The Common Foundation of Neo-Logicism and the Frege-Hilbert Controversy

This dissertation is submitted for the degree of Doctor of Philosophy

Miss Fiona Teresa Doherty

Trinity Hall

University of Cambridge

8^{th} March 2017

Abstract

In the first half of the thesis I investigate David Hilbert's early ontology of mathematics around the period 1899–1916. Hilbert's early views are of significant philosophical interest and have been largely ignored due to his later, more influential work. I suggest that, in this period Hilbert, can be understood as an early structuralist. In the second half of the thesis, I connect two important debates in the foundations of mathematics: Hale and Wright's neo-Fregean logicism and the Frege-Hilbert controversy. Using this connection, I adapt Frege's objections to Hilbert and apply them to Hale and Wright's account. By doing this, I show that the neo-Fregean logicists have long abandoned the Fregean element of their program in favour of a structuralist ontology. I conclude that our ontological conception of what exists in mathematics and what it is like constrains the foundations we use to characterise mathematical reality.

Declaration

This thesis is 74,479 words in length. The word count, including footnotes and references, falls within the range specified by the Degree Committee of the Faculty of Philosophy. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University.

Citation Conventions

The following citation conventions will apply throughout this thesis.

When there are two dates given in a citation the first date refers to the date of the original publication of the text and the second date refers to the translation edition. For example, with Frege (1903/1971*a*) the original publication of the German "*Über die Grundlagen der Geometrie*" is given by the first date (1903) and the date of the English translation, in this case "Collected Papers on Mathematics, Logic and Philosophy" by McGuinness, is given by the second date (1984). The item detailed in the bibliography will be the source of the English translation which is used in the text.

We will need to refer to the correspondence between Frege and Hilbert and identify the individual letters. Therefore, each letter has been separately included in the bibliography. A citation to a letter includes the original date and the translation, all of the page numbers are from the English translation of the same edition.

We will also need to refer to Hilbert's *Grundlagen der Geometrie*. The standard first German edition of his work was published in 1901. However, since our interest in this text will be restricted to the conversation between Frege and Hilbert we will refer to an earlier edition of this book called the *Festschrift*. The reason is that it is this early edition which Frege would have read since they correspond about the work prior to 1901. Frege refers to Hilbert's work as "your *Festschrift* on the foundations of geometry" (Frege, 1899/1980, 34). This early edition was based on lectures on Euclidean geometry which Hilbert delivered in the winter of 1898-1899. He wrote the monograph as a memorial address to mark the unveiling of the Gauss-Weber monument in 1899 and it was published in *Festschrift Zur Feier Der Enthüllung Des Gauss-Weber-Denkmals in Göttingen, Volume 1* (Leipzig: 1899). Nothing much will hang on using this edition rather than the standard, one but for the sake of historical consistency it is worth citing the older monograph.

Acknowledgements

The first thanks must be due to my supervisor Michael Potter, not in the very least for never giving up on me. I thank my examiners, Alex Oliver and Patricia Blanchette for their hard work in producing insightful and rigorous corrections. I also owe a debt to Gary Kemp and Peter Sullivan, without whose generosity, mentoring, and philosophical example I would never have gotten this far. I am humbled and grateful for all the discussions, support and friendship which I have been given by Tim Button, Shyane Siriwardena, Ali Boyle, Adam Bales, Georgie Stratham, Daniel Billinge, Chloé de Canson, Alexander Greenberg, Lucy Campbell, Chris Cowie, Lukas Skiba (the 'famous five'), Claire Benn, Alison Fernandes, Fredrik Nyseth, Henry Taylor, Maxime Lepoutre, James Hutton, Alex Grzankowski, Kyle Mitchell, Carlo Rossi, Owen Griffths, Brian King, Craig French, Daniel Brigham, Matthew Hasler, Helen Holmes, Alex Hall, Bob Hale, Vera Flocke, Sarah Broadie, Patrick O'Hare, Robert Wilson, Josephine Salverda, Adam Rieger, and Joe Dewhurst and his tomatoes. A special thanks must be made to Felipe Contatto. I could not have done without the support of the Fisher House family, or of Adams Road. I thank Peter Wright, Fr. Gerard Tartaglia, Bebe Geen, Rosemary Summers, Sr. Ann Swailes, Decca Muldowney, Di Trevis and Robin Cunnah for everything they have done for me. I owe thanks to the audiences of conferences where I have developed and presented my work, most notably: *Serious Metaphysics Group* in Cambridge; *Mathematical Aims Beyond Justification* in Brussels; *Emergence of Structuralism and Formalism* in Prague; and *Filmat Philosophy of Mathematics* in Italy. I thank the AHRC and the Cambridge Faculty of Philosophy for the incredible financial opportunity given to me. I thank my parents Robert and Margaret, who worked hard to fund my MPhil. I will forever be in debt to them for the selfless support and love they have given me all my life. Finally, I thank my family for somehow carrying me through: my sisters Megan, Clare, and Miriam; my brothers Sam and Mark; my aunt Jackie; my uncle Francis; and my Granny Teresa and Granny Margaret, for being my home and my foundation.

Dedicated with the best of my love to Margaret Murray

1928–2016

Abstraction's might a boon is found
While man does keep it tamed and bound;
Awful its heav'nly powers become
When that its stops and stays are gone.

*Gottlob Frege 1906/1971*c*, 123*

Contents

Introduction

Since Euclid, the axiomatic method has been extremely influential in mathematics. This method can seem philosophically neutral if we think of axioms as providing a merely descriptive characterisation of the most fundamental feature of a field of mathematics – such as geometry or analysis – and allowing us to reason about this field more rigorously. However, the relationship between an axiomatisation and what it axiomatises is rather more complicated and is of great philosophical interest. In particular, one's conception of what kind of thing an axiom is turns out to be deeply interconnected to one's conception of what kind of thing a mathematical object is. Thus, adopting a particular conception of an axiom will influence the mathematical ontology we characterise by means of those axioms; and adopting a particular conception of what there is in mathematics will constrain the conceptions of axioms suitable to characterise this reality. More concretely, if we think – like Frege – that mathematics has a fixed subject matter of mathematical objects, then axioms must be collections of thoughts which refer to these objects. In contrast, if we take Resnik's view that mathematics is the 'science of structures', then axioms must be capable of characterising and referring to the right kind of structures.

The guiding concern of this thesis is to explore this interconnectedness between an axiomatisation and the mathematical reality it characterises. It will do so indirectly, first by looking at the Frege-Hilbert disagreement over the nature and purpose of axioms, and then by exploring the supposed alternative to axioms given by abstraction principles. In both cases our central concern will be with the conception of an axiom (or abstraction principle) which is being proposed and the corresponding ontology of each view.

The main aim of the first half of the thesis will be to explore Hilbert's early ontology – since Hilbert's ontology is so closely connected to his methodology. Chapter 1 will look at Hilbert's well-known ontological slogan that consistency is enough for existence in mathematics.[1] Chapter 2 will connect this to the Frege-Hilbert controversy. Chapter 3 will show that Hilbert can be understood as having an implicit structuralist ontology. This will allow us to connect Hilbert's structuralist position with his insistence to Frege that his axioms be understood as implicit definitions of the geometric primitives.

The aim of the second half of thesis will be to do the same with Hale and Wright's neo-Fregean logicism and then to connect the two debates in such a way as to present a new objection to Hale and Wright. Chapter 4 will adapt a point made by Frege in his correspondence with Hilbert into an objection to Hale and Wright. Chapter 5 will consider how Hale and Wright might avoid this objection and it will conclude that the only clear way to avoid the objection is be for them to adopt the structuralist ontology we associated with Hilbert's early years. Chapter 6 will argue that Hale and Wright have already been forced to implicitly adopt this ontology without calling attention to it. This is perhaps unsurprising as to abandon Frege's ontology is to abandon the project of providing a *Fregean* logicism which can at the same time preserve the commitment that "...number words are to be understood as standing for independent objects" and answer Frege's question: "How, then, is a number to be given to us, if we cannot have any idea or intuition of it?" (Frege 1884, §62).

[1] An early version of chapter 1 is forthcoming in *Logique et Analyse*.

Chapter 1

Hilbert's Principle

1.1 An introduction to Hilbert's Principle

David Hilbert is the best-known proponent of the striking thesis that all that is required for existence is *consistency*. Hilbert articulates this view in his famous address to the International Congress of Mathematicians "Mathematische Probleme" 1900/1996*b*. He also sets it out in his lecture "Über den Zahlbegriff" 1900/1996*a* and in a letter he writes to Frege in 1899. In the letter we find the first and most famous formulation of his position:

> You [Frege] write "From the truth of the axioms it follows that they do not contradict one another". It interested me greatly to read this sentence of yours, because in fact for as long as I have been thinking, writing and lecturing about such things, I have always said the very opposite: *if arbitrarily chosen axioms together with everything which follows from them do not contradict one another, then they are true, and the things defined by the axioms exist.* For me that is the criterion of truth and existence (Hilbert 1899/1980*d*, 39-40).

The emphasised extract from Hilbert's letter has received much attention. On the basis of it, a general principle for mathematical ontology has been attributed to Hilbert which I call *Hilbert's Principle*:

Hilbert's Principle: In mathematics, if it is consistent for something to exist then it does exist. [1].

It is important to note that Hilbert's Principle is not a summery of the quote from Hilbert's letter. It is the attempt to extract a *thesis* from Hilbert on the basis of his remark. A lesser-known proponent of the same view is Poincaré who asserts in his paper "Mathematics and Logic" that "...in mathematics the word exist can have only one meaning, it means free from contradiction" (Poincaré 1912*b*, 454).[2]

As a general approach to ontology, such a principle is unintuitive and highly unparsimonious. Even taking into account the restriction to mathematics, the view is controversial. Consistency is very plausibly a necessary condition for the existence of mathematical entities, but why should it be considered a *sufficient* one? To answer such a question, we must be careful to understand the context in which Hilbert's Principle is given and not to assess it in a philosophical vacuum.[3] This would be unproductive because Hilbert's Principle is not, by itself, a fully-fleshed out thesis. For example, it tells us nothing of what is meant by consistency, or by what means consistency is to be secured, or what kinds of things are established to exist. Because of this, no proper assessment of Hilbert's Principle can be reached before reconstructing Hilbert's actual contention. Thus, the concern of this chapter will be neither to attack *nor* to defend Hilbert's view, but to discover it. As such, the guiding question of the chapter will be as follows:

Qu. What does Hilbert mean by Hilbert's Principle?

To answer this question, I propose that we begin with what is commonly regarded as a bad answer. Namely, that Hilbert's Principle is an anticipation of the completeness theorem. I will henceforth call this the *misguided reading* of Hilbert's Principle.

[1]Hilbert is even occasionally attributed with nothing beyond this naive formulation, see for example in (Brown 2005, 105) and (Pudlák 2013, 602)

[2]Poincaré also makes this claim in his papers "The New Logics" 1912*c*, "The Latest Efforts of the Logisticans" 1912*a*, and in his book "Science and Method" 1952.

[3]Hilbert is even occasionally explained as having nothing beyond this naive formulation, see for example in Brown (2005, 105) and Pudlák (2013, 602).

In what follows, we will give ourselves the task of asking whether there is *any* truth to this bad answer, and of articulating precisely what is misguided about it. This will require attention to many considerations which will very nicely pave the way for us to develop an alternative, historically informed, good answer to (Qu).

1.2 The misguided reading of Hilbert's Principle

In this section, we will aim to give the misguided reading a fair hearing. To do this, I will offer a defence of the misguided reading against its most damaging problem and conclude that even with a rigorous defence, the misguided reading is untenable. However, in defending the misguided reading we will have extracted a germ of truth from it, which we will use to develop a new reading in the sections that follow.

I should note here that the focus in this chapter will be on Hilbert's remark in the context of his contemporaneous writings around 1900. The next chapter will develop the priority reading by introducing the Frege-Hilbert controversy, which forms the immediate context of Hilbert's quoted remark. For now, let us carry on to our critique of the misguided reading.

1.2.1 What is misguided about the misguided reading?

As we have mentioned, on misguided reading, Hilbert's Principle is an anticipation of the completeness theorem and evidence that Hilbert assumed the completeness of his system. We can formulate this answer to (Qu) in the following way:

> **Misguided Reading of Hilbert's Principle:** If a set of sentences is consistent, then there exists a model which satisfies them.

The first defining feature of this reading is that consistency is understood proof-theoretically. In other words, it is a relation holding between *sentences* under a *closed* *specified deductive system*. The second feature is that existence is understood as the existence of a *model* for those sentences – where a model is defined as a pair consisting in a domain and an interpretation function. The domain is a (non-empty) set of objects and the interpretation function is a function which maps the names of the language to objects in the domain; n-place predicate terms to a set of n-tuples from the domain; and n-place function expressions to functions. This provides truth values for all the sentences of the language.

So, the misguided reading imposes on Hilbert a modern understanding of syntactic consistency and a modern understanding of semantic completeness. It suggests that when Hilbert tells Frege that the consistency of the axioms guarantees the existence of what they define, what he means is that *syntactic* consistency guarantees the existence of a model, and in making such a claim Hilbert is implicitly appealing to the *semantic completeness* of his system.

Moriconi (2003) has pointed out an immediate problem with this answer: this was not the conception of completeness that Hilbert had at the time. In 1900, Hilbert spoke of the completeness of an axiomatisation in the sense that the deductive closure of the axiomatisation must recapture all of the intuitively known truths of, for example, geometry. Moriconi claims that this is what Kreisel means when he stresses that the problem of semantic completeness goes beyond the Hilbertian perspective (2003, 131).[4]

It is even more important to bear in mind that at the time of formulating his principle in the correspondence with Frege, Hilbert had not yet invented *proof theory*. His first presentation of proof theory was in the lectures he gave in Hamburg, as late as 1921 (cf. Seig 1999). Although it was Hilbert who invented proof theory – and, along with it, the proof-theoretic conception – it should not be assumed that he had a proof-theoretic understanding of consistency twenty years before.

Since I am presenting the misguided reading as universally unpopular, I should

[4]Moriconi (2003) argues in his paper that Hilbert does not *assume* the completeness of his system, but uses his completeness axiom to *discharge* any existential assumptions and in this way reduces existence to consistency.

mention that there are a few remarks in the literature which come close suggesting this view, although there is no one to my knowledge who defends it. For example, Resnik, the seminal expositor of the Frege-Hilbert controversy, comments as follows:

> ...[Hilbert's Principle] can be updated and even proved as a version of the completeness theorem: every deductively consistent set of sentences has a model (Resnik 1974, 134).

However, here Resnik stops short of claiming that *Hilbert* would have thought of himself as anticipating completeness. Rather, Resnik seems to be claiming that the best way for us now to understand Hilbert's Principle is as a version of the completeness theorem. This contention is doubtful – for reasons we will later encounter – but it is not, by itself, the misguided reading. More recently, Shapiro has remarked:

> ...Hilbert said that (deductive) consistency is sufficient for 'existence', or, better, that consistency is all that remains of the traditional, metaphysical matter of existence. This much continued into the Hilbert program. If we restrict ourselves to first-order axiomatisations, then Gödel's completeness theorem does assure us that consistency implies existence. The theorem is that if a first-order axiomatisation is consistent, then it has a model: there is a system that makes the axioms true. So perhaps Hilbert's claim about consistency foreshadowed the completeness theorem (Shapiro 2005, 71).

This offhand remark of Shapiro's seems closer to the misguided reading but even here Shapiro does not clearly say that the foreshadowing is in the mind of Hilbert rather than in our perspective as we look back on his remark in light of Gödel's results.

To generalise, I think that nearly all remarks in the literature which approach the misguided reading can be explained by reading Hilbert's later famous and influential work back into his early work. After all, Hilbert invented much of the modern equipment which now seems so intuitive. What we must keep in mind, however, is that Hilbert's later invention of proof theory and mathematical definition of consistency were momentous advances which *changed* the way in which consistency and completeness *could* be conceived of. What we are concerned with is Hilbert's view when he made these advances, and it is wrong to assume that Hilbert held a single position throughout his development when in fact he changed his mind at various stages. Indeed, at the time of 1899-1905 Hilbert was not yet even a formalist, and was deeply sympathetic to Russell's logicist project (Hilbert 1918/1967*a*, 153).[5] More relevantly, Hilbert's proof-theoretic understanding of consistency and semantic conception of completeness appear much later in his writings, and he provides no definition of either, in or around 1899.

What makes the misguided reading implausible, then, is that any proponent of it must claim that Hilbert already had a proof-theoretic understanding of consistency and conception of semantic completeness, at this early stage.

1.2.2 In defence of the misguided reading

To address the burden of proof which we have just outlined, we would require an appeal to evidence from texts contemporary with Hilbert's Principle which indicate that early Hilbert had the relevant conceptions of consistency and completeness. Here, I seek to construct such a defence of the misguided reading by employing all the textual sources which I believe the proponent of the misguided reading might appeal to in support of the argument that Hilbert had a proof-theoretic understanding of consistency in this early period. In doing so, I am already giving more attention to the misguided reading than is really necessary. However, this exercise will serve as a useful explanatory strategy for setting out a nuanced picture of Hilbert's early views, so that we can use this picture as a starting point from which to develop an answer to (Qu) which is sensitive to Hilbert's remarks across all of his relevant work circa 1900.

One substantive piece of evidence for the misguided reading comes from the correspondence with Frege. Hilbert sent some papers to Frege, one of which is known to be

[5] See Sieg (2009) and Ferreirós (2009) for more on Hilbert's early logicist sympathies.

an offprint of his famous lecture "Mathematische Probleme" 1900/1996*b*. In his reply, dated September 1900, Frege notes that some parts of Hilbert's lectures gave him the impression that Hilbert had discovered a new method of proving consistency.

> It seems to me that you believe yourself to be in possession of a principle for proving lack of contradiction which is essentially different from the one... you apply in your *Festschrift*. If you are right in this it could be of immense importance though I do not believe in it as yet... It would help to clear up matters if ... you could formulate such a principle precisely and perhaps elucidate its application by an example (Frege, 1900/1980*b*, 46-50).

This passage suggests that there is a possibility that Hilbert had invented proof theory as early as 1900. However, by itself all the letter establishes is that Hilbert gave Frege the *impression* of having another approach. Furthermore, it is unclear whether this approach qualifies as proof-theoretic, or offers *another* alternative to the model-theoretic approach of Hilbert's *Festschrift*.

In order to investigate this further, we must broadly characterise what would make a conception of consistency distinctively proof-theoretic. The proof-theoretic approach, is the idea of investigating the properties of strings of symbols (or – less strictly speaking – *sentences*), rather than the propositions or truths they express. Furthermore, those sentences are considered under an explicit system of rules which dictate the legitimate inferences which can be made between the sentences. I take these two elements to constitute the distinctive characteristics of the proof-theoretic approach. Thus, to establish that Hilbert's alternative approach to proving consistency was indeed proof-theoretic, we require corroborating evidence that he was in possession of the following characteristic elements:

(A) Consistency is a relation between sentences, or, certain strings of symbols in a formal language, i.e. a language which has a finitely specifiable set of formation rules.

(B) The relata of consistency operate under a *deductive system*; i.e. the formal language has a *specified* set of *inference rules*.

The evidence for (A) comes from an important passage in Hilbert's *Festschrift*: the introduction to §9 (1899/1971). In §9, Hilbert proves the (relative) consistency of his axioms and first exhibits model-theoretic reasoning. The introduction to this section is very significant because it is one of the few places in which Hilbert explicitly discusses what he aims to establish with his consistency proofs. Ajdukiewicz quotes the relevant part of the introduction in order to lend support to his own syntactic definition of consistency:

> Consistency is conceived by Hilbert in the way it was defined by us, since he writes: "The given axioms are not inconsistent i.e. it is not possible to derive logically from them a sentence contradicting any of the axioms" (Ajdukiewicz 1996, 23).

Here, it seems that Hilbert does indeed speak of sentences as the relata of the consistency relation. Strictly speaking, what Hilbert calls a sentence is a deductive consequent of an axiom, rather than an axiom, but we can assume he thought the relation of logical consequence holds between the same kinds of relata. This gives good contemporary textual support that Hilbert had (A), i.e. that he already thought of consistency as a relation between sentences.

There is also evidence that Hilbert had (B), i.e. that his conception of consistency was proof-theoretic because he understood consistency as holding between sentences *under a set of formally specified deductive rules*. This evidence comes from the content of Hilbert's famous address "Mathematische Probleme", which he gave in Paris to the International Congress of Mathematicians in 1900. He is known to have sent an offprint of this lecture to Frege (Gabriel et al. 1980, 49, IV/7. ft. 1). In this lecture Hilbert offered a proof sketch along the lines of a proof-theoretic approach:

> Now I am convinced that we must succeed in finding a direct proof that arithmetical axioms are free from contradiction, if we carefully work through

> the known methods of inference in the theory of irrational numbers with that aim in view and try to modify them in a suitable manner (Hilbert 1900/1996*b*, 50, ft. 4).

It is significant that here we have the idea of using a collection of inferential methods to attempt to modify – or rather, to articulate the deductive consequences of – a system of axioms. Further, Hilbert implies that there is a way to survey all "known" methods of inference in a field of mathematics. If the available inferential methods are known, this suggests they are finitely specifiable and thus that they will admit of a formal specification. Here we also see a hint of the notion of a closed deductive system where Hilbert localises these inferential methods to a particular theory: they are the methods used *in the theory of irrational numbers*. Altogether, this gives evidence that Hilbert had (B), the idea of a closed system of deductive rules which are formally specifiable. Moreover, Hilbert is advocating that such a system of rules should be the means by which we investigate consistency – in particular that the axioms of arithmetic should be investigated by checking whether they would lead to contradiction under any of the known methods of inference in the relevant theory.

However, it is not until Hilbert's 1904/1967*b* address "Über die Grundlagen der Logik und der Arithmetik" that he makes more explicit allusions as to how this investigation is to be carried out. Here he outlines a method of establishing the consistency of arithmetic directly by translating the mathematical proofs into a formal language and then taking the formal language itself as the object of study. The aim of this approach, he tells us, is to provide a proof that a formal contradiction could never be derived in the system (Hilbert 1904/1967*b*, 135). Here we see the two elements of the proof-theoretic conception (A) and (B), coming together. Truly, this is a recognisable sketch of the proof-theoretic method.

This section has presented the contemporaneous textual evidence that Hilbert had a modern proof-theoretic understanding of consistency around 1900. What it shows – I think – is that the misguided reading is not so obviously misguided as it first appeared to be. It seems that in 1900 Hilbert was already anticipating much of the apparatus that he would later be famed for inventing. In the next section I will argue that a careful re-examination of the best evidence shows much of it to be inconclusive. However, it will remain the case that there is an element of truth in the misguided reading – which we will aim to extract.

1.2.3 Critiquing the defence of the misguided reading

In this section I will argue that, although the evidence I have presented to support the misguided reading appears very strong, it is not enough to establish that Hilbert had a modern proof-theoretic understanding of consistency around 1900.

Let us first return to the evidence from the correspondence with Frege. What the correspondence makes clear is that as early as 1900, Frege had the impression that Hilbert had a method for proving consistency which was distinct from the model-theoretic method in *Festschrift*. It is also apparent that Frege was sceptical of this method. As we already noted about this source, in order to show that Hilbert's idea for establishing consistency was in fact the proof-theoretic method, we must refer to other textual sources.

Evidence that Hilbert had thought of (A), consistency as a relation between sentences, came from an appeal to the introduction of §9 of Hilbert's *Festschrift*. However, the important quote used by Ajdukiewicz is actually misleading. If we return to the primary text we see that what Hilbert actually says is:

> Die Axiome der fünf in Kapitel I aufgestellten Axiomgruppen stehen miteinander nicht in Widerspruch, d.h. es ist nicht möglich, durch logische Schlüsse aus denselben eine Tatsache abzuleiten, welche einem der aufgestellten Axiome widerspricht. Um dies einzusehen, genügt es, eine Geometrie anzugeben, in der sämtliche Axiome der fünf Gruppen erfüllt sind (Hilbert 1899/1971, §9).

Ajdukiewicz has mistranslated "tatsache" as "sentence" when it is the ordinary word for *fact*.[6] This striking mistranslation is explained by the more general problematic tendency to read back central elements of Hilbert's later and influential work into his early writings, in particular his formalism and his proof theory. A more faithful translation of Hilbert's introduction is the following:

> *The axioms of the five groups of axioms laid down in chapter 1 do not stand in contradiction to each other, i.e. it is not possible to derive, from the axioms, through logical reasoning (Schlussfolgerung), a fact (Tatsache) which contradicts one of those axioms that were laid down. To see this it is sufficient to present a geometry in which all of the axioms of the five groups are satisfied* (Hilbert 1899/1971, §9, translation mine).

If we pay attention to the vague and vacillating terminology employed by Hilbert to refer to his axioms, it becomes clear that Hilbert's conception of whether his axioms are syntactic or semantic is just as vague and vacillating. Most of the time Hilbert refers to his axioms simply as "Axiome"; in §9 above, he refers to them as facts. Importantly, he speaks of reinterpreting his axioms, which implies they are syntactic. However, in the correspondence he sometimes slips into calling them concepts (e.g. Hilbert, 1899/1980*d*, 42) and in his lectures even talks of "thought-objects" which are not themselves syntactic but are "denoted by a sign" (Hilbert 1904/1967*b*, 131). In short, there is no textual evidence to show that at around 1900 Hilbert had already made the leap to (A) and thought of the proper relata of consistency as a mere string of symbols. Rather, Hilbert's conception was still ambiguous since he saw no need to be precise about whether or not an axiom was a strictly formal entity.

We also considered evidence that Hilbert thought (B) – that the relata of consistency operate under a formally specified deductive system. This came from Hilbert's two addresses: "Mathematische Probleme" 1900/1996*b* and "Über die Grundlagen der Logik

[6]This is not a question of a difference in translation; any translation of *tatsache* will render it as more than a syntactic notion.

und der Arithmetik" 1904/1967*b*. It is true that in "Mathematische Probleme" Hilbert speaks of being convinced that it is possible to provide a proof of the consistency of the axioms of arithmetic without appeal to the existence of the arithmetic primitives. He suggests that this can be done by an examination of the axioms – in particular, by checking whether any inconsistency arises from applying all known methods of inference to the axioms. However, this is insufficient to infer that Hilbert could specify a deductive system *formally*. What Hilbert says here is *compatible* with the proof-theoretic method – and (B) in particular – but by itself it is too meagre to *constitute* (B). In other words, what Hilbert delivers in this address is a manifesto, and not a formally specified deductive system.

In the 1904/1967*b* address "Über die Grundlagen der Logik und der Arithmetik", Hilbert gives a much more substantive account of how a syntactic consistency proof is to be carried out. What he presents there can certainly be regarded as a sketch of the proof-theoretic method. However, taking into account some other aspects of Hilbert's view at the time, it is clear that Hilbert straightforwardly lacked the tools to realise this sketch. Most importantly, around 1900 Hilbert did not yet have a rigorous logical formalism.[7] Held back by this lack (and also in part by Poincaré's objection that Hilbert's proof sketch required a circular appeal to induction) Hilbert did not return to his work on the foundations of mathematics until 1917 and did not present his proof theory until the 1920s.[8] Thus, since Hilbert lacked a rigorous logical formalism, he would not have been able to specify a deductive system *formally* in 1904, which is to say that he lacked (B).

The case for the misguided reading is undermined by two simple and uncontentious points. The first is that Hilbert's conception of the relata of a negative consequence relation like consistency was not purely syntactic (which undermines that he had A).

[7] cf. Zach 2016. Peckhaus (1991) argues that the reason for this was that Hilbert's conception of logic was algebraic – which made it difficult for him to conceive of formalising the axioms of mathematics.

[8] See (Sieg 1998, 5) and Hilbert (1922/2013). For more on the chronology of proof theory see Zach (2016) and von Plato (2016).

The second is that he lacked a logical formalism (which undermines that he had B). Blanchette – for one – observes both of these uncontentious points,

> Hilbert had not yet specified a syntactic deductive system and does not view logical deduction as formal symbol-manipulation (Blanchette 1996, 321, ft. 8).

When we bring these observations to bear on what Hilbert suggests in *Über die Grundlagen der Logik und der Arithmetik*, they show that in 1904 Hilbert may have offered a sketch along proof-theoretic lines, but he was not in a position to realise that sketch, precisely because he lacked (A) and (B), which we have taken to be the characteristic elements of the proof-theoretic approach.

Therefore, there is no conclusive evidence that Hilbert had a proof-theoretic understanding of consistency around 1900, in so far as we take the proof-theoretic understanding of consistency to be characterised by the following:

(A) Consistency is a relation between sentences, or, certain strings of symbols in a formal language, i.e. a language which has a finitely specifiable set of formation rules.

(B) The relata of consistency operate under a *deductive system*; i.e. the formal language has a *specified* set of *inference rules*.

However, there is a closely related notion of syntactic consistency, which – although it falls short of our modern one – is nevertheless significant because Hilbert seems to have an explicit understanding of it, in this early period. What we have seen is that Hilbert did not yet think a proof of consistency required checking for inconsistency in the *strings of symbols* which could be derived from the axioms by a *formally specified deductive system*. Nevertheless, he did think a proof of consistency required checking for inconsistency in all of the *facts/sentences* which could be deduced from the axioms by all available *logical reasoning*. Clearly, the latter is still a species of syntactic consistency.

Furthermore, it shares and even anticipates some of the central features of the proof-theoretic conception. As such, it is evidence that Hilbert had already understood some key ingredients of his later proof-theoretic approach. Thus, let us call this latter kind of syntactic consistency, *proto-proof-theoretic* consistency.

Perhaps the most philosophically important feature of Hilbert's proto-proof-theoretic conception of consistency, is that – whether or not he had the formal tools to realise it – Hilbert already thought there was *some* way of proving the consistency of his axioms without making appeal to existential assumptions (this feature will form the basis of the new reading of Hilbert's Principle). Furthermore, Hilbert already has the idea that the way to go about this is to *somehow* identify the legitimate inferences in a field of mathematics and work through them, checking whether the axioms yield any inconsistency. Hilbert had already made another important conceptual advance; that of investigating the consistency of some axioms by turning the axioms – and the rules which governed them – into the objects of study. So that, as with any other branch of mathematics, we could offer a formal proof of the properties of this system. With this approach, meta-mathematics was born.

In conclusion, an examination of Hilbert's contemporaneous writings gives us reason to be uncomfortable with the fact that the misguided reading of Hilbert's Principle interprets Hilbert as having a modern understanding of consistency and completeness. However, the examination also showed that Hilbert does have an early prototype of the proof-theoretic conception around 1900. Paying attention to this prototype, we saw that it is tantamount to the broad brush strokes of a proof-theoretic approach, although Hilbert did not yet have the tools to realise the defining features ((A), (B)) of a modern proof-theoretic proof.

