the realm of "logic" is far wider than what today is regarded as belonging to Higher-order logic, let alone First-order logic (remember Hempel!). So it seems that at least *a large part* of "model-theoretic metatheory" (or ancient counterparts thereof) could be developed within the "logic" of, say, *Principia Mathematica*.

However, what *is* an important difference is that, whereas on the modern account *set theory* itself is capable of being "reinterpreted" (just like any other axiomatized theory), the "background-theory" of Frege, i.e. *logic*, is *not*. Frege's logic was meant to be, as was stated at the beginning of this introduction, a *meaningful formalism*, a fully interpreted "universal language". An important lesson to be drawn from this is that metatheory is, from a modern point of view, essentially *relative*. There is no vantage point from which metatheoretical investigations could be judged *absolutely*. Modeltheoretic consistency- or independence proofs for instance are always relative to the set-theoretical metatheory in which they are framed.

---

[60] In Carnap's lecture notes for instance, extensions are not mentioned as beloning to the realm of "logic". See [10].

[61] The independence claim regarding σ, α and τ for instance is guaranteed by an axiom stating the existence of some infinite set. If only *finite* models were available in our metatheory, α were *indeed* a consequence of τ and σ.

Frege, on the other hand, *did* believe in such an "absolute" vantage point (or so it seems), and it is precisely fixed by the universal laws of *logic*.

1.5 Introductory remarks to the papers collected in this Dissertation

So let me summarize what kind of metatheoretical questions I think Frege "could have asked".

As it should be obvious from what has been said in the previous sections, I think that a lot of what today counts as "metatheory" could have been done by Frege, given some terminological adjustments. Metatheoretical questions that seem to lie within the bounds of Frege's "conceptual scheme" include

1. *Completeness* and *soundness* of deductive systems of restricted parts of what Frege would have regarded as belonging to the realm of logic.
2. Questions of "applied metatheory": In particular investigations relating to "formal axiomatics", such as questions of *independence*, *consistency* and *categoricity* of axioms (or "conditions").[62]

Of course, as I have said earlier, in claiming that Frege "could have been engaged" in such questions, it is *not* thereby said that he would have attributed to them the *significance* a modern logician might attribute to them. But this does not seem to be the issue when Frege is said to be "universalist" and therefore unable to be concerned with such questions *in principle*. The completeness of a given formalization of deductive consequence of first-order logic for instance, is only interesting to the extent that it shows that different concepts of consequence are "in harmony". But the fundamental importance that is sometimes attributed to completeness seems to stem from the assumption that the notion of *semantic consequence* is somewhat *more basic*. And further, that the notion

[62]It must be noted that this is true only *to a certain extent*: many interesting metatheoretical questions are concerned with axiomatic systems that are *not finitely axiomatized*. Standard first-order *set theory* (ZFC) and first-order *arithmetic* are cases in point. It is at least not *entirely clear* at this point how Frege would have handled *axiom schemes* like the first-order induction-scheme. It seems that, from a Fregean point of view, generality should always be expressible by means of object-linguistic quantifiers. But axiom-schemes require quantification in the (or *a*) *metatheory* over syntactical items (formulas or sentences) of the *object theory*.

of semantic consequence is adequately captured by the set-theoretically defined concept of *model-theoretic consequence*. But all of this can (and *has*) been doubted, as I have mentioned earlier. Frege, to state the most obvious, would not be very happy to base the notion of logical consequence on a theory of *sets*. It is well known that, even before Russell's letter came, Frege was very suspicous of *sets*. From his point of view, *sets* had to be construed as *extensions of concepts* if they were to be intelligible at all. It is not entirely clear what Frege's position with regard to sets was after learning of Russell's paradox, but he seems to have upheld the view that sets had to be construed as extensions of concepts, even though talk about extensions had to be restricted in some way. Anyways, the view that one could define a notion of logical consequence in terms of sets (or extensions) *and* believe it to be more fundamental than, say, derivability by forms of inference that are accepted as *logical*, would, I think, not come to his mind.

But there are still open problems when it comes to the questions of independence and consistency of *proper* axioms, as Frege conceives of them. As it has been shown, the independence of "axioms" understood as conditions would have posed no problems for Frege. Note though that, on Frege's account, it is not obvious that in showing the independence of the *conditions* corresponding to some set of proper axioms we have thereby shown the independence of the proper axioms *themselves*. The problem of what is *meant* by a "genuine axiom" and "independence" as applied to genuine axioms, is not even *posed* yet. So the questions still remain:

- What, according to Frege, is to be understood by a "genuine axiom"?
- What is meant by "independence" if applied to *genuine* axioms and how, according to Frege, can it be *shown* (if it can be shown at all) that a given set of genuine axioms is independent?
- What is meant by "consistency" if applied to *genuine* axioms and how, according to Frege, can it be *shown* (if it can be shown at all) that a given set of genuine axioms is consistent?

The articles collected in this dissertation are concerned exactly with these questions.

The focus on Hilbert as providing the area of friction (as well as providing points of contact in his later writings) in this introduction turns out to be no coincidence. For it is

precisely in his engagement in critising Hilbert's methodology in his *Grundlagen der Geometrie 1899* (which is part of what has later been called the "Frege-Hilbert-Controversy") where Frege develops his ideas on this issue. In a series of articles, dating from 1903 to 1906, titled *Über die Grundlagen der Geometrie*[63], Frege first tries to elucidate the problems that, according to him, Hilbert's methodology bear. Frege felt himself forced to write these articles because Hilbert refused to publish their correspondance, which had lasted for a couple of months, starting with a letter from Frege from December 1899 and ending with a short letter from Hilbert, dating from September 1900. After having set out what he thinks is the core of Hilbert's method, Frege goes on to review why he considers this method to be flawed if applied to genuine axioms. In the last section of his 1906 article, Frege then elucidates how, according to him, independence proofs regarding genuine axioms should be handled. Essentially, Frege's suggestion is that a "new science" has to be established in order to investigate the question of independence of genuine axioms. The articles contained in this dissertation take exactly this at their starting point.

The first article is concerned with a particular interpretive issue concerning Frege's *new science*, viz. the fact that according to Frege genuine axioms are *thoughts*, i.e. intensional entities. It is argued that this might create substantial problems for Frege, some of which Frege might have been well aware of. Remember that Frege's formal systems in his *Begriffsschrift* as well as in the *Grundgesetze* are firmly *extensional*, so an adequate treatment of independence of genuine axioms creates the need to provide an account of such entities. Further it is argued that Frege must have been aware of this problem, and that this was one of the reasons why he was somewhat reluctant towards his own proposal concerning independence proofs, set out in the last part of the 1906-article.

The second article is the core of this dissertation and presents a more detailled reconstruction of what Frege's "new science" might have looked like, had he spent more effort in spelling it out. It is shown that his proposal bears – in its essence – important points of contact with 20th century logical theory. In particular, Frege's "axiomatic" approach to metatheoretic questions (like independence of axioms) seems to suggest itself as a forerunner of the axiomatic treatment of metatheoretical concepts as exemplified by

[63]Collected and translated in English in [40].

axiomatic theories of *truth*. In fact, Frege's proposal even *presupposes* an axiomatization of truth if fully regimented proofs of independence should be possible.

The last article is an outgrowth of this engagement with Frege's axiomatic approach to metatheory. It is concerned with a certain technical point with regards to axiomatic theories of truth and is meant to provide a contribution to the contemporary discussion.[64]

References

[1] Antonelli A., May R. 2000. 'Frege's new science', *Notre Dame Journal of Formal Logic*, **41** (3), 242-270

[2] Bernays P. 1942. Review of 'Ein Unbekannter Brief von Gottlob Frege über Hilberts erste Vorlesung über die Grundlagen der Geometrie. by Max Steck', in *The Journal of Symbolic Logic*, Vol. **7** (2), 92-93

[3] Blanchette P. 'Frege on Formality and the 1906 Independence-Test', forthcoming in: Link G. (ed.), *Formalism and Beyond: On the Nature of Mathematical Discourse*, Ontos Press

[4] Blanchette P. 2007. 'Frege on consistency and conceptual analysis', *Philosophia Mathematica*, **15** (3), 321-346

[5] Blanchette P. 1996. 'Frege and Hilbert on consistency', *The Journal of Philosophy* **93** (7), 317-336

[6] Boolos G. 1985. 'Reading the *Begriffsschrift*', reprinted in Boolos G.: *Logic, Logic, and Logic*, Cambridge MA: Harvard University Press 1998, 115 - 170

[7] Boolos G. 1996. 'On the proof of Frege's theorem', reprinted in Boolos G.: *Logic, Logic, and Logic*, Cambridge MA: Harvard University Press 1998, 275 - 290

[8] Boolos G. 1996. 'The Standard of Equality of Numbers', reprinted in Boolos G.: *Logic, Logic, and Logic*, Cambridge MA: Harvard University Press 1998, 202-219

[64] See [26] for an extensive treatment of axiomatic theories of truth.

[9] Carnap R. 2000. (ed. Bonk T), *Untersuchungen zur allgemeinen Axiomatik*, Darmstadt: Wiss. Buchgesellschaft

[10] Carnap R. 2003 (eds. Reck E., Awodey S., Gabriel G.), 'Frege's lectures on Logic: Carnap's student notes 1910-1914', Illinois: Carus Puplishing Company

[11] Demopoulos W. 1994 'Frege, Hilbert, and the Conceptual Structure of Model Theory', *History and Philosophy of Logic* **15**, 211-225

[12] Dummett M. 1976. 'Frege on Independence and Consistency', in Schirn M. (ed.): *Studies on Frege I: Logic and Philosophy of Mathematics*, Stuttgart: Friedrich Frommann Verlag, Günther Holzboog GmbH & Co, 229-242

[13] Dummett M. 1991. *Frege. Philosophy of Mathematics.* London: Duckworth

[14] Etchemendy J. 1990. *The concept of logical consequence.* Harvard, MA: Harvad University Press

[15] Frege, G. 1879. *Begriffsschrift.* in I. Angelelli (ed.): *Begriffsschrift und andere Aufsätze*, Hildesheim: Georg Olms Verlag [2007]

[16] Frege, G. 1903. 'Über die Grundlagen der Geometrie', *Jahresbericht der deutschen Mathematikervereinigung* **12** (1903) reprinted in English in [40]

[17] Frege, G. 1906. 'Über die Grundlagen der Geometrie', *Jahresbericht der deutschen Mathematikervereinigung* **15** (1906), reprinted in English in [40]

[18] Frege G. 1892-1902. *Grundgesetze der Arithmetik. Begriffsschriftlich abgeleitet. Band I und II.*, Paderborn: mentis Verlag 2009

[19] Frege G. 1976. *Wissenschaftlicher Briefwechsel*, Hamburg: Felix Meiner Verlag

[20] Frege G. 1960. *Foundations of Arithmetic*, revised edition, transl. by J. Austin, Harper & Brothers, New York

[21] Frege G. 1984. *Collected Papers on Mathematics, Logic and Philosophy*, Basil Blachwell Publisher Ltd, Oxford, Brian McGuinnes (ed.)

[22] Frege G. 1969. *Nachgelassene Schriften.* Hamburg: Felix Meiner Verlag

[23] Gödel K. 1986. *Collected Works. Vol. 1.* Oxford University Press, New York. Clarendon Press, Oxford. Feferman S., Dawson J., Kleene S., Moore G., Solovay R., Heijenoort J. (eds.)

[24] Goldfarb W. 2005. 'Frege's conception of logic', in Reck E. and Beaney M. (eds.), *Gottlob Frege: Critical Assessments of Leading Philosophers*, New York: Routledge

[25] Goldfarb W. 1979. 'Logic in the Twenties: the Nature of the Quantifier', Journal of Symbolic Logic **44**

[26] Halbach V. 2011. *Axiomatic Theories of Truth*, Cambridge: Cambridge University Press

[27] Hallett M. 2010. 'Frege and Hilbert', in Ricketts T., Potter M. (eds). *The Cambridge Companion to Frege*, New York: Cambridge University Press 2010, 413-464

[28] Heck R. 1993. 'The development of Arithmetic in Frege's *Grundgesetze der Arithmetik*'. *Journal of Symbolic Logic*, **58** (2), 579-601

[29] Heck R. 1995. 'Definition by Induction in Frege's *Grundgesetze der Arithmetik*', in W. Demopoulos (ed.), *Frege's Philosophy of Mathematics*, Cambridge MA: Harvard University Press, 1995, pp. 295-333

[30] Heck R. 1997. 'Grundgesetze der Arithmetik, I, §§ 29–32, *Notre Dame Journal of Formal Logic*, **38**, 437–74.

[31] Heijenoort J. 1967. 'Logic as Calculus and Logic as Language', *Synthese* **17** (1), 324-330

[32] Hempel C. 1945. 'On the Nature of Mathematical Truth', *The American Mathematical Monthly* **52**, 543–56

[33] Hilbert D. 1899. *Grundlagen der Geometrie*, Leipzig: Teubner Verlag [1923]

[34] Hilbert D., Bernays P. 1934. *Grundlagen der Mathematik*. Volume 1, Berlin: Springer

[35] Hintikka J. 1988. 'On the Development of the Model-theoretic Viewpoint in Logical Theory', *Synthese* **77**, 1-36

[36] Hintikka J. 2011. 'What is the Axiomatic Method?', *Synthese* **183** (1), 69-85.

[37] Hodges W. 1997. *A shorter model theory*. Cambridge University Press

[38] Kaye R. 1991. *Models of Peano Arithmetic*. New York: Oxford University Press

[39] Kambartel F. 1976. 'Frege und die axiomatische Methode. Zur Kritik mathematikhistorischer Legitimationsversuche der formalistischen Ideologie', in Schirn M. (ed.): *Studies on Frege I: Logic and Philosophy of Mathematics*, Stuttgart: Friedrich Frommann Verlag, Günther Holzboog GmbH & Co, 215-228

[40] Kluge, Eike-Henner W. (ed.). 1971. *On the Foundations of Geometry and Formal Theories of Arithmetic*, Yale University Press, New Haven and London

[41] Korselt A. 1903. 'Über die Grundlagen der Geometrie', *Jahresbericht der Deutschen Mathematiker-Vereinigung*, **12**, reprinted in English [40]

[42] Kunen K. 1980. *Set Theory. An Introduction to Independence Proofs*, Amsterdam et. al.: North Holland

[43] Kutschera F. 1989. *Gottlob Frege*, Berlin: Walter de Gruyter & Co

[44] Linnebo O. 2004. 'Frege's proof of referentiality', *Notre Dame J. Formal Logic Volume*, **45** (2), 73-98

[45] McGee V. 1996. 'Logical Operations'. *Journal of Philosophical Logic* **25**, 567-580

[46] Peckhaus V. 1990. *Hilbertprogramm und Kritische Philosophie. Das Göttinger Modell interdisziplinärer Zusammenarbeit zwischen Mathematik und Philosophie*, Göttingen: Vandenhoeck & Ruprecht

[47] Prawitz D. 2005. 'Logical Consequence from a Constructivist Point of View'. in S. Shapiro (ed.), *Oxford Handbook of Philosophy of Mathematics and Logic*, Oxford University Press, 671-695

[48] Quine. W. 1950. *Methods of Logic*. New York: Holt

[49] Quine. W. 1970. *Philosophy of Logic*. Cambridge, MA: Harvard University Press [1986]

[50] Reck E. (ed.) 2002. *From Frege to Wittgenstein: Perspectives on Early Analytic Philosophy*, Oxford: Oxford University Press

[51] Reck E. and Beaney M. (eds.) 2005. *Gottlob Frege: Critical Assessments of Leading Philosophers*, New York: Routledge

[52] Resnik M. 1974. 'The Frege-Hilbert Controversy', *Philosophy and Phenomenological Research* **34** (3), 386-403

[53] Ricketts T. 1997. 'Frege's 1906 Foray into Metalogic', reprinted in Beaney M., Reck E. (eds.): *Gottlob Frege: Critical Assessments of Leading Philosophers*, Vol. 2, New York: Routledge 2005, 136-155

[54] Robinson A. 1966. *Non-standard analysis*. Amsterdam: North-Holland Publishing Co.

[55] Stanley J. 1996. 'Truth and Metatheory in Frege', reprinted in Beaney M., Reck E. (eds.): *Gottlob Frege: Critical Assessments of Leading Philosophers*, Vol. 2, New York: Routledge 2005, 109-135

[56] Schiemer G., Reck E. 'Logic in the 1930s: The Rise of Model Theory'. (forthcoming)

[57] Schiemer G. 'Carnaps early semantics'. *Erkenntnis*. (forthcoming)

[58] Schiemer G. 2012. 'Carnap's Untersuchungen: logicism, formal axiomatics, and metatheory'. In R. Creath (ed.) *Rudolf Carnap and the Legacy of Logical Empiricism*. Springer: Berlin

[59] Shapiro S. 1991. *Foundations without Foundationalism. A case for Second-order Logic.* Oxford: Clarendon Press

[60] Tappenden, Jamie. 1997. 'Metatheory and Mathematical Practice in Frege', *Philosophical Topics* **25** (2), 213-264

[61] Tappenden, Jamie. 2000. 'Frege on Axioms, Indirect Proof, and Independence Arguments in Geometry: Did Frege reject Independence Arguments?', *Notre Dame Journal of Philosophy* **41** (3), 271-315

[62] Tarski, A. 1977. *Einführung in die mathematische Logik* (fith edition), Göttingen: Vandenhoeck & Ruprecht

[63] Tarski, A. 1956. *Logic, Semantics, Metamathematics*, Oxford: Clarendon Press

[64] Wehmaier K. 1997. 'Aspekte der Frege-Hilbert-Korrespondenz', *History and Philosophy of Logic* **18**, 201-209

[65] Weiner J. 1990. *Frege in Perspective*, Ithaca, N.Y.: Cornell University Press,

2 Remarks on independence proofs and indirect reference

Abstract.[65] In the last two decades there has been increasing interest in a re-evaluation of Frege's stance towards consistency- and independence proofs. Papers by several authors deal with Frege's views on these topics. In this note I want to discuss one particular problem, which seems to be a main reason for Frege's reluctant attitude towards his own proposed method of proving the independence of axioms, namely his view that *thoughts*, i.e. intensional entities are the objects of metatheoretical investigations. This stands in contrast to more straightforward interpretations, which claim that Frege's hesitancy is mainly due to worries concerning the logical constants or what counts as a logical inference.

2.1 Introduction

In the last two decades there has been increasing interest in a re-evaluation of Frege's stance towards consistency- and independence proofs. Papers by Tappenden ([31] and [32]), Blanchette ([2] and [3]), Hodges ([17]), Antonelli and May ([1]) and others deal with the question whether Frege was able to generate independence results or whether – for whatever reason – he was not. The reasons for believing the latter range from general issues concerning Freges conception of logic (centered around the distinction between *logic as language* vs. *logic as calculus* of the classical [13]) to more specific ones, such as Blanchette's analysis-problem.

The most obvious reason for assuming that Frege was able to generate independence results is that Frege actually provided – with some reservations – a method for proving the mutual independence of axioms in his 1906-article *On the foundations of geometry*. The rest of this paper will be organized as follows: First of all I will give a brief sketch of the Frege-Hilbert controversy and Frege's *new science* ('new science' is what Frege calls the theory in which he wants to prove the independence of *real* axioms — as opposed to Hilbertarian 'Pseudo-axioms'). Next I will argue that the question of independence proofs within the new science and Frege's theory of indirect reference, as outlined in his

[65]This paper has been accepted for publication in *History and Philosophy of Logic*. Date of acceptance: 7th June, 2012.

1891, belong together. This will also partially explain Frege's reservations concerning his proposal.

2.2 The Frege-Hilbert Controversy

As is well known, Frege criticized Hilbert's methodology for proving the consistency and mutual independence of the axioms of geometry for various reasons, which are closely tied to his general conception of what axioms and definitions really are, or should be taken to be.[66] As the main points are fairly well known I will just give a brief sketch of Frege's main criticisms of Hilbert and how he eventually arrives at his own proposal for proving independence.[67]

The first criticism is aimed at Hilbert's doctrine that axioms can be used to define 'implicitly' the basic concepts that occur in these axioms. As Frege understands the axiomatic method, axioms cannot define anything because axioms and definitions serve quite different purposes. If we want to axiomatize some piece of knowledge, we lay down a set of sentences which are known to be true and which cannot be proven from more basic truths. Definitions on the other hand do not have the purpose of expressing unprovable truths but are meant to give a meaning to a heretofore meaningless sign. Definitions therefore cannot be axioms (although they function in inferences as if they were). On the other hand axioms obviously cannot be definitions in Frege's strict sense of 'definition', which includes eliminability and conservativeness.

The second criticism is closely related to the first one: As already mentioned, for Frege axioms are truths which cannot be proven. Therefore, if we lay down some set of sentences expressing such basic truths we are considering *real* sentences which are either true or false and which express – what Frege calls – 'thoughts'. In proving his independence results Hilbert on the other hand is not concerned with just *one* interpretation of his axioms, but with *various* ones. That is, Hilbert is considering his system of axioms as laid down

---

[66]The 1906 article is in fact adressed against Alvin Korselt, a defender of the Hilbertarian methodology. However, it is obvious from the short correspondence between Frege and Hilbert, that the main points of Frege's criticisms are directed at Hilbert too.