The next section will develop a reading of Hilbert's Principle which respects the chronological development of Hilbert's thought and methodology and begins with the insight that early Hilbert had a proto-proof-theoretic approach.

1.3 A new reading of Hilbert's Principle

Uncovering the extent of Hilbert's early conception of the proof-theoretic approach has provided us with a concrete means of denying that Hilbert can be understood as anticipating the completeness theorem in 1900. More importantly, it has provided us with an understanding of how much of the proof-theoretic conception Hilbert had already developed. This latter insight will form the backbone of our understanding of Hilbert's Principle, which will be set out in this section. If this is the backbone, then the remaining skeleton will be provided by taking into account the immediate context of Hilbert's Principle and – in particular – the fact that Hilbert's remark is elicited as a *response* to Frege.

Before we begin, let us remind ourselves that Hilbert's Principle is not a direct interpretation of Hilbert's quoted remarks; rather, what is at stake is the explanation of *why* Hilbert makes the remarks that he does, i.e. the answer to (Qu).

1.3.1 Hilbert's Principle as a response to Frege

The kernel of the new reading of Hilbert's Principle – and the answer to (Qu) – is that Hilbert meant to bring one of the central features of his proto-proof-theoretic approach into contrast with Frege. In particular, Hilbert meant to contrast his conception of the relationship between the consistency of the axioms and the existence of the theory's primitives with Frege's understanding of that relationship.

To unpack this answer, let us begin with the uncontroversial observation that whatever explanation is given of Hilbert's contention, it must be one which coheres with the context in which Hilbert makes his remark in the first place. As such, we cannot analyse Hilbert's Principle using an isolated remark. In that regard, it is vital to recognise that Hilbert formulates this principle *as a comment on a remark made by Frege*. When this is taken into consideration, I believe it can throw much light on Hilbert's controversial principle.

The first thing we should observe is that Frege explicitly restricts his attention to mathematics when he says, "I should like to divide up the totality of *mathematical* propositions into definitions and all the remaining propositions..."(Frege 1899/1980*e*, 36, emphasis mine). If Hilbert intended his principle to apply in other domains, he would have had to cancel the restriction implicit in the discussion, but Hilbert does not do this. Remembering that Hilbert does nothing to indicate that such a principle would apply beyond the mathematical realm means that Hilbert can avoid the more simplistic counterexamples to Hilbert's Principle – granted we set aside the problem of demarcating the domain of mathematics. Localising the principle may soften its dissidence, but even in the mathematical case it is still unclear what it means to say that consistency is the *criterion* for existence.

This brings us to the second feature, which becomes clearer when we bear in mind that Hilbert's remark is intended as a comment to Frege. In it, Hilbert explains to Frege the way to understand his (and allegedly Cantor's) consistency proofs. The full remark made by Hilbert is as follows:

> You [Frege] write "From the truth of the axioms it follows that they do not contradict one another". It interested me greatly to read this sentence of yours, because in fact for as long as I have been thinking, writing and lecturing about such things, I have always said the very opposite: if arbitrarily chosen axioms together with everything which follows from them do not contradict one another, then they are true, and the things defined through the axioms exist. For me that is the criterion of truth and existence. The proposition 'Every equation has a root' is true, and the existence of a root is proven, as soon as the axiom 'Every equation has a root' can be added to the other arithmetical axioms, without raising the possibility of contradiction, no matter what conclusions are drawn. **This conception** is indeed the key to an understanding not just of my *Festschrift* but also for example of the lecture I just delivered in Munich on the axioms of arithmetic, where I prove or

> at least indicate how one can prove that the system of all ordinary real numbers exists, whereas the system of all Cantorian cardinal numbers or of all alephs does not exist – as Cantor himself asserts in a similar sense and only in somewhat different words (Hilbert, 1899/1980*d*, 39-40, emphasis mine).

In order to explain his proofs to Frege, Hilbert does not offer a proof sketch; instead he offers a particular *conception* of the relationship between consistency and existence. We saw that as part of Hilbert's proto-proof-theoretic approach, he already had the idea of establishing the consistency of axioms by a demonstration that no inconsistency would result from any application of the legitimate inferences which could be made from the axiom set. That is to say, he had the idea of establishing the axiom set's consistency, without appeal to any existential assumptions. In the context of disputing Frege's remark, the significance of the idea of establishing an axiom set's consistency in this way is that it introduces an *alternative* to the traditional way of establishing the consistency of the axioms. Frege's quoted remark endorses this traditional way of establishing consistency by appeal to the *truth* of the axioms (thus assuming the existence of the primitives referred to by the axioms). Therefore, I think that Hilbert's intention in his (full) remark is to contrast the features of his proto-proof-theoretic conception of consistency with the approach he found in Frege.

As such, the key to understanding Hilbert's remark is that his emphasis is not on anticipating a technical result, nor on advocating a generalisable *ontological* principle; it is to demonstrate the advantages of his alternative and fruitful conception of *consistency* in mathematics. Let us be quite precise about what this conception is. What Hilbert is bringing into contrast with Frege (and presenting as an advantage of his proto-proof-theoretic approach to consistency) is a particular understanding of the *relationship* between the consistency of an axiom set and the existence of the axiom set's primitives.

This answer to (Qu) avoids interpreting early Hilbert as having an understanding of consistency and completeness that outstrips his methods of proof, but it nevertheless accommodates the fact that he had already made progress towards a proof-theoretic understanding of consistency. However, if we are to develop these considerations into a full and satisfying answer to (Qu), we must examine in detail Hilbert's early conception of the aforementioned relationship between an axiom set's consistency and the existence of the axiom set's primitives.

1.3.2 The priority reading

Let us begin by returning to (Qu) *what does Hilbert mean by Hilbert's Principle?* I think we can make the contention of Hilbert's Principle explicit by first articulating the following two conditions:

1. There is no non-circular way to establish the existence of x which does not rely on the consistency of y.

2. There *is* a non-circular way of establishing the consistency of y which does not rely on the existence of x.

There is a lot is packed into (1) and (2) here. We can make things a bit clearer by distinguishing two further conditions. For two concepts A and B, it can be established that some x falls under A *directly* if there is a way of establishing that x falls under A which does not make any reference to B. Further, A and B are *connected* if there is a way of establishing that x falls under B using an appeal to the fact that x falls under A. For example, let A be the concept of corresponding with letters and B be the concept of disagreeing and let x and y pick out the relevant pair of two distinct German mathematicians. Then A is *connected* to B because one can establish that the mathematicians disagree by appeal to their writing letters to each other. Further, A can be established *directly* because one can establish that the mathematicians were corresponding with letters without appeal to their disagreeing. Let us label these two further conditions as follows:

Connect. For any x falling under A, there is a way of establishing that y falls under B by substantive appeal to the fact that x falls under A.[9]

Direct. For any x falling under B, there is a way of establishing that y falls under A without making any appeal to the fact that x falls under B.

If both (Connect) and (Direct) hold, then there is a non-circular way of establishing A which does not rely on B. So, in the case of our example, there is a non-circular way of establishing that two people wrote letters to each other which does not rely on them being in disagreement. If (Connect) or (Direct) do not hold then there is *not* a non-circular way of establishing A which does not rely on B. Again, in our example, there is not a non-circular way of establishing that Frege and Hilbert are disagreeing which does not rely on their writing letters to each other.

To introduce some terminology to make this relationship easier to think about, let us say that a *conceptual priority relation* holds between A and B if there is an asymmetry between (Connect) and (Direct) when these variables are reversed. That is to say that it is not the case that (Connect) and (Direct) are satisfied for A and B in the same way they are satisfied for B and A. So, in our example, the concept of corresponding is conceptually prior to the concept of disagreeing. I will henceforth refer to the conditions (Connect) and (Direct) above as the *priority schema*. From the beginning, it is important to emphasise that the particular species of conceptual priority which will concern us is not an epistemic, semantic, or ontic relation. To say that two concepts stand in a conceptual priority relation is not to say that one concept is *reducible* to the other or is in any way *contained* within the other or that it can only be *understood* using the other. I mean only to say that there is an asymmetry between the way in which the concepts are related such that a proof of one requires an appeal to the other, but not vice versa. Essentially, what is doing the work of determining the priority in this relationship between concepts

[9] It ought to be specified that A is doing *substantive* work in the proof because we can gerrymander any proof to introduce and eliminate an appeal to concepts which are irrelevant to the argument of the proof.

is the *order to proof*.[10] In some cases, of course, this ordering may evidence that the relevant concepts also stand in an epistemic, semantic, or ontic priority relation, but in what follows we will remain agnostic regarding these more common conceptual priority relations.

Applying all this apparatus to the relevant case, Hilbert's claim is – very roughly – that consistency is direct and is connected to existence, i.e. (2), but that existence is not direct and not connected to consistency, i.e. (1). Which is to say that the reverse instantiations of the schema are asymmetric. Consistency will thus count as the prior concept in virtue of its being both direct and connected. It is worth repeating that what we have said here using (Connect) and (Direct) is equivalent to what we have said using (1) and (2), which is equivalent to how we are understanding a conceptual priority claim. The priority schema merely provides a very detailed way of saying that Hilbert's conception of the relationship between consistency and existence is that these metamathematical concepts are conceptually interconnected and that – with respect to proof – the entry point to the conceptual circle is *consistency*. It is consistency which can and should be used to prove existential statements in mathematics.

In characterising Hilbert's remark in this way, however, we have been talking very loosely. As Hilbert does not claim straightforwardly that the criterion for existence is consistency, but rather that his criterion for existence is given by the principle that the consistency of some "axioms" suffices for the existence of "what those axioms define". He says nothing to indicate that the priority relation of consistency over existence holds *in general* but only with respect to these two instantiations of x and y. In order to speak more exactly, therefore, we must return to the priority schema and not only instantiate A and B with consistency and existence, but also x and y with *axioms* and *what those axioms define*, respectively. In the case under discussion by Frege and Hilbert, the relevant axioms are those of Hilbert's re-axiomatisation of Euclidean geometry and those axioms define the primitive geometric terms such as "point", "line", "congruence" etc.[11]

[10] For reasons which will later become apparent, *proof* here should be understood in more general sense than mathematical proof.

[11] Note that Hilbert's remark allows for an even more general formulation of (Connect) and (Direct) than

Taking this into account yields the following four conditions:

Connect (1). There is a way of establishing the consistency of the axioms using substantive appeal to the existence of the geometric primitives.

Direct (1). There is a way of establishing the existence of the geometric primitives without making any appeal to the consistency of the axioms.

Connect (2). There is a way of establishing the existence of the geometric primitives using substantive appeal to the consistency of the axioms.

Direct (2). There is a way of establishing the consistency of the axioms without making any appeal to existence of the geometric primitives.[12]

Using these two instantiations of the priority schema, we can say more precisely that Hilbert would accept Connect (2) and Direct (2) and reject Connect (1) and Direct (1). In the next chapter, after a throughout exposition of the Frege-Hilbert controversy, I will present textual evidence in support of this. For now, let us step back and consider what this detailed characterisation of Hilbert's contention amounts to.

Most simply, it shows that Hilbert's Principle can be explained using a species of priority claim. Again, it is worth emphasising that this priority claim has nothing to do with one concept *grounding* another or being *reducible* to another or *contained* within the other; it is merely concerned with the concepts having some particular asymmetric

the one above. I have restricted attention to the geometric case merely because it is the one of relevance to the Frege-Hilbert controversy. But it is clear that the fate of this particular claim has implications for the fate of the more general priority claim that Hilbert supports, namely: the consistency of *any* axiomatisation is conceptually prior (in terms of proof) to the existence of *what those axioms define*.

[12]Note that although what Hilbert actually says is that his concern is with the existence of what the axioms *define*, Frege denies that the axioms are definitions and is instead concerned with the referents of the geometric primitive expressions, such as 'point' and 'line'. To neutralise this difference we can say that both are concerned with the existence of the primitives, whether these primitives are understood to be secured by definition, or, as the referents of the non-logical expressions in a sentence whose meaning is fully determinate.

relation to each other. The asymmetry is in the fact that the axioms' consistency can be proven without appeal to the primitives' existence, but a proof of the primitives' existence needs an appeal to axioms' consistency. This gives us a new way of answering the question (Qu.): *what does Hilbert mean by Hilbert's Principle?* This new understanding – which I call the priority reading – is as follows:

> **Priority Reading of Hilbert's Principle:** Consistency of the axioms is conceptually prior to the existence of the geometric primitives.

So where Hilbert talks to Frege of the conception that is needed to understand his proof, he is speaking about a conception of these concepts (consistency of an axiom set and existence of an axiom set's primitives) as interrelated in mathematics with regard to order of proof.

1.3.3 Reflections on the priority reading

The initial interest in Hilbert's controversial principle came about because of its centrality to Hilbert's philosophy of mathematics. Across the next three chapters we will see that using the priority reading to better understand and articulate the contention behind Hilbert's Principle will not only help us to explain Hilbert's early ontological position but also the elusive point of dispute in the Frege-Hilbert controversy. In the remainder of this chapter we will carefully articulate the difference between the priority reading and the misguided reading, and then make some suggestive remarks about what our discussion has so far brought to light about Hilbert's early ontological position. The latter is an issue we will return to at the end of chapter 3.

As a result of the priority reading, the dissonance of Hilbert's controversial remark appears somewhat softened. First and foremost, Hilbert's Principle is a local and not a general (implausible) ontological principle. Furthermore, it is localised not only to the domain of mathematics but to the special relationship between axioms and the mathematical reality they characterise. However, this is not to suggest that Hilbert's conception of the relationship between consistency and existence is immune from criticism as it is eminently contestable that the consistency of a set of axioms merits inference to the existence of the primitive objects of that theory, especially if one is a Platonist about mathematical objects. Nevertheless, it is clear that Hilbert's Principle is not as easy to refute as a naïve reading takes it to be, and that in order to critique it we must first draw out the subtlety and motivation of the background conception which Hilbert is advocating.

An obvious and immediate concern which we might have about the priority reading is that it is entirely compatible with the misguided reading. After all, the claim that Hilbert was anticipating the completeness theorem is compatible with the claim that Hilbert thought the consistency of an axiom set was conceptually prior to the existence of the primitives of that axiom set. If Hilbert already had the conception of completeness and of consistency required to formulate the completeness theorem then of course he would think of the consistency of an axiom set as being established directly and as being connected to the existence of a model. This is worrying because the priority reading looks to be in danger of being subsumed by the very reading it is intended to improve upon.

This worry actually highlights the distinguishing claim of the priority reading: that Hilbert conceived of the consistency of an axiom set as prior to the existence of its primitives, and that *this is the extent of his conception at the time*. This contention is supported by the textual evidence we have considered, since the evidence shows that Hilbert *did* have an early prototype of the proof theoretic method but that in 1900 he still fell short of the kind of proof-theoretic conception which he would need to formulate the completeness theorem. In virtue of his proto-proof-theoretic conception, Hilbert would have thought that there was a direct method of establishing the consistency of an axiom set and that this was connected to the existence of the axiom set's primitives. But this is different from the claim that Hilbert already had the proof-theoretic conception of consistency, as characterised by (A) and (B).

This divergence between the misguided reading and the priority reading makes all the difference – not just in terms of historical accuracy, but also in terms of the philosophical significance of Hilbert's Principle. For, if the misguided view where correct, then Hilbert's Principle as an anticipation of completeness, is undermined by the fact that Hilbert failed to anticipate Gödel's results. Curtis Franks (treading dangerously close to the misguided reading) makes this observation:

> As a doctrine of mathematical existence, [Hilbert's Principle] is doubly dubious... As Gödel would emphasize, it is careless to define existence in this way, because the validity of that inference depends on the completeness of the underlying logic. Among the reasons that a contradiction might be underivable from a set of axioms is the possibility that the logic used is too meagre to fully capture the semantic entailment relation. In the case of first-order theories, consistency does indeed imply the existence of a model, but the incompleteness of higher-order logic with respect to its standard semantics leaves open the possibility of consistent theories that are not satisfied by any structure at all (Franks forthcoming, 4).

While Gödel's completeness theorem shows that the misguided reading of Hilbert's Principle is true in the first-order case, Gödel's incompleteness theorems show that the contention is false in the second-order case. Thus, the misguided reading would render Hilbert's Principle weak and philosophically uninteresting because Gödel's results show it to be restricted only to first-order cases and because it was superseded by the completeness theorem itself.

The importance of the priority reading is not only to avoid historical inaccuracy, but also to make the philosophical significance of Hilbert's Principle more apparent. Drawing out some preliminary philosophical implications, we can observe that – since an axiom set's consistency is the very criterion for the existence of the set's primitives, Hilbert would maintain that all things defined by a consistent set of axioms must exist (at least, in the restricted domain of mathematics). To marry axiom-consistency and primitive-existence in this way means that it is incoherent to think of the existence of anything which can be defined only by an inconsistent set of axioms; and it is incoherent to think of things which can be defined by a consistent set of axioms and which do not exist. If we understand existence in mathematics as aligning with possibility, and possibility as aligning with consistency, then Hilbert can be understood as proposing a kind of *maximalism* with respect to mathematics. On this view, everything that *could exist* (mathematically) *does exist* – and what could exist is given by what is it consistent to define. In making this suggestion we must be careful about issues around defining mathematical possibility. Whether mathematical possibility it is understood as a kind of logical possibility or as a kind of conceptual possibility, the very formulation of a maximalist view in mathematics is made difficult by problems akin to the paradox given by the universal set.[13] By the end of chapter 3 we will see that a better way of understanding Hilbert's metaphysics of mathematics is that he supports an intuitive kind of mathematical structuralism. Nevertheless, when Hilbert's Principle is better understood in isolation it opens up the door to many interesting positions regarding the ontology of mathematics.

There is another aspect of Hilbert's Principle that is of philosophical interest. This is the sense in which adopting a suitable conception of consistency and existence is part of a successful proof – or at least – is a necessary prerequisite of a proof. Hilbert tells Frege that it is his *conception* which is the kernel of his consistency proof. He speaks of the required conception with a striking subjectivity and detachment. He does not tell Frege that he is wrong or misguided, but merely notes with interest that his own conception is "the very opposite" to Frege's and that *his* is the conception needed to understand his consistency proofs and also Cantor's remarks. In the passage below I have highlighted the subjective way Hilbert speaks of his conception:

> It *interested me greatly* to read this sentence of yours, because in fact for as long as I have been thinking, writing and lecturing about such things, *I have always said the very opposite*: if arbitrarily chosen axioms together with everything which follows from them do not contradict one another, then they are true,

[13]This difficulty in formulation echoes the difficulties in formulating the second principle of plenitude in set theory, where we are tempted to say that there exist all the collections that are possible (see Potter 2004, 56).

> and the things defined through the axioms exist. *For me* that is the criterion of truth and existence... *This conception* is indeed the key to an understanding not just of my *Festschrift* but also for example of the lecture I just delivered in Munich on the axioms of arithmetic, where I prove or at least indicate how one can prove that the system of all ordinary real numbers exists, whereas the system of all Cantorian cardinal numbers or of all alephs does not exist – as *Cantor himself asserts* in a similar sense and only in somewhat different words (Hilbert, 1899/1980*d*, 39-40, emphasis mine).

I think what Hilbert's detachment here shows is a certain deliberateness about the conception he uses in his thinking, writing, and lecturing. Hilbert does not speak as if his conception is the only available one or because he takes it to be the only coherent and correct one. Rather, Hilbert's conception of the relationship between axiom-consistency and primitive-existence is *part of his proof* in the sense that *his very conception is part of his advance*. What we can take from this is that – for Hilbert – the way in which we understand and elucidate meta-mathematical concepts for use in mathematics is not arbitrary, nor is it intended to capture our intuitions or common usage. The conception which we adopt is instead primarily constrained by how fruitful that conception is in facilitating proofs and understanding. So that, if the conception is fruitful in facilitating a proof then the success of the proof legitimates the conception in the same way that an erroneous proof might call the underlying conception into question. The conception will be further enhanced by the other methods used in carrying out the proof. For example, we have characterised a proof-theoretic conception of consistency by appeal to two facets of its methodology: (A) that it is a relation between strings of symbols in a language that (B) has a set of specified formation rules and a set of specified inference rules. The methodology-first approach which I am attributing to Hilbert is already widespread (in no small part due to his influence) and is already typically used to characterise proof-theoretic consistency, and indeed model-theoretic consistency. Quite simply, the former relation is the kind of consistency that is established by proof-theoretic means, and the latter is the kind of consistency that is established by model-theoretic means.

These reflections serve to highlight that recovering the contention behind Hilbert's Principle is important not just for historical accuracy but also in order to recover the philosophical interest of that contention.

Conclusion

The guiding question of this chapter was (Qu.) *what does Hilbert mean by Hilbert's Principle?* When we take into account the time period and context in which Hilbert states his famous principle, we see that he already had an early proto-proof-theoretic conception of consistency, although it is implausible that he was anticipating the completeness theorem because, given the development of his methodology around 1900, this is simply too sophisticated a position for Hilbert to have held. Taking care to preserve Hilbert's early insights while not reading too much modern formal equipment into his early work allowed us to begin to recover the philosophical views which paved the way for his ground-breaking discoveries in logic and which remain an interesting contribution to the ontology of mathematics.

Thus we developed the priority reading from the ashes of the misguided reading. The principle insight of the priority reading is that Hilbert's Principle is not an anticipation of a formal result, nor is it a mere conditional statement or a statement of the necessary and sufficient conditions for existence. The answer to (Qu.) is that Hilbert is advocating the most important philosophical aspect of his proto-proof-theoretic conception to Frege: that axiom-consistency and primitive-existence are interconnected – not in the sense of linguistic or conceptual synonymy, but in the sense that one is conceptually prior to the other. This conceptual priority is best understood in terms of an asymmetry which issues from the fact that consistency admits of formal proof in a way that existence does not. Hilbert's intention is to make questions of mathematical existence tractable and rigorous by means of a conception of the relationship between the two meta-mathematical concepts whereby an axiom set's consistency is used to guide the fruitful investigation of mathematical reality.

Chapter 2

The Frege-Hilbert Controversy

Introduction

Between the 1^{st} of October 1895 and the 7^{th} of November 1903, Gottlob Frege and David Hilbert exchanged a number of letters regarding Hilbert's *Grundlagen der Geometrie* disputing what an axiomatisation requires in order to successfully characterise the theory of geometry. This has become known as the Frege-Hilbert controversy. The substance and nature of Frege and Hilbert's disagreement is notoriously elusive; both struggle to understand the other's view and to clearly articulate their own due to their implicit and diametrically opposed understandings of the fundamental nature and purpose of the axiomatic method. The fruits of their exchange lie not in the explicit faring of one view over another, but in how their incompatible conceptions force both to articulate and argue for what is most central to their purpose and method.

The previous chapter unpacked the contention behind Hilbert's Principle by appealing to the broader context in which Hilbert made his remark; specifically referencing Hilbert's contemporaneous writings and his fuller remark. The first aim of this chapter will be to show that the priority reading we have developed is coherent with the immediate context of Hilbert's remark – namely – the Frege-Hilbert controversy. We will then go on to show that the priority reading and the Frege-Hilbert controversy turn out to be mutually enhancing. Explaining how the priority reading coheres with the Frege-Hilbert controversy will lead us to a more sophisticated development of the priority reading; this improved reading will in turn prove to be a useful way of characterising and identifying the elusive nature of the disagreement between Frege and Hilbert.

In the first section, we will expound the Frege-Hilbert controversy. Since the Frege-Hilbert controversy will be a recurring topic of this thesis, it will be useful for this section to provide a very detailed exposition of the dispute. However, for the sake of clarity the exposition will be non-linear. And, to avoid repetition, the exposition will non-exhaustive.

2.1 The axioms and aims of Hilbert's *Grundlagen der Geometrie*

In the winter of 1898-1899 Hilbert delivered a series of lectures on Euclidean geometry which led to his influential work *Grundlagen der Geometrie*. Hilbert continued to amend the precise content and use of his axioms as well as their ordering for the book's next seven editions. As already noted, we will refer to an early version of Hilbert's monograph called the *Festschrift*.

In order to understand the objections which Frege raises against the approach of Hilbert's *Grundlagen der Geometrie* we must first set out the aims and achievements of Hilbert's seminal work.

2.1.1 The aims of the *Festschrift*

Hilbert's seminal work not only advanced the field of geometry; it also entrenched the axiomatic method in the methodology of mathematics and logic alike. In demonstrating the fruitfulness of his formal conception of axioms, his work established the foundations for our modern conception of properties, such as consistency and independence as well as providing the now standard fully axiomatised version of Euclid's original system and the model-theoretic approach. Hilbert introduces his *Festschrift* as follows:

> Geometry, like arithmetic, requires for its logical development only a small number of simple, fundamental principles. These fundamental principles are called the axioms of geometry. The choice of the axioms and the investigation of their relations to one another is a problem which, since the time of Euclid, has been discussed in numerous excellent memoirs to be found in the mathematical literature. This problem is tantamount to the logical analysis of our intuition of space. The following investigation is a new attempt to choose for geometry a *simple* and *complete* set of *independent* axioms and to deduce from these the most important geometrical theorems in such a manner as to bring out as clearly as possible the significance of the different groups of axioms and the scope of the conclusions to be derived from the individual axioms (Hilbert 1899/1971, xi).

The task of unpacking the logical consequences of a set of axioms and the inquiry into the properties of those axioms, Hilbert here takes to be one in common with Euclid. To understand the starting point of Hilbert's investigation, it will prove helpful to briefly return to Euclid's *Elements*.

In his masterpiece – used as a textbook for around 2,000 years – Euclid attempted to derive 465 geometric theorems from 5 axioms, 23 definitions, and 5 common notions. His common notions are essentially logical axioms which work towards specifying the rules of a deductive system. His definitions fall into two kinds: those which allow us to determine which figures have certain properties (like congruence) and those which explain the meaning of the five primitives: point, line, lies on, congruence and between. For example, "A point is that of which there is no part", "A line is length without breadth" (Euclid 1956, book I). Further work on Euclid's systematics revealed that not all of his theorems follow directly from his axioms. They rely on legitimate but nonetheless inexplicit appeal to features of diagrams and intuitions in order to guarantee, for instance, that circles intersect, that any line contains at least two points, and his appeal to superposition, such as in the fourth construction.

Hilbert's aim was to fully rigorize Euclid's system, to identify a set of axioms which can comprehensively characterise geometry so that even the implicit reasoning that Euclid uses in some of his proofs is encoded explicitly in the axioms. The rigorization of geometry is philosophically very significant. Rigorizing Euclid's proofs frees our knowledge of geometric truth from a reliance on the empirical contingent properties of diagrams or from Kantian intuitions of space. Hilbert is aware of this in his introduction where he identifies the task to be one of *analysis of a logical nature* and which has as its subject matter our "intuition of space". It is true that Hilbert's more explicit concern in his *Festschrift* is with a mathematical investigation into important and interesting geometric theorems and with the relationship between the axioms (such as the independence of the parallel postulate). He makes clear to Frege that,

> ... if we want to understand each other, we must not forget that the intentions that guide the two of us differ in kind. It was of necessity that I had to set up my axiomatic system: I wanted to make it possible to understand those geometrical propositions that I regard as the most important results of the geometrical inquiries: that the parallel axiom is not a consequence of the other axioms, and similarly Archimedes' axiom, etc. I wanted to answer the question whether it is possible to prove the proposition that in two identical rectangles with an identical base line the sides must also be identical, or whether as in Euclid this proposition is a new postulate. I wanted to make it possible to understand and answer such questions as why the sum of the angles in a triangle is equal to two right angles and how this fact is connected with the parallel axiom. That my system of axioms allows one to answer such questions in a very definite manner, and that the answers to many of these questions are very surprising and even quite unexpected, is shown, I believe, by my *Festschrift* ... (Hilbert, 1899/1980*d*, 38).

However, Hilbert is not unconscious to the philosophical significance of his work. He immediately continues:

> Of course I also believe I have set up a system of geometry which satisfies the stricter demands of logic, and this brings me to the proper answer to your letter (Hilbert, 1899/1980*d*, 38).

It will be Hilbert's philosophical aims that are of concern to us in what follows, since it is these aims which bring him into conflict with Frege.

2.1.2 The axioms of *Foundations of Geometry*

Hilbert's axiomatisation of geometry uses twenty axioms and six primitive notions of which there are three sortal concepts (point, line, plane) and three relations (between, lies on, and congruence). He introduces these primitives as follows:

> Let us consider three distinct systems of things. The things composing the first system, we will call *points* and designate them by the letters A, B,C,...; those of the second, we will call *straight lines* and designate them by the letters a, b, c,...; and those of the third system, we will call *planes* and designate them by the Greek letters α, β, γ,...We think of these points, straight lines, and planes as having certain mutual relations, which we indicate by means of such words as "are situated'," "between", "parallel", "congruent", "continuous", etc. The complete and exact description of these relations follows as a consequence of the *axioms of geometry* (Hilbert 1899/1971, §1).

Hilbert noticeably avoids giving any intuitive definition of the primitives in the vein of Euclid's explanations of his primitives as "length without breath", etc. Indeed, all he asks us to assume about the primitive objects is that they are distinct and that the three specified relations hold between them. In place of an explanation of the three relations, he claims that they are "described" as a consequence of a complete set of axioms of geometry, which is to say that their description follows, perhaps in some indirect way, from the entire set of axioms in which they feature.

Hilbert further organises his axioms into groups in which the task of completely and exactly describing the primitives is shared between them. The five groups are as follows:

- The axioms of connection [I] which serve to "connect" the "three distinct systems of things" which are to be called points, straight lines and planes.

- The axioms of order [II] which describe the relation "between"; the axiom of parallels [III] also known as Euclid's axiom or the parallel postulate, concerning the eventual intersection of lines given that the lines satisfy suitable angles with respect to each other.

- The axioms of congruence [IV] which describe the congruence relation in which line segments stand to each other.

- The axiom of continuity [V] also known as the Archimedean axiom, which introduces the idea of continuity.

Hilbert says of the five groups: "Each of these groups expresses, by itself, certain related fundamental facts of our intuition" (Hilbert 1899/1971, §1). Thus, for him, the primitive terms require no supplementary explanation since they are characterised by the axioms. Further, the meaning of the primitives is explained in entirety by an identifiable sub-group of the axiom set. Each of these groups, Hilbert claims, captures something which we have knowledge of via intuition, but he remains deliberately unspecific regarding what kind of facts these are. We also have the claim that the groups fully express these facts and so it follows that the axioms are capable of replacing our intuition as a different route to knowledge of these facts. In other words, Hilbert's system rigorises our intuitive grasp of Euclidean space.

What is clear from this is that while Hilbert traces the aims of his project back to Euclid, he self-consciously departs from him in the method of introducing and conceiving of the primitive terms. The focus of theoretical attention is very much shifted away from the meaning of the primitive terms of geometry and onto the axioms themselves. For Hilbert, we do not incorporate into our axioms our intuitive conception of geometric objects. Rather, we become acquainted with geometric objects through a new epistemic route carved out by unpacking a set of axioms which encode their relation to each other and in this way indirectly provide us with a complete and exact description of the foundational elements of geometry.

2.1.3 The methodology of Hilbert's *Festschrift*

Having settled the meaning of the primitives in his *Festschrift* and set out his twenty axioms, Hilbert then begins his investigation into the relations between them. In particular, he establishes the provability and independence of important geometric theorems from his axioms, the consistency of the complete set of his axioms, and the consistency of various sub-groups, and the independence of the parallel postulate and Archimedean axiom.

Let us begin by setting out how Hilbert establishes the consistency of his axioms. We once again return to the important §9 of Hilbert's *Festschrift*. As Blanchette explains, here Hilbert establishes the consistency in two stages: first, he uses a background theory to construct an interpretation of his axioms, then he proves that so interpreted the sentences are true, i.e. they are theorems of the background theory (Blanchette 1996, 320).

To better understand the work involved in Hilbert's first stage, consider his axioms of connection. Recall that this group of axioms are supposed by Hilbert to "establish a connection between the concepts indicated above; namely, points, straight lines, and plane". The second axiom of connection is as follows:

I, 2. *For every two points there exists at most one line which lies between those points.*

To construct an interpretation, Hilbert assigns a meaning to the non-logical terms in the axiom ("point", "line" and "lies on") using the domain Ω: a fragment of the real numbers. These are assigned as follows:

- *P is a point*; assigned the set of pairs $\langle x, y\rangle$ from Ω.
- *L is a line*; assigned the set of ratios $[u : v : w]$ from Ω.[1]

[1] Where u and v are not both equal to 0.

- *P lies on L*; assigned the set of pairs $\{\langle x, y\rangle, [u : v : w]\}$ from Ω, such that $ux + vy + w = 0$.

The second axiom of connection is thus reinterpreted as:

> [I, 2.$^{\Omega}$] *For any pair of pairs of real numbers $\langle\langle a, b\rangle, \langle c, d\rangle\rangle$ there is at most one ratio of real numbers $[e : f : g]$, such that both $ae + bf + g = 0$ and $ce + df + g = 0$.*

This provides us with a good sense of Hilbert's two-step method for one of his axioms. For the proof, Hilbert must systematically reinterpret all of his twenty axioms and show them all to be true with respect to the domain Ω and the interpretation given above. Hilbert provides a proof sketch to this effect and then concludes:

> From these considerations, it follows that every contradiction resulting from our system of axioms must also appear in the arithmetic related to the domain Ω. The corresponding considerations for the geometry of space present no difficulties. If, in the preceding development, we had selected the domain of all real numbers instead of the domain Ω, we should have obtained likewise a geometry in which all of the axioms of groups I–V are valid. For the purposes of our demonstration, however, it was sufficient to take the domain Ω, containing on an enumerable set of elements (Hilbert 1899/1971, §9).

Here we have Hilbert recognising the availability of alternative interpretations. Furthermore, he acknowledges in the first sentence that what has been proven is a *relative* consistency result. Hilbert specifically claims is that if there is a contradiction in his axioms, there must also be a contradiction in the arithmetic of domain Ω. From this we can infer that if the theory of arithmetic is consistent then so are Hilbert's axioms. Blanchette provides an explicit reason for why Hilbert's reinterpretation method is able to secure the consistency of a set AX of geometric axioms, relative to the consistency of a background theory B.

> If the set AX were inconsistent, then it would logically imply a contradiction. But as logical implication is independent of the specific meanings of such terms as "point" and "line," AX would continue to imply a contradiction under its reinterpretation. But that is just to say that a set of theorems of B would imply a contradiction, hence that B itself would be inconsistent (Blanchette 2014, §2).

Thus, Hilbert reduces the question of his axiom set's consistency to the question of the consistency of the arithmetic of domain Ω. He further observes that it could likewise be reduced to the question of the consistency of the theory of the real numbers.

The last chapter has argued that Hilbert's early methodology fell short of his later proof-theoretic method. It is therefore worth observing that this proof sketch exhibits all of the essential features of the modern model-theoretic method. The only difference is that the specification is very informal, such that the domain and interpretation function are not defined set-theoretically. Nevertheless, the basic mechanism of the proof is that if the reinterpreted axioms are shown to all be true relative to the background theory then they are shown to be *true in a model*. From the truth of the axioms in a model we can infer their consistency, since all that varies across the different models is the interpretation of the primitives. For instance, we can infer from the fact that sentences like [I, 2^{Ω}] are true in the theory of arithmetic related to Ω , that sentences like [I, 2] do not lead to contradiction. The legitimacy of this inference – as we have already noticed – relies on the assumption that the background theory is itself consistent and what Hilbert concludes is that "every contradiction resulting from our system of axioms must also appear in the arithmetic related to the domain Ω" (Hilbert 1899/1971, §9). So, what is precisely established by Hilbert's two-stage maneuver is the following: the question of the consistency of his axioms is reduced to the question of the consistency of the background theory used in the reinterpretation. In short, Hilbert's proof sketch exploits the truth of the reinterpreted axioms in the background theory to establish the model-theoretic consistency of the axioms. Thus, Hilbert presents the first model-theoretic proof in his *Festschrift* in all but detail.

With respect to independence, an axiom α, included in an axiom set Σ, is independent of that set if it cannot be derived from its members. In this sense, a consistency proof is just a special kind of independence proof that shows the sentence $\alpha \wedge \neg\alpha$ is independent of an axiom set. In §10 of *Festschrift* Hilbert exploits this relationship in the opposite direction, demonstrating the independence of axioms by consistency proofs. For, if the axiom is independent of Σ then since it cannot be derived from the members of Σ there should be no problem replacing it with its negation. Thus Hilbert demonstrates the independence of α from Σ by testing whether negating the axiom would lead to contradiction, i.e. by testing whether the set $\Sigma - \{\alpha\} \cup \{\neg\alpha\}$ is consistent.

2.2 The dispute of the Frege-Hilbert controversy

Now that we have met Hilbert's *Festschrift* we can look at how Frege mounts an objection to Hilbert by articulating the consequences of Hilbert's approach to axioms. This section will draw out the features of Frege's most emphatic charge against Hilbert: that his 'axioms' are definitions, not axioms. The next section will consider Frege's argument that Hilbert's axioms are not even suitable as definitions and the subsequent section will briefly tie in the final and most familiar strand of the controversy: whether or not consistency is enough for existence.

2.2.1 Hilbert's muddle

Frege's first and most prolonged criticism of Hilbert is that Hilbert interchangeably employs the terms "definition, explanation and axiom". Frege is empathic that these must be kept distinct; to his mind Hilbert has blurred the terminology for the three categories and in doing so he has also blurred together their various properties and so illegitimately conferred features unique to one category of expressions onto another. Let us draw out Frege's objection to Hilbert's axioms being called definitions, explanations and finally even axioms.

Definitions: Frege begins his critique in the letter of December 1899 by lecturing Hilbert – rather condescendingly – on the difference between definitions, explanations and axioms, beginning with definitions. He puts in the mouth of his colleague Thomae the condition that definitions should provide a "characteristic mark" whereby what they define can be recognised. The adequacy condition echoes Frege's requirement for concepts to have "sharp boundaries", a condition which Frege considers necessary for concept-expressions to successfully refer. This vague adequacy condition on a definition is ambiguous between requiring a definition to provide the means by which we recognise the sense of the *definiendum* or requiring it to provide the means by which we recognise the reference of the *definiendum*.

Frege puts to Hilbert that both he and Thomae agree that his definitions fail to meet this adequacy condition since they cannot "recognise whether the relation *Between* obtains" on the basis of what Hilbert has told them about the relation. What *has* Hilbert told them? As we saw in the previous section, Hilbert claims in his introduction that it is his axioms which provide a complete and perfect characterisation of the primitive concepts. He also claims that he regiments his axioms into groups according to which of the primitives they "fully" characterise. The "between" relation is given by the second group of axioms – the axioms of order. Here are the first three axioms of order along with his introduction of them:

> The axioms of this group define the idea expressed by the word "between", and make possible, upon the basis of this idea, an order of sequence of the points upon a straight line, in a plane, and in space. The points of a straight line have a certain relation to one another which the word "between" serves to describe. The axioms of this group are as follows:
>
> [II, 1.] *If A, B, C are points of a straight line and B lies between A and C, then B lies also between C and A.*
>
> [II, 2.] *If A and C are two points of a straight line, then there exists at least one point*

> *B lying between A and C and at least one point D so situated that C lies between A and D.*
>
> [II, 3.] *Of any three points situated on a straight line, there is always one and only one which lies between the other two* ... (Hilbert 1899/1971, §1).

Frege objects to Hilbert's claim that the group of axioms define the relation described by the word "between" since his axioms provide no means by which to recognise the instantiation of the relation. But it is not until later in the correspondence that Frege develops this charge into a substantive objection. This objection will be very important in Chapter 5, so we will expound it here in some detail.

Most succinctly, Frege's objection is as follows: rather than providing a full characterisation of the relation *between*, the axioms in Hilbert's groups might be satisfied by various relations, or by none at all. Frege puts this objection to Hilbert in the letter dated January 1900, using an algebraic simile;

> Your system of definitions is like a system of equations with several unknowns, where there remains a doubt whether the equations are soluble and, especially, whether the unknown quantities are uniquely determined (Frege, 1900/1980*b*, 45).

We can understand Frege's point here in the following way. Hilbert's definitions determinately establish neither the sense nor the reference of the primitive terms. For example, if *between* is to be understood as referring to a relation which satisfies the axioms of order, this leaves it open whether there is more than one relation which might satisfy the axioms, in the same way that $x + y = 7$ for different values of x and y. It also leaves open whether there is any such relation since for all we are told by Hilbert's definitions, the axioms might be analogous to the equation $x+4 = x+2$. If the primitive terms may not refer and do not refer determinately then, for Frege, it stands to reason that Hilbert's definition has not secured the reference of the primitive expressions.

Frege makes the same objection from another angle where he notes in the same letter:

> Given your definitions, I do not know how to decide the question whether my pocket watch is a point (Frege, 1900/1980*b*, 45).

Again, the thrust of the objection is that Hilbert's definitions do not determinately characterise his primitives. However, there is something particularly interestingly and important about this formulation of the objection: it is a version of the Julius Caesar problem. In particular, it claims that the definition does not provide the resources for distinguishing the primitives from other objects. This link will become important in the latter half of this thesis when the first formulation of this objection is raised against the neo-Fregean logicist.

By Frege's objection that Hilbert's apparent definitions provide no characteristic mark by which to recognise the primitives of geometry, Frege articulates an adequacy condition on successful definition whereby a definition must determinate the sense of the term being defined. To successfully define a word, its sense must be secured unambiguously and the definition must provide the means by which we can distinguish its reference from other objects (in the case of singular terms) and other concepts (in the case of predicates). According to Frege (and his quick survey of his colleagues), Hilbert's axioms do not meet this requirement and so cannot legitimately be called definitions.

Axioms: On Frege's view, theories are composed of the thoughts expressed by sentences. Since distinct sentences can express the same thought, the same theory can be expressed in different languages or by distinct sentences of the same language. This feature of theories cannot be accounted for by a conception of a theory as a collection of sentences. Frege says it is unpalatable to consider "what is audible or visible" as having the metamathematical properties of consistency and independence, or indeed of truth and falsity (Frege 1903/1971*a*, 274).

For Frege, axioms are thoughts which have been established as true. He explains his conception of axioms most explicitly to Hilbert in one of his letters:

> I call axioms propositions that are true but are not proved because our knowledge of them flows from a source very different from the logical source, a

> source which might be called spatial intuition. From the truth of the axioms it follows that they do not contradict one another (Frege, 1899/1980*e*, 37).

Therefore, the set of sentences with which we begin our axiomatisation must express thoughts which we know to be true through such a means as spatial intuition, prior to the axiomatisation of the theory. Frege seems to be thinking of the axioms of geometry and not of the source of our knowledge of axioms in general, but we could generalise his claim to apply to all axioms.

A conception of axioms as mere sentences is compatible with Hilbert's procedure and is potentially reconcilable with Frege's, if Hilbert's set of sentence-axioms express a set of Frege's thought-axioms. After all, sentences do not drop out altogether on the Fregean conception. Sentences are the machinery we have for manipulating the intangible thoughts they express. The point Frege takes issue with is that Hilbert's axioms fail to express *any* thoughts. This is because for a sentence to express a thought it must have parts which express a sense since – for Frege – the thought expressed by a sentence is made up of the senses of the parts of the sentence and the way they are put together. However, not all of the parts of Hilbert's axioms are senseful (in particular the expressions for the geometric primitives) and so Hilbert's axioms are inept to express determinate thoughts. Rather, such sentences are incomplete; partially determined; capable only of expressing a thought if senses are assigned to their primitive terms (as they are in the interpretation). Frege first puts this point to Hilbert with his complaint that Hilbert does not define his primitives in the way of Euclid:

> The explanations of sects 1 and 3 are apparently of a very different kind, for here the meaning of the words 'point', 'line', 'between' are not given, but are assumed to be known in advance. At least it seems so. But it is also left unclear what you call a point. One first thinks of points in the sense of Euclidean geometry, a thought reinforced by the proposition that he axioms express fundamental facts of our intuition. But afterwards (p.20) you think of a pair of numbers as a point. I have my doubts about the proposition

> that a precise and complete description of relations is given by the axioms of geometry (sect. 1) and that the concept 'between' is defined by the axioms (sect. 3)... (Frege, 1899/1980*e*, 36).

It is an important kind of failure for Frege that Hilbert's primitive expressions have not been given a determinate meaning prior to their use in the axioms . This is because – for him – the formal relations between sentences reflect the logical relations between thoughts. The very purpose of logic is to uncover this "logical linkage of truths" (Frege 1906/1971*b*, 302).

A Fregean axiom set is a set of foundational truths which form the basis upon which a theory can be constructed. Hilbert's semantically indeterminate axioms do not express thoughts and thus cannot aid this endeavour. Indeed, since thoughts are the primary truth bearers according to Frege, Hilbert's axioms are not even capable of being true or false. It is clear that the sentences which Hilbert classes as axioms are entirely distinct from Frege's canonical conception of axioms as 'true thoughts'.

Explications: Frege separates out explications, explanations, and elucidatory propositions from definitions and axioms. The former kinds of propositions explain the meaning of certain terms but unlike the latter kinds of propositions they are not required to do so determinately. Furthermore, explications, explanations, and elucidatory propositions are strictly used prior to the construction of a theory.

Frege explains to Hilbert that such propositions as explications must not be counted as part of mathematics proper but should be referred to "the antechamber, the propaedeutics" (Frege, 1899/1980*e*, 35). He gives a far clearer exposition of their role in his later 1903 article "Foundations of Geometry":

> My opinion is this: We must admit logically primitive elements that are indefinable. Even here there seems to be a need to make sure that we designate the same thing by the same sign (word)... Since definitions are not possible for primitive elements something else must enter in, I call it explication. It

> is this therefore that serves the purpose of mutual understanding among investigators, as well as of the communication of the science to others. We may relegate it to a propaedeutic (Frege 1903/1971*a*, 281).

Establishing the axioms as true prior to the theory requires that we establish a determinate sense for the primitive terms as well as for the logical vocabulary. This is the role of elucidatory propositions: to coordinate our understanding of the primitives in the antechamber of an axiomatisation without attempting to define the potentially indefinable. Elucidations are different from axioms in that they concern the meaning of a word which is not already assigned a determinate sense and they are different from definitions in that the explanation which they give is happily ambiguous and imprecise. An elucidation will serve its purpose so long as it brings about the desired "mutual understanding" required for more than one scientist or thinker to develop a theory.

Hilbert occasionally uses the term "explanation" to speak of his axioms. Frege objects to this since in this case Hilbert is employing the process of explanation in the course of the axiomatisation rather than in the antechamber. He expresses the difficulty he has with Hilbert's practise by saying:

> I was surprised to learn that all the axioms of groups I to V are to be taken as supplements to the explanation in sect. 1. Accordingly, this explanation, with all that belongs to it, fills all of your chapter I and interlocks with many other explanations and theorems contained in the chapter. I confess that this logical edifice strikes me as mysterious and extremely imperspicuous (Frege, 1900/1980*b*, 44-45).

The substance of Frege's criticism here is that employing explanations prior to an axiomatisation, and employing supplementary explanations in the course of the theory undermines the logicality of the foundation. This criticism shows that Frege is concerned with Hilbert's imprecision because of the effect it has on Hilbert's logical foundation. In the next section we will spell out this point more fully by characterising the philosophical objection which Frege is bringing to Hilbert beyond pointing out the difference in their respective terminology for axioms, definitions, and explanations.

2.2.2 The substance of Frege's terminological objection

We are now in a position to reconstruct the philosophical significance of the terminological objections which we have seen Frege make to Hilbert.

Frege thinks that Hilbert has blurred together three very different kinds of sentences: definitions, which determinately characterise the sense of some unidentified term; axioms, which are the foundational truths of a theory expressed by determinate sentences; and explanations, which indeterminately coordinate the sense of some unidentified – perhaps indefinable term – prior to construction of the theory. By failing to keep these categories distinct Hilbert has wrongly taken there to be sentences which, like axioms, play the role of foundational sentences from which the theory is constructed but which, like definitions and explanations, contain expressions which are not yet meaningful. Further – as with explanations – the axioms do not assign a precise meaning to the primitives but – as with definitions – they are part of the theory and not established prior to it, in the antechamber.

Thus, Hilbert's terminological imprecision has caused him to blur together features unique to each kind of sentence into one impossible kind of sentence which can at the same time be used as an axiom of a theory, an explanation of the primitives, and a definition of their meaning. It is such an impossible kind of sentence that fulfils these three processes at once which Hilbert uses as a foundation for geometry. This provokes Frege's condemnation of Hilbert's project as ... "on the whole a failure, and in any case that it can be used only after thorough criticism" (Frege, 1900/1980*d*, 90).

Frege's categorisation of the different sentences untangles the muddle and presents Hilbert with the following problem: if Hilbert's primitive expressions are meaningful, then they must employ some prior meaning, such as the meaning given to them by Euclid. But in this case they are not definitions or explanations because otherwise they would be circular, presupposing the meaning they purported to define or explain. However, if the primitive expressions are *not* meaningful then there are parts of a sentence which do not yet have a sense and, as a result, the sentence cannot express an axiom since the sense of all the parts of a sentence must work together in order to form a thought. We can formulate this as a dilemma:

> **Frege's Dilemma**: Either the primitive expressions in the axioms are laid down as meaningful (in which case the axioms are not definitions), or they are not (in which case they are not axioms).[2]

2.2.3 Hilbert's reply to the terminology objection

We now turn to Hilbert's responses to Frege's charges, in reverse order of our presentation of them.

Hilbert briefly dismisses Frege's issue with calling his sentences explanations. He concedes that his explanations may be rephrased as definitions without any non-superficial amendment to his system and gives an example of doing so for the introduction to his axioms of order (Hilbert, 1899/1980*d*, 39).

Hilbert reiterates many times to Frege that his conception of an axiom is simply different from Frege's and entirely legitimate. Hilbert contends that the indeterminacy of his axioms is not a defect of his theory but rather "a tremendous advantage" (Hilbert, 1899/1980*d*, 41 ft.*). It is advantageous because it makes his axioms capable of being multiply re-interpretable and so of characterising commonalities across infinity many structures. In this feature lies the very power of Hilbert's methodology. It is clear from this that Hilbert would reject the first horn of Frege's dilemma – that the primitive expressions in his definitions already have meaning. Thus, Hilbert's definitions are not circular; rather, they deliberately dispose of the pre-theoretic meaning of primitives. This pushes Hilbert onto the second horn of the dilemma, and to this point we will return presently.

[2]Thanks to Owen Griffiths who first gave me the idea of formulating this as a dilemma.

Hilbert's repeated attempts to convince Frege that his alternative conception of an axiom is well-motivated, as opposed to a mere muddle, proves futile. He eventually lashes out at Frege in his penultimate letter:

> In my opinion, a concept can be fixed logically only by its relations to other concepts. These relations, formulated in certain statements, I call axioms, thus arriving at the view that axioms (perhaps together with propositions assigning names to concepts) are the definitions of the concepts. I did not think up this view because I had nothing better to do, but I found myself forced into it by the requirements of strictness in logical inference and in the logical construction of a theory (Hilbert, 1900/1980*c*, 51).

Hilbert claims that he was forced to adopt his conception of axioms by the rigour demanded by his logicist motivations. The requirements which he elliptically makes reference to here are most plausibly those of avoiding Russell's paradox.[3] Hilbert was already familiar with the paradoxes which would infect Frege's basic laws and cause him to conclude that there is no secure route to the logical primitives. In his last letter to Frege, Hilbert even makes reference to the contradiction and relates this explicitly to his conception of axioms:

> Your example at the end of the book was known to us here; I found other even more convincing contradictions as long as four or five years ago; they led me to the conviction that traditional logic is inadequate and that the theory of concept formation needs to be sharpened and refined. As I see it, the most important gap in the traditional structure of logic is the assumption made by all logicians and mathematicians up to now that a concept is already there if one can state of any object whether or not it falls under it. This does not seem adequate to me. What is decisive is the recognition that the axioms that define the concept are free from contradiction (Hilbert, 1903/1980*b*, 51-52).

[3]Thanks to Michael Potter for this observation.

In chapter 3 we will have more to say about this motivation for Hilbert's conception of axioms. For now, we need only observe that Hilbert's frustration with Frege is understandable; Frege thinks Hilbert's conception of axioms is derived from an imprecise use of terminology whereas Hilbert understands his particular conception to be carefully motivated by the need to expunge the primitives from the very real danger of contradiction. Hilbert explicitly identifies avoiding this danger as the motivation underlying his methodology – which is not a muddle, but a self-concious attempt to secure an alternative route to the logical primitives.

With regard to definitions, Hilbert strongly resists Frege's adequacy condition that a successful definition provides a characteristic mark that can secure a determinate sense and reference on successful definitions. To begin with, this might be a plausible conception of the process of explicit definition, but what Hilbert is claiming is that his axioms provide an implicit definition of the primitive geometric concepts. Further, Hilbert straightforwardly rejects Frege's requirement that a definition determinately establish reference to a unique set of basic elements since on his concept of definition the basic elements of a theory are wholly unimportant to the formal properties of consistency and independence. Hilbert thus explicitly endorses landing on the second horn of Frege's dilemma – that the primitive expressions in his axioms are not made determinately meaningful. However, he denies that this prevents them from being suitable axioms. Hilbert goes so far as to say that as long as the system of elements satisfies the axioms, these basic elements could be "love, law or chimney sweeps".

There are two related reasons for this indifference. The first is Hilbert's conception of a theory in general, not as a set of truths regarding a fixed subject matter but as a "scaffolding or a schema of concepts together with their necessary relations" (Hilbert, 1899/1980*d*, 40). Underlying this understanding of theory is the second reason: Hilbert does not want to inherit any problematic features of our intuitive conception of the geometric primitives into his theory. For him, Frege's method of fixing the meaning of the primitives prior to the theory is as dangerous as Frege thinks Hilbert's ambiguous terminology is. Indeed, Hilbert identifies this point as the cardinal point of misunderstanding between himself and Frege, insisting:

> I do not want to assume anything as known in advance; I regard my explanation in sect. 1 as the definitions of the concepts point, line, plane – if one adds again all the axioms of groups I to V as characteristic marks. If one is looking for other definitions of a 'point', e.g., through paraphrase in terms of extensionless, etc., then I must indeed oppose such attempts in the most decisive way; one is looking for something one can never and because there is nothing there; and everything gets lost and degenerates into a game of hide-and-seek (Hilbert, 1899/1980*d*, 39).

As such, Hilbert decisively opposes the use of explication to coordinate the meaning of the primitives. No such meaning exists. It is the assumption that the meaning of the primitive expressions can and should be known in advance which causes vagueness and risk of contradiction to leak into the otherwise secure foundation. The sense of the primitives should be understood as being given by the theory rather than the theory being beholden to them.

In general, Hilbert defiantly resists the charge that his axioms are a muddle or that he has landed in a dilemma. Perhaps on Frege's understanding of definitions, explanations, and axioms, Hilbert's foundational sentences have features properly unique to all three processes – but because of Hilbert's alternative conception of an axiom, such sentences are perfectly legitimate. Further, conceiving of them in the way Hilbert does is not unmotivated. That there can be sentences which play the role of foundational axioms and, at the same time, serve to indeterminately define the primitive non-logical expressions occurring in them; not only lies at the heart of Hilbert's powerful reinterpretation method, but is also an attempt to avoid the paradoxes which seem to have been known to him.

As Hilbert notes to Frege, although he is imprecise in his terminology this is because "definitions (i.e. explanations, definitions, axioms) must contain everything required for the construction of a theory *and only this*" (Hilbert, 1899/1980*d*, 41, emphasis mine).

Here, then, we have two very different conceptions of what kind of sentences are suitable as axioms, and of definitions: Frege wants the meaning of the primitives and the truth of the sentences to be confined entirely to the antechamber of the theory. Hilbert wants nothing to be assumed in advance lest the theory be infected by contradiction. Frege wants the axiomatisation to function for science, unfolding and rigorizing our current primitive geometric concepts. Hilbert wants to leave our vague and potentially contradictory primitive concepts behind and build our geometric concepts anew, on higher ground.

2.2.4 Frege's diagnosis of Hilbert's axioms

In this way, Hilbert places the onus on Frege to say why Hilbert's alternative conception of an axiom is illegitimate. Frege accepts the onus and presents an argument to Hilbert. He points out that even if we were to accept the legitimacy of having definitions as axioms, Hilbert's axioms define concepts distinct from the primitive concepts that Hilbert requires. Looking at this argument will bring us closer to the substance of the disagreement between Frege and Hilbert and it will also form the backbone of the argument against Hale and Wright's neo-Fregean logicism which will be given in chapter 4.

To identify why Hilbert's axioms do not successfully deliver the primitives, Frege employs the powerful machinery of his hierarchy to make the following objection:

> The characteristic marks you give in your axioms are apparently all higher than first level; i.e., they do not answer to the question "What properties must an object have in order to be a point (a line, plane, etc.)?", but they contain, e.g., second-level relations, e.g., between the concept point and the concept line. It seems to me that you really want to define second-level concepts but do not clearly distinguish them from first level ones (Frege, 1900/1980*b*, 49).

Here we have the important distinction between first-level and second-level concepts: first level concepts are those which are saturated by objects which are of level zero; second-level concepts are those which are saturated by first level concepts. The primitives ‘point’, ‘between’, etc. are all first-level concepts since geometric objects fall under them. Frege points out to Hilbert that if his definitions successfully define anything they do not define these first-order concepts because they are only apt to define second-level concepts.

Frege, here, is making a point that is applicable in general to the elusive process of bestowing meaning on a term by the process of implicit definition. By this process we articulate (in this case) a second-level concept which the first-level concepts satisfy. The higher-level concept is built from the determinate terms in the sentence; it is a structure which we can hold up to some lower-level concepts in order to verify that they fit the structure and so qualify as an interpretation.

Applying this observation to Hilbert’s axioms shows that although his axioms include predicates for first-level concepts, those predicates are not the *definienda* of the definition. In other words, the role of the first-level predicates in Hilbert’s axioms is not to gain determinate sense and reference for first-level concepts, they instead function as markers for the argument places of a single higher-order predicate – the true *definiendum* of Hilbert’s axioms. As Resnik puts the same point:

> Hilbert’s axioms contain both these first-order predicates and second-order predicates, namely, quantifiers. But the first-order predicates ‘point’, ‘line’ and so on, may not only be viewed as variables; they can also be construed as marking argument places in the second-order predicates of Hilbert’s axioms. Thus, the conjunction of these axioms can be construed not as defining several first-order predicates but as defining a single second-order relational predicate (Resnik 1980, 112).

Drawing on this, Frege’s point may be put most succinctly as the observation that Hilbert’s axioms define a higher-level predicate in which his non-logical primitives are not furnished with meaning but instead function as variables. These are replaceable by various sets of first-level concepts in different interpretations. As such, the six primitive terms contribute to Hilbert's axioms, not by picking out first-level concepts as we would antecedently read them, but as variables which stand for the argument places in a single second-order relational predicate composed of the conjunction of all of Hilbert's twenty axioms for all his various groups. What is thus defined by Hilbert's axioms is a second-level six-place relational expression. We could also read the expression as a second-level one-place concept expression with a complex six-tuple argument. The latter reading is slightly more natural, given that Frege speaks of Hilbert as defining a higher-order *concept*, but nothing beyond ease of exposition hangs on the issue.[4]

Again, Frege does not express his point in terms of a dilemma, but we can characterise Frege's general diagnosis here by adapting what I have called Frege's dilemma to be particular to purported cases of implicit definition:

> **Frege's Dilemma for Implicit Definition**: Either your *definiendum* has a prior meaning and so your definition is circular, or you have an *explicit* definition of a concept one level higher than your *definiendum*.

This more succinct characterisation of Frege's general diagnosis will prove useful when we do the work of transplanting the spirit of Frege's objection onto the neo-Fregean logicist project.

In conclusion, by this laying bare of the process of implicit definition Frege shows Hilbert's axioms to be incapable of fulfilling the purpose which Hilbert sets out for them

[4]The alternative readings here are in my voice and not in Frege's. It is a separate interpretive issue what Frege would have thought of these two candidates. Oliver (1994) has argued that Dummett was wrong to say that Frege would have rejected concepts with complex plural arguments. In support of his position he cites passages like:

> In the inequality $3 > 2$... we can also regard '3 and 2' as a complex subject. As a predicate we then have the concept of the relation of the greater to the smaller (Frege 1882/1980*f*, 101).

In a similar way, it seems that we could write Hilbert's axioms into a predicate referring to the concept of *six primitives related so as to form a Euclidean geometry*. However, Oliver suggests this is not available to Frege for other reasons. In particular, because Frege's saturated/unsaturated division allows for no complexity in the saturated part, i.e. in the argument of a concept. Without any complexity, Frege is not able to order the primitives using n-tuples, as we have done (Oliver 1994, 82).

– which is to provide a complete and exact description of the primitive concepts. As such, Frege considers his diagnosis of the definitional capacity of Hilbert's axioms to be a decisive objection to the legitimacy of using such a conception of axiom within the framework of Hilbert's project.

2.2.5 Hilbert's reaction to diagnosis

Hilbert's reaction to Frege's diagnosis is not the one expected by Frege. Frege takes himself to be raising an objection to Hilbert when in fact Hilbert is happy to endorse the second horn of what we have called Frege's dilemma.