[67]For an introduction to the Frege-Hilbert controversy see Blanchette's entry in the *Stanford Encyclopedia of Philosophy* http://plato.stanford.edu/entries/frege-hilbert/. For a general discussion of the significance of the Frege-Hilbert dispute see for instance [4], [25], [18] and [33].

in his *Grundlagen der Geometrie* (*1899*) as being *formal* in the sense of being open to re-interpretation. As he famously writes to Frege:

> But surely it is self-evident that every theory is merely a framework or schema of concepts together with their necessary relations to one another, and that the basic elements can be construed as one pleases.[68]

For Frege this is plainly impossible. A proper sentence either has a determinate sense (that is, expresses a particular thought) or it does not. For the purpose of science it is also necessary that a sentence has a 'meaning' (in Frege's technical usage of the word) which means that it is definitely either true or false.[69] In short: To speak of 'uninterpreted sentences' is for Frege simply a contradiction in terms. Frege correctly points to the fact that Hilbert is using the word 'axiom' not in its *traditional*, but a *novel* sense, and one which – in Frege's opinion – lacks the clarity which he expects such a basic concept to have. As he puts it at the end of the 1906-article with an eye to independence proofs:

> As long as the word "axiom" was used as a heading only, a fluctuation in its reference could be tolerated. Now, however, since the question of whether an axiom is independent of others has been raised, the word "axiom" has been introduced into the text itself and something is asserted or proved about what it is supposed to designate.[70]

Now the only sense that Frege can make of Hilbert's method of reinterpretation and his insistence that the basic concepts 'point', 'line', etc. of geometry do not mean anything *specific* is that they function as *variables*. For Frege a sign either has a particular meaning which cannot be altered randomly or it is a letter which serves to lend generality to a

---

[68] [20], p. 13

[69] Recall that Frege introduced the notion of a *thought* in his landmark paper 'Über Sinn und Bedeutung', where he sets out his semantical theory in order to solve a puzzle with regard to the relation of identity. According to this theory two dimensions of semantic value can be made out, the level of *reference* ('Bedeutung') and the level of *sense* ('Sinn'). That is, each expression of a language properly so called, has a determinate sense as well as a reference. The relation between sense and reference is that each expression has a sense which determines or 'picks out' a unique referent in a particular way. Now, according to Frege, a *thought* is just the sense of a declarative sentence, and it's referent is one of the two truth values.

[70] [20], p. 111

judgement. So although the 'pseudo-propositions' '$x < 1$' and '$x > 2$' do not express thoughts and are neither true nor false, the sentence '$(\forall x)($ if $x < 1$ then $x > 2)$' *does* express a particular thought and has a determinate truth value (at least if we take the signs '$<$', '$>$', '1' and '2' as having their usual meaning).[71]

So what Hilbert is in fact proving in the eyes of Frege with his method of reinterpretation is not that the real axioms of Euclidean geometry – which have a determinate meaning – are mutually independent, but that certain *second-level concepts* are mutually independent. To be more explicit: Let Geo^- stand for the conjunction of the axioms of Euclidean geometry save the axiom of parallels PA. Then, according to Frege, in replacing each primitive predicate (such as 'point', 'line', etc.) with a variable of the appropriate type, we arrive at a formula $Geo^-(P, L, ...)$ determining a the second-level concept which applies to (sequences of) first-level concepts and relations. Similarly, starting from PA we arrive at a second-level concept $PA(P, L, ...)$. Now, from Frege's viewpoint, what Hilbert has shown is that the second-level concept defined by $PA(P, L, ...)$ is independent of the second-level concept defined by $Geo^-(P, L, ...)$ in the sense that the quantified conditional $\forall P \forall L ...(Geo^-(P, L, ...) \rightarrow PA(P, L, ...))$ is not logically valid by exhibiting a *counterexample*, i.e. a sequence of meaningful first-level concepts and relations, which, if substituted accordingly for the variables, yields a true antecedent and a false consequent.[72] But, as we shall see shortly, this has no obvious bearing on the question of the independence of the *genuine axioms* of Euclidean geometry as Frege conceives of them.

In the third section of his 1906 paper on geometry Frege turns to his own account of how independence proofs should be handled.

The first thing to mention is that Frege states very clearly what he thinks *are* and what *are not* the objects of investigation when we ask ourselves if some axiom is independent of others. Frege says that we are concerned with *thoughts* and not the *sentences* which express these thoughts. As he writes:

> When one uses the phrase 'prove a proposition' in mathematics, then by the word 'proposition' we clearly mean not a sequence of words or a group of signs, but a thought: something of which one can say that it is true. And similarly,

---

[71] [20], p. 99

[72] [20], pp. 83-91. For a classical exposition of this reconstruction of Hilbert see [25] and [18].

> when one is talking about the independence of propositions or axioms, this, too, will be understood as being about about the independence of thoughts. [...] We have to distinguish between the external, audible or visible which is supposed to express a thought, and the thought itself. [...] no one wants to predicate this independence of what is audible or visible.[73]

That is, for Frege sentences are just 'chalk on the board' or 'ink on the paper', they only have physical properties.[74] So when Frege speaks of 'axioms' he is not talking about *sentences* but about the *thoughts* they express and the relation of logical dependence is therefore – strictly speaking – a relation between thoughts. Note that on this account the sentence 'The axiom of parallels is independent from the rest of the Euclidean axioms' is not just expressing a determinate thought, but also has thoughts *as its subject matter*. So unlike other mathematical theories, which have points, numbers or sets as their objects, the new science deals with *thoughts*.

Secondly, Frege elucidates what he holds that 'independence' – as applied to real axioms – should be taken to be by referring to what he calls a 'logical step'. By a 'logical step' he means the following: let Ω be a set of thoughts. If a thought A can be obtained from Ω by a *logical inference*, then we can form a new set of thoughts Ω' by adding the thought A to the set Ω. Now a thought G is said to be *dependent* on Ω, if by a finite sequence of such logical steps we eventually arrive at a set of thoughts Ω'', of which G is a member. If this cannot be done, then G is said to be *independent* of Ω.[75] (In what follows I will understand 'the thought A is inferable from the thought B' to mean 'the thought A can be obtained by a finite chain of logical inferences from the thought B'). This explains, why for Frege it is not enough to establish the negation of the conditional $\forall P \forall L ... (Geo^{-}(P, L, ...) \rightarrow PA(P, L, ...))$ – for in this quantified conditional *inferences* or something like inferential relations between thoughts are not even mentioned.

Thirdly, the new science stands on a par with other axiomatic theories like geometry

---

[73] [20], p. 101

[74] As far as the 1903 - 1906 papers are concerned, Frege never distinguishes clearly between expression-*types* and expression-*tokens*. Throughout the 1906-paper the word 'proposition' ('Satz') is used in the meaning of sentence-*token*. Frege, however, *was* aware of the distinction, as a letter to Dingler shows. See [10], p. 35. In any case, according to Frege, the relation of independence applies neither to sentence-types nor to sentence-tokens.

[75] [20], p. 104

in that it has its own axioms and basic concepts. Frege writes:

> Now we may assume that this new realm has its own specific, basic truths which are as essential to the proofs constructed in it as the axioms of geometry are to the proofs of geometry; and that we also need these basic truths especially to prove the independence of a thought from a group of thoughts.[76]

It may also be assumed that like any other theory, the new science too should be formalizable within Frege's 'Begriffsschrift' (or some extension of it). In fact, this was the very reason for introducing a 'Begriffsschrift': it should provide a framework, in which *every* piece of knowledge could be expressed and it should provide precise syntactical rules which guarantee that in a proof no step can occur, which is not in accordance with accepted forms of inference. So unlike Hilbert, Frege does not rest content with establishing the independence of axioms informally. Quoting again from Frege's 1906-paper:

> As it stands, we remain completely in the dark as to what he [Hilbert] really believes he has proved and which logical and extralogical laws and expedients he needs for this.[77]

As in every part of logico-mathematical discourse, the axioms and rules of inference needed in proofs ultimately have to be laid down explicitly.

These are the main points which have to be kept in mind when talking about Frege's new science.

2.3 The *new science* and *indirect reference*

As I pointed out in the previous section, according to Frege, the new science has thoughts as its objects in just the same way as number theory has numbers and geometry points and lines as its objects. As Frege puts it:

> How can one prove the independence of a thought from a group of thoughts? First of all, it may be noted that with this question we enter into a realm

---

[76] [20], p. 106
[77] [20], pp. 111-112

> that is otherwise foreign to mathematics. For although like all other disciplines mathematics, too, is carried out in thoughts, still, thoughts are otherwise not the object of its investigations.[78]

I also indicated that like any other science the new science must have its basic truths and that these truths must be expressible in Frege's system of logic (or some extension). Frege mentions three such axioms. He states the first two explicitly and elucidates the last one (which he calls an 'efflux of the formal nature of the logical laws'[79]) just informally. The first two are the following:

> (NS1) If the thought G follows from the thoughts A, B, C by a logical inference then G is true.
>
> (NS2) If the thought G follows from the thoughts A, B, C by a logical inference then each of the thoughts A, B, C is true.

(NS2) codifies Frege's conviction that something can be inferred only from premises that are true, whereas (NS1) states that everything so inferred must likewise be true.[80]

First of all note that Frege is talking about *logical inferences*. Clearly, Frege has no *general* account about what counts as a ('genuine') *logical* inference, but it seems fairly obvious, that the concept of logical inference is closely tied to the syntactically defined concept of *derivation*. *At least* the kind of deduction-rules and laws Frege actually states in his *Begriffsschrift* or his *Grundgesetze der Arithmetik* (save the infamous basic law V) should count as codifications of genuine logical inferences and logical laws. It has to be kept in mind however, that *inferability* is a relation between *thoughts*, whereas

---

[78] [20], p. 106

[79] [20], p. 107. In fact, this last axiom is the key-axiom of his new science. Roughly speaking, it states that a logical proof is invariant under substitutions of the non-logical vocabulary. By means of this new axiom it should be possible to prove the independence of the axiom of parallels from the rest of the Euclidean axioms by finding a series of concepts and relations which yield true sentences $Geo^{-'}$ when substituted for the geometrical concepts in the actual axioms of Euclidean geometry Geo^{-} (safe the axiom of parallels) and a false sentence PA' when substituted for the geometrical concepts in the axiom of parallels PA. If the axiom of parallels PA were provable from the rest of the Euclidean axioms Geo^{-}, then, by the new law, PA' would be provable from $Geo^{-'}$, and hence true by (NS1) and (NS2) (*soundness*) (and the fact that every sentence in $Geo^{-'}$ is true). But this contradicts the falsity of PA'.

[80] So (NS1) and (NS2) imply the 'soundness' of the relation of 'being logically inferable from'. [20], p. 107

derivability is a relation between *sentences*. So the fact that Frege admits that he has no general criterion about what counts as a genuine logical inference shows that Frege was aware of the possibility that there might be *logical inferences* that are not represented by any deduction rules he states for his formal systems and that there might be *logical truths*, which are not derivable within these systems.[81] After all, the purpose of inventing the *Begriffsschrift* was a rather specific one: to deduce arithmetic from logical principles alone — there was no need for Frege to come up with an *exhaustive* list of logical principles.

In any case, the relation of 'being logically inferable from' appears in the axioms (NS1) and (NS2) and hence has to be introduced either as a *primitive* relation or as *defined* by other more basic concepts of the new science. So let '$Bew(\xi, \zeta)$' stand for the expression 'ζ is inferable from ξ', where the Greek letters 'ξ' and 'ζ' mark the argument places of this two-place relational concept. '$Bew(\xi, \zeta)$' expresses a relation which applies to *thoughts*. But we have yet to explain how the application of a predicate to a thought is to be understood. This is where Frege's theory of indirect reference might be invoked. Recall that Frege developed his theory of indirect reference in order to deal with *opaque contexts*, created for instance by verbs expressing – what are now called – *propositional attitudes* (to believe, to know, to hope, etc.). Opaque contexts have the peculiar property that they seem to create problems for the *principle of extensionality*, which states that in replacing a subexpression in a more complex expression with a co-referential expression, the referent of the complex expression remains the same. If, for instance, in the true sentence 'Mary believes that Vienna is the capital of Austria' the subexpression 'Vienna is the capital of Austria' is replaced with the co-referential expression 'Kuala Lumpur is the capitol of Malaysia', we might yield the falsehood 'Mary believes that Kuala Lumpur is the capitol of Malaysia'. Frege's solution to such apparent counterexamples to the principle of extensionality is that he determines that in opaque contexts an expression does not have it's *usual reference* (it's 'gewöhnliche Bedeutung') but it's *indirect reference* (it's 'ungerade

[81] 'In particular, we will find that this final basic law [the law of the 'efflux of the formal nature'] which I have attempted to elucidate by means of the above mentioned vocabulary still needs more precise formulation, and that to give this will not be easy. Furthermore, it will have to be determined what counts as a logical inference and what is proper to logic.' [20], p 110f. It seems to me though that once it is determined 'what counts as a logical inference', it should be possible to capture these inferences by means of syntactical rules.

Bedeutung'), which, according to Frege, is just it's *direct sense* (it's 'gewöhnlicher Sinn'). Therefore, the given example is not a counterexample to the principle of extensionality after all, because *with respect to the given context*, the two sentences 'Vienna is the capitol of Austria' and 'Kuala Lumpur is the capitol of Malaysia' are *not* co-referential, for they do not have the same *direct sense* (they do not express the same *thought*).[82]

Now going back to our question how the inferability-predicate is to be applied to thoughts: if the context '$Bew(\xi, \zeta)$' is *opaque*, this will have the effect that sentences occurring within this context will not have their *direct* but their *indirect reference*. And this in turn will have the consequence that the relation expressed by '$Bew(\xi, \zeta)$' will apply to the *thoughts* expressed by these sentences just as intended. But that the context '$Bew(\xi, \zeta)$' should indeed be considered as an opaque context from the Fregean point of view can plausibly be seen from the following 'counterexample' to the principle of extensionality:

1. It is inferable from ($1 + 1 = 2$ and $2 + 1 = 3$), that $(1 + 1) + 1 = 3$[83]
2. Vienna is the capital of Austria = $(1 + 1) + 1 = 3$
3. It is not inferable from ($1 + 1 = 2$ and $2 + 1 = 3$), that Vienna is the capital of Austria

Obviously there is no significant logical connection between the truth of 'Vienna is the capital of Austria' and the truth of '$(1 + 1) + 1 = 3$'. So what has been said above about Mary applies here too: If we are interested in the truth of 'It is inferable from ($1 + 1 = 2$ and $2 + 1 = 3$), that $(1 + 1) + 1 = 3$', then, according to Frege, we are not concerned with the truth of '$1 + 1 = 2$ and $2 + 1 = 3$' or '$(1 + 1) + 1 = 3$' *at all*, but rather with the

---

[82]For an exposition of Frege's theory of indirect reference, see his [9]. For a thorough discussion see chapter 9 of Dummett's [4].

[83]This must not be confused with the sentence ' "$(1 + 1) + 1 = 3$" is inferable from "$1 + 1 = 2$ and $1 + 2 = 3$" ' where the relation of inferability is construed as a relation between *syntactical* objects, namely *sentences* (sentence-*types*). As I mentioned above, for Frege the relation of inferability is a relation which applies to *thoughts*. But in ' "$(1 + 1) + 1 = 3$" is inferable from "$1 + 1 = 2$" and "$2 + 1 = 3$" ' the relation of inferability applies to the *sentences* '$(1 + 1) + 1 = 3$' and '$1 + 1 = 2$ and $2 + 1 = 3$' of which " '$(1 + 1) + 1 = 3$' " and " '$1 + 1 = 2$ and $2 + 1 = 3$' " are *names*.

question *what* is inferable. We are interested in the *thoughts expressed by* '$1 + 1 = 2$ and $2 + 1 = 3$' and '$(1 + 1) + 1 = 3$'.[84]

We can see more clearly now why the Fregean treatment of independence proofs requires serious changes to be made in the Begriffsschrift, as presented in his *1879* or his *1893*: Like any other axiomatic theory the new science too must have as basic truths some *general* statements. One might be tempted to regard e.g. (NS1) as an axiom *scheme* in the following way:

> (NS1') '$Bew(G, H) \to H$' expresses an axiom, whenever we put *particular* sentences at the places of 'G' and 'H' respectively.

The letters 'G' and 'H' in the axiom scheme '$Bew(G, H) \to H$' are *schematic* letters which can be replaced by *particular* sentences yielding *particular* axioms. But this does not seem to be the correct reading of what Frege has in mind with his (NS1). As mentioned earlier, the kind of signs which can occur in some specific theory are limited to the signs with a *determinate meaning* (such as the one-place predicate expression 'ξ is a point' in geometry or the two-place predicate expression 'ζ is inferable from ξ' in the new science) and *variables*. But schematic letters seem to be neither. According to Frege generality should always be expressed by means of *quantifiers*.[85] So it seems that the correct formalization of (NS1) should be something like this:

---

[84]The fact that in the context in question extensionality apparently fails, does not show that thoughts *must* be considered as the objects to which the predicate '$Bew(\xi, \zeta)$' applies (and not the *sentences* expressing these thoughts). The decision, that sentences are not the objects to which '$Bew(\xi, \zeta)$' applies has already been made. The point is rather that there is a natural explanation of this apparent failure within a Fregean view on logic, language and semantics.

[85]This view could be challenged on the ground that Frege did use schematic letters in the exposition of his formal systems. In his *Begriffsschrift* for instance Frege uses letters 'a' or 'b' to state axiom schemes like '$a \to (b \to a)$' (for the propositional fragment of his system) or letters like 'f' to mark a function as in '$(\forall a) fa \to fb$'. Clearly Frege uses 'schematic letters' in this sense. (In fact, some scholars like Warren Goldfarb think that even this use of schematic letters is foreign to Frege and e.g. '$a \to (b \to a)$' should be read as '$(\forall a)(\forall b)(a \to (b \to a))$'. see his [12]) But this sense of 'schematic' has to be distinguished from the sense in which this word is used in the above mentioned axiom scheme of the 'new science'. Schematic letters of the former kind are necessary for the exposition of a formal system of logic like the propositional calculus or quantification theory. They are used to mark propositional (or quantificational) structure but they do not occur in sentences of some specific theory like geometry or number theory (obviously '$\forall x(x < 1 \vee a)$' for instance is not a well-formed sentence of arithmetic).

(NS1") $(\forall G)(\forall H)(Bew(G,H) \rightarrow Tr(H))$

But this is a new situation: Up to now we have just explained how we can talk about *particular* thoughts, namely by noting that '$Bew(\xi,\zeta)$' provides an *opaque* context. By means of Frege's theory of indirect reference we were able to explain our ability to *refer* to thoughts and the apparent failure of extensionality within this context. But so far nothing has been said about *quantifying into* such contexts.[86]

There are several problems to be solved in order to make sense of Frege's proposal concerning independence proofs, but some of the most crucial ones have to do with the cases of *oratio obliqua* that occur within metatheoretical investigation and the fact that, according to Frege, quantification over *senses* is required in order to state *general* principles governing his new science.

Although Frege doesn't mention the problems of intensional logic explicitly in his 1906 article, I think this might have been one important reason why he raised some doubts about the feasibility of his own proposal and not *just* specific worries concerning the 'logical constants' or the question what counts as a 'logical inference' or 'logical law'. This is not to say that these are *not* problems for Frege. On the contrary, as he puts it in the 1906-article:

> In particular, we will find that this final basic law [the law of the 'efflux of the formal nature'] which I have attempted to elucidate by means of the above mentioned vocabulary still needs more precise formulation, and that to give this will not be easy. Furthermore, it will have to be determined what counts as a logical inference and what is proper to logic. If, following the suggestions

---

But (NS1) above is *indeed* meant to express a particular basic truth of the 'new science' -– a truth about *all thoughts*.

[86] Also note that in regarding (NS1) as a *quantified* sentence, the truth predicate occurs *ineliminably* on pain of ungrammaticality. It is precisely with (referential) quantification that the truth predicate becomes necessary. One might circumvent the use of a truth predicate by invoking *substitutional*, rather than *referential* quantifiers. But most commentators seem to agree that Frege's first-order quantifiers must be interpreted referentially. So substitutional quantifiers do not present an alternative, for *thoughts* clearly seem to be first-order entities. Note though that nothing in the following depends on the question if Frege was *aware* of the fact that an ineliminable truth predicate is needed in order to state general principles of the new science.

> above, one wanted to apply this to the axioms of geometry, one would still need propositions that state, for example, that the concept *point*, the relation of a point's lying on a plane, etc. do not belong to logic. These propositions will probably have to be taken as axiomatic.

So it is true that Frege regarded the problem of delineating 'what belongs to logic' as a pressing one, but it seems to me highly unplausible that this was his *only* worry.[87] Recall that he explicitly states what he thinks is 'new' in the *new science*: the difference between the *new science* and its older siblings like geometry or number theory lies in the *subject matter* of these sciences. The *new science* is *about* thoughts in just the same way as geometry is *about* points and arithmetic *about* numbers.

Frege's *axiomatic* approach together with his commitment to *thoughts* as the subject matter of metatheoretical investigation and the requirement that the new science should be general therefore presupposes a fully developed 'intensional logic' and most importantly, it presupposes a *semantic theory*, which explains how *quantification* works when intensional objects come into play. Moreover, given Frege's continual complaints throughout the 1903-1906 papers about Hilbert's deviant use of the word 'axiom' and his expressed point of view, according to which an 'axiom in the traditional sense' is a *thought*, it seems to me highly unlikely that Frege was not *aware* of the fact that dealing with such entities requires far-reaching amendments to his Begriffsschrift. To repeat the central point: 'with this question [the question of independence of thoughts] we enter into a realm that is otherwise foreign to mathematics. For although like all other disciplines

---

[87]The problem of the logical constants and 'what is proper to logic' has been particularly emphasized in [26]. The problem is of course that it may create *undergeneration* if we declare a form of inference (or some notion) to be logical if 'in fact' it is not. If we take a particular formalization of logic – say the logic of *Begriffsschrift* – as a basis for a definition of the inferability-relation, we may 'miss' some logical inferences. That is, it might happen that we declare some thought not to be inferable from some set of thoughts although 'in fact' they are. (As a matter of fact, *overgeneration* cannot be ruled out either, as the dramatic case of Basic Law V showed.) On the other hand, *Antonelli and May 2000* have proposed a method to define the logical constants and the concept of logical truth by adapting Frege's 'permutation argument', a method which, according to Antonelli and May, could have been known and accepted by Frege. From this they conclude that Frege's expressed skepticism concerning independence proofs was not warranted, for they seem to believe that this was the *only* problem Frege saw with respect to independence proofs.

mathematics, too, is carried out in thoughts, still, thoughts are otherwise not the object of its investigations'. So it seems to me that the fact that Frege has not given these matters enough thought, might to some extent explain his cautious stance towards independence proofs.