By the time Frege delivers his diagnosis to Hilbert their correspondence is already trailing off (from Hilbert's side). Hilbert, in a short reply, reiterates to Frege that a concept may only be logically established by giving its connections to other concepts, repeating the point he first expresses with the metaphor that a theory is a scaffolding of concepts. In saying this, Hilbert points Frege to the key to diffusing his 'objection'. As we have seen, it is essential to Hilbert's method that his axioms be suitable for reinterpretation and Frege's diagnosis makes it clear how distinct sets of first-level concepts can satisfy the axioms if what the axioms define is in fact a higher-level concept rather than a particular set of first-level concepts. Frege's diagnosis gives us insight into the powerful mechanism of Hilbert's method: it is *because* his axioms define a concept of a higher level that they are able to characterise structures and properties of infinitely many systems.

To his own mind, Hilbert has characterised the primitives as far as he could need to; he does not require a definition of first-order concepts since his definitions provide the higher-level conceptual scaffolding he needs. Recall that Hilbert cares not whether the basic elements of the theory are points or chimney sweeps; Frege is mistaken that Hilbert intends to give a definition of first-level concepts, as Hilbert cares nothing for them beyond whether or not they satisfy the second-level concept defined by his axioms.

From Hilbert's point of view, what Frege understands as an objection is *only* a diagnosis, and rather than giving reason to support the illegitimacy of Hilbert's axioms, the application of Frege's hierarchy elucidates Hilbert's conception of a theory and provides insight into the way in which his axioms can be said to both define the primitives and leave them open to interpretation.

2.3 Hilbert's Principle in the context of the Frege-Hilbert controversy

Having laid out in detail the main strands of the controversy as well as the text which Frege and Hilbert are disagreeing about, this section will place the last chapter's explanation of Hilbert's Principle within this context.

2.3.1 The misguided reading in context

It was useful to articulate what was wrong with the misguided reading in order to uncover the contention of Hilbert's Principle. In the same way, it will also prove useful to articulate why the misguided reading is incongruent with the context of the Frege-Hilbert controversy, before absolving the priority reading of that same charge.

Recall that, on the misguided reading, Hilbert is interpreted as remarking that consistency implies existence because he is assuming that his system is complete and so that the consistency of his axioms implies the existence of a model. In light of our recent exposition it should strike us immediately that if this is the case Frege's remark is very hard to make sense of. That is to say, Frege is not talking about completeness when he tells Hilbert that the consistency of an axiom set is trivially secured by the requirement that an axiom set is comprised of thoughts which are true. As we have seen, however, Hilbert intended his remark as a comment on Frege's remark. On the misguided reading, it seems hard to make sense of this and there is a risk that when confronted with the context of Hilbert's Principle the misguided reading might have to implausibly maintain that Hilbert was just confused when he told Frege that he believed "the exact opposite" regarding the criterion of consistency and existence. Unless, of course, Frege actually *was* somehow implicitly concerned with completeness. This is, in fact, Dummett's interpretation of Frege's position in the Frege-Hilbert controversy. Dummett sets out four statements which he uses to characterise Frege's view:

(i) The consistency of a theory requires proof;

(ii) the consistency of a theory does not imply the existence of a model for it;

(iii) anyway, the only way of proving the consistency of a theory is to provide a model for it; and

(iv) even if it were possible to prove the consistency of a theory by some other means, what matters is the existence of a model and not the bare formal consistency.[5]

By means of (ii) alone we have a suitably direct conflict with the misguided reading of Hilbert's Principle. Thus, if Frege can be understood in this way then there is a way for the misguided reading to make sense of the context of Hilbert's Principle.

Blanchette articulates why it cannot be correct to interpret Frege as Dummett does. For various reasons, we will return to the point she makes here in later parts of the chapter. Her argument is as follows:

> If Dummett means simply that for Frege it is important that the words and sentences used in mathematics express senses and refer to determinate referents, then there is little here with which to take issue. But it is misleading to express this view in terms of models. For to give a *model* of a sentence is to give an interpretation on which that sentence is true. And there is nothing in Frege that indicates that the meaningfulness of a false sentence turns on the existence of some *re*interpretation on which the sentence expresses a truth. As Dummett says, to understand a sentence is in part to understand

[5] These are taken directly from Dummett (1976, 229).

> the conditions under which it would be true; but this is an understanding of the conditions under which the thought actually expressed would be true, not an understanding of what *other* actually true thought the sentence might express. Dummett's claim, then, that Frege's "procedure is exactly the same as the modern semantic treatment of the language of predicate logic" and that the "modern distinction between the semantic (model-theoretic) and the syntactic (proof-theoretic) treatments of the notion of logical consequence ... is implicit in his writing" ibid., p. 81) cannot be right. Frege's procedure is certainly *semantic*, in that it is concerned with the meanings, and not simply with the syntax, of sentences. But it is not semantic in anything like the modern model-theoretic sense, in which the emphasis is on reinterpretations of non-logical terminology (Blanchette 1996, 333-334).

The core of Blanchette's argument here is that the notion of a model cannot do the same work for Frege as in our modern semantic conception. This is because it is essential to building a model that we use reinterpretable sentences, and it is entirely alien to propose that thoughts could be expressed by sentences whose parts did not have a determinate sense. Furthermore, if we were to reassign the sense of some words in a sentence to build a new sentence, the new sentence would express a distinct thought (since this thought would be composed of distinct constituent senses). Thus, for Frege, we can learn very little about the target thoughts we wish to examine by considering some *other* thoughts which are related to the target thoughts only by the fact that they can be expressed by sentences with the same logical vocabulary. If we take this into consideration, it once again becomes extremely strained to read Frege as having a concern with completeness – implicit or otherwise.

Since it will present a recurring worry, it is worth being clear why this is problematic for the misguided reading. If Frege and Hilbert are not both concerned with completeness, then it seems that the reading loses the ability to offer a good explanation of their disagreement. All that a proponent of the misguided reading could suggest is that Frege and Hilbert are having a misunderstanding – a merely verbal dispute. I take it as entirely fundamental to providing any satisfactory explanation of the Frege-Hilbert controversy that their disagreement is a substantive one, and is best understood as issuing from a very direct, important, and deliberate disagreement between Frege and Hilbert – albeit one which is difficult to articulate. I take it that if a substantive characterisation is available then it is preferable and explanatorily superior to one which dismisses the disagreement as a mere linguistic confusion.

2.3.2 The priority reading in context

Let us now consider whether we can provide such a substantive characterisation by means of the priority reading. After all, in the first chapter we developed the priority reading by drawing it out of the immediate context of Hilbert's remark. As such, this reading is already better placed to cohere with the wider context of the controversy. However, we will here articulate how it does so and how it avoids the threat of interpreting Frege and Hilbert's disagreement as a merely verbal dispute, as we have seen to be the case with the misguided reading.

Recall that – according to the priority reading – Hilbert claims by his principle that consistency is conceptually prior to existence in the sense that:

1. There is no non-circular way to establish the existence of x which does not rely on the consistency of y.

2. There *is* a non-circular way of establishing the consistency of y which does not rely on the existence of x.

Most simply put, the priority reading avoids the threat of a merely verbal disagreement by showing that Frege's position in the correspondence can be understood by the *the very opposite conceptual priority claim*. That is to say, Frege takes it to be the case that:

3. There is no non-circular way to establish the consistency of y which does not rely on the existence of x.

4. There *is* a non-circular way of establishing the existence of x which does not rely on the consistency of y.

Where (3) is just the the denial of (2) and (4) is just the denial of (1).

Recall the four conditions which we used to characterise Hilbert's conceptual priority claim:

Connect (1). There is a way of establishing the consistency of the axioms using substantive appeal to the existence of the geometric primitives.

Direct (1). There is a way of establishing the existence of the geometric primitives without making any appeal to the consistency of the axioms.

Connect (2). There is a way of establishing the existence of the geometric primitives using substantive appeal to the consistency of the axioms.

Direct (2). There is a way of establishing the consistency of the axioms without making any appeal to existence of the geometric primitives.

We can use these same conditions to characterise Frege's conceptual priority claim in such a way that will bring Frege into direct conflict with the characterisation we gave for Hilbert. In particular, the next section will establish that Frege would accept both (1)s and deny both (2)s and Hilbert would accept both (2)s and deny both (1)s.

In this way we will be able to explain the substantive disagreement between Frege and Hilbert as concerning *the order of conceptual priority between consistency and existence*, thus illustrating that their views are the "very opposite" in this regard.

2.4 Textual support for the four priority conditions

Here we consider Frege and Hilbert's position on each of the four relevant conditions which we identified as characterising a conceptual priority claim.

Connect (1): ***There is a way of establishing consistency of the axioms using substantive appeal to the existence of the geometric primitives.***

In order to best explain Frege and Hilbert's position on this first condition, we will initially focus on exploring a tension in Hilbert's early work. Let us begin with a simplified explanation of the tension: Hilbert formulates his principle that consistency should be used to establish existence in the context of a dispute with Frege centering on Hilbert's *Festschrift*. However, as we have seen, the methodology of the consistency proofs in *Festschrift* exploits the existence of a model (roughly speaking) to show the consistency of the axiom set. In this case, Hilbert's methodology fits squarely under what we might call *Frege's* Principle: that the only way to prove consistency is to assume truth and existence. Thus, Hilbert's Principle is quite vividly incongruent to the actual methodology he employs. Any satisfying reading of Hilbert's Principle must therefore be able to diffuse this tension so as to avoid attributing to Hilbert a profound and implausible confusion between his explicit commitments and his actual methods of proof.

The priority reading appears to be particularly vulnerable to such a tension because it incorporates the progress Hilbert had made towards a proof-theoretic conception. A similar position is taken up by Seig where he claims,

> Hilbert formulated consistency as a syntactic notion in "Über den Zahlbegriff", and also in *Grundlagen der Geometrie*. That does not mean, however, that he sought to prove consistency by syntactic methods: the (relative) consistency proofs given in *Grundlagen der Geometrie* are all straightforwardly semantic, using arithmetic models, although information about the possibility or impossibility of proofs is extracted from the semantic arguments (Sieg 2009, 4).

How we are to bridge the gap between Hilbert's syntactic formulation and semantic methodology underlies the tension we will be concerned with.

Let us first remind ourselves of Frege's approach to consistency proofs. Frege explains to Hilbert that the existence of the relevant mathematical objects secures the truth of the axioms referring to them, from which we can infer the axiom's consistency. Indeed, Frege repeatedly tells Hilbert that he can think of no *other* way to prove the consistency of some axioms than with an appeal to the existence of the relevant primitives.

> What means have we of demonstrating that certain properties... do not contradict one another? The only means I know is this: to point to an object that has all those properties, to give a case where all those requirements are satisfied. It does not seem possible to demonstrate the lack of contradiction in any other way (Frege, 1899/1980*d*, 43).

The relevant properties in question are of course those set out by Hilbert's axioms – the six primitive geometric concepts: *point*, *line*, *plane*, *betweenness*, *lies on*, and *congruence*. Frege's point is that to show Hilbert's axioms are consistent we must point to an object falling under these concepts, i.e. we must point to a geometry.

First of all, it is clear that what Frege says here is sufficient to constitute a strong endorsement of Connect (1). Not only is there a way of establishing the consistency of an axiom set by means of the existence of the geometric primitives, but – according to Frege – this is the *only* way.

Let us now ask what would establish the consistency of Hilbert's axioms, according to Frege. Notice that Frege formulates his approach with respect to properties (which we can interpret as Fregean concepts). This is not accidental, but intricately connected to Frege's diagnosis of Hilbert's axioms. With respect to Frege's diagnosis, we saw in §2.1.4 that Hilbert self-identified as falling on the second horn of what we have called:

> **Frege's Dilemma for Implicit Definition**: Either your *definiendum* has a prior meaning and so your definition is circular, or you have an *explicit* definition of a concept one level higher than your *definiendum*.

Recall, Frege's observation is that Hilbert's axioms define a higher-level predicate in which his non-logical primitives are not furnished with meaning but instead function as *variables*. As such, the primitive concepts are replaceable by various sets of first-level concepts in different interpretations. Hilbert's six primitive terms contribute to his axioms, not by picking out first level concepts as we would antecedently read them, but by standing for the argument places in a single second-order relational predicate composed of the conjunction of all of Hilbert's twenty axioms across each of his five groups. What is thus defined by Hilbert's axioms is a second-level concept expression with a single complex six-tuple argument.

We can connect this with what Frege says in the above quotation by extracting a general characterisation of the method by which Frege thinks consistency must be proven:

> **Frege's Principle**: To establish the consistency of a concept, we must exemplify something falling under that concept, i.e. an object (if the concept is first-order) or a lower-order concept (if the concept is above first-order).

This gives us the link between Frege's method and Hilbert's axioms. By Frege's method we can show that the second-order concept defined by Hilbert's axiom system is consistent by exhibiting a system of (six) first-order concepts suitably related such that they perform the role of the primitives. That is to say, the concepts fall in the range of the variables in the second-order predicate expression. This precise point is observed by Dummett,

> Now Hilbert is concerned with the consistency and independence of his axioms. How, then, on Frege's view, can such a second-level relation be proved consistent? Frege's answer is: precisely by finding particular first-level concepts, relations and ternary relations which stand in that second-level relation – the exact analogue, at the second level, of finding an object which falls under a first-level concept, or finding two objects which stand in a first-level relation (Dummett 1976, 6).

Most simply put, Hilbert's axioms can be established as consistent by pointing to some first-order concepts and relations which satisfy them.

The tension emerges when we observe that this is precisely what Hilbert does in his *Festschrift*. As we have seen, §2.1.2 he takes first-order concepts from (a fragment of) the theory of real numbers and shows them to satisfy his axioms. Having done so he concludes that his axioms are consistent (relative to the theory of the Reals) (Hilbert 1899/1971, 30). Blanchette observes that Hilbert does precisely this.

> As we have seen, Hilbert's consistency proofs demonstrate the consistency of the property defined by a set of sentences when the non-logical terms are read as purely schematic. They do so in a way that corresponds to Frege's own views about such consistency demonstrations, since they exhibit something (in this case a Σ-structure of first-level arithmetical concepts and relations) that satisfies the defined property. As Frege would put it, Hilbert's consistency proofs demonstrate the consistency of a complex second-level relation and hence of its defining conditions (Blanchette 1996, 329).

Herein lies the tension: Hilbert's consistency proof in his *Festschrift* exploits the existence of the real numbers to establish the consistency of the (higher-order concept defined by) the axioms, and this approach seems to fall squarely under Frege's Principle.

The way that the priority reading avoids this tension is by its careful formulation of the condition (Connect). Precisely what is claimed by Connect (1) is that there is a way of establishing consistency of the axioms using substantive appeal to the existence of the *geometric* primitives. The essential point to observe is that nowhere in Hilbert's proof does he appeal to the existence of the *geometric* primitives; he instead only appeals to the existence of the *real numbers*, i.e. the primitives of the domain Ω. Therefore, although it is true to say that Hilbert provides a way of establishing the consistency of his axioms by substantive appeal to the existence of *some* primitives, it does not follow that these are the *geometric* primitives. The tension arises on the priority reading only if we think of the priority schema as a way of understanding Hilbert (and Frege's) Principle as being concerned with the relationship between the consistency of axioms and the existence of primitives when the contention behind the priority reading is that Frege and

Hilbert's concern is more specific: What these interlocutors mean to disagree about is the relationship between the consistency of an axiom set and the existence of the primitives *of that very axiom set*. If this is taken into account the methodological tension we have laid out cannot get off the ground since Hilbert's proofs in his *Festschrift* do not exploit the existence of the primitives of his axiom set, i.e. the geometric primitives.

Indeed, if Hilbert were to think that the geometric objects could be used to establish consistency, then there are two things he could say with regard to how the existence of the geometric primitives is to be secured in the first place. Either he could say that their existence is established by the consistency of the axioms (which would be straightforwardly circular), or he could say that the existence of the geometric primitives is established by some other means, out-with the axiomatisation. The latter is Frege's view which we saw him to articulate in his later 1903 article "Foundations of Geometry" and where he tells Hilbert that propositions such as explications must not be counted as part of mathematics proper but should be referred to "the antechamber, the propaedeutics" (Frege, 1899/1980*e*, 35).

We also saw that Hilbert is very careful to distance himself from such a conception whereby the meaning of the primitive elements is established prior to the axiomatisation by means of some semantic co-ordination facilitated by explication. He emphasises this clearly Frege:

> I do not want to assume anything as known in advance... If one is looking for other definitions of a 'point', e.g., through paraphrase in terms of extensionless, etc., then I must indeed oppose such attempts in the most decisive way; one is looking for something one can never find because there is nothing there; and everything gets lost and degenerates into a game of hide-and-seek (Hilbert, 1899/1980*d*, 39).

Here Hilbert makes reference to Euclid's definition: "A point is that of which there is no part" (Euclid 1956). He denounces the assumption that the meaning of the primitive expressions can and should be known in advance as problematic. He decisively opposes the use of explication to coordinate the meaning of the primitives because no such meaning exists (further to the meaning given implicitly by the theory). For Hilbert, the meanings of the primitives should be understood as being given *by* the theory rather than the theory being beholden to them.

Given that Hilbert was so opposed to determining the primitive geometric expressions outside of the axiomatisation, it follows that Hilbert did not think that the existence of the geometric primitives could be exploited to establish the consistency of his axioms. This is because, to prove the consistency of the axioms we would require some independent means of characterising the primitives. However, if there is nothing to be said of the geometric primitives independently of the axiomatisation, then an appeal to their existence would presuppose the legitimacy of the axiomatisation and thus be unsuitable for establishing the consistency of the axioms in the first instance.

In conclusion, it is quite clear that Hilbert would be unhappy with Connect (1) and that his proofs in *Festschrift* do not rely on it. It is also clear that Frege explicitly endorses such a principle.

Direct (1): ***There is a way of establishing the existence of the geometric primitives without making any appeal to the consistency of the axioms.***

The above reflections on Connect (1) bleed into the relevant considerations for Direct (1). Frege's idea of fixing the reference of the relevant primitives in the antechamber annexed to the front of the theory entails that there *is* a way of establishing the existence of the geometric primitives which is independent of any appeal to the axiom's consistency. And indeed, this prediction is consistent with what we saw Frege indicate to Hilbert where he claims axioms are true propositions and that:

> ...our knowledge of [geometric axioms] flows from a source very different from the logical source, a source which might be called spatial intuition (Frege, 1899/1980*e*, 36).

Although Frege does not detail this "source" any further, it cleanly qualifies as a means of securing the truth of the axioms, and thus the existence of the geometric primitives referred to by the non-logical parts of the sentences expressing the axioms. Furthermore, this source is a means which is independent of the axiom's consistency. As such, it is clear that Frege would accept Direct (1).

The reasons we saw for Hilbert's rejection of Connect (1) also underlie the reasons that Hilbert would reject Direct (1). We saw that Hilbert spoke out vehemently against any attempt to appeal to a characterisation of the primitives external to the one given by the axiomatisation itself, for the reason that there *is* no such characterisation to be given since "...only the whole structure of axioms yields a complete definition" (Hilbert, 1899/1980*d*, 40). On Hilbert's view, Frege's reliance on a source involving some Kantian-esque intuition will not result in rigorous referential coordination; instead, the investigators will be trapped in the antechamber in a game of hide-and-seek in which nothing can be found because nothing has been hidden.

It is worth bearing in mind that first and foremost Hilbert had a *mathematical* interest in the foundations of geometry. For him, the only acceptable means of 'establishing' the existence of the primitives is by *proving* their existence. Even intuitively speaking, existence is not the kind of concept which admits of formal proof. Frege himself admits that he has no way of establishing existence *mathematically*. Remember that he calls his axioms "true but not proved".

Frege and Hilbert thus agree that there is no way to *prove* the existence of the geometric primitives directly and disagree as to whether their existence can be established by *other* means. Frege admits an appeal to Kantian intuition while Hilbert rejects any means except from proof as logically unhygienic. For these reasons, Frege would accept Direct (1) while Hilbert would deny it.

Connect (2): ***There is a way of establishing the existence of the geometric primitives using substantive appeal to the consistency of the axioms.***

Connect (2) is the most immediate fit with Hilbert's remark. Indeed, it is hard to understand any sense in which consistency is "the criterion of truth and existence" which does not entail this condition. In a later letter, Hilbert makes the same point in another way:

> As I see it, the most important gap in the traditional structure of logic is the assumption made by all mathematicians up to now that a concept is already there if one can state of any object whether or not it falls under it. This does not seem adequate to me. What is *decisive* is the recognition that the axioms that define the concept are free from contradiction (Hilbert 1900/1980*c*, 51-52, emphasis mine).

Rather than admit a non-mathematical appeal to pointing at things, Hilbert presents an approach which aligns the existence of the primitives with the consistency of the axiom set which characterises them. In this way he hopes to make questions of existence in mathematics tractable and rigorous. Recall Hilbert's example of an equation's root: he says that the existence of the root is "proven, as soon as" it is demonstrated that the axiom 'every equation has a root' can be added to the other axioms without contradiction. In this way, for Hilbert, an appeal to the consistency of the new axiom set proves the existence of the new primitive (the root), and does so immediately. We can assume that Hilbert means for Frege to infer from this that the existence of the geometric primitives follows immediately from the consistency of Hilbert's axioms. In this strong sense, then, Hilbert endorses Connect (2).

Frege, however, does not endorse Connect (2). When he writes a reply to Hilbert's Principle, Frege emphasises to Hilbert that, "our views are perhaps most sharply opposed with regard to your criterion of existence and truth" (Hilbert, 1900/1980*b*, 47). Frege then presents Hilbert with an example where one attempts to infer the existence of a god from the deductive consistency of the properties ‘A is an intelligent being’, ‘A is omnipresent’, and ‘A is omnipotent’ (Frege, 1900/1980*b*, 47). Frege then asks the following:

> Could we infer from this that there exists an omnipotent, omnipresent, intelligent being? I don’t see how! The principle would go something like this:
>
> If (generally, whatever A may be) the propositions
>
> A has the property ϕ
>
> A has the property Ψ
>
> A has the property χ
>
> together with all their consequences do not contradict one another, then there exists an object that has all of the properties ϕ, Ψ, χ.
>
> This principle is not at all evident to me; and if it were true, it would probably be useless (Frege, 1900/1980*b*, 47).

It is clear that Frege is using the theological example to attempt a *reductio ad absurdum* of Hilbert’s position. The particular assumption of Hilbert’s which Frege attempts to expose as illegitimate is the principle which he generalises above in which one moves from the consistency of a set of propositions to the existence of what they characterise. We can conclude from this that Frege took it to be absurd that the mere consistency of an axiom set could establish that its primitives existed. While the consistency of the axiom set could perhaps form *part* of a proof to establish a necessary condition on the primitives’ existence, it could never function as a sufficient condition; i.e. as a substantive ground from which to infer the existence of the primitives which the axioms characterise.

Frege goes on to point out that if the consistency of the axiom set were used as even a partial appeal from which to establish the primitives’ existence, this too would be illegitimate because it would be circular. As we have seen already, Frege tells Hilbert that he does not know of any way of proving consistency of the axioms except by appeal to existence of the relevant primitives. This makes going on to establish the existence of primitives by employing consistency of axioms an entirely circular venture. Frege articulates precisely this point to Hilbert:

> Is there some other means of demonstrating lack of contradiction besides pointing out an object that has all the properties? But if we are given such an object, then there is no need to demonstrate in a roundabout way that there is such an object by first demonstrating lack of contradiction (Frege, 1900:47).

From this we can conclude that at the time of writing to Hilbert, Frege would not have agreed that there is a way of establishing the primitives' existence using a substantive appeal to the consistency of the axioms, i.e. he would have rejected Connect (2) where Hilbert would have accepted it.

Direct (2): ***There is a way of establishing the consistency of the axioms without making any appeal to existence of the geometric primitives.***

The key observation for Hilbert with respect to Direct (2) is that (as we have seen in detail) Hilbert's early consistency proofs establish the (relative) consistency of his axioms by appeal to the existence of the real numbers (roughly speaking). Since the construction of a model that uses distinct primitives is an alternative to using an appeal to the geometric primitives to establish the consistency of the axioms, Hilbert would be forced to accept Direct (2) solely on the grounds of his model-theoretic proof.

The key observation for Frege's position on Direct (2) is that Frege would not accept that from the existence of the arithmetical model we can infer anything about the consistency of the geometric axioms. For Frege, the problem with Hilbert's model-theoretic reasoning is that the content of Hilbert's reinterpreted primitives is relevant to the consistency of the geometric thoughts in a way that it is *not* relevant to the consistency of mere sentences.

Let us unpack this last point a little slower. As is now familiar, Hilbert's methodology includes reinterpreting the primitive expressions. For Frege, this amounts to changing the sense of the primitive expressions in the sentences expressing the axioms. Since the original thoughts and the reinterpreted thoughts include distinct constituent senses, and since a thought is determined by the senses of the parts of the sentence which expresses them, and the way those parts are put together – it follows that the original thoughts are distinct from the reinterpreted thoughts. In the case of Hilbert's proofs, the reinterpreted arithmetical thoughts are distinct from the original geometrical thoughts. Since Frege thinks the task at hand is to establish the consistency of the *geometrical* thoughts, reasoning about arithmetical thoughts in this way is merely to change the subject.

We have already set out the related objection which Blanchette brings to Dummett's portrayal of Frege as manifesting implicit model-theoretic reasoning. If we return to that passage we can see that Blanchette is making the same point as the one which concerns us here:

> ...there is nothing in Frege which indicates that the meaningfulness of a false sentence turns on the existence of some *re*interpretation on which the sentence expresses a truth... to understand a sentence is in part to understand the conditions under which it would be true; but this is an understanding of the conditions under which the thought actually expressed would be true, not an understanding of what *other* actually-true thought the sentence might express (Blanchette 1996, 333).

Thus, the truth of the arithmetical thoughts would certainly, and trivially, demonstrate the consistency of the arithmetical thoughts as well as the consistency of the logical schemata which they have in common with the geometric thoughts. However, the truth of the arithmetical thoughts is not enough to demonstrate the consistency of any distinct set of thoughts – in particular, any geometrical thoughts. This is because inconsistency could be lurking in the unanalysed content of the geometric primitives. At any rate, this possibility is not excluded by demonstrating that some distinct thoughts with some distinct primitives are free of any inconsistency. Therefore, while it might be true that the arithmetical thoughts and the geometrical thoughts are connected in the sense that one is a reinterpretation of the other, this connection is not enough to infer that they are equi-consistent according to Frege.

Therefore, the consistency of Frege's thought-axioms cannot be established by Hilbert's early model-theoretic reasoning. We can conclude from this that Hilbert's reason for accepting Direct (2) cannot be Frege's reason.

Frege is aware that Hilbert thinks there *is* a way to prove consistency which is a non-circular alternative to the method which we have called Frege's Principle. He tells Hilbert:

> It seems to me that you believe yourself to be in possession of a principle for proving lack of contradiction which is essentially different from the one I formulated in my last letter and which, if I remember right, is the only one you apply in your *Foundations of Geometry*. If you were right in this, it could be of immense importance, though I do not believe in it as yet... (Frege 1900/1980*d*, 49-50).

Frege leaves the possibility of an alternative to Frege's Principle quite open ended and acknowledges the significance of such a discovery. However, he also clearly states that he does not at the moment believe there is such an alternative for proving consistency. This is all we need to conclude that at the time of the correspondence Frege would have rejected Direct (2).

All in all, we can conclude that Hilbert's model-theoretic approach commits him to Direct (2), and Frege's scepticism that there could be an alternative to using a demonstration of the existence of the geometric primitives to establish the consistency of the axioms means that he is committed to denying Direct (2).

2.4.1 Frege's 1906 independence test

Six years later, when Frege goes on to publish his second article, "Foundations of Geometry" (1906/1971*b*) Frege himself proposes an alternative means of establishing consistency. His proposal is a test for independence but of course Frege would have the means of establishing consistency, if he could establish independence (and vice versa) since he could prove that the axiom set in question was independent of a contradiction. The proposal he makes in 1906 thus deserves attention as it helps shed light on Frege's position regarding a priority claim which we have just characterised.

At the start of addressing the question of whether it is possible to prove the independence of a "real axiom" from a set of "real axioms", Frege (1906/1971*b*, 107) sets out the following laws:

L1. If the thought G follows from the thoughts A, B, C by a logical inference, then G is true.

L2. 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.

Recall that, for Frege, "only true thoughts can be the premises of inferences" (Frege 1906/1971*b*, 107). This explains the formulation of the above laws as issuing from his conception of axioms as thoughts that are true, together with his less controversial assumption that logical inference is truth-preserving.

Frege then (informally) defines a mapping of a language into itself (he calls it a "translation" of "vocabulary"), in which:

- Non-logical expressions are mapped to expressions which refer to the same type (i.e. an object or a concept) and the same order (i.e. first-level, second level, etc.) of referent.
- Logical expressions are mapped to themselves.

Furthermore, Frege asks us to assume that the language being used is "logically-perfect" in that each term has a reference and each term expresses a determinate sense.

With this in place, Frege then outlines his test for independence:

> Let us now consider whether a thought G is dependent upon a group of thoughts Ω. We can give a negative answer to this question if, according to our vocabulary, to the thoughts of group Ω there corresponds a group of true thoughts Ω', while to the thought G there corresponds a false thought G'. For if G were dependent upon Ω, then, since the thoughts of Ω' are true, G' would also have to be dependent upon Ω' and consequently G' would be true (Frege 1906/1971*b*, 110).

Let us unpack this approach. We begin with the vocabulary of the language, i.e. the set of premise sentences Ψ, expressing the set Ω and the conclusion sentence F, expressing G. Under Frege's mapping, Ψ is mapped to Ψ', which expresses the set Ω' and F is mapped to F', which expresses G'. For Frege, the thoughts in Ω' must be true "in order to be premises". Then we have it that the thought G is independent of Ω if there is a mapping such that G' is false.

Why does Frege say this is the case? Most simply, because if G was dependent on the thoughts in Ω there would be no way to produce a mapping (of the sort Frege outlines) in which the thoughts of Ω' are true while G' is false. This is because the relevant mapping preserves the sense of the logical terms, so that G' and G have their logical schemata in common. As a result, if there was a purely logical route by which G could be derived from Ω then *the very same logical route* would be available to enable us to derive G' from Ω', in virtue of the common logical schemata. However, since G' is false, G' cannot have been derived from Ω' at all. This is in virtue of L1: if G' is false then (modus tollens) the thought G' *does not follow from* Ω' *by logical inference*. Therefore, although there must be some non-logical means in virtue of which G can be inferred from Ω, G cannot be a *logical* consequence of Ω and so is *independent* of Ω. Frege thus concludes:

> ... with this we have an indication of the way in which it may be possible to prove independence of a real axiom from other [real] axioms (Frege 1906/1971*b*, 110).

The 1906 independence test shows that some time after his correspondence with Hilbert, Frege came to think there is an alternative to Frege's Principle when it comes to establishing consistency. It is worth asking how this later position would have interacted with the priority claim.

To begin with, it does not seem to affect Connect (1), for Frege never suggests that this proposal should be understood as the *only* means of establishing consistency, and so there is no indication that he would revise his view that there is a way of establishing the consistency of the axioms using substantive appeal to the existence of the geometric primitives. Similarly, the proposal for an independence test does not seem to have any bearing on Direct (1), i.e. on whether or not Frege thinks that an appeal to spatial intuitions means that there is a way of establishing the existence of the geometric primitives without making any appeal to the consistency of the axioms.

With respect to Connect (2), we said that Frege would have denied that there is a way of establishing the existence of the geometric primitives using substantive appeal to the consistency of the axioms. We used two of Frege's objections to Hilbert in support of this. First, that the consistency of a set of axioms could never establish the existence of a god. Secondly, that since the only way to establish the consistency of the axioms would be via the existence of the primitives, this would be viciously circular. Since the independence test provides a way of establishing the axioms' consistency which does not rely on an appeal to the existence of the primitives but only to some laws concerning thoughts, a language and a mapping, this would defuse Frege's latter objection by avoiding the circularity. However, Frege only adds the latter objection as supplement to the theological reductio and there is nothing in Frege's new proposal to suggest that he no longer thinks it absurd to ever infer existence from mere consistency. As such, Frege would probably still have denied Connect (2).

The only position which I believe Frege would revise in light of his new proposal is the last – Direct (2). As we have just noted, the independence test makes no appeal to the existence of the referents of the non-logical terms and, as we have observed, if it can establish independence then it can provide a means of establishing the consistency of the axiom set. Therefore, it seems that Frege would no longer deny that there is a way of establishing the consistency of the axioms without making any appeal to the existence of the geometric primitives.

In virtue of changing his position regarding Direct (2), Frege would no longer clearly support a priority claim. Recall that a priority claim requires that a proponents position regarding Connect and Direct is asymmetric. In 1906, it seems that Frege's position on the consistency of the axioms and existence of the primitives is only asymmetric with respect to Connect. What this shows is that Frege's commitment to the priority claim and in particular to Direct (2), is not deeply entrenched but one that he was willing to reconsider.

However, Frege himself emphasises that he only means to sketch such a means of establishing independence and that much more would have to be said to realise such a test. Frege also acknowledges – though rather lightly – what appears to be the main difficulty with the test: its essential mechanism relies on a demarcation of the logical from the non-logical. Furthermore, not only does Frege never return to the task of developing the test, he never again mentions it in his later writing.

It should also be noted that a proposal which Frege makes six years later which revises some of his views and which he never returns to, does not undermine our claim that *at the time of his correspondence with Hilbert* Frege would have denied Direct (2) and thus can be understood as holding a priority claim. After all, we are concerned with providing an understanding of what was at issue in the Frege-Hilbert controversy in which Frege is clear that he was pessimistic about the possibility of any test for consistency and independence which did not fall under Frege's Principle.

Nevertheless Frege's later (albeit short-lived) optimism about developing such a test is of interest because it gives us an insight about which elements of his priority claim he took to be negotiable. The other interesting feature of the 1906 independence test is that the *kind* of alternative which Frege considers is strikingly close to the model-theoretic method which we saw in Hilbert and which we saw Frege denounce in the correspondence. This is because, the set of sentences in the image of Frege's mapping will be roughly the same as the set of fully interpreted sentences of Hilbert's models. Let us therefore ask what the difference is between the two tests.

One immediate difference seems to be that, where Frege reinterprets sentences using a mapping between languages, Hilbert does so by specifying the referents of the new primitive expressions. However, this difference does not seem to be substantive. I think the most important difference is that Frege reasons solely about sentences which express determinate thoughts and thus whose parts all express a determinate sense. Whereas the latter feature is precisely what Hilbert's axioms lack before they are interpreted. This difference is important – from Frege's perspective – because it means that the new test is consistent with Frege's ardent objections to Hilbert's axioms and Frege's insistence on the limitations of Hilbert's methodology. The difference is significant, because it shows that it is not so much the *re*interpretation strategy that Frege takes issue with, but rather the idea of the axioms themselves being in any way *uninterpreted* or schematic. Frege's own test avoids this by mapping sentences expressing determinate thoughts to distinct sentences expressing distinct – but nevertheless determinate – thoughts. Hilbert's method, by contrast, is to establish consistency by interpreting initially schematic sentences, so that only in a model can the fully-interpreted sentences be said to express a fully-determinate thought. We have seen that, for Hilbert this is not an oversight. Hilbert insists to Frege that his axioms should be understood schematically or as having their primitives implicitly defined by the theory.

I would now like to articulate Hilbert's model-theoretic methodology as far as possible within Frege's framework. This will help us articulate what exactly holds Frege back from recognising the power and potential of the model-theoretic approach.

To do this, we must first characterise Hilbert's conception of the relata of consistency in Fregean terms. In other words, we must ask the following question: what can

Hilbert's schematic axioms be said to express, if not determinate thoughts? To answer this, let us again consider the mechanism of Hilbert's proof. Since the primitives will vary across different models, it must be that nothing in their meaning can influence the consistency of the axioms. What is in common between the axiom and its reinterpretation is the logical form of the sentence; in the sense that [I, 2^{Ω}] is an instance of the schema displayed in [I, 2]. This is true at the sentential level but it is also true at the level of thoughts.

We now come to the critical question: what would it take for Hilbert's approach to establish consistency by Frege's lights, given the approach of the 1906 independence test? The answer is that we would have to relax Frege's semantic doctrine that *only a sentence whose parts have a determinate sentence can express a thought*. This is not to suggest that this doctrine is not important to Frege's philosophy of language or even that Frege should reject this doctrine. Rather, what I am suggesting is that this is the true point of issue that Frege has with Hilbert's method, and this can be seen by the fact that if we relax the doctrine Hilbert's method can be made acceptable within Frege's framework.

Let us spell out the consequences of giving up Frege's doctrine that only a determinate thought can express a sentence. It follows we can conceive of a schematic sentence such as [I, 2] expressing a thought *form* in common with [I, 2^{Ω}]. Henceforth, I shall call this a schematic thought:

df. A schematic thought is what is expressed by a schematic sentence.

Schematic thoughts, let's say, are related to determinate thoughts in the way that incomplete sentences are related to sentences. For this reason, establishing that a determinate thought has a certain property allows us to infer something about a schematic thought if that determinate thought is an instantiation of the schematic thought. Let us take our example again: the truth of [I, 2^{Ω}] establishes the consistency of [I, 2] in virtue of the fact that the thought expressed by [I, 2^{Ω}] is an instantiation of the schematic thought expressed by [I, 2].

By allowing schematic sentences to express some species of thoughts, this view diverges from Frege's in an important way: that is, with respect to the contribution of the primitives. For Frege, the relata of the consistency relation are thoughts like [I, 2^{Ω}] but for Hilbert the relata of consistency are schematic thoughts like [I, 2]. This is another way of saying that for Hilbert the primitive terms do not contribute to the consistency of his axioms but for Frege they do. What is at issue is whether it is the form of a thought – that which is expressed by the non-logical vocabulary – that is the primary object of study and whether its consistency is invariant under re-interpretation.

An important observation to make is that what we have called a schematic thought, Frege called an unsaturated sense, or, the sense of a concept word. For Frege, the referent of a schematic thought is a concept since every incomplete expression refers to a concept (see, Frege 1980*g*).This connects to Frege's diagnosis that Hilbert's axioms define a higher-order concept; that the primitive terms in his axioms do not contribute by picking out first-level concepts but by representing the argument places in a six-place relational predicate which expresses a higher-order concept. This fits nicely with where Hilbert says things like:

> But it is surely obvious that every theory is only a scaffolding or schema of concepts together with their necessary relations to one another... (Hilbert, 1899/1980*d*, 40).

Blanchette has already noticed that Hilbert's proofs establish this kind of consistency:

> Say that the complex property defined in this way by a set of partially interpreted sentences is consistent just in case it is not self-contradictory, i.e. just in case some series of concepts or sets could have that property... Though Hilbert does not speak explicitly of property-consistency, his proofs immediately show consistency in this sense (Blanchette 1996, 322).

In the Fregean framework we are operating in we can approximate *property*-consistency here to *concept*-consistency. Blanchette also acknowledges that the relevant concept must be higher-order.

In conclusion, what Frege's flirtation with an independence test brings to light is that the defining difference between Frege and Hilbert (at the time of the correspondence) concerns the true relata of consistency and independence. Frege thinks that what needs to be established as consistent are determinate thoughts; whereas Hilbert does not much care about whether he is understood as establishing the consistency of properties, facts, schematic thoughts, or concepts. In the next chapter, we will return to this issue and we will consider what it reveals about Frege and Hilbert's background philosophical assumptions.

2.5 The two priority claims

In the first chapter, we characterised a conceptual priority claim to be exhausted by Direct and Connect. We also said that the claim that one concept is conceptually prior to another is made when and only when there is an asymmetric view of Connect and Direct, when the relevant concepts are reversed. It is now clear that (during the time of their correspondence) both Frege and Hilbert meet such a criterion: Frege would accept Direct (1) & Connect (1) and deny Direct (2) & Connect (2). Inversely, Hilbert would deny Direct (1) & Connect (1) and accept Direct (2) & Connect (2). Recall that the difference between Connect (1) & Direct (1) and Connect (2) & Direct (2) is precisely that the relevant concepts *consistency of axioms* and *existence of geometric primitives* are reversed in the respective conditions. Therefore, Frege and Hilbert can both be understood as advocating conceptual priority claims. On the one hand we have:

> **Hilbert's Principle:** Consistency of the axioms is conceptually prior to existence of the geometric primitives.

For Hilbert, the asymmetry between the consistency of the axioms and existence of the geometric primitives issues from the fact that consistency is the property which admits of proof and is therefore adept to act as the decisive condition for existence in mathematics. As such, the consistency of the axioms can be proven without appeal to the existence of the geometric primitives. However, a proof of the existence of the geometric primitives is given (immediately) by an appeal to consistency of the axioms.

Frege, on the other hand, holds that:

> **Frege's Principle:** The existence of the geometric primitives is conceptually prior to the consistency of the axioms.

For Frege, the asymmetry between the consistency of the axioms and existence of the geometric primitives issues from the fact that the only means of establishing consistency is to identify the set of primitive geometric objects which have the properties ascribed to them by the axioms. Thus, as far as Frege can see in 1900, while the existence of the geometric primitives can be established by appeal to spatial intuition, the consistency of Hilbert's axioms cannot be established without an appeal to the geometric primitives.

Concluding remarks

The priority reading can do more than cohere with the context of the Frege-Hilbert controversy; it can help us understand the controversy. It is useful because, we now find ourselves in a position to offer an explanation of the point of contention between Frege and Hilbert by employing the proposed characterisation of Frege and Hilbert's priority claims. Most simply put, Frege and Hilbert are supporting the very opposite priority claims, such that they are in deliberate disagreement over *the order of conceptual priority between the consistency of the axioms and the existence of the geometric primitives*.

We have even gone some way towards articulating what motivates each proponent to defend their particular priority claim: namely, their conception of the correct tools that should be used to guide a logico-mathematical investigation. For Frege, the investigation is to be guided by our extra-logical knowledge of a domain of true thoughts (determinately expressed by a set of sentences). For Hilbert, the investigation should be guided by the most successful methods of ensuring consistency by means of proof, regardless of whether it is true thoughts, properties, facts, or concepts which such a proof establishes as consistent.

Chapter 3

The Deeper Disagreement and Hilbert's Ontology

Introduction

After the previous chapter, it seems that the priority reading straightforwardly coheres with the context of Hilbert's Principle. However, this chapter will demonstrate that if we consider the priority reading together with the insight given by Blanchette's seminal work on the Frege-Hilbert controversy, then this no longer becomes quite so straightforward. First we will consider whether a tension between the priority reading and the context of Hilbert's Principle re-emerges in another place. By addressing this tension we will ultimately develop a much fuller picture of the priority reading and – in consequence – of the Frege-Hilbert controversy and Hilbert's early ontological position in the philosophy of mathematics.

3.1 Blanchette on Frege and Hilbert on consistency

In "Frege and Hilbert on Consistency" 1996, Blanchette novelly distinguishes Hilbert-consistency from Frege-consistency. She argues that Frege and Hilbert's conceptions of consistency are distinct such that the consistency proofs given in Hilbert's *Festschrift* establish Hilbert-consistency but not Frege-consistency. This dichotomy is relevant to the priority reading because – as we will shortly spell out in detail – it undermines the substantive point of disagreement between Frege and Hilbert. On the priority reading, Frege and Hilbert disagree over whether consistency or existence is conceptually prior. However, if by consistency Frege means Frege-consistency and Hilbert means Hilbert-consistency, then they are no longer in direct conflict over the order of priority between two concepts. Instead, Frege takes existence to be conceptually prior to Frege-consistency and Hilbert takes Hilbert-consistency to be conceptually prior to existence and the threat of a verbal disagreement emerges again. Indeed, for all the priority reading has said it may be that Frege and Hilbert are in agreement that Hilbert-consistency is conceptually prior to existence which is in turn conceptually prior to Frege-consistency, in which case Frege and Hilbert would not be disagreeing at all. It is clear that the priority reading must have something to say on this matter, and we will develop this in §3.3. For now, let us set out Blanchette's work so that we can explain how it threatens the priority reading (§3.2).

3.1.1 Frege-consistency

Blanchette defends Frege against the view that he was entirely wrong to say that Hilbert's proofs do not establish consistency or independence. She argues that Frege has an alternative, valuable conception of the relations of logical consequence, consistency and independence which are distinct from our modern understanding. In making a distinction between two kinds of consistency and disentangling their ambiguous use in the correspondence she considers her argument to be providing a defence of Frege. Although Hilbert's proofs unambiguously establish consistency and independence by our modern standards, they do not establish these meta-mathematical properties as Frege understands them. Blanchette concludes that although Frege was wrong to say that Hilbert's work was on the whole a failure, and although he did not anticipate the power of Hilbert's methods, he was not wrong to point out that Hilbert's proofs do not establish consistency and independence as he thought these concepts *should* be understood.

Blanchette points out that, according to Frege, a proof manipulates sentences by a set of formal rules in such a way as to reveal the relations between the thoughts expressed by those sentences. This is just to say that – for Frege – the aim of a well-designed deductive system is to mirror the relations that hold between thoughts by restricting which inferences between sentences are legitimate. In this way, a purely formal process gives us an insight into the laws of thought. On this picture, the legitimacy of a formal inference between sentences in a proof can be taken to evidence the existence of a logical relation between the thoughts expressed. The relations that hold between sentences Blanchette calls *derivability relations* and the relations that hold between thoughts Blanchette calls *provability relations* (Blanchette 1996, 323-324). Thus, if we can derive a sentence A from the members of a set of sentences Σ. then we can conclude that the thought expressed by A is provable from the thoughts expressed by the members of Σ.

According to Blanchette, Frege's concern as he writes to Hilbert is whether Hilbert's proofs do enough to establish that the thoughts expressed by those sentences are consistent. As we have seen already, Frege is right in thinking that Hilbert's proofs do not do enough, since Hilbert's axioms are schematic and thus incapable of expressing Fregean thoughts. However, Blanchette goes further by arguing that even if Hilbert's axioms *did* express Fregean thoughts, Hilbert's proofs do not establish the consistency of these thoughts.[1]

3.1.2 Hilbert-consistency

Blanchette specifies Hilbert-consistency in the following way:

[1]Note that the question of whether or not Hilbert's model-theoretic proofs establish the consistency of *any* thoughts, is a separate question from whether they can be said to establish the consistency of any *schematic* thoughts/Fregean concepts. In what follows we will agree with Blanchette's negative answer to the former, while consistently maintaining a positive answer to the latter, in line with the discussion of §2.2.1.

> When no contradiction is deducible from a set via such syntactically specifiable rules, we shall say that the set is *syntactically consistent*. Assuming the consistency of the background theory, then, Hilbert's consistency proofs establish the syntactic consistency of sets of sentences (Blanchette 1996, 320).

Note that Blanchette attributes a syntactic notion of consistency to Hilbert. In particular, she is in agreement with what we said regarding both features characteristic of a proof-theoretic conception of consistency:

(A) Consistency is a relation between sentences, or, certain strings of symbols in a formal language, i.e. a language which has a finitely specifiable set of formation rules.

(B) The relata of consistency operate under a *deductive system*; i.e. the formal language has a *specified* set of *inference rules*.

With respect to (A), Blanchette talks about Hilbert specifying sentences and so having a broadly syntactic conception of the relata of consistency. With respect to (B), Blanchette is careful to note that, Hilbert had not yet specified a set of inference rules formally – so all that can be said is that he thought the rules of inference were specifiable *in principle*.

> To clarify: Hilbert has not at this point specified a syntactic deductive system, and does not view logical deduction as formal symbol-manipulation. He does however view logical deduction as independent of the meanings of the non-logical (here, the geometrical) terms, which makes his implicit principles of deduction syntactically specifiable, though not explicitly so specified (or specified at all, for that matter) (Blanchette 1996, 320 ft. 8).

This means that Blanchette more or less interpreting Hilbert as maintaining what we have called a proto-proof-theoretic conception of consistency.

Blanchette's own aim in distinguishing Hilbert-consistency from Frege-consistency, is to show that Hilbert's proof certainly establishes Hilbert-consistency but does not establish Frege-consistency. Since Blanchette is in broad agreement with the way we have already characterised Hilbert's early conception of consistency, we can now skip straight to this aim.

3.1.3 Frege-consistency and Hilbert's proofs

We have observed several times already that Hilbert's proofs do not establish consistency by Frege's lights, if only because Hilbert's axioms are inept to express thoughts. Blanchette provides a far more detailed examination of this issue. Her discussion also makes sense of why Frege is reluctance to compromise on the doctrine that only a complete sentence can express a determinate thought.

Blanchette argues that, for Frege, the reason why the syntactic consistency of sentences can never be enough to evidence the consistency of thoughts is that the relation between sentences and thoughts is one-to-many. That is to say that different sentences can express the same thought (but each determinate sentence will express only one thought). Blanchette points out that it is essential to Frege's logicist project that sentences of entirely different syntactic structures can express the same thought. She asks us to consider the thoughts expressed by the following sentences:

a. Every cardinal number has a successor.

b. $(\forall x)(Nx \supset (\exists y)Syx)$.[2]

According to Frege, these sentences express the same thought despite their distinct syntactic structures. Sentence (a) is the kind of sentence which Frege aims to give a logical treatment of in *Grundlagen*. Sentence (b) is a formalisation of sentence (a) where N is read as cardinal number and S as a two-place successor relation.

[2]These sentences are taken directly from Blanchette (1996, 331).

In this way, the seemingly distinct sentences are connected by expressing the same thought. This can be discovered by conceptual analysis – in this case of the concept of number. Although they express the same thing, the sentences each exhibit different features of what they express; their distinct syntax reflects different aspects of the thought's structure. Since different sets of sentences with different syntax can express the same set of thoughts, then a proof which establishes a feature of one set of these sentences exhibiting one aspect of the thought's structure will not *always* be enough to establish that the thought has this feature. That is because there are other sets of sentences which are relevant to those thoughts and these sentences may have entirely different formal properties which reveal aspects of the thoughts that are not revealed by the initial set of sentences.

Put this in another way: consistency is a negative property, it says of some sentences or thoughts that we *cannot* derive a contradiction from them. In other words, it is the *lack* of contradiction that establishes consistency. This makes *in*consistency, by contrast, a positive metamathematical property: it says of some sentences or thoughts that we *can* derive a contradiction. Thus, to establish that some thoughts are inconsistent, we only require one inconsistent set of sentences expressing them to evidence the possibility of deriving a contradiction. To establish consistency, however, we must reason about all the sets of sentences expressing a thought and establish that contradiction is not something that can be derived from any of the sets. Establishing the consistency of a single set of sentences expressing the relevant thoughts will not be enough.

Given that this is the case, establishing that one set of sentences is consistent only shows that a contradiction between the thoughts – if there is one – has not *yet* been detected. Importantly, if the thoughts are inconsistent it is detectable in the sense that they could be expressed by a set of sentences which *are* syntactically inconsistent and thus reflect the thought's inconsistency. A syntactic proof is a proof pertaining to a specific set of sentences. As such it is incapable of excluding the possibility of the syntactic inconsistency of another co-expressing set of sentences. There is always a danger lurking that this distinct set of inconsistent co-expressing sentences reveals that the thoughts were inconsistent all along and that this inconsistency is only concealed by the kind of syntactic complexity in the initial set of sentences. Blanchette gives the example of the open-sentence pair:

(i) x is a natural number; x has no successor.

She points out we can use a "Hilbert-style" proof to show the syntactic consistency of this pair of sentences, but by Frege's lights they are inconsistent. Their inconsistency is revealed by analysing the concept of number as Frege does:

(ii) $(\forall x)(Nx \supset (\exists y)Syx)$

Elsewhere, Blanchette uses the simple example:

C. Jones had a nightmare; Jones didn't have a dream.

'Jones', 'nightmare' and 'dream' can easily be reinterpreted to give two true sentences and thus demonstrate their syntactic consistency; but they do not establish consistency for *Frege* since the inconsistency of the sentences lies not in the syntax but in the unanalysed concept of a nightmare. Appropriate conceptual analysis yields the following:

D. Jones had a bad dream; Jones didn't have a dream.[3]

The sentences in the pair (D) express the same thoughts as the sentences in the pair (C) but can be shown to be syntactically inconsistent. The inconsistency of the sentences in (D) evidences that the thoughts expressed are inconsistent and thus that the thoughts expressed in (C) are inconsistent. Thus, the syntactic consistency proof of (C) does not do enough to rule out the existence of another pair of sentences which show the thoughts expressed to be inconsistent.

Thus, we have it that a mere proof of syntactic (or proto-proof-theoretic) consistency, cannot by itself establish Frege-consistency.

[3] Blanchette (2014)

Proving Frege-consistency

Towards the end of her article, Blanchette articulates a natural question: how, according to Frege, could we establish this negative property of thought consistency by *formal* means? We have seen that in the correspondence, Frege recommends establishing consistency by first establishing truth and existence (Frege's principle) and that the only means of first securing truth and existence for the geometric primitives is a non-formal appeal to spatial intuition. We have also seen that, at a later point, Frege attempts to answer such a question himself by sketching out a suggestion for an independence test which employs a formally specifiable means of reinterpreting sentences expressing determinate thoughts. Blanchette provides an answer to this question, and this answer helps to further explain why the Hilbert-consistency of an axiom set does not imply the Frege-consistency of the axiom set. As such, it will be helpful to set out Blanchette's answer before moving on.

Blanchette points out that in order to establish Frege-consistency by *formal* means we would have to show that all sentences capable of expressing the thoughts were syntactically consistent. This would give enough evidence to conclude that the thoughts themselves were consistent. She then makes the point that Frege is right to be pessimistic about the prospect of such a proof. For one, due to the compositionality of language we are always able to formulate new sentences capable of expressing any given set of thoughts *ad infinitum*. For example:

- Jones had a bad dream.
- Jones had a bad dream or he had a bad dream.
- Jones had a bad dream or he had a bad dream or he had a bad dream.

Thus, any treatment could not take each set of sentences as a separate case, but would have to exploit a generalised characterisation of all sentences capable of expressing the thoughts. Since the sentences in these sets could be of entirely different complexity and syntax, there are only two things they all have in common. The first is that they are not entirely unrelated to each other but are connected by conceptual analysis. The second is that they all express the same thoughts.

We cannot exploit an appeal to the thoughts because the only way we can get to a thought is via a sentence. Thus, we cannot appeal to the characteristics of some thoughts in order to identify all the sentences capable of expressing them. Since we cannot access thoughts directly, any attempt to do so would only privilege one set of sentences over the others.

Neither can we usefully exploit an appeal to conceptual analysis. This is because analysis can similarly continue *ad infinitum* and in all directions. Depending on our theoretical interests, we can *analyse* a syntactically simple sentence into a more complicated one, or vice versa. Both will display different aspects of the thought's structure. Blanchette maintains that Frege does not believe in a level of analysis which alone is privileged in revealing the 'true' structure of the thought. Even the sentences in a formal language – although they have a more perspicacious syntax – do not give a direct model of the thought's syntax.

The best we could do by appeal to analysis is to gather a large finite number of sentences which, for various theoretical purposes, count as analyses of each other and so express the same thoughts. We could then provide a large finite number of consistency proofs of these sentences. It is clear this will not be enough to establish that the thoughts are Frege-consistent until we close off the possibility of any further analysis. To this point, Blanchette makes the observation that even if analysis did terminate, we still could not prove consistency by merely syntactic means, as Hilbert does. This is because to establish consistency we would not only need a syntactic consistency proof but also the further claim:

> ... that the analysis is finished. And this, of course, is a claim that turns not just on the syntactic form of the sentences ... but on the content of their non-logical terms. It is the kind of claim, in short that is indemonstrable via a Hilbert-style consistency proof (Blanchette 1996, 335).

Let us call this the totality claim, i.e. the claim that our set of sentences exhaust the analysis of the initial sentences, and so constitute all the sets of sentences capable of expressing a given set of thoughts. Blanchette's point is that if we were to attempt to establish consistency by appealing to an endpoint of analysis then we must establish the totality claim alongside our syntactic consistency proofs. However, the totality claim cannot be established by mere syntactic means since it is a claim about the *content* of the sentences; it says of some sentences that their constituent expressions do not preserve their meaning if replaced by any further expressions than those which have already been revealed by analysis.

In fact, Blanchette's point here can be extended further: to appeal to conceptual analysis *at all* in order to characterise all of the relevant sentences which co-express some thoughts, is already to go beyond a syntactic appeal. Before we reach the totality claim, we have already gathered the different sentences which various analyses have yielded – the potential totality, as it were. However, the process of analysis by which we have gathered them is not a syntactic one either. Take again the simple example of an analysis of the concept *nightmare* yielding the concept *bad dream*. What allows us to make this substitution has very little to do with syntax. It is legitimised by the fact that both expressions either express the same sense or refer to the same concept. In either case, an appeal is made to their sense either as an ends in itself or as a mode of presentation of a concept. The sense of these expressions constitutes part of the thought expressed by "Jones had a bad dream". In the end, therefore, the actual mechanism of conceptual analysis has no advantage over an appeal to thoughts. Thus, appealing to a part of a thought will encounter entirely the same problem we saw arise for the strategy of appealing to the whole thought: just as sentences constitute our only and inherently limited guide to thoughts expressions constitute our only and inherently limited guide to parts of thoughts, and to concepts.

In conclusion, neither an appeal to conceptual analysis nor to the thoughts they co-express can give us a generalised characterisation of all the relevant sentences which need to be examined in order to infer the consistency of the thoughts. We have seen that this is because both strategies require some appeal to the thoughts themselves, which cannot be provided by a syntactic proof.

By these considerations, Blanchette definitively establishes that mere syntactic consistency does not imply Frege-consistency. Taking Blanchette's diagnosis into account, we can make sense of Frege's uncharitable view of Hilbert's proofs. We have seen that Hilbert's proofs neglect determinate geometric *thoughts* and that Frege has good reason to be pessimistic while writing to Hilbert that there could be a way of establishing the Frege-consistency of any *thoughts* whatsoever by purely syntactic means.

3.2 Another threat of a merely verbal disagreement

We can now set out how Blanchette's dichotomy between Frege-consistency and Hilbert-consistency comes to bear on the priority reading.

We have seen in detail that, on the priority reading, Frege and Hilbert are characterised as forwarding opposite priority claims. Most generally, Frege maintains:

> **Frege's Principle:** Existence is conceptually prior to consistency.

While Hilbert insists, to the contrary that:

> **Hilbert's Principle:** Consistency is conceptually prior to existence.

Now we see that the priority reading can no longer straightforwardly characterise Frege and Hilbert as having opposite priority claims. Instead, their principles should be more faithfully understood in the following way:

> **Frege's Principle:** Existence is conceptually prior to Frege-consistency.

> **Hilbert's Principle:** Hilbert-consistency is conceptually prior to existence.

Since these two kinds of consistency are distinct such that establishing Hilbert-consistency does not entail Frege-consistency, it seems that Frege and Hilbert are not disagreeing after all about the order of priority between the existence of the primitives and consistency of the axioms. We can make this point vivid by identifying a further priority claim which neither of them reject:

> **Frege-Hilbert Principle:** The Hilbert-consistency of the geometric axioms is conceptually prior to the existence of the geometric primitives *which is in turn* conceptually prior to the Frege-consistency of the geometric axioms.

I am not suggesting that either Hilbert or Frege could be read as advocating such a principle; only that neither of them appear to say enough to deny such a claim and therefore that the seemingly opposing priority claims we characterised are in fact *compatible*. In this way, we lose the priority reading's characterisation of the point of disagreement between Frege and Hilbert in their correspondence.

Of course, one way in which the priority reading might try to avoid this threat of a verbal disagreement is to reject Blanchette's dichotomy between the two kinds of consistency and insist that Frege and Hilbert have the same notion of consistency in their sights. However, I do not think that attempting to ignore or reject Blanchette's work is a real option for the priority reading; not merely because it would be difficult for the reading to undermine the explanation which Blanchette sets out – but also because I think we can generate a similar kind of problem for the priority reading internally. This can be done by observing the priority reading's account of Hilbert's early proto-proof-theoretic conception in contrast to Frege's thought-consistency. In this sense, Blanchette's point is a sophisticated articulation of assumptions which the priority reading is already committed to. Thus, the priority reading has no choice but to address this issue by demonstrating that it is able to substantively characterise the disagreement between Frege and Hilbert.

In providing this alternative characterisation, the priority reading must deal with the fact that Frege and Hilbert employ the same word to refer to distinct kinds of consistency; which means the threat of reducing Frege and Hilbert's disagreement to a merely verbal debate has returned even stronger. This is indeed a threat; I simply do not believe that Frege and Hilbert's dispute over consistency would have been resolved if they had merely come to terms with their divergence in meaning (for instance, if Blanchette had been around to explain it to them). It is true that Frege and Hilbert's priority claims are not the opposite of each other in precisely the neat way that the priority reading earlier proposed. Furthermore, it is true that Frege and Hilbert's disagreement is to *some* extent a verbal dispute over the meaning of 'consistency'. However, these facts do not entail that the disagreement is a *merely* verbal one and, in the next section, I will argue that Frege and Hilbert should be understood as having a *deliberate* verbal dispute, and that their lack of disambiguated terminology is entirely well-motivated.