2.4 Conclusion

The conclusion we reached might at first glance seem rather poor; after all, Frege makes clear from the outset that the relation of *dependence* (or *inferability* as I called it) is a relation that holds between *thoughts* and not the sentences expressing these thoughts. As thoughts are intensional objects, the claim that independence proofs require some kind of intensional logic seems to be trivial. Nevertheless, I think it is worth stating this fact explicitly because most of the discussion on Frege's proposed method for proving independence seems to neglect it.[88] The fact that Frege views thoughts as the objects of metatheoretical investigation is often treated as an inessential peculiarity which need not be taken seriously. But this view cannot be sustained if we take into account that Frege wants to establish the new science as an *axiomatic* theory with its own axioms and basic concepts. If the new science is worked out in this way, it will become apparent that Frege's conviction that thoughts are its objects is highly non-trivial.

Another thing to remark concerns the role of the theory of indirect reference with regard to some of the problems that occur in connection with the new science. Remember that the theory of indirect reference is capable of explaining how *reference* to thoughts is achieved in *particular* metatheoretical statements. Furthermore, by invoking the theory of indirect reference we were able to give an account of the apparent failure of extensionality in the context '$Bew(\xi, \zeta)$'. It was central to this task to recognize that '$Bew(\xi, \zeta)$' provides an opaque context and is therefore subject to the theory of indirect reference. From this it becomes apparent that there is a kind of 'systematic stringency' in Frege's stance towards metatheory. Hence, we may take the theory of indirect reference to elucidate how Frege's conviction that *thoughts* are the objects of metatheoretical statements fits in with his more general views on logic and language.

---

[88] A notable exception is provided by [2] and [3]. I think an explicit formulation of Frege's views on intensional logic might also throw some light on Blanchette's 'analysis-problem'.

Nevertheless, the theory of indirect reference as developed by Frege himself is not sufficient to explain the *generality* involved in (at least *some*) statements of the new science. In fact, in the presence of quantifiers it seems to create serious problems, even if (NS1") above is dropped in favour of some *schematic* version. The problems that occur are analogous to the well-known problems with the interpretation of sentences like '$(\exists x)(x$ is human and Martin believes that x is taller than 3 metres)' where the first occurrence of the variable 'x' lies outside the context 'Martin believes that ξ', whereas the second occurrence lies within this context. Similarly, cases like '$(\exists x)(x$ is a number and $Bew(AR, x < 1))$' (where AR stands for some arithmetical theory), where it is not clear what the range of the variables should be taken to be, cannot be ruled out from the outset. If we expand the theory of indirect reference to include quantification, it will become difficult to give a satisfactory semantical theory which accounts for such sentences. Recall that the theory of indirect reference makes the reference of an expression *context-dependent* and this will carry over to variables that are bound by quantifiers.[89]

A second way to deal with intensional entities would be to drop the theory of indirect reference altogether and adopt a *method of direct discourse* along the lines of Church's 'logic of sense and denotation', where *reference* to intensional entities is achieved by new expressions of the kind 'the sense of the expression ξ'. This strategy, which Frege seems to have favoured later, (as a letter to Russell suggests[90]) has the advantage that reference is no longer context-dependent, for we no longer use the *same* signs for *different* referents in transparent vs. opaque contexts. But there are still huge obstacles in making full sense of this suggestion.[91]

It should also be borne in mind that, according to Frege, there is no *prima facie* problem in considering *thoughts* as objects of investigation. This is clear for instance from

---

[89] Again, the problem seems not to arise if we assume that the first-order quantifiers are interpreted substitutionally rather than referentially, for then a sentence like '$\exists x(\phi(x) \wedge \psi(x))$' (where $\phi(x)$ is a transparent and $\psi(x)$ an opaque context) is true just in case there is a *name* 'a' in the language in question such that '$\phi(a) \wedge \psi(a)$' is true, and this case can be dealt with in a 'standard way' by the theory of indirect reference.

[90] In the letter from December 28, 1902 Frege writes: 'Eigentlich müsste man ja, um Zweideutigkeit zu vermeiden, in ungerader Rede besondere Zeichen haben, deren Zusammenhang mit den entsprechenden in gerader Rede leicht erkennbar wäre.' [10], p. 236

[91] See [19] for a formal development of Fregean ideas along these lines.

his reaction to Russell's 'second paradox' (nowadays called 'Russell-Myhill-Paradox'), the paradox concerning sentences expressing the 'logical product' of some class of propositions.[92] Frege does not find anything particularly 'wrong' with considering classes of propositions (or 'thoughts' as he would say) or sentences expressing the 'logical product' of such a class. It is just that he is not satisfied with Russell's *presentation* of the paradox, which, in Frege's eyes, lacks the stringency of the notorious 'first paradox' concerning the class of all classes not being members of themselves.[93] The reason for this is not hard to find: It is a paradox about *intensional entities* and Frege has no idea what a 'proof' involving such entities might look like[94], i.e. he has no idea of how the paradox (if it *is* one at all) might even be *formulated* properly. That is, the problem is *not* that Frege thinks that quantification over senses is somehow 'weird'. It's just that he hasn't so far thought through the matter carefully enough. And this surely remains the case until his 1906 paper on geometry.

So in any case, all the problems of intensional logic afflict metatheoretical reasoning — a fact that Frege must have been aware of. Recall that Frege is very concious in declaring intensional entities as the objects of metatheoretical investigation.[95]

---

[92]The paradox arises if we consider the class of all propositions: we can now consider arbitrary subclasses and with any such class of propositions we can correlate a sentence which says that every proposition in this class is true. Any of these propositions is in its correlated class or it is not. Now consider the class A consisting of all and only the propositions which are not in their correlated class. The proposition which is correlated with the class A is then in A if and only if it is not. For more on this paradox see Myhill's classical [22] and [19].

[93]See Frege's correspondence with Russell in [10], pp. 230-242

[94]Frege expressly asks Russell by which form of inference ('Schlussweise') exactly he got to his 'second paradox'. See [10], p. 237.

[95]Although my main aim in this paper was to give an explanation – or at least a *partial* explanation – of Frege's reluctance to fully endorse his method of proving the independence of genuine axioms *in his 1906-paper on geometry*, it seems plausible to assume that if Frege had worries about intensionality in *1906*, these worries would have persisted until *1910*. Hence keeping this point in mind might also throw some light on his notorious remark in his notes on Jourdain as well: 'The unprovability of the axiom of parallels cannot be proved. If we do this apparently, we use the word "axiom" in a sense quite different from that which is handed down to us.' ('Die Unbeweisbarkeit des Parallelenaxioms kann nicht bewiesen werden. Wenn man es scheinbar thut, gebraucht man das Wort 'Axiom' in einer von der überlieferten ganz verschiedenen Bedeutung. (Vergleichen Sie meine Aufsätze *Ueber die Grundlagen der Geometrie* im 15. Bd. d. Jahresber. der Deutschen Mathematiker-Vereinigung.)' ([10] p. 119)) Commentators have often been baffled about this remark, for it seems to be in direct opposition to the fact that, at the end of his 1906-article, Frege actually presented a proposal how to prove the independence of

So in conclusion we have the following: although there was no need to introduce intensional logic for Frege's logicist project of deriving number theory from logical principles alone, there *is* a strong need to do so for an adequate treatment of metatheoretical questions (according to Frege's standards) like the question of the independence of axioms from one another. What has to be introduced, therefore, is a *deductive system* and *semantics* for quantified intensional logic to deal with the opacity occurring in metatheoretical investigations. Moreover, as I have tried to argue, because of Frege's emphatic commitment to *thoughts* as bearers of the (in-)dependence-relation, it seems highly implausible to assume that Frege was not *aware* of this fact.

It is plain that Frege's conception of *metatheory* as developed in his *Über die Grundlagen der Geometrie* is rather different from the current one. Here I am not trying to answer the question if the Fregean conception of metatheory could (at least to a certain extent)

---

genuine axioms. Tappenden on the other hand has tried to accomodate this passage with the 1906-article. His main argument is that Frege is not rejecting independence proofs *tout court*, but only independence proofs which rest on the possibility of *supposing an axiom to be false*, which, according to Frege, would be incorrect. Tappenden further argues that this reading would be suggested by the context surrounding the 'Jourdain-sentence', which is about Frege's view that something can be inferred only from premises which are *known to be true*. (See his [32]). Although I think that Tappenden is right on this point, it seems to me that this is only *half* of the truth. I agree that the context surrounding the remark suggests that in this passage Frege is indeed concerned only with *a certain type* of independence proofs, independence proofs which are not even concerned with 'genuine axioms' from Frege's point of view. On the other hand, one would expect that, if Frege in 1910 fully endorsed his 1906-method to prove the independence of genuine axioms, he would set things right and show how independence of genuine axioms *could* be proved according to him. But he only reiterates his 1906 diagnosis that independence proofs involving the supposition of an axiom to be false rest on a misuse of the word 'axiom' and refers back to his 1906-article. But note that in doing so, he also seems to refer back to the *doubts* expressed therein. Remember that the issue of independence proofs remained unsettled in the 1906-paper. Hence, even if Frege in the quoted passage only rejects a certain kind of independence proofs, this does of course *not* imply that he now fully endorsed his 1906-proposal. If, as I argued, problems concerning intensional objects were among the worries in 1906, then his 1910-reference to the 1906-paper seems to refer to the *same* worries. For neither did Frege anything in the direction of delineating 'what belongs to logic' nor did he make any serious attempts to make explicit his views on (intensional) semantics. And indeed his expressed remark that 'if we do this apparently, we use the word "axiom" in a sense quite different from that which is handed down to us' seems to hint exactly in this direction, for again, it was a central point of the 1906-article that axioms are *thoughts* and hence proving independence of axioms is proving something about *thoughts*.

be 'reconciled' with the current one or if – for whatever reason – Frege was 'blocked' from doing metatheory properly. The question if Frege was or was not able to do 'genuine metatheory' therefore remains untouched by what has been said in this paper.

References

[1] Antonelli A., May R. 2000. 'Frege's new science', *Notre Dame Journal of Formal Logic*, **41** (3), 242-270

[2] Blanchette P. 2007. 'Frege on consistency and conceptual analysis', *Philosophia Mathematica*, **15** (3), 321-346

[3] Blanchette P. 1996. 'Frege and Hilbert on consistency', *The Journal of Philosophy* **93** (7), 317-336

[4] Dummett M. 1973. *Frege: Philosophy of language*, London: Duckworth

[5] Frege G. 1879. *Begriffsschrift und andere Aufsätze*, edited by Angelelli I., Hildesheim: Olms [1964]

[6] Frege G. 1893-1903. *Grundgesetze der Arithmetik*, Hildesheim: Olms, [1998]

[7] Frege G. 1903. 'Über die Grundlagen der Geometrie', *Jahresbericht der Deutschen Mathematiker-Vereinigung*, XII, 1903, translated and edited by Eike-Henner W. Kluge, *On the Foundations of Geometry and Formal Theories of Arithmetic*, New Haven and London: Yale University Press, 1971

[8] Frege G. 1906. 'Über die Grundlagen der Geometrie', *Jahresbericht der Deutschen Mathematiker-Vereinigung*, XV, 1906, translated and edited by Eike-Henner W. Kluge, *On the Foundations of Geometry and Formal Theories of Arithmetic*, New Haven and London: Yale University Press, 1971

[9] Frege G. 1892. 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik*, reprinted in Textor M. *Funktion - Begriff - Bedeutung*, Göttingen: Vandenhoeck & Ruprecht, 2002

[10] Frege G. *Wissenschaftlicher Briefwechsel*, edited by Gottfried Gabriel et. al., Hamburg: Felix Meiner Verlag [1976]

[11] Greimann D. 2007. 'Did Frege really consider truth as an object?', *Grazer Philosophische Studien* **75**, 125-148

[12] Goldfarb W. 2005. 'Frege's conception of logic', in E. Reck and M. Beaney, *Gottlob Frege: Critical Assessments of Leading Philosophers*, New York: Routledge 2005

[13] Heijenoort J. 1967. 'Logic as Calculus and Logic as Language', *Synthese* **17** (1), 324-330

[14] Hendricks et. al. (eds). 2004. 'First-Order Logic Revisited', Berlin: Logos Verlag

[15] Hilbert D. 1899. *Grundlagen der Geometrie*, Leipzig: Teubner Verlag [1923]

[16] Hintikka J. 1988. 'On the Development of the Model-theoretic Viewpoint in Logical Theory', *Synthese* **77**, 1-36

[17] Hodges W. 2004. 'The Importance and Neglect of Conceptual Analysis: Hilbert-Ackermann iii.3', in Hendricks et. al. (eds): 'First-Order Logic Revisited', Berlin: Logos Verlag

[18] Kambartel F. 1976. 'Frege und die axiomatische Methode. Zur Kritik mathematik-historischer Legitimationsversuche der formalistischen Ideologie', in Schirn M. (ed.): *Studies on Frege I: Logic and Philosophy of Mathematics*, Stuttgart: Friedrich Frommann Verlag, Günther Holzboog GmbH & Co, 215-228

[19] Klement K. 2002. *Frege and the Logic of Sense and Reference*, New York & London: Routlegdge

[20] Kluge E. (ed.) 1971. *On the Foundations of Geometry and Formal Theories of Arithmetic*, New Haven and London: Yale University Press

[21] Korselt A. 1903. 'Über die Grundlagen der Geometrie', *Jahresbericht der Deutschen Mathematiker-Vereinigung*, XII, translated and edited by Eike-Henner W. Kluge, 'On the Foundations of Geometry and Formal Theories of Arithmetic', New Haven and London: Yale University Press, 1971

[22] Myhill J. 1958. 'Problems Arising in the Formalization of Intensional Logic', *Logique et Analyse* **1**, 78-83

[23] Reck E. (ed.) 2002. *From Frege to Wittgenstein: Perspectives on Early Analytic Philosophy*, Oxford: Oxford University Press

[24] Reck E. and Beaney M. (eds.) 2005. *Gottlob Frege: Critical Assessments of Leading Philosophers*, New York: Routledge

[25] Resnik M. 1974. 'The Frege-Hilbert Controversy', *Philosophy and Phenomenological Research* **34** (3), 386-403

[26] Ricketts T. 1997. 'Frege's 1906 Foray into Metalogic', *Philosophical Topics* **25** (2), 169-188

[27] Sluga H. 2007. 'Truth and the Imperfection of Language', *Grazer Philosophische Studien* **75** (1), 1-26

[28] Sluga H. 2002. 'Frege on the Indefinability of Truth', in E. Reck (ed.), *From Frege to Wittgenstein: Perspectives on Early Analytic Philosophy*, Oxford: Oxford University Press

[29] Stanley J. 1996. 'Truth and Metatheory in Frege', *Pacific Philosophical Quarterly* **77** (1), 45-70

[30] Textor M. (Ed.) 2002. *Funktion - Begriff - Bedeutung*, Göttingen: Vandenhoeck & Ruprecht

[31] Tappenden J. 1997. 'Metatheory and Mathematical Practice in Frege', *Philosophical Topics* **25** (2), 213-264

[32] Tappenden J. 2000. 'Frege on Axioms, Indirect Proof, and Independence Arguments in Geometry: Did Frege reject Independence Arguments?', *Notre Dame Journal of Philosophy* **41** (3), 271-315

[33] Wehmaier K. 1997. 'Aspekte der Frege-Hilbert-Korrespondenz', *History and Philosophy of Logic* **18**, 201-209

3 Frege's *On the Foundations of Geometry* and Axiomatic Metatheory

Abstract.[96] In a series of papers, dating from 1903 - 1906, Frege criticizes Hilbert's methodology of proving the independence and consistency of various fragments of Euclidean geometry in his *Foundations of Geometry*. In the final part of the last paper, Frege makes his own proposal how the independence of genuine axioms has to be proved. According to Frege, independence proofs require the development of a "new science" with it's own basic truths. The main purpose of this paper is a reconstruction of this "new science" and an examination of possible problems surrounding Frege's proposal. The strategy for this will be twofold: the reconstruction should draw attention to important connections to 20th century logical theory, while staying as close as possible with Frege's own views. The paper is organized as follows: in the first two sections the main points of the Frege-Hilbert Controversy are set forth and some issues surrounding the problem of independence proofs are discussed. Section 3 will contain an informal presentation of Frege's proposal, whereas section 4 sets out a more detailed reconstruction of what Frege's "new science" might have looked like. The concluding section is devoted to a discussion of Frege's general strategy of proving the independence of genuine axioms.

3.1 The *Foundations of Geometry*: Frege and Hilbert on Independence proofs

Hilbert's *Foundations of Geometry* (1899) is often – and rightly – seen as a landmark in the development of the so called *axiomatic method*. One of the main innovations in Hilbert's *Foundations* is that metatheoretic issues such as the questions of consistency and independence of axioms are for the first time systematically treated in a way that has since then become standard. In a famous letter to Frege he writes:

> I was forced to construct my system of axioms by the following necessity: I

[96] This paper has been submitted for publication in *Mind*. Date of submission: 1st October 2012.

> wanted to provide an opportunity for understanding those geometric propositions which I consider to be the most important products of geometric investigations – that the axiom of parallels is not a consequence of the remaining axioms; similarily the Archimedean axiom; etc. I wanted to answer the question whether it is possible to prove the proposition that two equal rectangles having the same base line also have equal sides. In fact, I wanted to create the possibility of understanding and answering such questions as why the sum of the angles of a triangle is two right angles and how this fact is related to the axiom of parallels. ([28] p. 10)

Hilbert's perspective in this passage is of decidedly *metatheoretical* character in the sense that questions about what *can* and what can *not* be proved from some given set of axioms, are posed from a point of view *external* to geometrical investigations properly so called. The purpose of Hilbert's axiomatization of geometry in his *Festschrift* therefore was not just to provide a basis for geometry from which every geometrical truth could be proved, but rather it was from the very beginning aiming at *metatheoretical properties* of Euclidean geometry and subtheories of Euclidean geometry.[97] In a similar spirit he writes in a small paper titled *Über den Satz von der Gleichheit der Basiswinkel im gleichschenklingen Dreieck* (dating from the same period):

> Unter der axiomatischen Erforschung einer mathematischen Wahrheit verstehe ich eine Untersuchung, welche nicht dahin ziehlt, im Zusammenhange mit jener Wahrheit neue oder allgemeinere Sätze zu entdecken, sondern die vielmehr die Stellung jenes Satzes innerhalb des Systems der bekannten Wahrheiten und ihren logischen Zusammmenhang in der Weise klarzulegen sucht, dass sich sicher angeben lässt, welche Voraussetzungen zur Begründung jener Wahrheit notwendig und hinreichend sind. ([22] p. 119)

At the heart of Hilbert's methodology lies his consequent — what has since become to be known as — *modeltheoretic* approach to axiom systems. Geometric axioms are no longer seen as true propositions which are immediate from our spatial intuition, but are

[97] This has been emphasized, among others, by Hintikka in his [23] and [24].

now considered to be like *conditions* in being satisfied by some interpretations and not by others. As Hilbert puts it in a famous letter to Frege:

> But surely it is self-evident that every theory is merely a framework or schema of concepts together with their necessary relations to one another, and that the basic elements can be construed as one pleases. If I think of my points as a system of other things, e.g. the system of love, of law, or of chimney sweeps ... and then conceive of my axioms as relations between these things, then my theorems, e.g. the Pythagorean one, will hold of these things as well. In other words, each and every theory can always be applied to infinitely many systems of basic elements. ([28] pp. 13-14)

Hilbert's modeltheoretic approach to axiomatic theories is the key for his independence proofs, for it is clear that, in order to prove the unprovability of some proposition ϕ from other propositions S, one cannot "go through" all possible proofs and check that none of them is actually a proof of ϕ using only propositions from the set S.[98] Hence, in the absence of prooftheoretical methods properly so called, the only way to prove an axiom ϕ to be independent from a group of axioms S is to produce a *countermodel*, i.e. an interpretation, in which every axiom in the group S is true, but ϕ is false. The conceptual presupposition for such a strategy seems to be that the *domain* of the theory S (i.e. the set of objects the theory is supposed to talk about) and it's *basic concepts* are free to be *reinterpreted*. Roughly this means that, although one may have an *intended interpretation* in mind when setting up the axioms, this intended interpretation is no longer privileged among other interpretations that might satisfy the axioms. Intuitions about an intended interpretation of some given discourse have only heuristic value in setting up the axioms and drawing attention to possibly fruitful applications, but they are irrelevant as far as the *logical content* of the thereby established axiomatic theory is concerned.

To get a feeling for Hilbert's method, let us look at the following example from Hilbert's *Festschrift*: Here, Hilbert wants to show that the *axiom of completeness* is independent from the rest of the axioms for Euclidean geometry. The *axiom of completeness* is a *maximal axiom* and (roughly) states that the system of things the theory talks about (i.e.