Blanchette herself, considers the dispute over consistency to be grounded in a substantive disagreement which she calls the 'deeper disagreement' between Frege and Hilbert. The most succinct explanation of which she gives in the following passage:

> The crucial point in Frege's criticism of Hilbert, however, is not a disagreement about particular analyses or the consequent failure of particular consistency and independence claims, but instead concerns the general methodology of consistency and independence proofs. Because for Hilbert the consistency of a set of sentences turns entirely on the overall structure they exhibit, while for Frege the consistency of the set of thoughts expressed turns additionally on the contents of the non-logical terms appearing in the sentences, Hilbert-consistency doesn't imply Frege-consistency (Blanchette 2014).

She identifies the central point of dispute as the fact that Frege considers the content of the non-logical terms to be relevant to consistency, when Hilbert does not. This point of divergence is presented as a wholly intentional one. For her, Frege and Hilbert self-consciously offer genuine alternatives for how consistency *should* be understood, and these are the two kinds of consistency that she articulates in her paper.

Although Blanchette provides an explanation of the crucial difference between Frege and Hilbert's conceptions, commandeering her characterisation will not yet deliver the priority reading from the woods. It is one thing to establish what the disagreement consists of, but another to establish that the disagreement is *deliberate*. In order to do the latter task, we must provide a well-motivated explanation of why Frege and Hilbert would deliberately advocate the particular conception of consistency which they do. We must also explain why they would *reject* their opponent's conception of consistency, which is to say that they would reject what we have called the Frege-Hilbert principle. Only this would suffice to recapture a genuine disagreement between Frege and Hilbert and, in doing so, provide a substantive characterisation of disagreement in the controversy.

To make it clear why this is the case, suppose we commandeered Blanchette's explanation of the difference between Frege and Hilbert's conceptions and used it as a characterisation of their underlying point of dispute. In this case, the threat of a merely verbal dispute re-emerges. Perhaps Frege and Hilbert were not aware that they had different views of the inferential relevance of the content of the non-logical terms, such that, if this had been exposed, the disagreement would have been settled as a mere difference in the meaning of 'content' with respect to the axioms. In this case, both Frege and Hilbert would conceded that the content of the non-logical terms was relevant to *thoughts* and thus to Frege-consistency but not to *sentences* or to Hilbert-consistency. In this case, neither would be rejecting what I have called the Frege-Hilbert principle.

One more remark before we turn to developing the priority reading: If Frege and Hilbert are merely offering compatible accounts of consistency which run in parallel with each other, then they are mistaken to believe they are talking of the same phenomenon, disagreeing, or even interacting with their opponent's account. If they are genuinely disagreeing, it must be that there is at least one common point of contact between Frege and Hilbert from which their accounts diverge in different directions – there must be *something* which Frege and Hilbert are disagreeing *about*. It is this point that we must find if we are to establish the nature of the disagreement between Frege and Hilbert and establish it as something more than a verbal and terminological confusion.

3.3 The priority reading on the deeper disagreement

In this section we will show how the priority reading can avoid the threat of interpreting Frege and Hilbert as having a merely verbal dispute, and we will do so by developing what Blanchette has called the "deeper disagreement" between Frege and Hilbert. I do not think that, in the face of this threat, the priority reading should be abandoned. Even if Frege and Hilbert's priority claims are not directly in tension with each other, I will argue that the reason that the interlocutors have for adopting one priority claim over the other *is* in direct tension. As such, we will develop the priority reading so that it can go deeper in order to offer a satisfying characterisation of the difficult and important issue at heart of the Frege-Hilbert controversy.

In the last section I suggested that Hilbert and Frege do not explicitly disambiguate what they mean by 'consistency' because each deliberately advocates their own conception as a superior candidate for the orthodox understanding of consistency. I then noted that, if this is the case, we are left asking what the difference is between the two conceptions and why the correspondents would advocate one over the other. We saw that Blanchette answers the former question: the essential difference between Frege-consistency and Hilbert-consistency is the significance of the content of the non-logical terms. What we require in order to satisfyingly answer the latter question is a well-motivated reason for Hilbert to insist on the irrelevance of the content of the non-logical terms to consistency, and for Frege to insist on their importance.

It will be the task of this section to give such a reason and in doing so provide an answer to the question of why Frege and Hilbert would reject one conception of consistency in favour of another – thus providing an explanation of their dispute whereby their admittedly verbal disagreement over the meaning of consistency is both deliberate and important.

In broad brush strokes, I will suggest that Frege and Hilbert concur on the usefulness and power of the axiomatic method but that they disagree as to how this method should best be realised. In particular, they disagree about which advantages of an axiomatisation take *priority* over other theoretical advantages.

Let us spell this out a little more. It is as yet unobvious how *this* disagreement provides a motivation for Frege and Hilbert to maintain the particular conceptions of consistency which they do. Some initial light can be shone here by considering the following connections. Generally speaking, whether one holds that the content of the non-logical terms in an axiom set is relevant to consistency depends upon the view one has of what an axiom set is, i.e. whether the members of an axiom set are sentences with a determinate meaning or not. One's view of what an axiom set is like, in turn, will depend upon one's view of the purpose of the axiomatic method in general, i.e. is it to investigate the relationship between truths which we already have, or to rigorize our theories and develop new truths. Depending on one's view of the purpose of the axiomatic method, then, one will have a different opinion about which theoretical advantages take priority over others; i.e. successfully capturing all the truths of a theory, or rigorizing the primitives concepts to ensure the theory's consistency.

In this way, Frege and Hilbert's disagreement over which theoretical pay-offs have priority, is the reason that their lack of disambiguated terminology is well-motivated and as such does not evidence a merely verbal dispute. Of course, the connection as we have articulated it here is only a sketch, so let us now turn our attention to fleshing it out in more detail. We will do so back to front, beginning with Blanchette's characterisation of the deeper disagreement between Frege and Hilbert. To explain the motivations and alternatives when it comes to deciding the contribution of the non-logical terms to an axiomatisation, we will have to delve deeper into the philosophical and mathematical projects that shape Frege and Hilbert's approach to the axiomatic method in the period around 1900. Once this has been set out it will become clear how this underpins their different conceptions of axioms and of consistency – both of which are now familiar. Finally, we will present the new priority reading in light of these considerations.

3.3.1 The purpose of the axiomatic method for Frege

Let us begin with Frege's approach. What we will see is that Frege's insistence on the relevance of the primitives for consistency issues from the philosophical reasons that commit him to the conception of an axiom as a *true thought*. Frege's position is thus well-motived by the philosophical project which he employs the axiomatic method to carry out.

To see this, let us begin by unpacking Frege's philosophical project. For Frege, any logical investigation is most fundamentally an investigation into "the logical linkage of truths" (Frege 1906/1971*b*, 50). Indeed, Frege makes it clear that the centrality of truth to his approach cannot be exaggerated:

> It would not perhaps be beside the mark to say that the laws of logic are nothing other than an unfolding of the content of the word 'true'. Anyone who has failed to grasp the meaning of this word – what it marks off from others – cannot attain to any clear idea of what the task of logic is (Frege 1897/1979, 128).

How is the task of logic – so conceived – to be carried out? In Frege's work, it is realised in part by a particular, foundational approach to analysis. Briefly put, Frege's analysis aims to investigate the nature of certain truths by first taking for granted *that they are truths*. Recall that a very revealing feature of his conception is his insistence that thoughts are secured as true *prior* to their use as axioms.

This approach and its motivations are best unpacked by Frege himself in *Begriffschift*:

> The most reliable way of carrying out a proof, obviously, is to follow pure logic, a way that, disregarding the particular characteristics of objects, depends solely on those laws upon which all knowledge rests. Accordingly, we divide all truths that require justification into two kinds, those for which the proof can be carried out purely by means of logic and those for which

> it must be supported by facts of experience. But that a proposition is of the first kind is surely compatible with the fact that it could nevertheless not have come to consciousness in a human mind without any activity of the senses. (Since without sensory experience no mental development is possible in the beings known to us, that holds of all judgements). Hence it is not the psychological genesis but the best method of proof that is at the basis of classification. Now, when I came to consider the question to which of these two kinds the judgements of arithmetic belong, I first had to ascertain how far one could proceed in arithmetic by means of inferences alone, with the sole support of those laws of thought that transcend all particulars (Frege 1879/1964, 5).

Frege places this explanation in the preface of *Begriffschift*, which shows he is self-concious of his own approach and the way in which it provides the frame for his project at large. He goes on to say that his guiding purpose – and the reason he found it necessary to invent his *begriffschift* (and along with it modern predicate logic) – was to aid his investigation into the grounds of certain "scientific truths". Frege classifies truths according to the kind of justification which is used in their 'best method' of proof. Following Kant, he delineates two categories of justification: experiential and logical. A truth can admit of more than one proof and it is clear that Frege is not interested in the contingent proof by which we happen to come to know some truth, or in the most explanatory or simple proof. Instead, he takes it to be obvious that if a scientific truth can be shown to have a proof which makes purely logical appeal – most probably in addition to having an impurely logical proof appealing to facts of experience – then this is enough to establish the logicality of the truth under consideration. Thus, to establish the logicality of the truths of arithmetic – as Frege intends to – the construction of a single proof will suffice so long as the proof derives the arithmetical truths from nothing other than the truths and laws of logic.[4]

Let us step back a moment to consider this approach. Its most striking characteristic

[4]For more on this see Sullivan (2004), Dummett (1981) and Dummett (1998).

is that it does not aim to *establish* truths, but instead to investigate truths which we already have. Frege notes in the same preface that "one should not mind the fact that there are no new truths in my work" (Frege 1879/1964, 6). Neither is it Frege's aim to establish, or indeed to falsify, the truth of the scientific propositions whose methods of proof are under investigation. To emphasise the point once again, what Frege's analysis questions whether it is possible to produce a proof of a scientific truth which proceeds solely along logical means and not whether the scientific proposition being proven *is a truth indeed*. This, Frege takes to be a matter which is resolved prior to the axiomatisation, in two senses. One is that, as we have seen, the meaning of the sentences expressing axioms is to be coordinated by the process of 'explication' in the axiomatisation's "propaedeutic". The other is that it is inherent in the nature of what an axiomatisation is that it deals in truths. In other words, an axiomatisation organises a body of truths by identifying a finitely specifiable set of fundamental truths which the rest of the truths can be derived from as theorems. One passage from Frege's "Foundations of Geometry" explains this principle very well:

> *Only true thoughts can be premises of inferences.* Therefore if a thought is dependent upon a thought-group Ω, then all the thoughts in Ω that are used in the proofs must be true (Frege 1906/1971*b*, 105).

Frege then goes on to consider that one can reason hypothetically when it seems that one is making deductions from a thought without accepting that the relevant thought is true (Frege clearly has Hilbert's approach in mind here). He says that in such a case it is not the same thoughts (i.e. it is not the members of Ω, the thought-group asserted as true) which are the premises of the inference but "certain hypothetical thoughts that contain the thought in question as antecedents". These antecedents will carry through to the conclusion and can only be omitted if they are "admitted as true". Most simply put, Frege's point is that conditional reasoning can only ever hope to yield a conditional conclusion. Thus, even in the conditional case *the premises of the inferences are true thoughts*; they merely happen to be conditional thoughts (Frege 1906/1971*b*, 105-6).

What must be understood about Frege's conception of an axiom is that there is a sense in which there is no risk of 'discovering' that one of our axioms which we assumed to be true was in fact false. In the worst case, we discover that the axioms we have adopted lead us to contradiction and must be revised. Even in this case, we discover the problem by first adopting the axioms as truths and deriving their consequents. This would lead us to abandon the defective axiom or axioms responsible, but falls short of making them "false". Frege underlines this point with a vivid comparison:

> A false axiom – where the word "axiom' is understood in the proper sense – is worthy of exhibition in Kastan's Waxworks, alongside a square circle (Frege 1906/1971*b*, 104).

Given the absurdity which Frege here points to, it is natural to ask on what grounds a set of thoughts – or "thought-group" – is initially adopted as true? Frege clearly thinks that we cannot investigate the nature of the axioms by an agnostic suspension of belief, but neither does he think that we should accept that any old thoughts are true in blind faith. Instead, Frege says that "axioms express basic facts of our intuition" (Frege 1903/1971*a*, 26). In this sense, axioms are truths which are known by a source of knowledge other than proof and external to the axiomatisation itself. Frege considers his conception of axioms to be the orthodoxy and lectures Korsolt thusly:

> ...the meaning of the word "axiom" may no doubt be called the traditional, Euclidean meaning. Axioms differ from theorems in that they are unprovable... In saying that modern mathematics no longer designates certain facts of experience with its axioms but at best indicates them, Mr Korselt bring the axioms of modern mathematics into contrast with those of Euclid (Frege 1903/1971*a*, 52-53).

Frege only ever makes suggestive remarks about such an alternative source of knowledge and says nothing at all about how it is to be recognised or made rigorous. What he does say is enough to make it clear that for him the axiomatic method is a method for organising and investigating truths which we already have. As such, an axiomatisation is only useful if there is some means by which its logico-deductive structure is, as it were, tethered to reality. This is the central role of an axiom; the axioms of geometry ensure that the theory of geometry is *about space*, just as the axioms of arithmetic ensure that the theory of arithmetic is *about numbers*, and so on. In other words, it is essential to Frege that an axiom set has a *fixed subject matter*, for it can only successfully ground the axiomatisation in reality if its members are "facts of experience" or "basic facts of our intuition" or "self-evident truths", as Frege variously categorises them.

Altogether, we can see that Frege defends a traditional understanding of an axiom whereby an axiom is a fundamental truth of a particular theory which has been established by some reliable means outwith the axiomatisation. This is what it means to say that Frege investigates the relation *between* truths rather than establishing *new* truths.

Given that this is so, it is natural to ask how such a relation between truths is to be investigated. The short answer is that it can be investigated with the help of the semantic perspicacity of Frege's *begriffschift* whilst following Elucid's canonical axiomatic method. This answer is quite general and it will be the same for geometry and arithmetic, despite the fact that Frege conjectures arithmetical truths to be logical truths in contrast with geometric truths which he claims requires appeal to experience. This is because, although both groups of truths turn out to be justified by different sources (corresponding to those Frege sets out in the preface to begriffshcift) that this is the case is established by the *same means*. For Frege, the real measure of an axiomatisation of any theory – including Hilbert's – is whether it aids a foundational investigation into the justification of *the* most basic truths of that theory.

How does all of this bare on our question of how and why Frege would adopt a particular conception of the semantic contribution of the non-logical terms in an axiom-expression? To answer this most directly, we return to the point which we have emphasised: for Frege an axiom is – most essentially – a *truth*. It follows from this that the goal of the non-logical terms in an axiom-expression is first and foremost to contribute towards expressing a truth by *themselves* expressing a sense. Frege puts the same point succinctly were he says:

> ...the expression of an axiom must contain no unknown sign, for otherwise it would express no thought at all (Frege 1903/1971*a*, 52-53).

Therefore, Frege's conception of the contribution of the non-logical primitives flows right out of his conception of an axiom as a thought that is true. How the non-logical primitives are equipped with the requisite sense, is another matter. Here, Frege is firm that the best we can offer is coordination of sense through the process of elucidation and that definition is not an option:

> My opinion is this: We must admit logically primitive elements that are undefinable....(Frege 1903/1971*a*, 59).

Frege later elaborates his issue with primitives being defined as the following:

> Once the explanation including the [axioms] has been posited, the latter may be asserted as true; however, their truth will not be founded on an intuition, but on the definition. And it is precisely because of this that no real knowledge is contained in them – something which undoubtedly is the case with axioms in the traditional sense of the word (Frege 1903/1971*a*, 27).

We are again led back to the same source to explain Frege's prolonged protest to Hilbert's claim that the primitives are defined by his axioms. That this approach undermines the central role of an axiom as that which tethers the axiomatised theory to the reality which it seeks to investigate.

Another way to put this important point is to observe that – from Frege's realist perspective – any investigation into some field of mathematics must start by adequately characterising the relevant objects of that field, whether these are points and lines or real numbers and ratios. This is only done when the axioms are truths, since in this case the primitive-expressions all express a sense and *determine a reference*. The axiomatisation becomes an axiomatisation of geometry, for example, when and only when its primitive terms successfully refer to *the* geometric objects and concepts.

We can conclude that Frege's insistence on the importance of the non-logical terms to consistency is very well-motivated by how he viewed the purpose of the axiomatic method. Frege's conception of the contribution of the non-logical primitives flows right out of his conception of an axiom as a true thought. From this point of view, the only assumption which cannot be discharged from an axiomatisation is the knowledge or 'intuition' of the fixed collection of objects which forms the proper subject matter of a theory.

3.3.2 The purpose of the axiomatic method for Hilbert

Just as with Frege, Hilbert's stance on the relevance of the content of the non-logical primitives is well-motivated and issues from his conception of the purpose of the axiomatic method.

The first question to ask is how Hilbert's use of the axiomatic method compares to Frege's use of the axiomatic method to model and study the logical relations between truths. There is no doubt at all that Hilbert shared Frege's reverence of Euclid's method; he describes it as the method which "deserves first rank" for the purpose of "the final presentation and the complete logical grounding of our knowledge" (Hilbert 1900/1996*a*, 14). In this statement about the purpose of the axiomatic method we see much in common with Frege. They share the same concern with logical rigour. Indeed, in their correspondence Hilbert stresses to Frege that this is an important feature of his axiomatisation, saying:

> ... of course I also believe I have set up a system of geometry which satisfies the strictest demands of logic (Hilbert 1899/1980*d*, 39).

The other striking feature in common with Frege is the aim to provide a foundation for the pre-theoretic – or at least pre-axiomatic – knowledge we have of an area of mathematics – in this case geometry.

As to whether Hilbert has the same concern to investigate the relations between truths, we must be slightly careful. Although Hilbert does in places speak of his axioms as truths, we have seen that this upsets Frege greatly because Hilbert's primitive terms have no pre-theoretical meaning and are implicitly defined. Hilbert explains this approach in his "Mathematische Probleme".

> When we are engaged in investigating the foundations of a science, we must set up a system of axioms which contains an exact and complete description of the relations subsisting between the elementary ideas of the science. The axioms so set up are at the same time the definitions of those elementary ideas, and no statement within the realm of the science whose foundation we are testing is held to be correct unless it can be derived from those axioms by means of a finite number of logical steps (Hilbert 1900/1996*b*, 447).

Thus, although both share a concern to investigate the foundational truths of a branch of mathematics, the particular truths they are concerned with are different. As Frege suggests to Hilbert in their correspondence, Hilbert's truths can be understood as the conditionals truths which have Frege's truths as their consequents Frege (1900/1980*b*).

We have seen already what Frege has to say about Hilbert's approach: to derive the truth of the theorems from a definition means that the theory can contain 'no real knowledge'. However, Hilbert presents this approach as something which *must* be carried out in any scientific investigation of our foundational knowledge. How are we to reconcile this difference of opinion? The crucial observation, I think, is that Hilbert has an entirely different conception of the relationship between an axiomatisation and the mathematical reality which it axiomatises.

We saw that Frege's sustained criticism of Hilbert's terminological imprecision regarding his axioms was motivated by his conception of axioms as a collection of truths. These truths fulfilled the essential role of tethering the axiomatised theory to the reality it was intended to rigorize. This was done by means of fixing the reference of the primitive expressions unambiguously to the relevant mathematical objects and relations, prior to the construction of the axiomatisation. Hilbert has a very different understanding of how an axiomatised theory is connected to pre-theoretic knowledge. On this point, I am in agreement with Hallett, who draws the following conclusion.

> In sum, then, the view is clear in Hilbert that geometry must relinquish finally any claim to offer in a straightforward way a description of the shape and/or the behaviour of bodies in space... As Hilbert says somewhat later in his notebooks, "The points, lines, planes of my geometry are nothing other than things of thought, and as such have nothing whatsoever to do with real points, lines and planes" (1908). Thus, geometry comes to be recognised as a product of the mind, a product largely independent of the outer world, an independence reflected, to repeat, in the example of Hilbert's work on the notion of congruence (Hallett 1994, 166-167).

Here Hallett quotes a very striking passage from Hilbert's notebooks. Indeed, it reads more like it was written by Frege as a criticism of Hilbert's ontology, rather than by Hilbert himself. For Frege, an axiomatisation which falls short of making reference to real points and lines and planes has failed in one of its fundamental purposes and cannot be called an axiomatisation of *geometry*. Clearly, this is not so for Hilbert, who instead claims that a point in an axiomatisation is a different animal altogether from the point of a mathematical realist. This interesting divergence shows us that the disagreement is not over whether there is any way of construing Hilbert's definition of points to refer to "real" points. Rather, what is at issue here is what kind of thing an axiomatisation should *aim* to re-capture.

It is clear that Hilbert thinks that the "things of thought" in his geometry are relevant, if not identical, to geometric reality. We must, however, ask for more detail regarding Hilbert's understanding of the nature of the connection between the former and the latter. Hallett's interpretation of Hilbert amounts to the very appealing view that Hilbert took the former to be an *abstraction* of the latter. Hallett substantiates his interpretation with his translations of a range of quotations from various passages in Hilbert's lecture notes. Two of these are of particular relevance to our discussion here. In the first quotation, Hilbert speaks directly of the relationship between the domain of knowledge and the axiomatic framework. He calls it a "mapping" and illustrates it as follows:

> Through this mapping, the investigation becomes completely detached from concrete reality. The theory has nothing more to do with real objects or with the intuitive content of knowledge. It is a pure thought construction, of which one can no longer say that it is true or false. Nevertheless, this framework has a meaning for knowledge of reality, in the sense that it presents *a possible form of actual connections*. The task of mathematics is then to develop this framework of concepts in a logical way, regardless of whether one was led to it by experience or by systematic speculation (Hilbert 1922/2013, 3, emphasis mine).

We might explain this view by saying that, for Hilbert, the *abstracta* are entirely distinct from the objects they are abstracted from. As such, there is a clean cut between the various sources of our mathematical knowledge and its axiomatised development. Nevertheless, the study of the properties of the *abstracta* have implications for the primitive objects because the *abstracta* encode certain possible conceptual connections which the primitive objects can enter into. Another way to describe this is to say that the axiomatic theory *models* reality (and of course this will turn out to be far more than a metaphor).

This explanation only goes so far. We must ask what Hilbert means by "a possible form of actual connections" and how this is apt to model reality when a clean cut has been made between *abstracta* and the objects of mathematics. The latter is a very large question indeed, but we can glimpse how Hilbert may have begun to answer it by the second relevant quotation Hallett picks out from Hilbert's lectures.

> In general we must state: Our theory furnishes only the schema of concepts connected to each other through the unalterable laws of logic. It is left to human reason how it wants to apply this schema to appearance, how it wants

> to fill it with material. This can happen in manifold ways. But whenever the axioms are satisfied, then the theorems must apply too. The easier the application and the more kinds of application there are, the better* the theory.
>
> *Any system of units and axioms which gives a complete description of the appearances is as justified as any other. Show nevertheless that the axiom system specified here is, in a certain respect, the only possible one (Hilbert 1894/2004, 60).

What this passage brings to the fore is Hilbert's focus on using an axiomatisation to recapture and investigate the primitive geometric *concepts* in contrast with Frege's focus on recapturing the primitive geometric *objects* and investigating *truths about those objects*. I think this difference lies at the heart of Frege and Hilbert's divergent conceptions of the proper use of the axiomatic method. We shall duly return to it in the following section.

The other point to notice is that Hilbert relegates the question of the connection between the axiomatisation and reality, to the question of the theory's application, i.e. whether some system of things, some "material", falls under the concepts defined by the axioms. This makes sense of where we saw Hilbert tell Frege in the correspondence that:

> But it is surely obvious that every theory is only a scaffold (schema) of concepts together with their necessary connections, and that the basic elements can be thought of in any way one likes. If in speaking of my points I think of some system of things, e.g. the system: love, law, chimney-sweep... and then assume all my axioms as relations between these things, then my propositions, e.g. Pythagoras' theorem, are also valid for these things. In other words: any theory can always be applied to infinitely many system of basic elements. One only needs to apply a reversible one-one transformation and lay it down that the axioms shall be correspondingly the same for the transformed things... But the circumstance I mentioned can never be a defect in

> a theory,* and is in any case unavoidable. However, to my mind, the application of a theory to the world of appearances always requires a measure of good will and tactfulness.
>
> *It is rather a tremendous advantage (Hilbert 1899/1980*d*, 40-41).

Whereas in Frege we have the idea of a theory as a set of thoughts, Hilbert describes a theory as a schema of concepts. I think the best explanation of what this amounts to comes from Frege's diagnosis that Hilbert's axioms define a six-place higher-order property. If we bare in mind Hilbert's remark that his points have nothing to do with "real" points, then Hilbert is happy to say that any system of first-order concepts which satisfies this property can be called a system of 'geometric' concepts, and any objects which fall under these first-order concepts can likewise be called 'geometric' objects.

If the axiomatisation is thought of as characterising, not the objects of geometry, but the connections between its concepts by defining a higher-order concept, then the relationship between axiomatisation and the axiomatised becomes quite different from what we saw in Frege. In particular, it is the independence of subject matter which Hilbert achieves in his geometry which marks his break from the traditional conception of an axiom as a fundamental truth, alla Euclid and Frege.

This important feature of Hilbert's new (and now dominant) axiomatics bears most directly on the question of what motivates Hilbert's view about the irrelevance of the content of the non-logical primitives to consistency. The answer is that since, as Frege puts it, the primitive expressions function as variables in defining a higher-order schema of concepts; what is shown to be consistent by a consistency proof is the *schema* which is common to the interpreted and uninterpreted axioms. The content of the primitives under a particular interpretation is entirely irrelevant to this schema. In this way, Hilbert's belief that the content of the primitive expressions is irrelevant to consistency issues from his conception of the primitive expressions in his axioms as inherently variable.

We can pause to draw out a general reflection here: how one attempts to show that the axioms of a theory are consistent depends on how one conceives of an axiom in the first instance. If the content of the primitive expressions is unfixed for the axioms then the content of these expressions will not bear on the axiom's consistency. Contrapositively, if the consistency of the axioms is sensitive to the content of the primitives it must be because the meaning and reference of these expressions is already fixed for the axioms themselves. We began by asking why Frege and Hilbert would adopt one view of the relevance of the content of the primitives to consistency, rather than another. Now we are asking why they would adopt one conception of axioms and their parts rather than another.

We saw that Frege took himself to be upholding a traditional conception whereby an axiom was a truth and also that his driving philosophical project was to explore the nature of 'old' truths to discover whether they were logical or relied on some appeal to experience. Let us now consider what motivates Hilbert's understanding of an axiom. In doing so, we should bear in mind that Hilbert's conception was, at the time, a self-concious break from tradition.

What is immediately striking in Hilbert's writings is that wherever Hilbert speaks of his conception of axioms, he implies that something compelled him adopt it. In the above quotation – for example – he speaks of his conception as 'unavoidable'; in his penultimate letter to Frege, Hilbert on the point of frustration, states that:

> In my opinion, a concept can be fixed logically only by its relations to other concepts. These relations, formulated in certain statements, I call axioms, thus arriving at the view that axioms (perhaps together with propositions assigning names to concepts) are the definitions of the concepts. I did not think up this view because I had nothing better to do, but I found myself forced into it by the requirements of strictness in logical inference and in the logical construction of a theory (Hilbert, 1900/1980*c*, 51).

Hilbert names the restriction that forces his adoption of a particular conception of axioms as the necessary "strictness" in building a theory so as to preserve the logical inferential relations. We can glean some clue of what Hilbert means by this from other passages in the correspondence where he elaborates on this point. For instance, Hilbert tells Frege that he was aware already of Russell's paradox and furthermore claims:

> I found other even more convincing contradictions as long as four or five years ago; they led me to the conviction that traditional logic is inadequate and that the theory of concept formation needs to be sharpened and refined. As I see it, the most important gap in the traditional structure of logic is the assumption made by all logicians and mathematicians up to now that a concept (A SET) is already there if one can state of any object whether or not it falls under it. This does not seem adequate to me. What is decisive is the recognition that the axioms that define the concept are free from contradiction (Hilbert 1900/1980*c*, 51-52).

So when Hilbert writes that his motivation is to ensure strictness in logical inference and theory construction, the background concern he seems to have in mind is that the axioms are *consistent*. It is important for the axioms to be consistent is because it is the axioms alone which are responsible for concept formation. Recall how Hilbert insists on a sharp divide between the pre-theoretic concepts and the concepts implicitly *defined* by the theory. It follows from this that since the axioms are the sole *definiens* of the concept, any potential inconsistency must be rooted in them. This seems to be why Hilbert insists the axioms are secure and that their consistency is the essential hallmark of a strict concept formation, which is logically rigorous in the sense that it is fortified against paradox.

The way Hilbert presents his conception here is very interesting. He speaks of it as issuing from his encounter with the set-theoretic paradoxes. Ferreirós (2009) gives a very convincing explanation of why Hilbert makes this claim. Ferreirós argues that Hilbert – at this stage – took for granted the principle of comprehension, in that he did not see any distance between defining a concept and establishing the existence of a set whose members were the objects falling under that concept.[5]

[5] It is undeniable that Hilbert took for granted some notion of a set in his *Festschrift*, but because this is so

In support of this plausible claim, he points out the following:

> As regards Hilbert, a simple fact suggests that Comprehension was deeply imbedded in his mind around 1900: he tends to use the words "concept" and "set" as synonyms (Ferreirós 2009, 20).

Thus Ferreirós explains, as a result of this underlying assumption Hilbert heeds the warning given by the set-theoretic paradoxes as a warning about rigorous concept formation. Since, for Hilbert, it is the axioms which are responsible for defining the primitive concepts, then it is the axiom's consistency which becomes the obvious focus of avoiding the paradoxes. Ferreirós summarises this in the following way:

> The circle of ideas in which Hilbert is moving should now seem much clearer. An axiom system can always be regarded as the definition of a concept, and a concept is always linked – via the comprehension principle – with a corresponding set. But, from the paradoxes, Hilbert learnt that it does not suffice to define a "concept" such as Euclidean space (or real number) with enough precision to be able to determine whether a given object falls under it or not. The concept of an aleph seemed definite, precise enough to establish whether a given object is or is not an aleph, but the totality of alephs was not a "mathematically existing" object (Ferreirós 2009, 26).

---

implicit in his writings it is hard to say exactly how much set theory he assumed. Ferreirós has been outspoken that, "Hilbert's early axiomatic work (e.g., in his arch-famous Foundations of Geometry) was deeply set-theoretic" (Ferreirós 2016, §4). On the other hand – as is pointed out by Dreben and Kanamori – Hilbert did not make any published mathematical contribution to developing set theory (Dreben & Kanamori 1997, 77). Not directly, at least – for, as they go on to point out:

> ...[Hilbert] fostered its development through his encouragement of Ernst Zermelo. Zermelo began his investigations of Cantorian set theory at Göttingen under Hilbert's influence. Zermelo soon found Russell's Paradox independently of Russell and communicated it to Hilbert. Zermelo then established the Well-Ordering Theorem in a letter to Hilbert... (Dreben & Kanamori 1997, 86-87).