[98] For a general discussion of the developement of model-theory see [7].

points, lines) cannot be extended while still satisfying the remaining axioms. (If added to the remaining axioms, the *axiom of completenss* therefore guarantees that the resulting system captures the "usual" structure of Euclidean space up to isomorphism.) To show that the axiom of completeness is independent along the lines indicated above, Hilbert reinterprets the primitive concepts of Euclidean geometry as follows:

1. "a is a point" is reinterpreted by "a is a pair (x, y) of *algebraic numbers*, i.e. numbers x and y that can be constructed by repeated applications of the four basic arithmetic operations together with the operation $|\sqrt{1+\xi^2}|$ from the number 1"
2. "b is a line" is reinterpreted by "b is the ratio $(u : v : w)$ of three such algebraic numbers"
3. "the point a is incident with the line b" is reinterpreted by "a is a pair (x, y), such that..., b is the ratio $(u : v : w)$... and $ux + vy + z = 0$"

It can be shown now that under this reinterpretation every axiom of Euclidean geometry, except the *axiom of completeness*, is satisfied. (Just add some non-algebraic number and make sure to "close off" under the four basic arithmetic operations and the operation $|\sqrt{1+\xi^2}|$.)

Before we can see more clearly what Frege's troubles with this kind of independence proofs was, let us first clearly state what is involved conceptually in independence proofs *à la* Hilbert.[99]

As far as Hilbert is concerned, an independence proof of an axiom ϕ from a set of axioms S is supposed to show that neither ϕ nor $\neg\phi$ are *provable* from S, i.e. that no sequence of logical inferences might bring us from S to ϕ or its negation.[100] This is done by doing two things: First by providing an interpretation I with respect to which all axioms in S as well as ϕ come out true, and second by providing another interpretation J with respect to which all axioms in S are true while ϕ is false.[101] Here an interpretation is specified by a set of objects D forming the domain of objects of the theory, together

[99] To keep things straight I will read Hilbert's method in a more or less straighforward, modern way and ignore certain deviations that seem to me irrelevant for the further discussion.

[100] This can be seen in various places, for instance [22], p. 24 and 26. or in his *Über den Zahlbegriff*, [22], p. 242.

[101] Note that Hilbert in his *Grundlagen der Geometrie* had already shown the "first half" of an independence proof for the *axiom of completenss*. This was done by showing that the axioms of Euclidean geometry are

with a specification of the denotations of the primitive concepts over the given domain D. Now, appealing to the informal concept of semantic consequence, ϕ is a semantical consequence of S if and only if ϕ is true in every interpretation in which all the sentences in S are true. Therefore, I witnesses that $\neg\phi$ is not a semantic consequence from S and J that ϕ is no semantic consequence of S either. Summing up, on this straightforward model-theoretic reading of Hilbert's independence proofs, he is relying on

1. the informal notion of *provability*
2. the informal notion of an *interpretation*, and the notion of *truth with respect to an interpretation*
3. the relation of *semantic consequence*, defined in terms of all possible interpretations and
4. the *soundness* of the intuitive notion of proof with respect to the informal semantic consequence relation

Hilbert, however, never adresses any of these presuppositions explicitly (at least during his dispute with Frege and some time after) but instead takes them to be part of ordinary mathematics. On Hilbert's view of axioms as *conditions*, the independence of axioms can be proved (and to this extent Frege's reconstruction captures the essential point of his 1900-conception of the axiomatic method) just like *any* universal statement is refuted: namely by giving a *counterexample*. To prove that the sentence "Every continuous function is differentiable" is false, just give an *example* of a continuous function which is not differentiable. The same goes for the proposition "Every affine plane is desarguesian" which can be shown to be false by providing an example of an affine plane in which Desargues Theorem (or the *condition* corresponding to it) does not hold. The choice of this example is no coincidence: those were the kind of questions Hilbert wanted to adress[102], questions that have been bothering geometers throughout the 19th century and which he apprehends as pertaining to logico-methodological or "foundational" issues only in a derivative sense. Hilbert's mathematical viewpoint is echoed in his remark:

---

consistent by exhibiting an analytic interpretation in which all axioms (including the *axiom of completeness*) are satisfied.

[102]See [39], [47], [46] and [18] for more on the mathematical background of the Frege-Hilbert dispute.

> So spielt denn in der neueren Mathematik die Frage nach der Unmöglichkeit gewisser Lösungen oder Aufgaben eine hervorragende Rolle, und das Bestreben, eine Frage solcher Art zu beantworten, war oftmals der Anlaß zur Entdeckung neuer und fruchtbarer Forschungsgebiete. Wir erinnern nur an Abels Beweis für die Unmöglichkeit der Auflösung der Gleichungen fünften Grades durch Wurzelziehen, ferner an die Erkenntnis der Unbeweisbarkeit des Parallelenaxioms und an Hermites und Lindemanns Sätze von der Unmöglichkeit, die Zahlen e und π auf algebraischem Wege zu konstruieren. ([22] p. 111)

I think the problems accompanying the problem of the independence of the axiom of parallels in this paragraph make it reasonably clear that Hilbert thinks of independence of axioms in a straighforwardly informal, mathematical way. The fact that Hilbert mentions the independence of the axiom of parallels in one breath with the transcendentality of e and π clearly suggests this reading and shows his inawareness of the conceptual presuppositions displayed by the items 1. - 4 above. Although Hilbert is very concious with respect to the metatheoretical character of his overall enterprise, he is less clear about what this might involve *exactly*. To make a long story short: Hilbert's approach at the time surrounding the appearance of his *Foundations of Geometry* seems to be that of a working mathematician, who is not too worried about the problem of making precise the basic ingredients of his methodology.

If Hilbert is viewed as a *revolutionary*, Frege, by contrast, can be considered a *conservative* with respect to the axiomatic method. Frege time and again charges Hilbert for his allegedly inappropriate use of the word "axiom". For Frege, an axiom "in the Euclidean sense" (a locution he uses over and over again) is a *true proposition* which cannot be proved. That is, on Frege's view, a proper axiom has a determinate content and says something about a *specific* domain. Therefore an axiom can *by fiat* not shown to be *false*. The notion of *being false in an interpretation* on the other hand, to which Hilbert alludes to, has to be construed quite differently, for, according to Frege, a proper language leaves no room for interpretation.[103] In Frege's eyes Hilbert's geometrical axioms are at best *Pseudo*-propositions, i.e. groups of signs that seemingly express particular thoughts but

[103]For more on Frege's and Hilbert's conceptions of language see [1], [7] and [18].

do so only *apparently*, because the concepts that occur in them (like "point", "straight line", "congruence" etc.) do not designate something specific. But for Frege not to designate something specific is not to designate at all. Frege consequently takes Hilbert to conceive of the primitive concepts of his axiomatization of geometry as *variables in disguise* and whenever a primitive concept is "reinterpreted", what's really going on is that, according to Frege, a variable is *instantiated* by a meaningful concept ([28] p. 81). Let P for instance stand for the axiom of parallels and Φ for the remaining axioms of Hilbert's axiomatization of Euclidean geometry. Following Frege's reconstruction of what he takes Hilbert to have in mind, we arrive at propositional functions $P(X, Y, ...)$ and $\Phi(X, Y, ...)$ corresponding to P and Φ respectively by substituting variables of the appropriate type for the primitive concepts "point", "straight line", etc. What Frege takes Hilbert to have proved then is that the universally quantified conditional

(A) $\forall X \forall Y ...(\Phi(X, Y, ...) \rightarrow P(X, Y, ...))$

is not valid by constructing a *counterexample*, i.e. a sequence of concepts $P', G', ...$ which yield a true antecedent and a false consequent when substituted respectively for the variables $X, Y, ...$ Hence, on Frege's recommended reading of Hilbert' method of *reinterpretion*, "axioms" should explicitly be conceived as *conditions* expressed by formulas containing free variables, yielding true or false propositions only when *meaningful* concepts are substituted for these variables.[104]

Summing up, it can be said that neither was Frege stubborn in his critique of Hilbert nor did he misunderstand what Hilbert was up to on a large scale: it's just that he did not have the same view of axioms and mathematical truth and — for reasons that hopefully will become apparent — could not agree with Hilbert's method (even in it's reconstructed form) as an adequate method to prove the independence of *genuine axioms*.

---

[104] As an aside it should be mentioned that this reconstruction of the axiomatic method (axioms as propositional functions; theorems as consequens-parts of conditionals etc.) was extremely common at least until the thirties. Much of the work on the axiomatic method done by Carnap in the late 20s for instance can be seen as a further development of Frege's principal reconstruction of Hilbert, incorporating even questions like that of *categoricity* or *completeness* into this framework. (See in particular his [5]). Even Tarski, one of the founders of modern logic, in his (admittedly *popular*) *Introduction to Mathematical Logic* proposed a reconstruction of axiomatics that is nearly identical to Frege's reading of Hilbert. See chapter VI of his [42].

Each of these points is fairly well known, but do not yet constitute a substantial critique of Hilbert. The only thing we have seen so far is how Frege reconstructs what he thinks Hilbert has shown and that he gives some recommendations concerning – what he thinks is – the proper usage of the word "axiom".[105] Frege is claiming – correctly of course – that Hilbert does not use the word "axiom" in the traditional but in a novel sense, one which lacks in his opinion the clarity he wishes such a basic concept to have. But so far no argument has been given why Hilbert's methodology should be so fundamentally flawed as Frege wants us to believe.[106] So if these were the only criticisms of Hilbert, Frege's 1906-paper should not have gotten the attention it, in my opinion, deserves.[107] To see what Frege's real troubles with Hilbert's methods were, recall that Frege was looking at Hilbert's *Festschrift* and it's methodology with the eye of the *logician* (or at least *one* eye). One has to remember that, at the time of the appearance of Hilbert's *Festschrift*, Frege had — unbeknownst to logicians, mathematicians as well as philosophers — already revolutionized logic and developed a sophisticated system of conceptual innovations concerning the basic notions of *proof* and *inference*. As I will try to show, Hilbert's omissions mentioned two pages ago are the target of a substantial critique of Hilbert and Frege's own proposal how independence *should* be proved in the case of genuine axioms will reveal how to circumvent these problems while sustaining a traditional view of axiomatics.

But before we go on to look more closely on Frege's suggested method, let me once again dwell on one point: for Frege, but *not* for Hilbert there is a blatant difference between *genuine axioms*, having a determinate meaning on the one hand, and — what Frege calls — *pseudo axioms* on the other hand. According to Frege, *pseudo axioms* are strings of signs which *seemingly* express particular thoughts but nevertheless contain variables and must therefore be conceived as *conditions*. In Hilbert's conceptual repertoire

---

[105]See for instance [8], [26], [35], [3] and [4].

[106]That Frege would have attacked a strawman in the case of the word "definition" seems to be even more obvious: Frege of course did not seriously believe that Hilbert wanted to "define" the concepts "point" etc. in the sense of *explicitly define*. Rather he could not accept his wide usage of the word "definition" just as he could not accept his usage of the word "axiom".

[107]Of course, even if these were in fact the only points Frege wants to call attention to, the paper would still provide an important contribution in evaluating the historical triumph of Hilbert's algebraic version of the axiomatic method. For, if nothing else, Frege at least clearly apprehends Hilbert's radical shift towards a new conception of axiomatics.

on the other hand, *genuine axioms* in the Fregean sense no longer occur: they are *replaced* by *pseudo axioms*. Axioms for Euclidean geometry or the natural numbers are now on a par with "axioms" for groups, lattices or topological spaces. As a quick look in a modern textbook shows, the mathematical community was glad in following Hilbert in this shift in mathematical nomenclature. I will not discuss the widely ramified consequences of this shift for the philosophy of mathematics (for clearly it is not just a *shift in terminology*), but I simply want to point to the fact that for Frege the difference exists and is important. As he puts it:

> It must be noted that Mr. Hilbert's independence proofs simply are not about real axioms, the axioms in the Euclidean sense; [...] Instead, Mr. Hilbert appears to transfer the independence putatively proved of this pseudo-axioms to the axioms proper, and that without more ado, because he simply fails to notice the difference between them. This would seem to constitute a considerable fallacy. ([28] p. 102)

From Frege's point of view Hilbert's axiom system can be conceived of in *both* ways, as a set of conditions *or* as expressing "axioms in the Euclidean sense". Both views have their merits (of course Frege would not recommend using the word "axiom" in the former case), and the questions of independence between genuine axioms and axioms in the algebraic sense are mutually connected — but they do not amount to the same thing according to Frege.[108] It is apparent from many of Frege's writings, that he had a clear picture of "algebraic" axiom systems. In a letter to Jourdain for instance, Frege writes:

> Wenn man z.B. Untersuchen will, was aus den Gesetzen $a+(b+c)=(a+b)+c$ und $(a+b)+c=(a+c)+b$ folgt ganz unabhängig von der gewöhnlichen Bedeutung des Additionszeichens, so sollte man das Wort "Addition" und ebenso das Zeichen "+" ganz vermeiden und die Gesetze so ausdrücken: $f(a,f(b,c))=$

[108] As we shall see, Frege thinks that in proving that the axiom of parallels is independent from the rest of the Euclidean axioms we have to consider judgements *about* inferences and truth. Frege's reconstruction of Hilbert on the other hand shows that this is not the case if independence is meant as Frege thinks Hilbert conceives of it, for on Frege's reconstruction of Hilbert, independence of the axiom of parallels from the other axioms is expressed by the sentence $\exists X \exists Y ...(G(X,Y,...) \wedge \neg P(X,Y,...))$, which does not mention inferences or truth at all.

> $f(f(a,b)c)$ und $f(f(a,b)c) = f(f(a,c),b)$. Der Buchstabe "f" dient nun dazu, die Betrachtung allgemein zu machen. ([12] p. 117)

This is of course just the view he advocates how Hilbert's "axioms" for geometry should be seen. The same view is indicated in a letter to Huntington, a well known protagonist of the so called *postulate theorists*, which were known for laying down axioms for large parts of mathematics.[109]

It is necessary to dwell on the distinction genuine/pseudo axioms, because even careful writers on this issue sometimes mix things up.[110] Keeping in mind this distinction is important not just for a faithful interpretation of Frege, but necessary in evaluating (and appreciating) his positive account on the problem of how to prove the independence of genuine axioms.

3.2 Preliminaries to Frege's "New Science"

Before looking closer at Frege's own proposal how independence proofs should be handled, it might be useful to look at some points surrounding this issue.

We have seen that one of the main targets of Frege's criticism of Hilbert is the fact that

---

[109] See his letter to Huntington in [12], p. 90.

[110] To support his critique of Tom Ricketts, Jamie Tappenden for instance cites the following remark from Frege's *Basic Laws Pt. II* ([11], p. 534), in which Frege takes back what he had claimed earlier concerning the independence of the clauses in his definition of a "Positivalklasse" in the main text:

> It should not necessarily have been stated that the independence of the stated conditions from one another could not be proven. It is of course conceivable that one could find classes of relations, to which every condition would apply but one, and that every condition would fail in one of the examples.

Tappenden comments on the passage as follows:

> Here, Frege indicates that the independence (in the sense of 'no derivation possible') can be demonstrated by producing a counter-interpretation. ([39] p. 216)

Keeping the distinction just explicated in mind, it seems clear to me that in this passage Frege is *not* talking about what he later calls *axioms in the Euclidean sense*, but about *conditions* ("Bestimmungen") in the sense just elucidated. The passage is about the clauses of his *definition* of a "Positivalklasse" and it seems to me that, according to Frege, the sense in which *clauses in a definition* can be said to be mutually *independent* must be construed quite differently from independence of genuine axioms.

according to Frege Hilbert's independence arguments lack the kind of stringency which he expects from an argument to count as a *proof*:

> As it stands, we remain completely in the dark as to what he [Hilbert] really believes he has proved and which logical and extralogical laws and expedients he needs for this. ([28] pp. 111-112)

In particular Hilbert's loose talk about interpretations, the appeal to the informal semantical consequence relation and the assumption of informal soundness (with respect to informal provability) seem to be what Frege has in mind here. It is obvious, at least as far as the 1906-paper is concerned, that Frege does not find anything particularly wrong with the *mathematical content* of Hilbert's arguments but rather with their *presentation* and the *language* in which they are stated (specifically the "interpretation-talk"). This is not only indicated by his remark that "the question may still be raised whether, taking Hilbert's result as a starting point, we might not arrive at a proof of independence of the real axioms" ([28] p. 103), but it is obvious from the idea lying behind his own proposal. That is, Frege is not so much worried about the *truth* of the independence results established by Hilbert, but with the *means* of establishing them. In particular, Frege, at least in the 1906-paper, does not show any qualifications that proofs of independence are somehow *impossible*.[111] What is really at stake here is the *form* that independence arguments concerning *real axioms* should take if they should count as genuine *proofs*. So in order to assess Frege's own proposal, something should be said about Frege's notion of *proof*.

It is a commonplace that the main aim of Frege's invention of the *Begriffsschrift* was the rigorization of the concept of *mathematical proof*. But it is a matter of controversy as to what exactly Frege was up to, if his motivation was driven mainly by *mathematical*

---

[111]This seems to be in contrast with some of his remarks dating before and after the 1906-proposal. In a letter to Liebmann from July 1900, he writes for instance: "Ich habe Gründe zu glauben, dass die gegenseitige Unabhängigkeit der *euklidischen* Axiome [emphasis by Frege] nicht bewiesen werden kann." ([12] p. 148) Another famous quote can be found in Frege's comments on Jourdain, where he explicitly states: "Die Unbeweisbarkeit des Parallelenaxioms kann nicht bewiesen werden." ([12] p. 119) I will not try to accomodate these remarks with Frege's 1906-proposal in this paper (which I think *could* be done). Instead I will focus on Frege's positive account of independence proofs in the 1906-article.

or by *philosophical* interest. What Jamie Tappenden has called "the myth" is the view that Frege's foundational project was a natural continuation of the rigorization of analysis starting with Cauchy and Weierstraß in the early nineteenth century. Just like Cauchy and Weierstraß were interested in laying solid foundations for real analysis, Frege, according to the "myth", was interested in laying solid foundations for number theory. Philip Kitcher on the other hand has argued that, unlike in the case of analysis, Frege with his attempt to secure the foundations of number theory did not answer any pressing needs of the mathematicans. The introduction of exact notions of limit, continuity, differentiability etc. by Cauchy and Weierstraß in the case of analysis was driven by the simple need of *consistency*. Intuivive analysis plainly lead to *contradictions*. The consistency of number theory on the other hand was never seriously in question. Kitcher concludes that Frege

> advanced an explicitly philosophical call for rigor [...] Instead of continuing a line of foundational research, Frege contended for a new program of rigor at a time when the chain of difficulties that had motivated the nineteenth century tradition had, temporarily, come to an end. ([27] p. 268)

Whatever the motives behind Frege's project were exactly, it seems to me that the Kitcher-view is right in emphasizing what Frege himself did not get tired to emphasize: that a central motive for introducing a *Begriffsschrift* was to delineate with absolute rigour the philosophically motivated border between what can be known *a priori* and what can be known only by appeal to *intuition*, between what is *analytic* and what is *synthetic*.

Important for our purposes is that whatever Frege's *goal* for inventing the Begriffsschrift was ultimately (besides the obvious one just mentioned), the *means* for establishing this goal was a rigorization of what is involved in mathematical proof. Frege repeatedly criticizes that all too often the mathematician is content when every step in a proof is "obvious" ("einleuchtend", [11] p. VIII), without checking what the source of this obviousness is. Frege's answer to this problem is well known: it was to devise a notion of proof that is defined in *purely syntactical terms*, so that every form of inference has its syntactical equivalent "on the paper". This enables one not just to check a proof by mechanical procedures, but also to evaluate on what premises a proof ultimately rests upon, for what was previously only in the thinking mind of the mathematician and only half-way expressed, becomes intersubjectively assessable. This is a commonplace, but a

commonplace worth recapitulating in the context of the Frege-Hilbert controversy, which is after all (at least in part) a controversy about *proofs of a particular kind*.

A second point has to be kept in mind: Frege criticizes Hilbert for taking *pseudo-propositions* (i.e. *conditions*) as axioms, that is, strings of signs which only apparently express thoughts, but which contain variables and consequently do not express thoughts at all. Frege concludes that such conditions cannot serve as axioms if the word "axiom" is taken in it's traditional sense. But Frege goes one step further: he is claiming that an axiom *is* a thought, i.e. it's not just that for a string of signs to be an axiom-candidate it is necessary to *express* a thought, but it is necessary for something to be an axiom-candidate even to *be* a thought. That is, he is claiming that in considering dependence or independence of genuine axioms we are concerned with thoughts as the *objects* of investigation. For Frege sentences are just the audible or visible *expression* of what is relevant to the question of (in-)dependence, and as such they only have physical properties. As Frege puts it:

> When one uses the phrase 'prove a proposition' in mathematics, then by the word 'proposition' we clearly mean not a sequence of words or a group of signs, but a thought: something of which one can say that it is true. [...] We have to distinguish between the external, audible or visible which is supposed to express a thought, and the thought itself. [...] no one wants to predicate this independence of what is audible or visible. ([28] p. 101)

This is an important point for Frege which will occupy us later on.

One last word on the mathematical-vs.-philosophical-reading of the motivations lying behind Frege's logicist project, which has been much investigated lately by Wilson, Tappenden or Hallett in their [47], [46], [39], [40], [41] and [18] with respect to Frege's *mathematical* background. I think it's of considerable importance to place Frege firmly in the context of his mathematical environment. Let me mention just one example Mark Wilson elaborates in some detail in two of his papers.[112] It is well known that in the course of his logicist project of deriving the basic truths of arithmetic from logic, Frege repeatedly considered so called *abstraction principles*. Abstraction principles are first discussed in

[112]See his [47] and [46] for more on "extension elements", as well as Halletts [18].

his celebrated *Foundations of Arithmetic* where he contemplates on defining the natural numbers by means of the following principle (nowadays called *Hume's Principle*)[113]:

(HP) $Nu(F) = Nu(G) \leftrightarrow F \approx G$

Here F and G are concept variables, the operator Nu stands for the number operator applying to concepts and the relation $\approx$ stands for the relation of *equinumerosity* between concepts which is straightforwardly definable in higher-order logic. HP therefore says that the number of F's is equal to the number of G's if and only if the F's and G's are equinumerous. Put slightly differently: HP states that concepts are counted as "equal" if and only if they are equivalent *modulo* equinumerosity, so there is no difference between the concepts "is a county of Austria" and "is a planet of the solar system" *with respect to cardinality*; all the differences are "abstracted away". Hence, if HP could be justified, the introduction of numbers as *objects* would be justified as well, for a term $Nu(F)$ could then be seen as a name of an abstract object "generated" by abstraction.

Now it is well known that Frege was not satisfied with HP alone as *defining* the number operator (and hence numbers as objects) for a peculiar problem now kwown as the "Julius Cesar problem".[114] What is important for us is the *context* surrounding the discussion of HP in the *Foundations of Arithmetic*. Frege spent a lot of effort in discussing a similar abstraction principle which lies at the heart of *projective geometry*, viz. the principle

(DP) $d(h) = d(g) \leftrightarrow h||g$

where g, h are variables for straight lines, d stands for the direction-operator and $||$ for the equivalence relation of parallelism between straight lines. DP therefore says that two lines have the same direction if and only if they are parallel. This abstraction principle was considered to be central for the foundations of projective geometry, for if it would fix the reference of the direction operator — which, according to Frege, it does *not* for the very same reason HP does not fix the reference of the number operator — it would once and for all resolve the then salient issue of the so called *points at infinity* (the "points"

---

[113] [13], p. 74.

[114] However, a justification *could* be given with the help of a "super-abstraction principle" encompassing all other abstraction principles, viz. *Basic law V*. With the aid of basic law V *any* abstraction operator O with respect to some equivalence relation R can then be defined by $O(a) := \{b : aRb\}$.

where parallel lines "meet"). Just like "number abstracts" $Nu(F)$ in the case of HP, "direction abstracts" $d(h)$ could be seen as names of certain abstract *objects* that are generated by identifying parallel straight lines. That is, *via* DP, talk about *points at infinity* could be shown to be reducible to talk about straight lines, for *points at infinity* could simply be *identified* with directions.

The point of this little digression is to make it vivid that Frege's ideas concerning his logicist program[115] must be seen in the light of the broader mathematical context, in particular with respect to developments concerning the foundations of projective geometry, which occupied the center stage of 19th century geometry. Even as early as 1972 in his doctoral dissertation Frege wrote about the problem of the points at infinity ([14] p. 1), and the problem of methodologically correct foundations of projective geometry occupied him throughout his career as a mathematician.[116]

So I think it's extremely important to look closely at Frege's mathematical environment to get a firm grasp of his thinking. Still, the relevance of such contextualization seems to be limited in the particular case of an evaluation of his attitude towards independence proofs and metatheoretic questions in general. The fifty-odd pages of his papers on the foundations of geometry for instance, as well as Frege's correspondance with Hilbert and others on the issue, make it apparent that Frege has no "mathematical" complaints about Hilbert's *Festschrift* in the sense that an appeal to a shared mathematical background could make the differences look less serious. Frege in fact rejects a whole lot of what (then) contemporary "mathematical practice" consists in. One has to take only a quick look at his *Basic Laws Pt. II* to get a feeling for his opinion on contemporary mathematicians. It is a continual complaint throughout his career that mathematicians are all too often too lazy about the foundations of their respective fields. So although Frege must be seen as a child of 19th century mathematics and has to be placed within it's tradition, he could arguably considered one of it's greatest foes too. For Frege "mathematical practice" is

---

[115]Recall that although insufficient to fix the reference of the number operator, HP still plays a major role in Frege's subsequent definition of the natural numbers.

[116]That Frege saw the methodologically correct foundation of projective geometry as an unsettled issue can be seen from his correspondence with Moritz Pasch, in particular Pasch's letter from the 17th of january 1905. (See [12] p. 173). Although Frege's part of the correspondence got lost, his main line of reasoning can be extracted, to some extent, from Pasch's letters.

not *sacrosanct*. On the contrary, according to Frege, 19th century mathematics was in a terrible shape, for it's core discipline *number theory* was built upon a rotten footing (mainly due to the bad influence of the "formal arithmeticians"). Again, when Frege says such things he is looking at things with the sharpened eye of the entirely new discipline he has re-invented, the discipline of *logic*.

3.3 The "New Science"

As I have emphasized earlier, Frege's concerns should not be seen as a full-scale rejection of Hilbert's arguments. At the time Frege wrote his paper on Hilbert's *Foundations of Geometry*, the independence of the axiom of parallels from the remaining axioms for instance was as certain as something could be. Frege is neither questioning the independence itself nor the possibility of *proving* this fact. What is at issue here is rather what kind of methods should be employed to obtain an acceptable *proof* of this well known fact. As we have seen, Frege blames Hilbert's independence proofs for either being misguided (in case they are meant to apply to genuine axioms) or irrelevant (because the word "axiom" is understood in the sense of "condition" and hence a proof of independence concerning such conditions has no obvious bearing on the genuine axioms corresponding to them). So the question is: what are the correct means to reach those well known results applying to *genuine axioms* according to Frege's standards?

Three committments concerning independence proofs are apparent from his outline in the final part of his 1906-paper. The first has already been mentioned in the second section: the kind of things we are concerned with, when we ask ourselves if some axiom is independent of others, are *thoughts*. As Frege puts it:

> What I understand by independence in the realm of thoughts may be clear from the following. I use the word 'thought' instead of 'proposition', since surely it is only the thought-content that is relevant, and the former is always present in the case of real propositions – and it is only with these that we are here concerned. ([28] p. 103)

Again, in Frege's terminology "propositions" ("Sätze") are just "marks on the paper" or "soundwaves", and as such they cannot be said to be (in-)dependent just like tables or chairs cannot be said to be (in-)dependent. It is just in virtue of it's expressing a *thought*

that a proposition becomes logically relevant in the first place.[117]

The second point concerns the question how independence of thoughts is defined – roughly: is it defined *semantically* in terms of *truth* or is it defined *proof-theoretically* in terms of *inferences*? This is Frege's answer:

> Let Ω be a group of thoughts. Let a thought G follow from one or several thoughts of this group by means of a logical inference such that apart from the laws of logic, no proposition not belonging to Ω is used. Let us now form a new group of thoughts by adding the thought G to the group Ω. Call what we have just performed a logical step. Now if through a sequence of such steps, where every step takes the result of the preceding one as its basis, we can reach a group of thoughts that contains the thought A, then we call A dependent upon the group Ω. If this is not possible, then we call A independent of Ω. The latter will always occur if A is false. ([28] p. 104)[118]

The talk about *logical inferences* makes it reasonably clear that Frege's conception of independence is meant in the sense of "Non-provability". It must be mentioned, however, that the concept of *being provable* employed here – that is, as applied to *thoughts* – must not be confused with the relation of *being derivable*, which applies to *sentences*. (More on this point later.)

The third point is related to an interesting general feature of Frege's strategy and reveals an important aspect every attempted interpretation of his stance towards independence proofs – and in particular every attempted interpretation of his 1906 paper – has to take into account. Let me quote the whole passage:

> We now return to our question: Is it possible to prove the independence of a real axiom from a group of real axioms? This leads to the further question: How can

---

[117] As far as the 1903 - 1906 papers are concerned, Frege never distinguishes clearly between expression-*types* and expression-*tokens*. Throughout the 1906-paper the word "proposition" ("Satz") is used in the meaning of sentence-*token*. (See for instance [28] p. 101.) Frege, however, *was* aware of the distinction, as a letter to Dingler shows. See [12] p. 35.

[118] Recall that for Frege axioms are *true thoughts*. Hence, if ϕ is an axiom, $\neg\phi$ will *trivially* be non-provable from some given set of axioms Φ. More generally, *no* false sentence (or "thought") can be provable from Φ, assuming that the notion of proof is *sound*.

> one prove the independence of a thought from a group of thoughts? First of all, it may be noted that with this question we enter into a realm that is otherwise foreign to mathematics. For although like all other disciplines mathematics, too, is carried out in thoughts, still, thoughts are otherwise not the object of its investigations. Even the independence of a thought from a group of thoughts is quite distinct from the relations otherwise investigated in mathematics. Now we may assume that this new realm has its own specific, basic truths which are as essential to the proofs constructed in it as the axioms of geometry are to the proofs of geometry, and that we also need these basic truths especially to prove the independence of a thought from a group of thoughts. ([28] p. 106)

Frege's fortrightness about his suggested method of proving independence in this passage is striking: it reveals that Frege wants to establish independence of real axioms in an *axiomatic framework*, i.e. by invoking *basic truths about thoughts*. As we shall see shortly the main reason for bringing up such basic truths is to provide links between the notions of *provability* and *truth* as needed in independence proofs. As I mentioned earlier, one byproduct of Frege's logicist project was the rigorization of the notion of mathematical proof, which was needed in order to trace back on which basic truths the truths of arithmetic ultimately rest upon. So if non-provability of some axiom of others should be capable of being proved, we have to ask the same questions for proofs of this kind too: what are the basic truths about *thoughts*, *truth* and *provability*, which are needed in order to prove non-provability of some genuine axiom from others?

In his outline Frege offers three axioms of what he calls the *new science*: let me quote the whole passage again:

> The basic truths of our new discipline which we need here will be expressed in sentences of the form
>
> If such and such is the case, then the thought G does not follow by a logical inference from the thoughts A, B, C.
>
> Instead of this, we may also employ the form:
>
> If the thought G follows from the thoughts A, B, C by a logical inference, then such and such is the case.
>
> In fact, laws like the following may be laid down:

If the thought G follows from the thoughts A, B, C by a logical inference, then G is true.

Further,

If the thought G follows from the thoughts A, B, C by a logical inference, then each of the thoughts A, B, C is true. ([28] p. 107)

(In what follows I will refer to the laws mentioned by Frege with NS_1 and NS_1' respectively.) As mentioned earlier, one thing that is missing in Hilbert's exposition of his independence arguments, even if his talk about interpretations, systems of things etc. were reformulated in an acceptable way, is a link between the notions of *truth with respect to an interpretation* and *provability*. One presupposition of Hilbert's arguments, if they are supposed to show independence in the sense of *non-provability*, is for instance that the notion of proof employed therein is *sound* with respect to the notion of semantic consequence, i.e. that everything provable from premises that are true in some particular interpretation, is itself true in that interpretation. Both "basic laws" above are meant to provide just this kind of missing link Hilbert would have needed as premises of fully regimented proofs even if the informal notions of *proof* and *semantic consequence* would have been spelled out.

Of course, the two laws Frege cites are clearly insufficient for independence proofs. What then, is needed additionally in order to be able to carry out "gapless proofs" of independence that will meet Frege's *Begriffsschrift*-standard of proof? To see what Frege has in mind, it should be recalled that Frege is clear about the fact that *mathematically speaking* we do the same thing in proving independence of *genuine axioms* as in the case of *conditions*, namely producing a *counter-interpretation*. It's just that, due to Frege's views on language, the same mathematical idea must be *implemented* differently in the case of genuine axioms.

To elucidate what he has in mind as a surrogate for the counter-interpretation method, he wants us to conceive of a language, a whole consisting of *meaningful expressions* (that is, expressions equipped with a fixed *sense* as well as a *reference*).[119] Think of the language of Euclidean geometry as an example of such a language. As Frege acknowledges,

[119]The relevant passage can be found in [28] pp. 107-110

one might follow Hilbert in expressing the axioms of Euclidean geometry by means of the primitive concepts "point", "straight line", "plane", "congruence" and "between". Besides these primitive concepts the language includes the logical apparatus consisting of variables, truth functional connectives and quantifiers. Frege then invites us to think of the expressions of this language as forming a list of more and more complex expressions built up from the primitive concepts by means of quantifiers and the truth functional connectives.

Now Frege's suggested new law amounts to the following: if we can couple each expression of such a list with an expression of another list, consisting of expressions of the same language such that

1. every expression is coupled with an expression of the same *grammatical category* and
2. the *logical* expressions are coupled with themselves

then every valid proof containing only expressions of the first list can be converted into a valid proof containing only expressions from the second list. To illustrate this simple idea, consider the following sequence:

1 Every human is mortal
2 Sokrates is a human
3 If Sokrates is a human then Sokrates is mortal
4 Sokrates is mortal

This sequence of sentences constitutes a valid proof of the thought that Sokrates is mortal from the premises expressed by the sentences 1 and 2. Now by replacing "human" with "prime number greater than 2", "mortal" with "odd" and "Sokrates" with "5" we get to the sequence

1' Every prime number greater than 2 is odd
2' 5 is a prime number greater than 2
3' If 5 is a prime number greater than 2 then 5 is odd
4' 5 is odd

which is again a correct proof of the thought expressed by the last sentence from the premises expressed by the first two sentences. Now this is indeed the simple idea lying

behind Frege's surrogate law that is needed for independence proofs: that a valid proof remains valid if non-logical terms are replaced by other non-logical terms, as long as the truth of the premises is not affected.

Note that Frege's talk about "lists of expressions" and "coupling expressions with expressions" can be straightforwardly restated in terms of a function t, "translating" expressions of the first list into expressions of the second. Thus, slightly generalizing on his key idea, Frege's point can be made more precise by defining a Frege-translation (henceforth F-translation) as a function t from a language L_1 to a language L_2 (possibly distinct from the first one[120]) which meets the following conditions:

1. For every primitive predicate P in L_1 there is some L_2-predicate ϕ_P, s.t. $t(P(x_1, ...x_n)) = \phi_P(x_1, ...x_n)$
2. $t(s_1 = s_2) = t(s_1) = t(s_2)$ for terms s_1, s_2
3. $t(\neg\phi) = \neg t(\phi)$ for every formula ϕ
4. $t(\phi \to \psi) = t(\phi) \to t(\psi)$ for formulas ϕ, ψ
5. $t(\forall x\phi) = \forall x t(\phi)$ for all formulas ϕ[121]

The clauses in this definition are straightforward: 3 for instance states that the F-translation of a negated sentence is the result of prefixing the negation sign to the translation of the original sentence. Similar for the other logical constants.

Now given some F-translation t, mapping expressions to expressions and thereby preserving their logical structure, we can correlate a function s_t with t, mapping *senses* of expressions to *senses* of expressions according to the F-translation t. Call such a function a *sense-translation* (S-translation).

With these stipulations at hand (and employing the notation ϕ^*, referring to the

---

[120] In the relevant passage Frege writes: "Imagine a vocabulary: not, however, one in which words of one language are opposed to ones of another, but where on both sides there stand words from the same language but having different senses." I am not entirely certain as to why Frege is here talking about *one* language, but as my main aim in the following is the broader one of highlighting links to certain concepts and methods of modern mathematical logic, I will simply skip this and other problems, at least if Frege's main line of reasoning is not distorted.

[121] Similarly for higher order quantifiers.

thought expressed by ϕ), Frege's new law, which he calls an "efflux of the formal nature of the logical laws" ([28] p. 107), can be stated as follows[122]:

> If t is an F-translation then (if ϕ^* is provable from the thoughts S^*, then $s_t(\phi^*)$ is provable from $s_t(S^*)$)

Now suppose we want to prove that the thought ϕ^* is independent from the thought S^*, expressing the conjunction of some set of sentences S. Suppose further that ϕ and S are formulated in a language L_1 and we are given an F-translation t mapping the primitive concepts of L_1 to concepts of some language L_2. As we have seen, t will induce a function s_t mapping ϕ^* and S^* to thoughts $s_t(\phi^*)$ and $s_t(S^*)$ respectively. Furthermore, suppose we could prove that $s_t(S^*)$ is true and $s_t(\phi^*)$ is false. Now if ϕ^* were provable from S^*, then, by the new law, $s_t(\phi^*)$ were provable from $s_t(S^*)$ and hence *true* by the soundness laws NS_1 and NS_1', which it, by assumption, is *not*. Contradiction. Hence ϕ^* is *not* provable from S^*.

A lot of things a left open by this sketch at this point. An often-read complaint about Frege's suggested method is that he had no account of "what belongs to logic". As we have seen, Frege's method relies on two things: First of all, *dependence* of (and hence *in*dependence of) axioms was defined in terms of *logical steps*, which themselves were defined via *logical inferences* and *laws*. So in order for his method to get off the ground, as he explicitly acknowledges, Frege would have had to delineate what counts as a logical inference and what the logical laws are. Secondly, his new basic law relies on the notion of an F-translation, which in turn was defined as a function on the vocabulary of a language which leaves *logical constants* fixed. So Frege should have been interested in delineating more precisely 1. what the logical constants are and 2. what counts as logical law/inference, for otherwise one would be left in the dark if his independence test yields correct results.[123]

---

[122]The "law of the efflux of the formal nature" (and Frege's suggested method to prove independence) has lately been discussed in it's relation to other informal metatheoretical principles such as *duality principles* in projective geometry, in particular by Jamie Tappenden in his [39]. In what follows I will reconstruct what I take to be Frege's central idea concerning the new basic law in a way that is closer to concepts and ideas that have become central in 20th century (mathematical and philosophical) logic.

[123]This point has been put forword especially by Tom Ricketts in his [37] pp. 149-150.

Although this question is important, I do not want to discuss it in any further detail and refer the reader to some of the relevant literature.[124] (In fact, in the definition of an F-translation given earlier it has been assumed that the logical constants are exhausted by the identity-sign, the truth-functional connectives and universal quantification.) The points I find most interesting in Frege's proposal are of a more general methodological character, so it seems to me that Frege's suggestion bears a lot of interesting questions even *modulo* the problem of "what belongs to logic properly".[125] Some of them shall be discussed in more detail in the following chapter.

3.4 Frege's New Science *explicit*

To get a better grip on what Frege is up to with his proposal, let us be a little bit more careful in what follows. So for the following we will fix some system K of axioms and rules comprising a codification of the "logical". That is, we assume that it has been delineated what the logical constants are and what laws and rules of inference are purely logical. For

---

[124]See Ricketts [37], Tappendens [39] for a reply and especially Antonelli/May's [1] for more on the problem of the *logical constants*.

[125]To delineate "what belongs to logic" is important though, for if no criterion of demarcation is provided, we are left in the dark about the results generated by Frege's (*or anyone else's*) method of proving independence. By way of a thought experiment, let us assume that someone would say that second order number theory would be exactly what "belongs to logic". If this were true, every number-theoretic statement would be a logical truth and hence dependent on the logical laws; if it were false, number-theoretic statements would be independent of the logical laws. So the problem is that overgeneration with respect to dependence may occur if we "falsely" take some law or expression to be "logical" which it "in fact" is not (as in the dramatic case of basic law V which Frege takes to be a logical law, but which allows us to prove *everything*), as well as undergeneration if we "falsely" take some law or expression not to be logical if it "in fact" is. The problem here lies of course in making sense of the locution "what is *in fact* logical". It seems to me that there is no straightforward way to establish what's "logical" without recourse to stipulation at some point or recourse to *philosophical* arguments such as Frege's early "everything can be counted"-rhetoric in favour of the logicality of the concept of number. There simply seems to be no criterion of "the logical" which meets everyones expectations. Hence one should not expect *Frege* to have such a criterion of demarcation when no one has. Unlike most of his contemporaries, Frege at least adresses the issue explicitly and it seems that he would have resolved this problem by *stipulating* axiomatically what's logical. "If, following the suggestions given above, one wanted to apply this to the axioms of geometry, one would still need propositions that state, for example, that the concept *point*, the relation of a point's lying on a plane, etc. do not belong to logic. These propositions will probably have to be taken as axiomatic."

the sake of definiteness take K to be Frege's system of his *Begriffsschrift*.[126]

Now if we look at the first two axioms Frege mentions (NS_1 and NS_1') it is apparent that three devices occur in them that do not show up in the system K (or in any of the systems Frege ever considers as providing a basis for logic). First, he uses the locution "the thought ...", secondly, the predicate "... is true" and thirdly, the relation "... is inferable from - - -"[127]. So in order to get gapless independence proofs we have to take care of these notions first. For this purpose let $\ulcorner . \urcorner$ represent the 1-ary function mapping an expression to the *sense* it expresses. Furthermore let $T(x)$ stand for the *truth predicate* and $Prv(x, y)$ for the *provability relation* which is determined by the axioms and rules of K. So if ϕ and ψ are sentences, $Prv(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner)$ is just an object linguistic expression of the sentence "The thought expressed by ψ is provable from the thought expressed by ϕ by means of the rules and axioms of K".

Some qualifications concerning the proposed reconstruction of Frege's new science should be made before we go on: the first relates to the "basic law" NS_1', which is just an expression of his often articulated view that something can be *proved* only from premises that are true. The reason for this is that, according to Frege, a genuine *proof* should establish the truth of it's conclusion, which of course requires the premises from which it proceeds to be true. But this piece of Fregean doctrine should not prevent us from construing the notion of *provability* in a way that dismisses the notion of truth altogether. True: Frege conceives of the possibility of regarding *proving* as some kind of "game" without paying attention to the truth of the premises — and openly *rejects* it! But the sole reason for doing so is precisely because we would be left in the dark as to the *truth* of thereby established conclusions.[128] To keep things straight we will therefore understand $Prv(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner)$ in a sense that *excludes* the necessary truth of ϕ, i.e. we will

---

[126]It seems that Frege, some time after receiving the letter from Russell, went back to 1879 and regarded his *Begriffsschrift*-calculus, something akin to *simple type theory*, as codifying "what belongs to logic". This is indicated for instance by Carnaps lecture notes, dating from 1910 - 1914. See [6].

[127]This locution is to be understood in the sense of "... is inferable from - - - by a finite sequence of logical inferences".

[128]See for instance his letter to Dingler, [12] p. 30. Note also that there is a blatant difference between claiming that proving theorems *is comparable* to playing some kind of game and the claim that proving theorems is *nothing more* than a game. Frege would not object to the former, but of course reject the latter claim. For more on the game analogy compare § 90 of the of his *Basic Laws Pt. II*.

view the provability relation $Prv(x, y)$ as *solely* depending on the notion of proof determined by the synatctical rules and axioms of K. This is not a serious departure from Frege, for "Fregean provability" could still be defined explicitly via $Prv(x, y)$ and the truth-predicate by $Prv(x, y) \wedge T(x)$.