Thus, Hilbert insists that concepts must be defined so as to determine whether an object falls under them or not; which is to say that concepts are not just given sharp boundaries but are also shown to be *consistent*. In this way, Hilbert understands concept regimentation as essential to avoiding paradox. In fact, Hilbert takes this to be so essential that he self-reflectively names the avoidance of paradox through rigorous concept formation to be what 'forces' him to adopt his view of axioms and of the axiomatic method in general.

I think that the following passage, Hilbert's 1904/1967*b* lecture "Über die Grundlagen der Logik und der Arithmetik", lends great support to Ferreirós argument. In it, we find the most extended articulation of what Hilbert found to be the most promising and the most problematic aspects of Frege's approach.

> *G. Frege* sets himself the task of founding the laws of arithmetic by the devices of *logic*, taken in the traditional sense. He has the merit of having correctly recognised the essential properties of the notion of integer as well as the significance of inference by mathematical induction. But, true to his plan, he accepts among other things the fundamental principle that a concept (a set) is defined and immediately usable if only it is determined for every object whether the object is subsumed under the concept or not, and here he imposes no restriction on the notion "every"; he thus exposes himself to precisely the set-theoretic paradoxes that are contained, for example, in the notion of the set of all sets and that show, it seems to me, that the conceptions and means of investigation prevalent in logic, taken in the traditional sense, do not measure up to the rigorous demands that set theory imposes. *Rather, from the very beginning a major goal of the investigations into the notion of number should be to avoid such contradiction and to clarify these paradoxes* (Hilbert 1904/1967*b*, 130).

Most strikingly, this is an instance of Hilbert speaking interchangeably of a concept and a set. Furthermore, Hilbert's implicit assumption that if a concept is well-defined then the corresponding set must exist is manifest in his inference that if a contradiction has arisen from assuming a set to exist then the original concept must not have been well-defined to begin with. Thus, Hilbert explains that Russell's paradox is the result of a lack of restriction, indeed of rigour, in the practice of defining concepts.

Hilbert ends his reflections on Frege's work by emphasising that one of the primary goals of a foundational investigation (we can assume this extends to an investigation of the notions of geometry) is to "avoid such contradiction" and to understand, anticipate, and avoid the threat posed by the set-theoretic paradoxes "from the very beginning".

In conclusion: conceiving of the contribution of the non-logical terms in an axiom-expression as uninterpreted (and thus re-interpretable) is the defining feature of Hilbert's model-theoretic approach. It marks his break from a traditional understanding of an axiom as a bedrock truths, and it is well-motivated by his flee from the set-theoretic paradoxes. For – on Hilbert's new approach – the only means of defining the concepts from which we infer the existence of the corresponding sets, is given by the resources of the theory itself. The theory, in turn, has its axioms vetted for consistency and relies on no pre-theoretic appeal to intuition or any other non-logical source of knowledge.

3.3.3 The final priority reading

Thus, both Hilbert and Frege's conception of axioms and the contribution of the non-logical terms in the axiom-expressions is well-motivated given their background mathematical and philosophical projects. This certainly establishes that each has a good reason for adopting the conception which they do. However, in order to say even more firmly that their disagreement on this issue is a deliberate one (so that they would reject the Frege-Hilbert principle), it is worth reflecting (more explicitly) why each would reject the other's position. Having done this, there will be no doubt that Frege and Hilbert have opposing conceptions of the purpose of the axiomatic method, and that they genuinely and deliberately disagree in their correspondence over the correct guiding theoretical priorities of axiomatising a theory. We will then be in a position to articulate the final priority reading.

We saw that the driving feature of Frege's approach was that a foundational investigation was an investigation of the logical linkages between a fixed set of truths. As such, the role of the axiom set was to connect the theory to the reality that it was constructed to investigate. Which is to say that an axiom's most important job was to capture the proper subject matter of an area of mathematics. We also saw that Hilbert was driven by avoiding the paradoxes of set theory by galvanising his account against any external appeal to intuitions and insisting on consistency as the essential hallmark of existence.

It is clear that Frege and Hilbert's approaches are opposed to each other: Frege wants an axiomatisation to function as an aid to science, unfolding and rigorizing our given primitive concepts by grounding them in a foundation of true thoughts. Hilbert wants to leave behind our vague and potentially contradictory primitive concepts and build our concepts anew, on higher ground. But how precisely are these different approaches in opposition? Let us briefly remind ourselves of the answer to this question. All of Frege's objections to Hilbert – even the terminological ones – can, at base, be understood as a worry about the fact that Hilbert's uninterpreted axioms divorce his theoretical foundation from the subject matter which it is their purpose to characterise. For Frege, geometry is about points and lines, as arithmetic is about numbers, and zoology is about animals. Since Hilbert's axioms are equally interpretable in these distinct background theories they have failed to qualify as distinctively *geometrical*. All of Hilbert's counter-objections to Frege can, at base, be understood as a worry about the risk of paradox that Frege's lax approach to concept formation exposes him to and the insistence that there is no other way to ensure the consistency of a theory other than to test the axioms for consistency and then refuse any appeal outwith these axioms when defining the primitive concepts.

Stepping back, then, we can observe that both Frege and Hilbert prioritise distinct and important theoretical advantages. With Hilbert's approach, we secure our theory from inconsistency by admitting no extra-theoretical content, and in doing so we risk irrevocably severing the link between our theory and the reality we mean it to characterise. With Frege's approach, we allow an alternative source of knowledge to provide the foundational truths of our theory and in doing so risk introducing inconsistency, due to the fallibility of this source.

Put in this light, the deeper disagreement between Frege and Hilbert has an interesting connection with one of the only points Frege and Hilbert agree on. In the very first letter Frege sends to Hilbert, he writes:

> The natural way in which one arrives at a symbolism seems to me to be this; in conducting an investigation in words, one feels the broad, imperspicious and imprecise character of language to be an obstacle, and to remedy this, one creates a sign language in which the investigation can be conducted in a more perspicacious way and with more precision. Thus the need comes first and then the satisfaction (Frege, 1895/1980*c*, 33).

Hilbert writes back praising Frege's view telling him that:

> I believe that your view of the nature and purpose of symbolism is exactly right. I agree especially that the symbolism must come later and in response to a need, from which it follows, of course, that whosoever wants to create or develop a symbolism must first study those needs (Hilbert, 1895/1980*a*, 34).

The "need" Frege talks about is to avoid the vagaries of natural language, and Hilbert generalises the point, adding that these needs are prior to the very development of the symbolism.

A very similar point seems to be true of the deeper disagreement which we have just articulated. Namely, that the purpose of an axiomatisation issues from a theoretical need which comes prior to the construction of a theory – whether it is to avoid paradox or whether it is to ensure reference to a determinate reality. We can capture this by transposing Hilbert's above remark:

> I agree especially that the [application of the axiomatic method] must come later and in response to a need, from which it follows, of course, that whosoever wants to create or develop [an axiomatisation] must first study those needs (adaptation of Hilbert, 1895/1980*a*, 34).

This adaptation is, of course, merely illustrative. The point is that Frege's employment of the axiomatic method and Hilbert's employment of the axiomatic method, satisfy different needs. As such, we can understand the deeper disagreement between Frege and Hilbert as a disagreement over which theoretical 'needs' are to be prioritised over others: preservation of *the consistency of a theory* or preservation of *the subject matter of a theory*. Not only does this establish that the Frege-Hilbert controversy is not founded upon a merely verbal dispute, it shows that the dispute is of philosophical importance for asking – more generally – about the relationship between a theory and reality, or between a theory and *what that theory is about*.

Initially, we used the priority schema to attempt to characterise the deeper disagreement between Frege and Hilbert by saying that they would support opposite instantiations of the schema:

Connect. For any x falling under A, there is a way of establishing that y falls under B by substantive appeal to the fact that x falls under A.

Direct. For any x falling under B, there is a way of establishing that y falls under A without making any appeal to the fact that x falls under B.

Now we can see that the difference in the priority claims that Frege and Hilbert would support is actually merely a manifestation of the real difference in a very different kind of priority: *priority* of theoretical virtue, rather than priority in the order of proof. This is the final priority reading. Whether one gives priority to a theory's consistency over its determinate reference to the primitives – or vice versa – will result in a very different understanding of the role of an axiomatisation; and thus of the semantic composition of a set of axioms; and thus of the whether the relation of consistency is sensitive to the content of the non-logical terms in the axiom-expressions or not. This underlying disagreement about theoretical priorities results in Frege and Hilbert supporting (if not precisely opposite) irreconcilable priority claims concerning the proper order of proof between the consistency of the axioms and existence of the primitives. Both Frege and Hilbert are deliberately advocating their conception of consistency as orthodox in order to thereby advocate their philosophical agenda concerning which are the most important features to safeguard in an axiomatisation.

In conclusion, the tension that is at the heart of the Frege-Hilbert dispute, is between preserving a theory's consistency and preserving its subject matter. Therefore, the final priority reading is able to answer (Qu.) in a way that not only coheres with the context of Hilbert's Principle but also provides a substantive characterisation of the elusive deeper disagreement in the Frege-Hilbert controversy.

3.4 Hilbert's early structuralism

Having found a way to articulate what was at issue between Frege and Hilbert, we cannot help asking the question: *who was right*? It turned out that Hilbert was certainly right to worry that Frege's approach left him vulnerable to paradox. What about Frege's philosophical worry that Hilbert divorced his theory from the proper subject matter of geometry? To answer this, we return to an aspect of Hilbert's approach which we touched upon earlier. Namely, that Hilbert used an axiomatisation to recapture and investigate the primitive geometric *concepts*, whereas Frege used an axiomatisation to recapturing the primitive geometric *objects*.

This presents Hilbert with another issue: if an axiomatisation is thought of as characterising, not the objects of geometry but the connections between its concepts, then what *can* Hilbert say about the mathematical objects falling under the defined concepts. How do we establish *their* existence? This brings us back to Frege's most forceful objection to Hilbert: that he cannot account for either the uniqueness or the existence of the objects falling under first-order mathematical concepts. In this section I will show that Hilbert has a broadly structuralist attitude towards Frege's worries. I will suggest that Hilbert *can* be read as a particular kind of non-eliminativist structuralist because this particular kind of structuralism will become relevant in later chapters. However, I stress from the very beginning that the particularities of Hilbert's structuralism remain up for debate – not least because Hilbert himself had no clear view of such philosophical details.

3.4.1 Non-eliminativist structuralism

Let us first set out the way in which Hilbert's conception of mathematical reality can be understood as structuralist in a way akin to the structuralist positions which have been read into Dedekind.[6] On the structuralist's approach, concepts (or structures) have

[6]Dedekind has been called a structuralist on the basis of passages such as:

> With reference to this freeing the elements from every other content (abstraction) we are justified in calling numbers a free creation of the human mind. The relations or laws which are derived entirely from the conditions $\alpha, \beta, \gamma, \delta$ in (71), and which are therefore always the same in all ordered simply infinite systems, whatever names may happen to be given to the individual elements (compare 134), form the first object of the *science of numbers* or *arithmetic* (Dedekind 1963, 67).

Reck has carried out the most sustained interpretive work on the issue of what kind of structuralism Dedekind held, if any. He concludes that Dedekind can be understood as a *logical* structuralist – a position closest to ante rem structuralism but distinguished from ante rem by its conception of objects as more ontologically robust than mere places or positions. Reck describes Dedekind's structural approach to defining arithmetic as logical in the sense that the truth of arithmetical statements can be given by pure logical reflection on the arithmetical "laws" governing the arithmetical objects. This is the case because all non-arithmetical properties of these objects have been "freed" by "abstraction" and Reck insists that this process of abstraction is – for Dedekind – a logical one. This means that numbers can be called "a free creation" in the sense that a simple infinity is:

> ...identified as a new system of mathematical objects, one that is neither located in the physical, spatio-temporal world, nor coincides with any of the previously constructed set-theoretic simple infinities (Reck 2003, 400).

Reck's reading thus rejects a psychological construal of Dedekind's talk of creation in favour of a reading whereby what we create by the process of abstraction, is something inherently logical (see Reck 2003, §11).

greater ontological importance than objects. That is to say that the existence and nature of the structuralist objects is grounded in the structure rather than the structure being grounded in the existence of the structuralist objects. Thus, rather than defining what it is to be a structuralist object by appeal to a background ontology, or by appeal to intrinsic properties of an object, the structuralist objects are characterised entirely by appeal to the concepts which they satisfy. What guarantees our link to the world of appearances is the inherent nature of the concept which we successfully characterise as being one which is capable of having objects falling in its range. After all, as Blanchette tells us, for a concept to be consistent is for it to be satisfi*able*.

Let us now consider what kind of structuralist position Hilbert can be said to adopt. We should bear in mind, of course, that Hilbert was not primarily interested in presenting a philosophical position. His structuralism was both implicit and very early in its development – not like the modern kinds of structuralism defended by Resnik and Shapiro. However, it will prove very useful to articulate on Hilbert's behalf what kind of structuralism he may have been moving towards.

Hilbert's structuralism is also noticed by Shapiro and Seig. Both call attention to the following passage:

> But it is surely obvious that every theory is only a scaffolding or schema of concepts together with their necessary relations to one another, and that the basic elements can be thought of in any way one likes. If in speaking of my points I think of some system of things, e.g. the system: love, law, chimney-sweep... and then assume all my axioms as relations between these things, then my propositions, e.g. Pythagoras' theorem, are also valid for these things (Hilbert, 1899/1980*d*, 40-41).

Here Hilbert seems to quite straightforwardly endorse an eliminativist structuralism (i.e. one which denies the existence either of the positions in the structure, or the structures themselves, or both). To say that different sets of objects can function as points is plainly to deny that points are objects in Frege's sense and to consider it misconceived to make any attempt to define or reconstruct geometrical objects further than they have been by the axioms. As we shall see in more detail later on, the most central feature of Frege's conception of an object is that singular terms "...are to be understood as standing for independent objects" (Frege 1884, §62). That is to say that objects are the referents of singular terms and these referents can always admit of different modes of presentation beyond those which were used to refer to it. To make the comparison with Hilbert more clear, let us employ the following characterisation of Frege's conception:

df. A *Fregean object* is an object which – with respect to some axioms – can have more properties than those that can be derived as logical consequences of its satisfaction of those axioms.

Frege finds Hilbert's above claim about love, law and chimney-sweeps extremely problematic. After all, for Frege, to apply a theory to infinitely many systems of objects is merely to collect infinitely many entirely distinct theories. One is about geometry, one is about the real numbers, one is about the most eclectic examples Hilbert could bring to mind, and so on. According to Hilbert, however, it is by looking for properties further to those that can be derived as logical consequences of an objects satisfaction of some axioms, that we enter into a fruitless game of hide-and-seek. Therefore, we can quite safely characterise Hilbert's position as eliminativist *with respect to Fregean objects*.

This is not the end of the structuralist story. Recall that earlier when talking about *abstracta* I suggested that Hilbert had a different conception of the proper subject-matter of a mathematical theory from Frege. For the sake of clarity, I will reserve the word 'object' for Frege conception and use 'basic element' or *abstracta* to talk of Hilbert's alternative. This being so, what are we to make of Hilbert's remark that the basic elements of a theory can be thought of in any manner so long as the axioms are satisfied? Or his remark that theories can be applied to infinitely many systems of the basic elements? In order to fully bring out Hilbert's view we will need to further articulate Hilbert's conception of *abstracta*.

The contrast between objects and basic elements is most vividly captured by Hilbert himself (as Hallett drew our attention to) where he remarks that the objects of his geometry have nothing to do with "real" points and lines and planes. In structuralist terms, we can understand Hilbert as saying that *an object of geometry is simply the reification of a role in a theory of geometry*. To contrast this with Frege's conception we can formulate the following characterisation:

df. A *Hilbertian basic element* is an object which – with respect to some axioms – has all and only those properties that can be derived as logical consequences of its satisfaction of those axioms.

To pinpoint the difference between the two definitions now in play, consider a Fregean object which happens to be exhaustively characterised by some axiomatisation. Despite having all and only those properties in virtue of which they satisfy the axioms this Fregean object is not a Hilbertian basic element. The fact that the theory characterises the object exhaustively is a *coincidence* but a basic element is exhausted by its role in the theory *by definition*. To check that the *object's* properties are exhausted by the axioms we would have to first check the properties which can be deduced from the axiomatisation; then, externally characterise the properties of the object; and finally compare the two to find any discrepancies. However, in the case of a *basic element* only the first step is required. In this way, the properties of an object are independent of the theory but the properties of a basic elements are entirely dependent on it.

The idea of a basic element, as we have characterised it, creates a split between the idea of an entity as part of an ontological inventory, and the use of an entity in a theory. This links back to the different theoretical goals we attributed to Frege and Hilbert. Frege wants an axiomatisation to relate appropriately to a universal domain of objects. Hilbert conceives of an axiomatisations as a "scaffold" of consistent concepts. So on Hilbert's view, his axioms define a higher-order concept which dictates how a set of lower-order concepts must be related to each other if they are to satisfy the axioms. By constraining the first-order concepts the higher-order concept also indirectly dictates how a system of *things* must be related to each other if they are to be used as the basic elements of the theory. In this way, the axioms of a theory *implicitly define* the basic elements.

I think it is safe to say that while Hilbert was an eliminativist about Fregean objects, he was not obviously an eliminativist about Hilbertian basic elements. Actually, I think that Hilbert *can* be understood as being non-eliminativist about these basic elements. To see this, we will consider the characteristics of a non-eliminativist object given by Linnebo. Linnebo (2008) identifies two claims which distinguish the non-eliminativist from a Platonist. He calls these:

i. **The Incompleteness Claim:** Mathematical objects are incomplete in the sense that they have no "internal nature" and no non-structural properties.

ii. **The Dependence Claim:** Mathematical objects from one structure are dependent on each other and on the structure to which they belong.[7]

Linnebo gives an example of Resnik endorsing the incompleteness claim:

> In mathematics, I claim, we do not have objects with an 'internal' composition arranged in structures, we have only structures. The objects of mathematics . . . are structureless points or positions in structures. As positions in structures, they have no identity or features outside a structure (Resnik 1981, 530).

And an example of Shapiro endorsing the dependence claim:

> The number 2 is no more and no less than the second position in the natural number structure; and 6 is the sixth position. Neither of them has any independence from the structure in which they are positions, and as positions in this structure, neither number is independent of the other (Shapiro 1997, 258).[8]

[7]Both of these are taken directly from Linnebo (2008, 3).

[8]For more on Shapiro's particular species of structuralism and its defence, see Shapiro (1997), Shapiro (1989), Shapiro (2000), Shapiro (2004).

With respect to the quotation we used to support Hilbert's eliminativism: Hilbert speaks of a system of *things*, i.e. he suggests the objects falling under the first-level concepts love, law and chimney sweeps might work together as a package to make a simple geometry. This is in line with the co-dependence of objects outlined in the dependence claim. Further, since a basic element is just Hilbert's way of speaking about the object positions in the scaffold of concepts, it is clear that there can be no sense in which the basic elements can exist independently of the scaffold. This point is also supported by the following passage:

> I do not want to assume anything as known in advance; I regard my explanations in sect. 1 as the definition of the concepts point, line, plane – if one adds again all the axioms of group I to V as characteristic marks. If one is looking for other definitions of a 'point', e.g.,through paraphrase in terms of extensionless, etc., then I must indeed oppose such attempts in the most decisive way; one is looking for something one can never find because there is nothing there; and everything gets lost and becomes vague and tangled and degenerates into a game of hide-and-seek (Hilbert, 1899/1980*d*, 39).

If there is nothing more to the geometric primitives than what is defined by the axioms this suggests that they are dependent on the structure defined by the axioms in the sense that Hilbert tells us that, beyond this structure, there is nothing there. The other important point being made is that the attempt to look for properties to characterise basic elements outside of the properties given by the axiomatisation is deeply misguided. The reason that it is misguided is that there are no further properties to be found: we merely enter into a game of seeking when nothing is hidden and the game cannot end. This satisfies the incompleteness claim since all the properties which the basic elements have will be those characteristic marks given by the structure defined by the axioms.

Thus, Hilbert's early structuralism can thus be thought of as *eliminativist* in that it denies the existence of Fregean objects. However, it can also be thought of as *non-eliminativist* in the sense that basic elements can be thought of as *non-eliminativist structuralist objects*.

We have said that one way to think of Hilbertian basic elements is as *abstracta*. It is worth pointing out that it is now obvious that this is entirely opposed to Frege's conception of *abstracta*. For Frege an object is the referent of a singular term and on this view *abstracta* are just as much objects as *concreta*. Therefore, it is not possible for Frege to say that an *abstracta* is merely a position in a structure or something falling in the range of a concept. Frege's universal domain must contain both the abstract objects and the concrete objects standing equally along side each other.

Paul Bernays – Hilbert's long term assistant and collaborator – describes how the existence of the *abstracta*, or basic elements, should be understood as nothing over and above the existence of the structure in which they feature:

> In mathematics we do not have such a precise marked difference in modality. For the mathematician's mode of reflection, the individual mathematical entity does not present itself as something that exists in a more eminent sense than the lawful lawlike relations. Indeed, one might say that there is no clear difference at all between a direct entity and a system of laws to which it is subject, since a number of laws present themselves by means of formal developments which, on their part, possess the character of the direct entity. Even systems of axioms may be considered as structural entity. In mathematics, therefore, we have no reason to assume existence in a sense fundamentally different from that in which we assume the existence of lawful lawlike relations (Bernays 1950/2002, 15).

What Bernays describes here is the fundamental structuralist tenet of the ontological primacy of the structure over the object or entity. This brings us to another question: how are we to ensure the existence of the structure? For this, we return to another kind of primacy: the primacy of the consistency of the axioms over the existence of what the axioms define. Altogether this gives us a picture of Hilbert's investigation into mathematical ontology which Bernays summarises in the following way:

> The viewpoint gained in this way places a mathematical reality face to face with a methodological framework constructed for the fixation of this reality. This is also quite compatible with the results of the descriptive analysis to which Rolin Wavre has subjected the relationship of invention and discovery in mathematical research. What is pointed out here is the intertwining of two factors: He points out that two elements are interwoven, *on the one hand the invention of concept formations, on the other hand the discovery of lawful lawlike relations between the conceived entities, and furthermore the circumstance that the conceptual invention is directed toward aimed at discovery* (Bernays 1950/2002, 19, emphasis mine).

This section has mostly been suggestive of Hilbert's implicit ontological position. I do not presume to have argued that Hilbert was a non-eliminativist about basic elements, or that he was an eliminativist about Fregean objects, or that he was a realist about mathematical concepts, for that matter. I think one can make a good case that Hilbert was an eliminativist about both Fregean objects and basic elements. However, at the end of chapter 5 we will find ourselves in a position where it is very useful to have spelled out the half-and-half eliminativist and non-eliminativist view. To keep this view in mind, a good illustrative comparison is given by how we might explain the relationship between the Higgs boson and the Higgs field: once the Higgs field is there, so are its excitations. These waves are, at the very same time, the particles we call the Higgs boson. In the same way, once the Euclidean structure is there, so are the basic elements. The basic elements are (to continue the metaphor) merely local manifestations of the structure.

3.4.2 Existence and uniqueness worry

The important point which we should take from the suggestive considerations of the last section is that if we understand Hilbert as maintaining an early kind of structuralism (regardless of what kind of structuralism) then we explain why Hilbert was not upset by Frege's twofold worry regarding uniqueness and existence. Here we will set out Hilbert's understanding of the two worries in some more detail.

Simply put, the uniqueness requirement is satisfied by relinquishing the incorrect assumption made by Frege that Hilbert is offering a reconstruction of geometric objects as Fregean objects. Rather, since basic elements have all and only those properties in virtue of which they satisfy the axioms every collection of objects satisfying the axioms merely provides a different 'way to think of' the basic elements of the theory.

Let us unpack this point more slowly. The first thing which is natural to ask is how Hilbert can hope to pick out the geometric primitives uniquely if all he has *defined* by means of his axioms are concepts. In addressing this we should note straight away that the concepts which are defined by Hilbert's axioms are not *irrelevant* to the geometric primitives. The higher-order concept is satisfied by sets of concepts (models) and under these concepts fall the relevant primitives. Thus Hilbert's *definiendum* is related to the geometric primitives, albeit indirectly.

This observation only goes so far. For, since the higher-order concept is satisfied by more than one model it seems to follow that there is more than one set of primitives capable of being called *the* geometric primitives on Hilbert's view. Before we even address the problem of how we can establish that the primitives *exist*, how is Hilbert to adjudicate between these sets of primitives in order to identify the target set of *geometric* primitives?

The key point to remind ourselves of here is that whereas Frege aims to establish the existence of primitives falling under the first-level concepts number, point, plane *as Fregean objects*; Hilbert is content to establish that they exist *as Hilbertian basic elements*. Quite simply then, Hilbert avoids Frege's uniqueness worry because characterising a basic element does not require characterising a determinate object, but merely a set of properties. In characterising these properties, Hilbert has provided the roles that a set of *abstracta* would have to fulfil in order to manifest the basic elements of a theory. In characterising these roles, Hilbert has specified a unique set of basic elements. All *abstracta* suitable for inhabiting the roles will be counted as manifesting the very same basic elements. In this way, Hilbert can characterise what it is to be the basic element *point* equally well using an abstraction of a real number pair or an abstraction of Frege's pocket watch; provided these meet the requirements demanded by the axioms.

It is now clear how the uniqueness requirement which Frege raises issues from his conception of a mathematical object as independent of an axiomatisation. On his view, we would require a check that the objects falling under the concepts defined by the axiom are *the* geometric objects. For Hilbert, however, so long as the higher-order concept is well-defined then all systems which satisfy the axioms will be equivalent with respect to being a Euclidean geometry (i.e. satisfying the higher-order concept defined by the axioms). The point doing all the work here is that a set of axioms can fail to uniquely determine some set of objects, but it cannot fail to uniquely determine a set of basic elements because basic elements – by definition – have their nature exhausted by the axioms. In this way, any axiomatisation uniquely determines a set of basic elements.

To avoid the ontological assumption which the uniqueness worry issues from, it is essential that we understand Hilbert to maintain (i) the incompleteness claim. Once we recognise this we can see that a set of axioms will always pick out a set of basic elements uniquely.

In a similar way, the existence requirement is avoided if we understand Hilbert to maintain (ii) the dependence claim. Since in this case, the question of the existence of the basic elements is just a question of the existence of the structure of concepts which they fall under; and, for Hilbert, this structure is secured by the consistency of the axioms which define it. In later chapters we will return to this fact as very important: that it is the structuralist aspect of Hilbert's conception in virtue of which he can hope to avoid Frege's most pertinent objections.

Concluding remarks

In this chapter, we have seen that the original priority reading only superficially appeared to be able to explain the context of Hilbert's Principle; the Frege-Hilbert controversy. Developing the priority reading in light of Blanchette's work led us to a more sophisticated understanding of the fundamental issue between Frege and Hilbert as a disagreement over whether the preservation of reference to the primitives of an axiomatisation takes priority over the preservation of consistency, or vice versa. We then saw that this was connected to their background ontological conception: Hilbert did not worry about ensuring consistency by divorcing his axiomatisation from what Frege took to be the proper subject matter of geometry. This is because Hilbert insisted that the subject-matter of his theory was not space, but *idealised* space. We emerged from these reflections with an improved final priority reading which was able to explain the contention behind Hilbert's Principle and was integrated with a fuller and satisfying explanation of the deeper disagreement of the Frege-Hilbert controversy.

In the last three chapters, we have explored the interplay between early Hilbert's methods and what he takes to exist. What has emerged from this is that Hilbert's *conception* of his axioms (as implicit definitions and thus as having re-interpretable parts) issues from his concern with avoiding paradox above all else and this influences his background ontological conception. We saw that Frege has a very different conception of an axiom and, as a result, a very different ontology.

In the next three chapters we will explore another foundational project: neo-Fregean logicism. I believe that making a connection between these two debates is extremely important – especially if we are to unpick the theoretical project which Hale and Wright are pursuing. I will show that neo-Fregean logicism has one central point in common with the Frege-Hilbert controversy: that Hale and Wright's conception of abstraction principles (an alternative to an axiom) is also the key to their ontology. Furthermore, I will explain why the conception of abstraction Hale and Wright use is closer to Hilbert's conception of an axiom than it is to Frege's conception of an axiom or of a basic law. I will argue that the ontology of the neo-Fregean logicist must be understood in light of this point.

Chapter 4

An objection to Neo-Fregean Logicism

Introduction

We have dedicated much time to understanding Hilbert's early ontology of mathematics. In this chapter I will connect Hilbert's foundational project to the foundational project of neo-Fregean logicism, as defended by Bob Hale and Crispin Wright. To do this we will use the simple observation that both projects attempt to axiomatise an area of mathematics with a foundation which employs implicit definitions.[1]

The overarching aim of the remaining chapters will be to show that this common use of implicit definitions produces commonalities in the mathematical ontology, and to argue that Hale and Wright have abandoned the distinctively Fregean element of their

[1] To my knowledge, only Gary Kemp has also seen that the Frege-Hilbert controversy can be used to understand Frege's own position on issues connected to neo-Fregean logicism. The relevant issue for him, however, is the Julius Caesar problem. He draws a very different conclusion from the Frege-Hilbert controversy from the one we will draw here. Kemp argues that the controversy shows that the Julius Caesar problem is a well-motivated problem for Frege which issues from his sharpness requirement for concepts (Kemp 2005).

project in order to make their logicist project viable. In doing so, they have implicitly adopted an ontological conception of mathematics which is in line with Hilbert rather than Frege.

Here, we will introduce Hale and Wright's neo-Fregean logicism and then to carefully transpose the details of Frege's objections to Hilbert onto Hale and Wright's account. In particular, we will consider how the diagnosis of Hilbert's axioms can be used to give a diagnosis of Hume's Principle. First, we will rehearse the details of Frege's objection to Hilbert §4.1 and then in §4.2 we will introduce the basic aspects of Hale and Wright's neo-Fregean logicism. In §4.3.2, we will consider two objections already brought against Hale and Wright before presenting a new objection in §4.3.3 based on the work of the previous three chapters. Finally, in §4.3.4 I will refute an obvious counter-objection.

4.1 Rehearsing Frege's diagnosis

In this first section, we will remind ourselves of the relevant features of Frege's diagnosis so that later we can apply Frege's critique of Hilbert's implicit definitions directly onto Hale and Wright's account.