Note that the object-linguistic concept of provability (i.e. the relation $Prv(x, y)$) as construed here is, by assumption, *dependent* on – although not *identical* to – the notion of *derivation* determined by the syntactically defined laws and rules of the given system K. That Frege would agree with this reconstruction of the provability relation might be seen as somewhat troublesome. Patricia Blanchette for instance writes:

> For Frege, on the other hand, the question of whether a given thought τ is independent of a collection Π of thoughts is the question of whether τ can be obtained by Π by a finite number of valid steps of logical inference. And while such a series of steps is straightforwardly a proof, it is not a proof in any particular formal system. Frege's question is not whether a given formula is derivable from a set of formulas in e.g. the system of the Begriffsschrift or Grundgesetze. It's rather the question of whether a thought follows via logical steps of logical inference from a collection of thoughts, independently of whether there is available for our use a good codification in a formal system of the inferential steps involved. ([2] p. 13)

To be sure: it is true that for Frege inferences (logical or otherwise) are not made in the "realm of the visible", i.e. by syntactical transformations. Synactically defined transformation rules are just the *expression* of such inferences from the Fregean point of view. But it is hard to find an essential difference between the notion of *provability* (as applying to *thoughts*) and the purely syntactical notion of *derivation* if the formal character of the logical laws and inferences in the realm of *thoughts* is to be upheld. Once the logical laws and valid forms of inference are delineated and the question what belongs to logic is settled (which we assume), there should be no question of whether these forms of inference are *codifiable* by syntactically specified rules: they surely must be! It seems odd to me (or to attribute such a view to Frege) to conceive of logical laws or rules of inference which do not correspond to one or another syntactical rule "on the paper", for on such a view proofs would seem to become *uncommunicable*. Blanchette is surely right

in claiming that for Frege a proof is not a proof in a *particular* system, but it seems to me that he would agree that different systems codifying the same priorly delineated realm of the logical, and thereby generating the same class of theorems (*modulo* notational differences), should count as equal. Hence if, for systems K, K', we have $\phi \vdash_K \psi$ if and only if $\phi \vdash_{K'} \psi$ for every pair of sentences ϕ, ψ, both K and K' are equally adequate in fixing the extension of the relation $Prv(x, y)$.[129]

A second point concerns the choice of the *primitives* of the *new science*, in particular the need of a truth predicate. As I see Frege's axioms, those are meant to state something about *all* thoughts and therefore should be stated explicitly by the use of *quantifiers*. I have no knock-down argument for this reading, but it seems to be suggested by most of his remarks concerning generality. One of these being his view that the principles/axioms of some theory should be *finite*, thereby excluding *axiom schemes* altogether.[130] So if it is granted that the axioms of the *new science* should be read as *quantified* statements, there seems to be no way to circumvent the use of a truth predicate on pain of ungrammaticality.

Keeping in mind these points the first axiom of the *new science* can be expressed by:

$(NS_1) \quad \forall x, y(Sent(x) \wedge Sent(y) \rightarrow (Prv(\ulcorner x \urcorner, \ulcorner y \urcorner) \wedge T(\ulcorner x \urcorner) \rightarrow T(\ulcorner y \urcorner)))$

where $Sent$ is a predicate that defines the class of all sentences of the language of the

---

[129] Another question still, is the question if Frege's notion of *provability* is in fact *formal* in any sense that excludes semantical features, a question Blanchette would deny. On her interpretation, Frege's concept of proof is *semantically laden*. This is particularly important for independence proofs, for although syntactical *derivability* might yield a *positive* test for provability of the corresponding thoughts, the fact that a sentence ϕ is *not* derivable (regardless of how this might be shown) from a sentence ψ does not yield such a test for *non*-provability of the corresponding thoughts. The reason for this is that the non-derivability of ϕ from ψ may depend on the *particular* expression of the thoughts expressed by ϕ and ψ, resp. That is, there might be sentences ϕ', ψ' expressing the same thoughts as ϕ and ψ, but for which ϕ' *is* derivable from ψ'. For more on this topic see Blanchette's [3], [4] and [2].

[130] See his [13] p. 6. That is, what is excluded is to treat e.g. his soundness axiom as a scheme $T(\ulcorner \phi \urcorner) \wedge Prv(\ulcorner \phi \urcorner, \ulcorner \psi \urcorner) \rightarrow T(\ulcorner \psi \urcorner)$ which yields a particular instance only if particular sentences are put at the places of the schematic metavariables ϕ, ψ. In fact, if "soundness" is treated schematically, it could be formalized by $\phi \wedge Prv(\ulcorner \phi \urcorner, \ulcorner \psi \urcorner) \rightarrow \psi$. It is precisely with (objectual) *quantification* that the truth predicate becomes ineliminable. Still, there is the possibility of viewing his quantifiers *substitutional* rather than *referential*. Another one would be to treat his axioms as *schematic rules of inference*, but it seems to me that both options are not particularly appealing from a Fregean point of view.

theory to which the *new science* is to be applied.

Note that, if the truth predicate occurs ineliminably and turns out to be undefinable, truths about this ineliminable notion must show up as further axioms. One must, for instance, be able to prove that everything provable from some set of genuine axioms must be *provably true*. Say, P stands for a sentence expressing the *Pythagoras theorem*: if the *new science* is applied to geometry, we should for instance be able to prove $T(\ulcorner P \urcorner)$. Furthermore, compositional axioms concerning the truth predicate must be available, that is, truths like[131]

$Tr_{\neg} \qquad \forall \phi \in Sent : T(\ulcorner \neg \phi \urcorner) \leftrightarrow \neg T(\ulcorner \phi \urcorner))$

or

$Tr_{\wedge} \qquad \forall \phi, \psi \in Sent : T(\ulcorner \phi \rightarrow \psi \urcorner) \leftrightarrow (T(\ulcorner \phi \urcorner) \rightarrow T(\ulcorner \psi \urcorner))$

Now the statement of the second basic law Frege mentions is somewhat more involved and reveals an important presupposition for the Fregean approach altogether: Recall that it amounts to the claim that the provability-relation is invariant under substitutions of the non-logical vocabulary. More precisely it was stated in terms of F-translations, that is, functions mapping the expressions ("words" in Frege's terminology) of some language L_1 to expressions of some language L_2 preserving logical structure. Such an F-translation will induce a function mapping *senses* of expressions to *senses* of expressions. This is important for Frege, for as we have seen, Frege thinks that *thoughts* are the kind of things that can be proved to be independent from one another. It is important though to realize that, although S-translations (relating senses to senses) are what is relevant to independence proofs, these S-translations are *parasitic* on the *syntactically* defined F-translations. There seems to be no obvious way to come up with an S-translation without providing an F-translation which induces it. Frege seems to be fully aware of this: in his informal presentation of the new law, "words" (*syntactical* items) are paired with

[131] It is well known that without restricting it in some way, such a theory of truth will be *inconsistent* due to the *liar-paradox*. In fact, the liar paradox was (and still *is*) a major driving force for the development of *consistent* theories of truth. (See Part I of [17] for further discussion of axiomatic theories of truth). Unfortunately there seem to be no hints in Frege's writings how the liar paradox is to be avoided or that he even considered it as a major problem at all.

"words"; it's just in virtue of the fact that we are dealing with *proper languages* (that is, *fully interpreted languages*) that thereby a pairing of *senses* with *senses* is induced.

Now the point I want to stress is rather simple: what is needed in order to even *formulate* the new law NS_2 (in fact even the first law NS_1) is a *theory of syntax*, which enables us to talk about linguistic items such as *variables*, *names*, *expressions* etc. and which tells us how the expressions of some given theory to which the new science is to be applied, are built up from the basic concepts by means of the quantificational apparatus. By using expressions denoting syntactical objects, such a theory has to provide the means to formulate and *prove* such basic facts regarding syntax. If, for instance, "$Sent(x)$" is a predicate true of all and only the sentences of some given language L, "P" a primitive concept of L, "$\circ$" stands for the concatenation-operation of signs, and the bar is used as a means to generate names for syntactical items, a theory of the kind required by Frege's method must prove, e.g., sentences like

$Sent(\bar{P} \circ \bar{x} \circ \bar{\rightarrow} \circ \bar{P} \circ \bar{x})$

and

$\forall x(Sent(x) \rightarrow Sent(\bar{\neg} \circ x))$

Given such a theory of syntax, one could then define an F-translation t by means of clauses like

$T_\neg \quad \forall x(Sent(x) \rightarrow t(\bar{\neg} \circ x) = \bar{\neg} \circ t(x))$

$T_\rightarrow \quad \forall x \forall y(Sent(x) \wedge Sent(y) \rightarrow t(x \circ \bar{\rightarrow} \circ y) = t(x) \circ \bar{\rightarrow} \circ t(y))$

Note that a theory providing this kind of facts regarding syntax is quite strong. It is not only "substantial" in a loose sense of "being about something", but in the sharp sense that it has *ontological*, as well as *ideological* import: it has ontological import because it is committed to an ontology of expressions and second, as is well known, it contains (and *is contained in*, as Gödel showed) an *arithmetical* theory of some kind.[132] [133]

---

[132] This has been shown by Quine in his [32].

[133] Although at this place it might seem somewhat anachronistic to speak of *ontology*, this point seems to me important. It is not clear if, at the time of writing the last series of papers on the foundations of geometry, Frege still held the view that there are *logical objects*, i.e. objects whose existence is implied by the basic laws of *logic*. But if he did *not* (and this seems to be indeed the case), there seems to be no way to construe the *new science* as

Furthermore, due to Frege's committment to *thoughts* (senses of sentences) as objects of metatheoretical investigation, a final presentation of Frege's views on the matter would have to include in addition a theory relating *expressions* to the *senses* they express, and should make apparent how the sense of a complex expression is composed of simpler senses according to its syntactical structure. Using "•" for the concatenation of senses (whatever this exactly means), compositional truths of the following kind should be among the consequences of the *new science*:

$S_\neg \qquad \forall x(Sent(x) \to \ulcorner \bar{\neg} \circ x \urcorner = \ulcorner \bar{\neg} \urcorner \bullet \ulcorner x \urcorner))$

$S_\to \qquad \forall x \forall y(Sent(x) \wedge Sent(y) \to \ulcorner x \circ \bar{\to} \circ y \urcorner = \ulcorner x \urcorner \bullet \ulcorner \bar{\to} \urcorner \bullet \ulcorner y \urcorner)$

Plainly, given Frege's committment to *thoughts* as objects of independence investigations, a *theory of sense* has to be given, not just for the specific reason of providing the necessary links between the syntactical F-translations and their corresponding S-translations on the level of sense: it is a simple lesson to be learned from Tarski's painstaking accuracy in setting up his truth definitions, that every theory of truth (and/or provability) has to provide a theory of the *objects* that are deemed true (and/or provable).[134] Obviously such a theory of sense is no less "substantial" than the theory of *syntax* that is presupposed in Frege's proposal.

Now, suppose all this has been provided and the class of F-translations (from a language L_1 to L_2) has been defined, say by TF. The new law that is needed in order to prove independence along the lines Frege advocates, would then look something like this:

$(NS_2) \quad \forall t \in TF : \forall x, y \in Sent : (Prv(\ulcorner x \urcorner, \ulcorner y \urcorner) \to Prv(\ulcorner t(x) \urcorner, \ulcorner t(y) \urcorner)$

Now let's pause for a moment and take a look at Frege himself. The reconstruction so far has brought to light a bunch of collateral committments that come together with Frege's proposal. So one question would be: what are the committments that Frege is *aware* of?

---

a part of "logic" (as it is insinuated for instance by Tappenden in his [39] p. 215), for the new science as required by Frege clearly implies the existence of certain objects, viz. *expressions*. And, as we shall see, still *more*.

[134] In the context of axiomatic theories of truth, this point has been emphasized, for instance, by Halbach. For further discussion on this point see the second chapter of his [17].

The answer to this question is somewhat stuttering, but at least one point seems to be clear from his 1906-paper: that Frege is aware of the fact that *some* kind of *semantic ascent* is needed in order to prove rigorously the independence of genuine axioms. This can be seen for instance from the passage already quoted:

> although mathematics is carried out in thoughts, thoughts themselves are otherwise not the objects of its consideration. Even the independence of a thought from a group of thoughts is quite distinct from the relations otherwise investigated in mathematics.

The same point seems to be at issue in his remark at the end of the 1906 paper where he writes:

> As long as the word "axiom" was used as a heading only, a fluctuation in it's reference could be tolerated. Now, however, since the question of whether an axiom is independent of others has been raised, the word "axiom" has been introduced into the text itself and something is asserted or proved about what it is supposed to designate.

Frege's dwelling on a clarification of what an "axiom" is and what the proper means of establishing independence of axioms are, must be understood in the light of Frege's awareness that by trying to prove independence of genuine axioms we have to step beyond the border of logical theory to *meta*-theory. Frege in the first place saw more clearly than Hilbert — although still somewhat obscure — that there *is* such a border and that *proving* things concerning this new field requires totally new concepts and methods.[135]

There is a final point which deserves closer attention. The new science as reconstructed here, conists of a theory of syntax together with a theory relating this syntax of expressions to the "syntax of sense" and axioms for truth and provability. But it is not clear at this point that such a theory suffices to prove rigorously the independence of genuine axioms. To be more precise, it is not clear if a new science comprising *only* of these

---

[135]Recall that this is not the case if "independence" is understood in the way Frege reads Hilbert, for on this type-theoretical reconstruction independence-claims are just an elliptical way of expressing the negation of a huge quantified conditional. It is just with Tarski's strict separation between *object-* and *metalanguage* that this crucial point had slowly been appreciated by the logicians.

items suffices for proving independence of the axioms of an *arbitrarily chosen* system of genuine axioms. The problem here is not specific to Frege's approach, but is also implicit in a *modeltheoretical* approach to independence proofs. Say, we want to show that ϕ is independent from the system of axioms (or "conditions") S. In order to do this we have to show that there exists an interpretation M (where $M = (D, I)$ for some set D and an interpretation function I, providing extensional meanings for the non-logical vocabulary), which satisfies all the axioms of S but which makes ϕ false. That is, we have to show

(Int) $\quad \exists M : (M \text{ satisfies } S) \text{ and } (M \text{ satisfies } \neg\phi)$

Now the question is simply: where do we get the needed models M, *witnessing* such an existential claim, from? The simple answer is: from wherever we please! An interpretation is provided by *any* set of objects together with interpretations for the non-logical vocabulary defined over this set. We just have to ensure that such a model satisfies S and $\neg\phi$. The default choice is of course to provide the required models by taking sets of *mathematical* objects, such as sets of *numbers*. One reason for this is that, in doing so, we can use *mathematical* methods to establish that M satisfies S and ϕ by standard mathematical reasoning. Even more customary is it to provide the needed models by resorting to some *set theory*. As it is possible to construct *any* mathematical model within set theory, set theory is sometimes said to be the proper framework for model theory and it is in this sense that model theory is sometimes regarded as *set theory in disguise*.

A similar question now arises for the translation-based method used by Frege as well. Recall that, if we want to show that some genuine axiom ϕ is independent from genuine axioms S, we have to show

$(Trans)$ $\quad \exists t : (t(S) \text{ is true}) \text{ and } (t(\neg\phi) \text{ is true})$

So in order for this method to work (which instead of the notion of *truth in an interpretation* or *satisfaction* uses the *absolute* notion of *truth simpliciter*), we have to make sure that the theory in which such a claim is to be proved, is sufficiently strong to provide the right kind of rtranslations, for it is *prima facie* not clear, if truths about the syntax of expressions and senses will suffice for this purpose.[136]

---

[136]In fact it *does* suffice in the case of *first-order* theories, but for a reason that is not entirely obvious. As has

To recap the situation so far: after rejecting Hilbert's informal independence proofs, which rest — from Frege's point of view — on a misconception of what axioms are, Frege eventually scetches how he would prove independence of genuine axioms. The two main features of this proposal are that proofs of independence proceed within an *axiomatic framework* and are based on *translations* instead of *interpretations*. It has been shown that, even if a solution to the problem of "what belongs to logic properly" were given, there would have been a bunch of collateral problems surrounding Frege's suggestion. For one thing, substantial theories of sense and syntax would have to be provided as well as a theory of truth and provability *plus* a theory providing enough translations ensuring the correctness of independence results generated by this method.[137]

As I hope I have made clear, Frege's axiomatic approach to independence proofs bears some interesting ideas and relates to some of the most important topics in 20th century logical theory. In the last section I will try to elucidate some of these connections a bit further.

3.5 Axiomatic Metatheory

The reasons for axiomatizing some part of discourse are manifold, at least on the modern, Hilbert-inspired conception of the axiomatic method. A concept may show up to be that fruitful in different areas of mathematics, that it deserves seperate treatment. Algebraic notions seem to be of this kind: the concept of *group* for instance turned out to be that useful in such different fields as combinatorics, geometry and number theory that an independent study of it was considered to be fruitful. This pattern is quite general: a bunch of structural properties show up in different areas of mathematics and are singled out as objects of an independent investigation. Of course, as the discussion in the first section should have made clear, Frege does not believe that this kind of practice of sorting

---

been mentioned earlier, every theory of syntax includes a basic theory of arithmetic. As a strenghtening of the downward Löwenheim-Skolem theorem by Hilbert-Bernays shows, the only interpretations we have to consider are interpretations where the domain is the set of natural numbers and the interpretations of the non-logical vocabulary are *definable* sets (sets of Tupels etc.) of natural numbers. A related point has been made by Resnik in his classic [35] with an eye to Quine's substitutional account of logical truth.

[137]Note that, if this theory is too weak, it might happen that we cannot find a translation witnessing the existential claim $Trans$, altough it might be possible to find such a translation in a *stronger* theory.

out structural properties has anything to to with *axiomatization properly so called*. One might call the properties a structure has to have in order to fall under concept *group*, "axioms"; yet, on the Fregean view, genuine axioms are something quite different.

Still, *some* of the reasons for axiomatizing some part of discourse, and in particular some part of *metatheoretical* discourse, are the same on both conceptions of the axiomatic method. Let's look at some of them.

Axiomatization has, for different reasons, been used as a tool for providing a framework to prove things about notions that apparently are *undefinable* (or undefinable in a *uniform* way) by means of more standard notions. One might look for instance at Gödel, who once considered to axiomatize the concept of *computability* in order to decide the notorious Church-Turing thesis.[138] The problem here is of course that one wants to relate an *exact notion* (f.i. *Turing-computability*) with the *informal notion* of *computability*. Altough axiomatization did not quite work out in this particular case, the *rationale* behind this move seems to me important and relevant also as a possible motivation for Frege's suggestion to prove independence within an axiomatic setting: axiomatization can sometimes provide a framework where hitherto undecidable statements (for instance due to lack of precision) turn out to be *provable* or *refutable*. (Compare this situation with Zermelos motivation behind his axiomatization of set theory, which was just to set the stage to prove the well-ordering theorem.)

Another example of an undefinable concept, without being *ambigous or vague* (at least in this particular context), is provided by the concept of *truth* for some given formalized language. One of Tarski's main conclusions at the end of his 1933 paper was that truth can be defined unambigously for (what he calls) *languages of finite order* whereas no such definition would be possible for the *languages of infinite order*. He remarks though that "even with respect to formalized languages of infinite order, the consistent and correct use of the concept of truth is rendered possible by including this concept in the system of primitive concepts of the metalanguage and determining its fundamental properties by means of the axiomatic method."[139] So Tarski (at this point) figured that the concept of truth was not capable of being *defined in a uniform way* for *any* given formalized language.

---

[138] See [16] for further discussion.

[139] See his classic [43] p.266

But, according to him, this should not prevent us from investigating the concept of truth by means of an adequate *axiomatization*.

A similar example is provided by Myhills suggestion to axiomatize the notion of *absolute (arithmetical) provability*.[140] Myhills key idea is rather simple: take some formalized arithmetical theory T as a starting point. By Gödel's first incompleteness theorem there is some canonical "Gödel-sentence" G which is neither provable nor refutable from the axioms of T. Still, one can show this very sentence to be *true* by means of "standard-reasoning". Hence G is – in a fairly natural sense – *provable* after all. Now take T' to be $T \cup \{G\}$: T' trivially proves G, yet T' will again contain some Gödel-sentence G', which is neither provable nor refutable from T'. But again, one can show this sentence to be true, hence G' is a *provable* arithmetical sentence. Take T'' to be ... What's important here is to notice that each of the arguments establishing the truth of G, G', etc. should count as a *proof*, though not a proof in a *particular* arithmetical system. On the other hand, on this conception, *arithmetical provability* does *not* reduce to arithmetical *truth* either, for there is a clear sense in which each of the G's is established by using *inferences*. Myhill then concludes that one should axiomatize the notion of *absolute arithmetical provability*, for there is no way to *reduce* it *uniformly* by reference to some *fixed* system T of arithmetic.

The reasons for resorting to axiomatization have of course their own, specific (and often technical) background, but I think two major motives for axiomatization can be extracted from the discussion so far, which seem to be relevant to Frege's proposal concerning independence proofs:

1. axiomatization may enable one to investigate concepts that are *not reducible to known concepts*

2. axiomatization may enable one to *prove* things that were hitherto neither provable nor refutable

Let's look at the first motive more closely. Recall, that on the suggested reconstruction of Frege's approach to independence proofs, three (purportedly *primitive*) notational devices were introduced: the truth predicate $T(x)$, a provability relation $Prv(x, y)$ and, due to Frege's seemingly idiosyncratic conviction that *thoughts* are the bearers of truth and

[140] See his [31].

provability, a notation for *senses*. Now the first question that comes to mind seems to be the following: why take these particular notions as primitives and not others? Well, I have no good answer to this question, for there is simply not enough evidence to assess it. Talk about thoughts and the truth predicate clearly seem to suggest themselves as irreducible on *any* reasonable account (from the Fregean point of view) of "reducibility",[141] but the case of the provability relation $Prv(x, y)$ seems to be less clear. The reason is this: as we have seen, a theory of syntax as well as a theory of sense has to be included in Frege's new science anyways, for otherwise his translation-based method of proving independence would not get off the ground. But given our assumption that the extension of the relation $Prv(x, y)$ is fixed by the *syntactically* defined system K, this implies that it might be possible to define the objectlinguistic expression $Prv(x, y)$ (applying to *thoughts*), given enough facts regarding syntax and senses.

It seems to me that this line of reasoning would be quite cogent to Frege, but I leave it to the reader to assess this question (and maybe take it as a *reductio* of the assumption that derivability in K fixes the extension of $Prv(x, y)$). Anyways, I don't want to push this line of reasoning any further, because what seems to me important is Frege's apparent conviction that *something* is involved in independence proofs that is *not straightforwardly reducible* to standard logico-mathematical notions (like *sets*, *functions*, *numbers*, *points*, etc.) and that this "something" has to do with the *semantic ascent* that independence proofs concerning *genuine axioms* require.

---

[141]Note that "reducibility" should, in the given context, not be reconstructed by appeal to interpretability (or something akin to interpretability) *alone*. Geometry for instance is straightforwardly interpretable in the theory of real numbers. Still, from the Fregean point of view, analysis belongs to *logic*, whereas geometry is the paradigm-example of a theory that is *not* reducible to logic. Reducability carries with it, for Frege, always some *epistemological* constraints, that is: a *reduction* of some part of discourse to some other part of discourse is given if it can be shown how knowledge regarding the former domain can be obtained through knowledge regarding the latter. Frege, at one point, regarded his logicist reduction as successful because he thought his reduction could explain how knowledge of the arithmetical truths could be gained through (purportedly) logical reasoning alone. It is in this sense that it seems to me that the basic laws of the *new science* are *not* (at least not *straightforwardly*) reducible to *logic* alone because the new science requires basic truths about *syntax* and *thougths*, the knowledge of which does *not* seem to stem from what Frege calls the "logical source of knowledge". Knowledge of the syntax and semantics of some given language (formal or otherwise) belongs to a quite different area which seems to be closer to *linguistics* than *pure logic*.

With regards to the second motive mentioned earlier (i.e. that axiomatization can provide a framework where hitherto undecidable propositions become provable or refutable) we can see that it is dependent on the first motive, for if there is something conceptually irreducible to independence proofs, then *truths* about these irreducible notions have to be taken as axioms *outright*. If talk about *thoughts* for instance is irreducible to talk about more "standard" logico-mathematical objects, and if some *truths* about thoughts are needed in order to prove some theorem, then truths about thoughts have to be taken as *axioms*. Underlying all of this is the more or less tacit conviction, that we *should* be able to prove things about metatheoretical notions like truth and proof. Hence, Frege's *Begriffsschrift*-standard of proof, together with his apparent conviction that something conceptually irreducible is involved in independence proofs, *forces* him to treat independence proofs in an axiomatic fashion.[142]

There is another important connection to modern mathematical logic which deserves closer attention. In their important book *Undecidable Theories* ([44]), Tarski, Mostowski & Robinson introduced the notion of *relative interpretability*, which is defined in terms of *syntactic translations*[143]. Here, a syntactic translation from a language L_1 into a language L_2 is a pair $(\delta(x), t)$ consisting of an L_2-formula $\delta(x)$ (a "domain-formula") and an effectively computable function t which maps every primitive n-ary predicate P of L_1 to some L_2-formula $\phi_P(x_1, ...x_n)$, such that the extension of t to *all* formulas of L_1 respects the logical constants, i.e.

1. For every primitive predicate P in L_1 there is some L_2-formula ϕ_P, s.t. $t(P(x_1, ...x_n)) = \phi_P(x_1, ...x_n)$
2. $t(s_1 = s_2) = t(s_1) = t(s_2)$ for terms s_1, s_2
3. $t(\neg\phi) = \neg t(\phi)$ for every formula ϕ
4. $t(\phi \to \psi) = t(\phi) \to t(\psi)$ for formulas ϕ, ψ

---

[142] As an interesting aside another peculiarity should be pointed out: it has been argued (for instance by Richard Heck in his [19]), that some of Frege's remarks in his *Basic Laws of Arithmetic* should count as "informal proofs" of the soundness of his formulation of logic. Now this reading becomes doubtful in the light of Frege's 1906-paper on the foundations of geometry, where it seems that he wants to take some kind of "soundness principle" (NS_1 above) as an *axiom* of the new science.

[143] For a thourough treatment of relative interpretability, compare [30].

5. $t(\forall x\phi) = \forall x(\delta(x) \to \phi)$ for formulas ϕ

An L_1-theory T_1 is then said to be *relatively interpretable* in an L_2-theory T_2 if there is a syntactic translation from L_1 to L_2, such that for any L_1-sentence ϕ we have

(NS_2') $\quad T_1 \vdash \phi \Rightarrow T_2 \vdash t(\phi)$

That is, a theory T_1 is relatively interpretable in T_2 if every translation of a T_1-theorem is a T_2-theorem. Note that the notion of a syntactic translation as described here is just a *syntactic counterpart* of the notion of an *interpretation* (or a *model*). An interpretation for some language L_1 is usualy conceived of as a pair (D, I) consisting of some nonempty *set* D and an interpretation function I, which assigns extensional meanings (defined over the domain D) to the non-logical vocabulary of L_1. A syntactic translation on the other hand does something similar: it provides a "domain" for the theory T_1 *relative to some theory* T_2 by the domain-predicate $\delta(x)$ (which should be T_2-provably non-empty just like the domain D of an interpretation should be non-empty) and it provides "interpretations" for the primitives of L_1, again, *relative* to the "background-theory" T_2 (and, of course, relative to the "domain" specified by $\delta(x)$).[144]

Now, how can we prove the independence of some axiom ϕ from some L_1-theory T_1 by invoking the notion of relative interpretability? Well, just find some *consistent* L_2-theory T_2, in which $T_1 \cup \{\neg\phi\}$ is relatively interpretable and a consistent L_2'-theory T_2' in which $T_1 \cup \{\phi\}$ is relatively interpretable! If $T_1 \cup \{\neg\phi\}$ is relatively interpretable in T_2, then $T_2 \vdash t(\neg\phi)$, i.e. $T_2 \vdash \neg t(\phi)$. So if ϕ were provable from T_1, then (because T_2 interprets T_1) $t(\phi)$ were provable from T_2, hence $t(\phi) \wedge \neg t(\phi)$ were provable from T_2, contradicting the assumed consistency of T_2. Similarily if $T_1 \cup \{\phi\}$ is relatively interpretable in T_2' then $\neg\phi$ cannot be provable from T_1 either.

Now obviously the notion of syntactic translation as defined above is (nearly) *exactly* what Frege seems to have in mind when he is elucidating his new basic law.[145] As we have seen, F-translations serve only the intermediate-purpose of providing functions mapping

---

[144] In a sense, models/interpretations can even be regarded as *special cases* of syntactic translations, viz. interpretations where the interpreting theory T_2 is some standard set theory.

[145] To be more precise, it is exactly the Fregean notion of translation if unrestricted quantifiers are used. As we have seen, the Fregean translation of a universally quantified sentence $\forall x\phi$ would be $\forall x t(\phi)$, whereas on the modern conception of a translation one would have to restrict the quantifier to some domain-predicate $\delta(x)$.

senses of expressions to *senses* expressions, for Frege considered *thoughts* to be kind of things we should be interested in when dealing with the question of independence. On the other hand, on Frege's account, those functions mapping senses to senses are *parasitic* on the corresponding syntactic functions mapping expressions to expressions. So in a sense the often heard complaint that Frege did not think in *semantic* (*model-theoretic*) terms but rather *syntactically* is both true and false (at least applying to the question at hand): it is true in the sense that 1. his translation-based method relies on "syntactical methods", i.e. syntactical objects ("words") stand at the center of his method and 2. informal independence arguments must be made explicit in order to count as genuine *proofs*. On the other hand, his approach is clearly model-theoretic *in spirit*, for again, translations are just "syntactified" models. Frege's method could then even be glossed in modern terminology: to show that a *genuine* axiom ϕ is independent of axioms T_1, show that $T_1 \cup \{\neg\phi\}$ is relatively interpretable in a *true* theory T_2 (the *truth* of T_2 of course implies T_2's consistency).[146]

The philosophical payoff – from the Fregean point of view – is of course that by invoking translation-terminology instead of (re-)interpretations Frege can give an account of independence proofs without giving up his *fixed interpretation conception* of language, i.e. his view that expressions of a language properly so called (formal or natural) come "immutably equipped"[147] with a *fixed sense* as well as a *fixed reference* which cannot be

---

[146]Recall though, that Frege does *not* argue like this. There is a difference between defining *relative interpretability* as above, and by stipulating that T_1 is relatively interpretable in T_2 if T_2 proves every translation of a T_1-*axiom*. To see where Frege departs from the modern approach via relative interpretability as defined in *this* way, suppose we were given some true theory T_2 and a translation t, such that $T_2 \vdash t(T_1)$ and $T_2 \vdash t(\neg\phi)$, i.e. $T_2 \vdash \neg t(\phi)$. Now one might argue that, as T_2 proves every translation of a T_1-*axiom*, it follows informally that T_2 proves every translation of a T_1-*theorem*. Therefore, if ϕ were provable from T_1, its translation $t(\phi)$ would be provable from T_2. But it is precisely this kind of appeal to informality that Frege wants to eliminate in his approach. Hence, Frege argues that *by the new law* NS_2, if $T_1 \vdash \phi$, then $t(T_1) \vdash t(\phi)$. From this, together with the assumption $T_2 \vdash t(T_1)$ and the transitivity of the provability-relation, we can infer that $T_2 \vdash t(\phi)$.

Note also that on Frege's view of axioms as *true thoughts*, the negation $\neg\phi$ of an axiom ϕ cannot be provable from the true thoughts in T_2 *trivially*. Speaking loosely (and anachronistically): for Frege, there is no need to "show" that there is a model in which $T_2 \cup \{\phi\}$ is true in order to show that $\neg\phi$ is *not* provable from T_2, for if $T_2 \cup \{\phi\}$ are genuine axioms, they will be true in their "intended interpretation".

[147]Antonelli/May in [1] p. 246

altered randomly (for whatever reason). In considering translations we no longer have to "change the meanings" or something alike – we are dealing with entirely *different languages*.[148] Hence, by relying on *translations* instead of *(re-)interpretations* one can in a sense *emulate* model-theoretic reasoning without being committed to a picture of language that is sometimes said to be prerequisite for semantically-minded independence proofs.

In conclusion I want to emphasize two points: first, Frege's suggestions (in particular in the last part of his 1906-paper) can be seen as an important step towards providing a framework for metatheoretical investigations that is compatible with a traditional conception of the "axiomatic method". Moreover, his axiomatic approach to independence proofs draws attention to the important observation that metatheoretical investigations do not take place in "vacuous space" and that recourse to "informality" is not a valid alternative either. Metatheoretical investigations have commitments too, and if done responsibly, one has to make plain what these commitments are.

Second, as I have tried to make plausible, the axiomatic way Frege approaches the question of independence is a direct ancestor of axiomatic approaches to metatheoretical concepts in 20th century logic, in particular attempts to axiomatize the concept of *truth*. Furthermore, some of the concepts employed by Frege have direct counterparts in modern logic (such as the notion of a *translation*). So although Frege's suggestions in his 1906-paper are rather sketchy, they nonetheless provide connections to important areas of 20th century mathematical and philosophical logic. And even though Frege does things quite differently from what we are used to from our modern logic textbooks, it still seems to be justified to raise the counterfactual question "What *could* a careful thinker like Frege have achieved, had he spent more effort in spelling out the details?" and answer it immediately: *quite a lot*.

---

[148]Note also, that there is a reading of Hilbert's independence proofs which takes it that Hilbert too had in fact translations in mind (rather than "reinterpretations"). See [18] for more on that issue.

References

[1] Antonelli A., May R. 2000. 'Frege's new science', *Notre Dame Journal of Formal Logic*, **41** (3), 242-270

[2] Blanchette P. 'Frege on Formality and the 1906 Independence-Test', forthcoming in: Link G. (ed.), *Formalism and Beyond: On the Nature of Mathematical Discourse*, Ontos Press

[3] Blanchette P. 2007. 'Frege on consistency and conceptual analysis', *Philosophia Mathematica*, **15** (3), 321-346

[4] Blanchette P. 1996. 'Frege and Hilbert on consistency', *The Journal of Philosophy* **93** (7), 317-336

[5] Carnap R. 2000. (ed. Bonk T), *Untersuchungen zur allgemeinen Axiomatik*, Darmstadt: Wiss. Buchgesellschaft

[6] Carnap R. 2003 (eds. Reck E., Awodey S., Gabriel G.), 'Frege's lectures on Logic: Carnap's student notes 1910-1914', Illinois: Carus Puplishing Company

[7] Demopoulos W. 1994 'Frege, Hilbert, and the Conceptual Structure of Model Theory', *History and Philosophy of Logic* **15** 211-225

[8] Dummett M. 1976. 'Frege on Independence and Consistency', in Schirn M. (ed.): *Studies on Frege I: Logic and Philosophy of Mathematics*, Stuttgart: Friedrich Frommann Verlag, Günther Holzboog GmbH & Co, 229-242

[9] Frege, Gottlob. 1903. 'Über die Grundlagen der Geometrie', *Jahresbericht der deutschen Mathematikervereinigung* **12** (1903) reprinted in English in [40]

[10] Frege, Gottlob. 1906. 'Über die Grundlagen der Geometrie', *Jahresbericht der deutschen Mathematikervereinigung* **15** (1906), reprinted in English in [40]

[11] Frege G. 1892-1902. *Grundgesetze der Arithmetik. Begriffsschriftlich abgeleitet. Band I und II.*, Paderborn: mentis Verlag 2009

[12] Frege G. 1976. *Wissenschaftlicher Briefwechsel*, Hamburg: Felix Meiner Verlag

[13] Frege G. 1960. *Foundations of Arithmetic*, revised edition, transl. by J. Austin, Harper & Brothers, New York

[14] Frege G. 1984. *Collected Papers on Mathematics, Logic and Philosophy*, Basil Blachwell Publisher Ltd, Oxford, Brian McGuinnes (ed.)

[15] Goldfarb W. 2005. 'Frege's conception of logic', in Reck E. and Beaney M. (eds.), *Gottlob Frege: Critical Assessments of Leading Philosophers*, New York: Routledge

[16] Shagrir O. 2006. 'Gödel on Turing on Computability', in Olenski A, Wolensik J, Janusz R. (eds.): *Church's Thesis after 70 years*. Ontos Verlag

[17] Halbach V. 2011. 'Axiomatic Theories of Truth', Cambridge: Cambridge University Press

[18] Hallett M. 2010. 'Frege and Hilbert', in [36], 413-464

[19] Heck R. 2007. 'Frege and Semantics', *Grazer Philosophische Studien* **75**, 27-63

[20] Heijenoort J. 1967. 'Logic as Calculus and Logic as Language', *Synthese* **17** (1), 324-330

[21] Hendricks et. al. (eds). 2004. 'First-Order Logic Revisited', Berlin: Logos Verlag

[22] Hilbert D. 1899. *Grundlagen der Geometrie*, Leipzig: Teubner Verlag [1923]

[23] Hintikka J. 1988. 'On the Development of the Model-theoretic Viewpoint in Logical Theory', *Synthese* **77**, 1-36

[24] Hintikka J. 2011. 'What is the Axiomatic Method?', *Synthese* **183** (1), 69-85.

[25] Hodges W. 2004. 'The Importance and Neglect of Conceptual Analysis: Hilbert-Ackermann iii.3', in [14], 129-153

[26] Kambartel F. 1976. 'Frege und die axiomatische Methode. Zur Kritik mathematikhistorischer Legitimationsversuche der formalistischen Ideologie', in Schirn M. (ed.): *Studies on Frege I: Logic and Philosophy of Mathematics*, Stuttgart: Friedrich Frommann Verlag, Günther Holzboog GmbH & Co, 215-228

[27] Kitcher P. 1984. *The nature of mathematical knowledge*, Oxford University Press, Oxford

[28] Kluge, Eike-Henner W. (ed.). 1971. *On the Foundations of Geometry and Formal Theories of Arithmetic*, Yale University Press, New Haven and London

[29] Korselt A. 1903. 'Über die Grundlagen der Geometrie', *Jahresbericht der Deutschen Mathematiker-Vereinigung*, **12**, reprinted in English [40]

[30] Lindström P. 1997. 'Aspects of Incompleteness', Berlin Heidelberg New York: Springer Verlag

[31] Myhill J. 1960. 'Some Remarks on the Notion of Proof', *The Journal of Philosophy* **57** (14), 461-471

[32] Quine W. 'Concatenation as a basis for arithmetic', *The Journal of Symbolic Logic* **11** (4), 105-114

[33] Reck E. (ed.) 2002. *From Frege to Wittgenstein: Perspectives on Early Analytic Philosophy*, Oxford: Oxford University Press

[34] Reck E. and Beaney M. (eds.) 2005. *Gottlob Frege: Critical Assessments of Leading Philosophers*, New York: Routledge

[35] Resnik M. 1974. 'The Frege-Hilbert Controversy', *Philosophy and Phenomenological Research* **34** (3), 386-403

[36] Ricketts T. Potter M. (eds). 2010. *The Cambridge Companion to Frege*, Cambridge University Press

[37] Ricketts T. 1997 'Frege's 1906 Foray into Metalogic', reprinted in Beaney M., Reck E. (eds.): *Gottlob Frege: Critical Assessments of Leading Philosophers*, Vol. 2, New York: Routledge 2005, 136-155

[38] Stanley J. 1996. 'Truth and Metatheory in Frege', reprinted in Beaney M., Reck E. (eds.): *Gottlob Frege: Critical Assessments of Leading Philosophers*, Vol. 2, New York: Routledge 2005, 109-135

[39] Tappenden, J. 1997. 'Metatheory and Mathematical Practice in Frege', *Philosophical Topics* **25** (2), 213-264

[40] Tappenden, J. 2000. 'Frege on Axioms, Indirect Proof, and Independence Arguments in Geometry: Did Frege reject Independence Arguments?', *Notre Dame Journal of Philosophy* **41** (3), 271-315

[41] Tappenden, J. 2007. 'The Riemannian Background to Frege's Philosophy', in *The Architecture of Modern Mathematics: Essays in History and Philosophy*, J. Ferreiros & J.J. Gray (eds.), Oxford: Oxford University Press

[42] Tarski, A. 1977. *Einführung in die mathematische Logik* (fith edition), Göttingen: Vandenhoeck & Ruprecht

[43] Tarski, A. 1956. *Logic, Semantics, Metamathematics*, Oxford: Clarendon Press

[44] Tarski, A. 1971. *Undecidable Theories*, Amsterdam: North-Holland Publishing Company

[45] Wehmaier K. 1997. 'Aspekte der Frege-Hilbert-Korrespondenz', *History and Philosophy of Logic* **18**, 201-209

[46] Wilson, M. 2010. 'Frege's Mathematical Setting', in [36], 379-412

[47] Wilson, M. 1992. 'The Royal Road from Geometry', *Nous* **26** (2), 149-180

4 Remarks on Compositionality and Weak Axiomatic Theories of Truth

Abstract.[149] The paper draws attention to an important, but apparently neglected distinction relating to axiomatic theories of truth, viz. the distinction between *weakly* and *strongly truth-compositional* theories of truth. The paper argues that the distinction might be helpful in classifying weak axiomatic theories of truth and examines some of them with respect to it.

The point I want to adress in this short note is concerned with an important, but apparently unnoticed distinction relating to weak axiomatic theories of truth, which have come to the fore in formal philosophy recently.

To motivate my point, recall that one of the problems that plagues the axiomatic theory of truth DT is the so called *generalization problem*. DT (sometimes labelled TB) ist the axiomatic theory of truth that consists solely of the instances of the so called *restricted T-scheme*:

$$(T) \qquad T(\phi) \leftrightarrow \phi$$

(T) is an axiom-*scheme*, i.e. one get's a particular instance of this scheme by replacing ϕ with a particular well-formed sentence of the base language (which, as it is customary in discussing axiomatic theories of truth, will be the language L_{PA} of arithmetic throughout this paper)[150]. The problem has already been mentioned in Tarski's seminal paper *The Concept of Truth in Formalized Languages* as early as 1933:

> A theory of truth founded on them [the T-Biconditionals; author] would be a highly incomplete system, which would lack the most important and most fruitful general theorems. Let us show this in more detail by a concrete example. Consider the sentential function '$x \bar{\in} Tr$ or $\bar{x} \bar{\in} Tr$'. If in this function

[149] This paper has been submitted for publication in *Journal of Philosophical Logic*. Date of submission: 2nd July 2012.

[150] In order to avoid paradox one has to restrict the T-scheme in one way or the other. This can be done by restricting the metavariable ϕ in (T) to sentences not containing the truth-predicate.

> we substitute for the variable 'x' structural-descriptive names of sentences, we obtain an infinite number of theorems, the proof of which on the basis of the axioms obtained from the convention T presents not the slightest difficulty. But the situation changes fundamentally as soon as we pass to the generalization of this sentential function, i.e. to the general principle of contradiction. From the intuitive standpoint the truth of all those theorems is itself already a proof of the general principle; this principle represents, so to speak, an 'infinite logical product' of those special theorems. But this does not at all mean that we can actually derive the principle of contradiction from the axioms or theorems mentioned by means of the normal modes of inference usually employed. On the contrary, by a slight modification of Th. III it can be shown that the principle of contradiction is not a consequence (at least in the existing sense of the word) of the axiom system described. ([7] p. 257)

To repeat his point: the problem is that, even though every instance of the *scheme*

$$(PC_S) \qquad \sim T(\phi) \vee \sim T(\sim \phi)$$

is derivable from DT, the corresponding *universally quantified statement*

$$(PC) \qquad \forall \phi \in L_{PA} :\sim T(\phi) \vee \sim T(\sim \phi)$$

is *not*.

Tarski concludes from this that an axiomatic theory of truth based *exclusively* on the T-Biconditionals (i.e. 'Convention T') cannot claim to be a satisfactory theory of truth. For Tarski, the derivability of the T-Biconditionals remains a *minimal* adequacy condition, but, as can be seen from the quote above, the derivability of the T-Biconditionals *alone* is not sufficient to warrant the adequacy of a theory of truth. More recently, Paul Horwich's minimalist theory of truth, which can be seen as a variant of DT if properly axiomatized, has been attacked for – among other things – a similar reason. Not being able to prove certain general statements concerning the truth predicate is therefore a major problem any weak theory of truth such as DT has to face.

One of the unfortunate consequences of the generalization problem is that on the basis of the T-Biconditionals alone it is not possible to prove certain facts about the truth predicate which are often taken to be central to the concept of truth (and semantics in general), namely that truth is *compositional*. Recall that the intuitive principle of compositionality is usually taken to consist in something like the following:

> For every complex expression e of some language L, the semantic value of e in L is determined by the syntactical mode of composition of e and the semantic values of the constituents of e in L.

Hence compositionality is, on the most natural reading of this principle, expressed by a sentence quantifying explicitly over expressions of some given language L (or in our case *Gödel numbers* of expressions). In the context of classical, typed axiomatic theories of truth the semantic values we are interested in primarily, are of course *truth* and *falsity*, and the modes of composition we have to consider for our arithmetical base language are *identity of terms* (the terms being composed of a constant $\mathbf{0}$, addition-, times- and the successor function), *negation*, *conjunction* and *universal quantification*. An axiomatic theory of truth is then said to be compositional, if – following the intuitive principle just sketched – the following universally quantified statements are provable[151]:

$TC1 \qquad \forall \phi \in L_{PA}^{atom} : T(\phi) \leftrightarrow T_0(\phi)$

$TC2 \qquad \forall \phi \in L_{PA} : T(\sim \phi) \leftrightarrow \sim T(\phi)$

$TC3 \qquad \forall \phi \forall \psi \in L_{PA} : T(\phi \& \psi) \leftrightarrow T(\phi) \& T(\psi)$

$TC4 \qquad \forall \phi(y) \in L_{PA} : T(\forall x \phi(x)) \leftrightarrow \forall x T(\phi(x))$

---

[151] In the following L_{PA}^{atom} stands for the class of atomic arithmetical sentences (i.e. equations) and $T_0(x)$ for the truth-predicate restricted to atomic sentences, which is definable in PA and for which Tarski's 'Convention T' can be shown to be satisfied in PA. Note also that $\forall x T(\phi(x)$ is to be understood as $\forall x T(\dot{sub}(\mathbf{n}, \mathbf{m}, \dot{nu}(x)))$, where n is the Gödelnumber of the formula $\phi(x)$, m is the gödelnumber of the variable x, $\dot{nu}(x)$ is the function mapping every natural number to the Gödelnumber of it's numeral and $sub(x, y, z)$ is the substitution function. Similar conventions apply to occurences of the negation- and conjunction sign within the scope of the truth predicate.

If these clauses are taken as *axioms*, what we get is a theory called TC (sometimes CT or $T(PA)$), one of the most studied theories of truth there is. Indeed, $TC1 - TC4$ are just the clauses Tarski used in his recursive *definition* of truth.[152] The compositional clauses also play a prominent role in Davidson's philosophy of language (see for instance his [1]).

Because of the generalization problem, DT does not prove *any* of these quantified statements and hence DT cannot be taken to capture compositionality in a *full-blooded sense*.[153]

DT however *does* prove something that looks very similar to these clauses: DT proves every instance of the *schemes* corresponding to the quantified sentences $TC1$, $TC2$ and $TC3$, that is, DT proves every arithmetical instance of the *schemes*

TC_S1 $\quad T(\phi) \leftrightarrow T_0(\phi)$ (for ϕ atomic)

TC_S2 $\quad T(\sim \phi) \leftrightarrow \sim T(\phi)$

TC_S3 $\quad T(\phi \& \psi) \leftrightarrow T(\phi) \& T(\psi)$

This can be seen very easily:

TC_S1: this is obvious, for even PA alone proves $T_0(\phi) \leftrightarrow \phi$ for every atomic sentence ϕ.

---

[152] This is strictly speaking not true, for Tarski defined the more general notion of *satisfaction* recursively in order to define the concept of truth.

[153] We just look at the negation-case: Assume $DT \vdash \forall \phi \in \mathcal{L}_{PA} :\sim T(\phi) \leftrightarrow T(\sim \phi)$; Now a proof of this statement can use only finitely many axioms DT_0 of DT, i.e. $DT_0 \vdash \forall \phi \in \mathcal{L}_{PA} :\sim T(\phi) \leftrightarrow T(\sim \phi)$. Let $T(\phi_1) \leftrightarrow \phi_1$, ... $T(\phi_n) \leftrightarrow \phi_n$ be a complete list of the T-Biconditionals used in the proof. Now define a model $(\mathbb{N}, E_T)$ (based on the standard-model of arithmetic $\mathbb{N}$) by including all (codes of) true sentences ϕ_i and the negations of all false sentences ϕ_i (for $1 \leq i \leq n$) into the extension E_T of the truth predicate. Then $(\mathbb{N}, E_T)$ makes all arithmetical axioms used in the proof true as well as all the T-Biconditionals. But in this model the quantified sentence $\forall \phi \in \mathcal{L}_{PA} :\sim T(\phi) \leftrightarrow T(\sim \phi)$ is clearly false. Hence $DT_0 \nvdash \forall \phi \in \mathcal{L}_{PA} :\sim T(\phi) \leftrightarrow T(\sim \phi)$ and therefore $DT_0 \nvdash \forall \phi \in \mathcal{L}_{PA} :\sim T(\phi) \leftrightarrow T(\sim \phi)$. Contradiction.

TC_S2: One instance of the T-Schema of DT is given by $T(\sim \phi) \leftrightarrow \sim \phi$, another one by $T(\phi) \leftrightarrow \phi$, which is equivalent to $\sim T(\phi) \leftrightarrow \sim \phi$. Putting these biconditionals together yields $T(\sim \phi) \leftrightarrow \sim T(\phi)$ as desired.

TC_S3: Similarily, for each pair of L_{PA}-sentences ϕ, ψ the following are instances of the T-schema: $T(\phi \& \psi) \leftrightarrow \phi \& \psi$, $T(\phi) \leftrightarrow \phi$ and $T(\psi) \leftrightarrow \psi$. Again, putting these together yields $T(\phi \& \psi) \leftrightarrow T(\phi) \& T(\psi)$.

■

Hence, one might hope that DT proves at least something *akin* to compositionality, namely, for every quantified statement $TC1$ - $TC4$ it's corresponding *scheme*. As Horsten, in his recent book *The Tarskian Turn* claims:

> It is important that each compositional truth axiom is expressed as a universally quantified sentence rather than as an axiom scheme. From the axiom scheme, the corresponding universally quantified sentence cannot be derived. But we have seen that each instance can be derived from DT, whereby the schematic version of TC is a consequence of DT. ([2]: 71; also to be found in [3]: 364)

However, this last claim is mistaken, for as can be seen, DT does not even prove (every instance of) the *scheme* corresponding to the quantifier clause $TC4$, i.e.: for every L_{PA}-formula $\phi(x)$ with exactly x free

$$TC_S4 \qquad T(\forall x \phi(x)) \leftrightarrow \forall x T(\phi(x))$$

Here is the short argument (which is similar to the argument establishing the non-provability of TC_2 from DT)[154]:

Let $\phi(x) \in L_{PA}$ be some formula for which we have $\mathbb{N} \vDash \forall x \phi(x)$ (take for instance the formula $x = x$). Now suppose $DT \vdash T(\forall x \phi(x)) \leftrightarrow \forall x T(\phi(x))$. Again, because a proof can only use finitely many axioms, there must be some finite subset DT_0 of DT, such that $DT_0 \vdash T(\forall x \phi(x)) \leftrightarrow \forall x T(\phi(x))$. Let

[154]I do not want to suggest that Horsten and Halbach are not aware of this fact. But it seems to me that they underestimate the significance of this point.

$T(\phi_1) \leftrightarrow \phi_1$

....

$T(\phi_n) \leftrightarrow \phi_n$

be a list of all the T-Biconditionals in DT_0: We will now construct a model $(\mathbb{N}, E_T)$, in which all these T-Biconditionals are true, but $T(\forall x \phi(x)) \leftrightarrow \forall x T(\phi(x))$ is not. This is done by including all true sentences ϕ_i, every negation of a false sentence ϕ_i (for $1 \leq i \leq n$) as well as $\forall x \phi(x)$ into the extension of the truth-predicate. In this model $T(\forall x \phi(x))$ is true, as is every instance of the T-schema listed above. But the universally quantified sentence $\forall x T(\phi(x))$ is *false*, because for this statement to be true, E_T would have to contain *every numerical instance* $\phi(\mathbf{0}), \phi(\mathbf{1}), \phi(\mathbf{2}), ...$, for this is exactly what $\forall x T(\phi(x))$ says. But, by definition, E_T is *finite*. Hence $DT_0 \nvDash T(\forall x \phi(x)) \leftrightarrow \forall x T(\phi(x))$ and therefore $DT_0 \nvdash T(\forall x \phi(x)) \leftrightarrow \forall x T(\phi(x))$. Contradiction.

■

Hence, DT does not capture a central feature of truth: that a universally quantified sentence is true if and only if all its *instances* are true, and it does so not even for the weak *schematic* form of this compositional clause. Although the compositional character of truth is somehow reflected in the case of the propositional connectives, the possibility of *quantifying into* the truth predicate is not reflected by the axioms of DT, which only apply to *whole sentences*. In a sense, the possibility to *quantify into* the truth predicate outstrips the proof-theoretical power of DT, i.e. DT cannot relate the truth of a quantified sentence to its numerical instances.

This motivates a weakening of the intuitive compositional principle, which, in the context of axiomatic theories of truth, takes the form of the requirement that every instance of the *schemes* $TC_S1 - TC_S4$ be derivable. A truth theory S satisfying this condition will be called *weakly truth compositional* (*wtc* for short) henceforth. By contrast, a theory may be called *strongly truth-compositional* (*stc* for short) if the quantified statements

$TC1 - TC4$ are derivable.[155]

I think being *wtc* is a desirable property of a weak theory of truth, for a theory satisfying this condition makes plain the *point* of the principle of compositionality, which does not lie in the quantifiers used to formulate it, but in the following: that the truth value of a complex sentence depends on its syntactical structure and the truth values of its sub-sentences (or as it is sometimes expressed: that truth 'distributes' over the logical operators), and *this* is arguably captured by the schemes as well.

One may compare the situation with the case of the *principle of induction* in arithmetic: straightforwardly, the principle of induction is expressed by a single sentence $\forall X(X(0) \wedge \forall x(X(x) \rightarrow X(S(x))) \rightarrow \forall x X(x))$ quantifying over *sets* of natural numbers. In standard first-order arithmetic, however, only quantification over natural numbers, not *sets* of natural numbers, is possible. Hence the best we can get in first-order arithmetic is the *scheme* $\phi(0) \wedge \forall x(\phi(x) \rightarrow \phi(S(x))) \rightarrow \forall x \phi(x)$ corresponding to the full-blooded principle of induction expressed by the quantified second-order (or set-theoretical) sentence. True: the induction-*scheme* does not capture the intuitive principle of induction completely, but only few would claim that the point of the principle of induction would be lost by weakening the full-scale principle of induction to the *scheme*. It seems to me that this situation is – at least in the main respects – analogous to the case of the principle of compositionality.

Now the question which immediately arises is: are there any interesting *wtc* theories of truth that are not *stc*? (Of course every *stc*-theorey is also *wtc*.)[156]

---

[155] A semantic theory may be called *strictly compositional* or simply *compositional*, if analogue conditions for the *satisfaction relation* $Sat(x, y)$ are derivable. Note also that although in this paper attention is restricted to *typed* theories of truth, there is no reason not to apply the distinction between *wtc* and *stc* theories of truth to untyped theories of truth as well.

[156] The theory comprising of *DT* plus all instances of the schema TC_S4 seems to be all too *ad hoc* to count as a well-motivated 'interesting' theory of truth.

Note also, that it does not seem to be a trivial matter if the theory consisting of the *schemes* corresponding to TC_1 - TC_4, call it TC_S, is 'interesting'. The problem with TC_S is that it is not obvious if it proves every instance of the T-scheme, thereby meeting Convention T. Call the schematic theory one gets from TC_S by allowing free variables to occur in its instances TC_S^P. Clearly TC_S^P proves every instance of the T-scheme $T(\phi) \leftrightarrow \phi$. This can be shown by proving the stronger claim that from TC_S^P every instance of the *uniform* T-scheme $\forall x(T(\phi(x) \leftrightarrow \phi(x))$, is derivable. This is proved by a straightforward induction on the complexity of $\phi(x)$. But it is not clear if this

As we have seen, DT fails to be *wtc*, for *not* every instance of TC_S4 is provable from DT. We do, however, get an interesting *wtc* theory of truth by slightly improving on DT. The axioms of this improved theory, which will be referred to as UDT (sometimes labelled UTB) and which is well known in the literature, consists in all instances of the schema:

$$(UT) \qquad \forall x(T(\phi(x)) \leftrightarrow \phi(x))$$

(UT) is a scheme with respect to *formulas*, not *sentences*, i.e. we get a particular instance of this scheme by inserting a particular formula $\phi(x) \in L_{PA}$ with exactly the variable x free. UDT is an interesting theory of truth to the extent that it meets the minimal requirement of proving all instances of the T-scheme, for DT is obviously a subtheory of UDT. Moreover, it is immediate that UDT is *wtc* too, because from (UT) it follows that every instance of the follwing scheme is provable:

$$\forall x T(\phi(x)) \leftrightarrow \forall x \phi(x)$$

which, together with the relevant instances of the scheme $T(\forall x \phi(x)) \leftrightarrow \forall x \phi(x)$ (which are already provable from DT), yields, as desired, every instance of the schema TC_S4.

So it seems that a case can be made for UDT as an attractive weak theory of truth, for it meets certain requirements that are sometimes considered necessary conditions for a theory of truth to be 'good'. In particular, UDT – being a supertheory of DT – is *minimally adequate* in the sense that it meets Convention T. Second, it is a *conservative extension* of the base theory PA, that is, it does not prove any new non-semantical facts (i.e. sentences not containing the truth predicate).[157] Also, conceptually, UDT is only a slight modification of and hence nearly as simple and elegant as DT. One might even say that someone who takes DT as his favourite theory of truth is thereby *commited* to UDT as well. After all, if a formula contains free variables, these may be regarded as *names* of

---

can be done for TC_S, where only *sentences* are allowed to occur in the axiom schemes TC_1 - TC_3.

[157] It has been argued by Shapiro in his [6] and Ketland in his [5] that an adequate *deflationist* theory of truth is bound to conservativity.

arbitrary objects of the intended domain. So if one is inclined to accept $T(\phi(a)) \leftrightarrow \phi(a)$, a being a name of an arbitrary object (as is the case if one accepts DT), one should also accept $\forall x(T(\phi(x)) \leftrightarrow \phi(x))$. Thus, in the light of it's conceptual similarity to DT it might even seem surprising that, unlike DT, UDT can claim to be *compositional* (in the sense of being *wtc*).

So it seems to me that it is worth further investigating *wtc* theories of truth, for it is often not a trivial matter whether a given minimally adequate theory of truth (in the sense of satisfying Convention T) is *wtc*.

References

[1] Davidson, D. 1984. 'Theories of Meaning and Learnable Languages'. In D. Davidson, *Inquiries into Truth and Interpretation*, 3 - 15, Oxford: Clarendon Press.

[2] Horsten, L. 2011. *The Tarskian Turn*. Cambridge, Mass.: MIT Press.

[3] Horsten L., Halbach V. 2011. 'Truth and Paradox'. In *The Continuum Companion of Philosophical Logic*, ed. Horsten L., Pettigrew R. 351 - 382. Continuum International Publishing Group.

[4] Horwich, P. 2010. *Truth-Meaning-Reality*. Oxford: Clarendon Press.

[5] Ketland, J. 1999. 'Deflationism and Tarski's Paradise'. *Mind* 108, 69 - 94.

[6] Shapiro, S. 1998. 'Proof and Truth: Through Thick and Thin', *The Journal of Philosophy* Vol. 95.

[7] Tarski, A. 1933. 'The Concept of Truth in Formalized Languages'. English translation in his *Logic, Semantics, Metamathematics* (1956), 152 - 278. Oxford: Clarendon Press.

5 Abstract (German)

Mit einigem Recht kann man sagen, dass sich moderne, formale Logik so gut wie ausschließlich mit *metatheoretischen* Aspekten formal-logischer Sprachen und mit Hilfe solcher Sprachen formulierter axiomatischer Theorien, beschäftigt. Metatheoretische Probleme wie *Korrektheit*, *Vollständigkeit* oder *Entscheidbarkeit* oder Fragen bzgl. *Konsistenz* oder *Unabhängigkeit* axiomatisierter Theorien, stehen im Zentrum des Interesses. Obwohl Frege einer der Hauptbegründer der modernen Logik war, hat sich in den letzten Jahrzehnten eine lebhafte Debatte über die Frage entwickelt, ob und inwieweit Frege überhaupt in der Lage war, sich derartige Fragen zu stellen. Einer einflussreichen Tradition zufolge würde Freges "universalistisches" Verständnis von Logik eine genuin metatheoretische Perspektive nämlich verhindern. Im Zuge dieser Debatte kam es in jüngerer Zeit zu einer Neubewertung von Freges Verhältnis zu Unabhängigkeits- und Konsistenzbeweisen bzgl. axiomatisierter Theorien. Zentral in dieser Diskussion ist die sogenannte "Frege-Hilbert-Kontroverse", ein wissenschaftlicher Streit über den Status der "axiomatischen Methode" zwischen Frege und dem Mathematiker und Logiker David Hilbert. Die in dieser Dissertation gesammelten Aufsätze sind als Beitrag zur Aufarbeitung dieser Kontroverse gedacht und beschäftigen sich in der Hauptsache mit Freges eigenem Verständnis von Unabhängigkeit und Konsistenz. Die Artikel beziehen sich größtenteils auf Freges eigenen Ansatz zu Unabhängigkeitsbeweisen, den er in seinem 1906 veröffentlichten Aufsatz *Über die Grundlagen der Geometrie* präsentiert. Freges Vorschlag besteht im Wesentlichen darin, eine "neue Wissenschaft" zu etablieren, innerhalb derer Fragen wie Unabhängigkeit von Axiomen verhandelt werden sollen.

Der erste Aufsatz "Remarks on Independence Proofs and Indirect Reference" beschäftigt sich mit einem bestimmten interpretatorischen Problem bezüglich Freges "neuer Wissenschaft", das sich aus seiner Festlegung ergibt, dass Axiome eine bestimmte Art *intensionaler Entitäten* ("Gedanken") sind. Der zweite Aufsatz, "Frege's *On the Foundations of Geometry* and Axiomatic Metatheory", der den Kern dieser Dissertation ausmacht, beschäftigt sich eingehend mit der Frage, wie Freges "neue Wissenschaft" genauer rekonstruiert werden könnte. Ich werde dort zum Schluss kommen, dass Freges Gedanken zur "neuen Wissenschaft" einige auffallende Bezüge zur modernen Beschäftigung mit axiomatischer Metatheorie, insbesondere axiomatischen Theorien der *Wahrheit*, aufweisen. Der

letzte Aufsatz “Remarks on Compositionality and Weak Axiomatic Theories of Truth” beschäftigt sich mit einem konkreten (teils technischen, teils philosophischen) Problem schwacher axiomatischer Wahrheitstheorien und ist als Beitrag zur aktuellen Debatte gedacht.

6 Acknowledgements

I am thankful to the *Logik Café* as well as the members of the *Wiener Forum für Analytische Philosophie*, in particular Leo Stadlmüller and Katharina Sodoma, for helpful discussions and suggestions. The third paper of this dissertation would not have emerged without the *Forum*'s decision to focus on formal theories of truth.

I am also grateful to Richard Heinrich, Sebastian Baldinger, Naomi Osorio-Kupferblum, Georg Schiemer and my colleague and friend Frederik Gierlinger for pointing to errors, shortcomings and infelicities (concerning both content and form) in earlier versions of the articles collected in this dissertation. As always, none of them can be blamed for the errors, shortcomings and infelicities that unquestionably have survived.

I am especially indebted to Esther Ramharter for many helpful comments, her advice as well as her consistent encouragement and support.

Last, but not least, I want to thank [*insert your name*]. I did not forget about your valuable help in writing this dissertation.

7 Curriculum Vitae

Günther Eder

University of Vienna

Department of Philosophy

Universitätsstrasse 7

1010 Vienna, Austria

Education:

2009 – 2012: Doctoral program at the Department of Philosophy, University of Vienna

2003 - 2008: M.A. University of Vienna (MA thesis: "ZFC vs. NFU: eine philosophische Untersuchung zur Mengenlehre")

2002: Graduation, HAK Bruck/Mur

Teaching:

Winter Term 2009/10 - Winter Term 2012/2013: Exercises in Logic

Winter Term 2012/13: Elementary Logic

Publications:

Günther Eder. 'Remarks on Independence Proofs and Indirect Reference', in *History and Philosophy of Logic* (forthcoming)