Recall that Frege believes that a lack of terminological clarity has left Hilbert in a muddle. According to Frege, Hilbert has blurred together three very different kinds of mathematical propositions: *definitions*, which determinately characterise the meaning and reference of a term; *axioms*, which are the foundational truths of a theory expressed by determinate sentences; and *explanations*, which indeterminately coordinate the meaning and reference of some unidentified (perhaps indefinable) term prior to the construction of the theory. In failing to keep these categories distinct Hilbert has wrongly taken there to be sentences which play the role of the foundational sentences from which the theory is constructed (like axioms), and which contain expressions which are not yet meaningful (like definitions and explanations). Further, Hilbert's axioms do not assign a precise meaning to the primitives (as with explanations) and are part of the theory and not prior to it (as with definitions). Thus, Hilbert's terminological imprecision has caused him to blur together features unique to each kind of sentence into one impossible kind of sentence which can at the same time be used to express an axiom, provide an explanation of the primitives and define the meaning of the primitive expressions.

In order to untangle this muddle and keep these three roles distinct, Frege sets out an objection to Hilbert. If his primitive expressions are meaningful, then they must employ some prior meaning – such as the meaning given to them by Euclid. But in this case they are not definitions or explanations because then they would be circular, presupposing the meaning they purported to define or explain. But if the primitive expressions are *not* meaningful, then there are parts of the sentence which do not yet have a sense. And if this is the case, the sentence cannot express an axiom, since the sense of all the parts of a sentence must work together in order to form a thought. In this way, Frege attempts to force a split between the role of axioms and the roles of definitions and explanations. I formulated Frege's point here as a dilemma:

> **Frege's Dilemma**: Either the primitive expressions in the axioms are laid down as meaningful (in which case the axioms are not definitions), or they are not (in which case they are not axioms).

Frege is right about the first horn of the dilemma. If Hilbert meant his primitives to be understood as Euclid explains them, then he cannot properly call his axioms definitions since they do not set out the geometric concepts but import them from somewhere else. However, this is not how Hilbert intends his axioms to be understood. Hilbert vehemently opposes the suggestion that his axioms rely on prior definitions or explanations of the kind which Euclid provides.

> I regard my explanations in sect. 1 as the definition of the concepts point, line, plane – if one adds again all the axioms of group I to V as characteristic marks. If one is looking for other definitions of a 'point', e.g., through paraphrase in terms of extensionless, etc., then I must indeed oppose such at-

> tempts in the most decisive way; one is looking for something one can never find because there is nothing there (Hilbert, 1899/1980*d*, 39).

Hilbert does not intend the primitive expressions to be meaningful prior to the axiomatisation but to be explained and defined by the axioms themselves. He is entirely happy to fall onto the second horn of Frege's dilemma. As we have explored in depth, Hilbert has a different conception of axioms according to which axioms are expressed by sentences containing semantic gaps where the primitive expressions feature. Because of this, Hilbert faces no dilemma at all: he can admit that his primitives are not meaningful without conceding that his sentences do not express axioms. What Frege takes to be an objection to Hilbert is merely an articulation of Hilbert's position.

Frege goes on to draw out the consequences of landing on the second horn of his dilemma, invoking his hierarchy.

> The characteristic marks you give in your axioms are apparently all higher than first-level; i.e., they do not answer to the question "What properties must an object have in order to be a point (a line, plane, etc.)?", but they contain, e.g., second-level relations, e.g., between the concept point and the concept line. It seems to me that you really want to define second-level concepts but do not clearly distinguish them from first level ones (Frege, 1900/1980*b*, 46).

Frege points out the problem with landing on the second horn of the dilemma. If you maintain that the axioms implicitly define the primitives, then you must conceive of the primitive expressions as functioning in the sentence as *variables*. So that, what Hilbert's axioms properly define is the second-order concept given by the conjunction of Hilbert's 20 axioms with the six primitives systematically replaced with variables. Intuitively, this defines a six-place higher-order concept, not the first-order concepts of *point* or *line*, etc. Using this diagnosis, we improved our dilemma by making it specific to cases of implicit definition:

> **Frege's Dilemma for implicit definition**: Either your *definiendum* has a prior meaning and so your definition is circular, or you have an *explicit* definition of a concept one level higher than your *definiendum*.

Frege's point that the concepts defined by implicit definition are one level higher than their *definiendum* is also a sophisticated development of the way in which Hilbert's axioms work, rather than an objection.

4.2 The basic tenet of neo-Fregean logicism

Let us now move on to expounding the doctrine of neo-Fregean logicism, with special attention to its use of implicit definitions. We will then be in a position to apply Frege's general observation to Hale and Wright. Note that I will speak interchangeably of neo-Fregean logicism and Hale and Wright's account.

The neo-Fregean logicist inherits Frege's logicist project to establish the thesis that the truths of arithmetic are truths of logic and thus that the subject matter of these truths are logical objects (see Hale & Wright 2001). Frege himself considers – and ultimately rejects – using the principle that the number of Fs are equal to the number of Gs when the Fs and Gs are in one-to-one correspondence, to secure the numbers as logical objects – settling instead on defining numbers in terms of extensions (Frege 1884, §63-64). Frege claims to have taken this principle from Hume; as such, Boolos subsequently named it "Hume's Principle". We can formalise it as,

$$\forall F \forall G(Nx : Fx = Nx : Gx \leftrightarrow \exists R(\forall x(Fx \rightarrow \exists! y(Gy \wedge Rxy)) \wedge \forall y(Gy \rightarrow \exists! x(Fx \wedge Rxy))).$$

We will henceforth use the following shorthand formulation:

Hume's Principle $\quad \forall F \forall G(Nx : Fx = Nx : Gx \leftrightarrow F \approx G)$,

in which '$\approx$' denotes a 1-to-1 correspondence between the sortal concepts F and G and '$Nx : Fx$' is read as, 'the (cardinal) number of Fs'.

The neo-Fregean logicists earn their prefix by their central claim that Hume's Principle can be used as the means by which the logicist thesis can be established. This claim is inspired by the fact that the Dedekind-Peano axioms for second-order arithmetic can be derived from Hume's Principle in full second-order logic. This result is now known as Frege's Theorem. Although Geach (1955) and Parsons (1965) both noticed that the Dedekind-Peano axioms could be derived from Hume's Principle, the conjecture did not gain much attention until Wright carried out most of the derivation in his "Frege's Conception of Numbers as Objects" 1983. Later, Richard Heck forcefully argued that:

> Careful examination of the proofs of the axioms of arithmetic in the *Grundgesetze* shows that all uses of value-ranges within those proofs are of one of the following three types:
>
> 1. The ineliminable use in the proof of Hume's Principle.
> 2. The use which allows the representation of second-level functions by first-level functions.
> 3. The formation of complex predicates to emphasize what is being proven.
>
> Except for those of the first sort, all uses of value-ranges are therefore easily, and uniformly, eliminated from Frege's proofs (Heck 1993, 583-584).

Heck concludes that the only "essential" use Frege makes of his ill-fated Basic Law (V) in *Grundgesetze* is to derive Hume's Principle (see Heck (1993, 581-584) and Heck (2011)).

The natural question to ask in light of Wright and Heck's work is whether Frege's logicist ambitions could be safeguarded by a seemingly minor amendment to his formalisation. The neo-Fregean logicists attempt to answer the question in the affirmative and to articulate what kind of logicism such an amendment would deliver. For the most part, the accusations put against the neo-Fregean logicists do not call Frege's Theorem into question, but rather question whether Hume's Principle is a suitable foundational principle.

It must be understood that the neo-Fregean logicists do not merely propose that Hume's Principle be laid down *as an axiom* from which we can derive the Dedekind-Peano axioms. It is the distinctive and controversial claim of Hale and Wright that so-called "abstraction principles" like Hume's Principle or the Direction Equivalence are superior to a set of axioms because they have the resources to introduce us to certain abstract objects without requiring an antecedent grasp of terms referring to such objects; as is required by the method of brute axiomatic stipulation (Hale & Wright 2001, 105-116, 189, 307-320). More precisely, abstraction principles out-class axiom sets in two regards: they are epistemically privileged and they are ontologically leaner. Laying down an axiom as true does not explain how we came to grasp the objects that it refers to; it does not provide an answer to Frege's question of how such objects are to be given to us, and so owes a kind of epistemic debt. Further, it does not provide a criterion which prevents us from positing whatever objects we like, by merely laying down an axiom which refers to them. Direct stipulation of the Dedekind-Peano axioms, for instance, entails the existence of infinitely many objects. It is in this privileged sense (which has the hope of vindicating a logicist thesis) that Hume's Principle is claimed by the neo-Fregean logicists to be a foundational principle.

We might put the point as follows: the guiding pursuit of the neo-Fregean logicists as inherited from Frege, is not merely to secure the numbers by any means, but to provide a logical route to them. Of course, if this route needs to appeal to Hume's Principle, then the status of Hume's Principle itself is important and whether or not a species of logicism can be vindicated hangs on this status. As Heck points out, however, we cannot straightforwardly assume Hume's Principle to be a logical law, lest we beg the very question of our enquiry.

> To suggest that we regard it as a fundamental law that we are justified in recognizing something common to two equinumerous concepts, and that accordingly logic allows us to transform a statement of equinumerosity into an identity of numbers would be blatantly to beg the question whether arithmetic is a branch of logic (Heck 2005, 177).66

Hale and Wright's goal is to secure both sides of Hume's Principle as involving purely logical notions, thus qualifying the principle as a whole to count as a logical law. Because of this, Hume's Principle cannot merely be laid down as a logical law (as Frege does with Basic Law (V)) on pain of abandoning the project of a logical reconstruction of the numbers.

Hume's Principle merits the privilege afforded to it because it is a definition. A definition, that is, of the number-operator "Nx...". This conception of Hume's Principle is not of some impassive privileged truth waiting to be grasped; instead, it is an active process by which "Nx..." is given a meaning by stipulation of the principle as true. This forcibly carves out the very reconstructive route to the numbers which neo-Fregean logicists have in their sights (Hale & Wright 2001, 117-150).

Let us be clear on what the neo-Fregean logicists consider to be the *definiendum* of Hume's Principle. Hume's Principle is taken to define the functional expression "Nx..." which takes first-order predicates to singular terms. It also defines the function referred to by "Nx..."; a function which takes concepts to objects (the numbers). We can pause here to ask what justifies this slide from defining a functional expression to securing the function it expresses.

For such a justification, the neo-Fregean logicists appeal to internal Fregean principles to re-orientate how the reference of the function is to be understood. In particular, they identify *the syntactic priority thesis*. This doctrine is attributed to Frege by both Wright (1983) and Dummett (1981) and holds that if an expression syntactically behaves like a singular term (in some range of true sentences) then it must refer. Of course, for Frege, if a singular term refers then it refers to an object. The thesis is one of priority in the sense that the syntactic features of a word are prior to the object which is its referent in determining the category of a singular term. This priority is taken by Wright and Dummett to issue from Frege's context principle; that only in the context of a proposition do words have sense (Frege 1884, §62). Winning "Nx..." as a singular term, then, is to win it as a term that refers to an object, thus achieving reference to numbers (as objects) from Hume's Principle.

The neo-Fregean logicists do not paint a more vivid picture of how they want to conceive of the mechanism of implicit definition than in the following passage.

> To invent a meaning, so conceived, is to fashion a concept: it is to be compared to making a mould and then fixing a certain shape-concept by stipulating that its instances comprise just those objects which fit the mould (or are of the same shape as something which does). There is a sense in which the shape – the bare possibility of matter so configured – existed all along. We did not create *that* possibility. But we did create a concept of that shape (whether or not we also fixed the meaning of a word to be associated with it). It would make no sense for someone who followed the performance to doubt that there is any such shape – we displayed the shape in fixing the concept of it. In rough analogy, we must so conceive implicit definition that – in the best case – it makes no sense to doubt that there is a meaning taken on by the defined expression, not because the meaning in question allows of independent specification but because it has somehow been *fully explained* in the very process that creates it (Hale & Wright 2001, 131).

Thus if Hume's Principle is conceived of as an implicit definition, the meaning of the N-operator is at the same time *created* and *displayed* by it. Since establishing the meaning of the functional expression "Nx..." is all that there is to settling its reference, and since the meaning of the functional expression is fully explained in the process that creates it (the process of implicit definition) then Hume's Principle, in this sense, *creates* reference to the cardinality concept $Nx\ldots$: the function which takes first-order concepts to the numbers. This is not to say that the implicit definition creates the cardinality *concept*. Wright and Hale make clear in the analogy (where they say that the possibility of the mould is not created) that the view is compatible with Frege's Platonism about concepts: what is made by the definition is a way of referring; a meaning; a mode of presentation of something that existed already.

Securing the meaningfulness of the N-operator in this way ensures that singular terms such as "$Nx : Fx$" are well-defined in that they are meaningful and referential.

By the syntactic priority thesis, if a singular term refers then it refers to an object and the particular objects referred to by such terms as "$Nx : Fx$" are supposed by Hale and Wright to be the numbers. The reference of such singular terms constitutes a "route" to the numbers in the sense that the numbers are presented as *the objects falling in the value-range of the N-operator*. In other words, in defining the N-operator we grasp what *sort of things* the operator maps to. Since the operator maps to numbers then this provides a way of understanding the numbers by grasping them as the elements of a given value-range. Thus, on the neo-Fregean logicist view, using implicit definitions to secure the meaning and reference of the N-operator (and thus the relevant singular terms) is tantamount to providing a logical route to numbers which relies only on an appeal to (full second-order) logic and definitions.

4.3 Frege's dilemma for the neo-Fregean logicists

We have seen that the goal of the neo-Fregean logicist project is to deliver the numbers as Fregean objects in such a way that establishes their logicality. Numbers are delivered via an abstraction principle which is superior to mere axiomatic stipulation. In general, on a Fregean framework, the route to an object is via a singular term, since for Frege what it is to be an object is just to be the referent of a well-defined singular term. Importantly, what is required of the neo-Fregean logicist is to provide an expression which does not merely have the shape of a singular term but which genuinely *functions* as a singular term which is apt to refer to numbers in the contexts in which it is identified.

As we have seen, Hale and Wright claim that such expressions can be harvested from Hume's Principle. In this section, I will argue that the candidate singular terms in Hume's Principle may have the appearance of singular terms but do not semantically contribute as singular terms. I will establish their true syntactic contribution to Hume's Principle by employing Frege's dilemma for implicit definition. Before this, however, we must isolate the two ingredients of Hale and Wright's account which are most relevant to the objection. Once this has been set out, we will look at two objections concerning singular terms which have already been made. At this point we will be in a good position to apply Frege's dilemma to Hale and Wright.

4.3.1 Two commitments of neo-Fregean logicism

Before applying Frege's dilemma this section will look in a little more detail at two of Hale and Wright's most basic commitments. First, that they require the presence of singular terms in Hume's Principle. Second, that they forward Hume's Principle as an implicit definition of number.

Here Wright and Hale describe how someone might come to grasp numbers with the use of Hume's Principle.

> What a recipient of [Hume's Principle] immediately learns is that whatever suffices for the truth of a statement of concept-equinumerosity is equally sufficient for the truth of the corresponding statement of number identity. However, she also understands that she is to take the surface syntax of number-identity at face value. She already possesses the general concept of identity and so is able to recognise that the expressions flanking the identity sign must be singular terms. Further, she already understands predicate variables, and so can recognise that 'Nx...x' must be being introduced as a function expression denoting a function from concepts to objects. From this she is able to gather that the objects in question simply are objects for whose identity it is necessary and sufficient just that the relevant concepts be equinumerous (Hale & Wright 2001, 117-8).

This quote makes clear the importance of taking the left-hand side of Hume's Principle to have the syntax which it appears to have. Here, and elsewhere in their writings, Hale and Wright call this the appeal to *surface syntax* (see Hale & Wright 2008, 2009). This appeal secures for Hale and Wright the recognition of a sign for the identity relation (given that the identity relation has already been grasped) and – as a result – the presence of singular terms on either side of the identity. Furthermore, it secures the recognition of a function expression which is apt to refer to the number function.

Hale and Wright's account of the process of abstraction seems to be in keeping with Frege's own. In Frege's discussion of Hume's Principle, he states that we begin with the general notion of identity and from it form the notion of *numerical identity*, so that a grasp of the identity relation is given prior to a grasp of Hume's Principle (Frege 1884, §64). For Hale and Wright this means we can understand identity as part of the *definiens* of Hume's Principle. When we recognise an expression for the two-place identity relation, then we recognise it in virtue of its syntax being suitable to refer to a relation we are already familiar with.

Furthermore – for Frege – canonical identity relations are flanked by singular terms. Thus, the recognition of an expression for the identity relation gives us good syntactic evidence that there are expressions featuring in Hume's Principe which are not merely the shape of singular terms ('the number of Fs', 'the number of Gs') but which are *functioning as* singular terms. Thus, Hale and Wright's appeal to surface syntax is well-motivated within a Fregean framework.

The second important feature of Hale and Wright's account is that they categorise Hume's Principle as an implicit definition. Hume's Principle cannot be considered an axiom since is put forward as a preferable *alternative* to an axiom. The only other categorisation which Hale and Wright have considered for Hume's Principle is that it is an analytic truth. Wright makes the point that however we articulate the notion of analyticity it will most likely be suitable to transmit across logical consequence. So that, if second-order consequence is a species of logical consequence then the analyticity of Hume's Principle will ensure the analyticity of a second-order consequence of Hume's Principle, namely, second-order arithmetic (Wright 2001, 307).

However, Hale and Wright eventually concede that although Hume's Principle might turn out to embody a species of analyticity which would secure a logicist result, the jury is out until philosophy delivers a sharper conception of "the status and provenance" of analytic truths (Wright 2001, 308). Wright addresses Boolos's worries about analyticity by noticing that the real question at issue concerns what the nature of our entitlement is to Hume's Principle and to whether it is true. To this he admits:

> A worked-out account of the notion of analyticity, in all its varieties, might well provide an answer to the question. But the answer the Neo-Fregean wants to give is not hostage to the provision of such an account. Let me rapidly recapitulate that answer. The Neo-Fregean thesis about arithmetic is that a knowledge of its fundamental laws (essentially, the Dedekind-Peano axioms) – and hence of the existence of a range of objects which satisfy them – may be based *a priori* on Hume's Principle as an explanation of the concept of cardinal number in general, and finite cardinal numbers in particular (Wright 2001, 321).

Here Wright demotes the analyticity of Hume's Principle as a secondary concern to whether Hume's Principle can give us *a priori* knowledge. Hale and Wright then maintain what they call *the traditional connection*, which is that *a priori* knowledge can be provided by *implicit definition*, and they argue that Hume's Principle is such an implicit definition.

Very broadly, their idea is that there is a type of implicit definition which creates a situation where the meaning of a term is so closely connected to the truth of the sentence in which it is embedded that it gives *a priori* knowledge. They conceive of implicit definition as broadly working in the following way: Take the partially interpreted sentence "$\#f$" where the expression "f" is the *definiendum* and the matrix has a determinate meaning and so forms the *definiens*. To define "f" by implicit definition we stipulate "$\#f$" to be true and thereby bestow on "f" the meaning that the term would have to have in order for it to systematically contribute along with "$\#$" to make "$\#f$" true. This is not to assume that "$\#f$" is true, but instead to introduce the meaning of "f" in a way which is sufficiently immediate and essentially connected to the truth of the sentence in which it appears to uphold the traditional connection. Hale and Wright outline the sufficient conditions for the connection to *a priori* knowledge as follows:

> To know both that a meaning is indeed determined by an implicit definition, and what meaning it is, ought to suffice for a priori knowledge of the proposition thereby expressed (Hale & Wright 2001, 126-127).

Thus the hope of Hale and Wright is that the process of implicit definition can give rise to a *species* of *a prioricity* – a kind of meaning-in-virtue-of-truth which, being won for Hume's Principle, is won for arithmetic also.

Therefore, it is central to Hale and Wright's account that abstraction principles are implicit definitions of the *abstracta* they introduce and that they are able to secure singular terms in the process of abstraction by appealing to the surface syntax of Hume's Principle.

4.3.2 Two criticisms

We have just seen that the purpose of the appeal to surface syntax is to provide evidence that there are expressions which don't just have the appearance of a singular term but are genuinely functioning as singular terms in abstraction principles. Let us call the genuine semantic contribution of the different parts of a proposition the *logical syntax* so that we may contrast it with what Hale and Wright are calling the *surface syntax*. Hale and Wright's claim, then, is that while the surface syntax does not always match the logical syntax, nevertheless the surface syntax gives us good evidence of the logical syntax of Hume's Principle. In particular, since the surface syntax of Hume's Principle involves singular terms, this gives us good evidence that the logical syntax of Hume's Principle involves singular terms. And establishing that the logical syntax involves singular terms is tantamount to establishing reference to logical objects.

In the next section, I will motivate an objection to this appeal to surface syntax purely from the basis of Hale and Wright's understanding of Hume's Principle as an implicit definition. I will argue that the surface syntax does not secure the logical syntax of abstraction principles precisely because abstraction principles are implicit definitions.

This objection has already been put to Hale and Wright, first by Dummett (1998) and most recently by Trueman (2014).

The purpose of this section will be to give a sense of Dummett and Trueman's objections and then discuss how Hale and Wright attempt to avoid them. Their counter-objection will give us greater insight into how Hale and Wright suppose the appeal to surface syntax to work.

Dummett has accused Hale and Wright of fixing the meaning of the identity sentence on the left-hand side of Hume's Principle in such a way as to render the parts of the sentence 'semantically idle'. He defines this notion in the following way:

> If the determination of the truth-value of [any sentence containing a term of that range] goes through the identification of the referent of the term, the notion of reference, as applied to it, is semantically operative; if not, that notion, even though legitimate is semantically idle (Dummett 1998, 385).

Dummett considers the left-hand side of abstraction principles to be semantically idle because its truth condition can be established by the right-hand side without any appeal to the reference of the terms. Thus, the meaning of the terms is fixed as a whole and not with respect to identification of the referent. If this wasn't the case then an identification of the numbers would be needed to establish the meaning of Hume's Principle. This point is an objection to Hale and Wright only in so far as categorising the terms on the left-hand side as semantically idle means that they are not functioning with the surface syntax they appear to be, and that their logical syntax is idle in the sense that it cannot muster the provision of reference to any objects. Dummett insists that when we introduce the meaning of a term the way in which we introduce this meaning determines whether or not that term is introduced as an identifiable unit contributing to the truth condition. In abstraction, we stipulate the meaning of a term by laying down a block truth condition using a bi-conditional which makes the sentence meaningful but does not bestow any meaning on the rest of the terms outside their particular combination with the other expressions in the sentence.

By what Trueman calls the 'sentential' model of stipulation, the left-hand and right-hand sentences of Hume's Principle are stipulated to have the same truth-value. Trueman argues that if Hale and Wright mean for us to understand implicit definition in this way, then the terms on the left-hand side do not function as replaceable semantic units. This is because the meaning of the left-hand side has been fixed as a block, and not by appeal to the *reference* of the expressions. Trueman draws the following lesson for establishing the meaning of sentences in general:

> When we fix a truth-value for a sentence, the way in which we do so settles what, if any, role the parts of that sentence play in determining that truth-value. So if we want 'Socrates' to appear as a name of Socrates in a given sentence, we are thereby restricted in the ways in which we are free to fix a truth-value for that sentence. We must do so in a way that assigns the appropriate role to 'Socrates': the fact that 'Socrates' refers to Socrates must have a knock on effect on the truth value of the whole sentence (Trueman 2014, 371-372).

Trueman's conception of the relation between meaning, reference, and truth is here very close to Dummett's. The name 'Socrates' is only established as semantically operative in a sentence if it goes through its reference. This can be understood along the lines of Dummett's charge of semantic idleness. Trueman presents Hale and Wright with what he takes to be a dilemma: either the meaning of '$Nx...x...$' is secured by sub-sentential stipulation but the truth of Hume's Principle is not guaranteed, or the truth of Hume's Principle is secured by sentential stipulation but the meaning of '$Nx...x...$' is semantically idle. Dummett and Trueman's common objection is that Hume's Principle by itself cannot ensure the expressions on its left-hand side make the necessary kinds of semantic contribution which Hale and Wright need them to. And as we have seen, what Hale and Wright need Hume's Principle to deliver is the genuine presence of singular terms in the logical syntax and not merely in the surface syntax of abstraction principles.

Wright has replied directly to Dummett's charge. He claims that the process which Dummett describes – of stipulating the truth condition of the left-hand side as a unit – is actually the process of *explicit* definition. Conceived of as part of an explicit definition, the left-hand side is merely an abbreviation – or notational variant – of the right-hand side. However, Wright claims the understanding of the process of *implicit* definition needed for abstraction is to conceive of the right-hand side as fixing the truth conditions of the left-hand side in equal partnership with the recognition of the surface syntax of Hume's Principle. He answers Dummett as follows:

> ...while the truth-conditions of such statements may indeed be given, via the contextual stipulations, as those of statements in which no such terms occur, it is necessary, in order to understand statements of the former kind, to *know* more than their truth-conditions may be so given: you have in addition to follow through the Fregean abstraction – to read the left-hand sides of the appropriate principles not merely as notional variants of the right-hand sides, but in a way which is constrained by their surface syntax and the familiar vocabulary they contain (Wright 2001, 271).

Wright's idea seems to be that features of the left-hand side are also appealed to as part of the definition. In particular, recognition of the identity relation and the syntax of the left-hand side supply a guide to the jobs – or semantic contributions towards a truth condition – which each expression on the left is apt to provide. Our grasp of the final truth condition of the left-hand side (given by the right-hand side) shows us which jobs would be required to achieve this truth condition. In this way, we are able to fix the semantic contribution of each of the terms on the left-hand side by allocating which expressions are suitable for doing the jobs which need done. Wright goes on to make the point that this is the sense in which abstraction principles do not merely establish truth conditions but *carve up their content* (Wright 2001, 272).

Wright's reply to Dummett reveals two important features of what he takes to be involved in an appeal to surface syntax. The first feature it reveals is that the syntactic evidence provided by such an appeal is more *substantive* than it initially appeared to be. The syntax is not merely a rough guide to what might be going on at the logical level; rather, *the recognition of the surface level of syntax replaces the usual appeal to reference in fixing the sense of a term*. The second feature revealed is that because of this, the appeal to surface syntax is more *important* than it initially appeared to be. The surface level of syntactic information is now essential to the mechanism of abstraction since, without it, abstraction principles are unable to carry out any content recarving.

Although neither Hale nor Wright respond explicitly to Trueman, their reply to Dummett gives us a good idea as to how they might do so. We have already noticed that the second horn of Trueman's dilemma takes up Dummett's semantic idleness point and so Wright's defence must – for better or worse – be the one he uses against Dummett. In fact, using Trueman's framework we can make the counter-objection which Wright makes to Dummett clearer: the correct model of implicit definition is one on which the sentential and sub-sentential models happen *at the very same time*. Thus there is no dilemma between two alternative understandings of definition, where neither can secure everything required of abstraction because the two conceptions of implicit definition *are not alternatives*. Rather, the meaning of the left-hand side is generated in part by its having the same truth value as the right-hand side and in part by employing its surface syntax to fix the meaning of the sub-sentential parts. During abstraction, we are to recognize the meaning as constrained both by the bi-conditional relationship between the left and right-hand sides and also by the syntax we are already familiar with on the left-hand side.

It is doubtful that Wright's strategy would satisfy either Dummett or Trueman, as the original objection was precisely that the syntactic appearance of a term is not enough to establish its semantic functioning – for that, the reference of the term is needed. To use Hale and Wright's own metaphor, it is not enough that the abstraction produces a mould (that is, the *appearance* of a singular term) because the mould must be somehow *filled*. Wright claims that in abstraction the syntactic appearance of a term can establish its semantic contribution and fill the mould. It comes down to whether one thinks that the reference of a term is absolutely necessary in order to establish the term as a legitimate semantic unit, or whether one considers abstraction principles as an exception to that rule, whereby the surface syntax can replace an appeal to reference in order to provide a novel kind of route to the term's meaning. It is difficult is to say what would decide the issue, but it is clear that Wright considers the burden of proof to lie with his opponents.

Rather than taking up this burden of proof, I will instead be concerned to provide another source of evidence for the logical syntax of Hume's Principle. Wright's defence still has an important gap which can be exploited. This is the gap between the surface syntax which is used to fix the meaning of the expressions and *the actual meaning which is established* with the help of the surface syntax. Of course, it is implicit in Wright's account that if the surface syntax is what is doing the work of fixing the semantic contribution of the terms (rather than the reference of those terms) then the semantic contribution which is fixed is precisely the one which can be directly read off the surface syntax. That is to say, if an expression has the shape of a singular term then this fixes its semantic contribution as that of a singular term. There is an important gap, however, between conceding that the surface syntax of '$Nx : Fx$' *helps* to fix its semantic contribution and inferring that this expression *is* contributing as its surface syntax suggests.

In the next section, I will employ Frege's dilemma for implicit definitions to provide alternative semantic evidence that the surface syntax does not mirror the logical syntax of Hume's Principle – even on Wright's account of implicit definition. Furthermore, I will establish that the mismatch between the two levels is problematic for Hale and Wright because it turns out that there are none of the singular terms they need to capture the numbers.

4.3.3 Applying Frege's dilemma for implicit definition

In this section I will adapt what I have called Frege's dilemma in order to apply it to Hale and Wright's account. In fact, I'll do it twice: once to sketch how the objection will be applied and show what it will affect, and once more to carefully transplant the target of the objection from Hilbert's axiomatisation to Hale and Wright's neo-Fregean logicist project.

A sketch

As we have seen, abstraction principles (like Hilbert's axioms) are implicit definitions. This fact is actually all we really need to apply to the dilemma:

> **Frege's Dilemma for implicit definition**: Either your *definiendum* has a prior meaning and so your definition is circular, or you have an *explicit* definition of a concept one level higher than your *definiendum*.

With respect to Hale and Wright's account, this implies that either the *definiendum* of an abstraction principle has a meaning – in which case definition by abstraction is a circular venture – or the *definiendum* does not have a meaning – in which case it acts as a variable which marks the argument place of a higher-order concept which the abstraction principle *explicitly* defines.

If Hale and Wright accept the first horn of the dilemma then all of the parts of an abstraction principle will already have a determinate meaning. Because of this, the abstraction principles will be laid down in the same way as axioms. It is clear that this undermines the privilege and purpose of abstraction principles.

If Hale and Wright accept the first horn of the dilemma then the abstraction principle is revealed to be functioning as an explicit definition rather than an implicit one, but it is an explicit definition of a distinct concept. Further, if the primitive expressions are acting as variables then this undermines the genuine presence of singular terms in the abstraction principle. In order to see precisely how the occurrence of singular terms is undermined we will have to more carefully apply the dilemma to Hume's Principle.

Transplanting the dilemma

To apply the dilemma we first need to identify what Hale and Wright consider to be the *definiendum* of Hume's Principle.

The most obvious candidates are the purported singular terms themselves. However, this suggestion is made awkward by the fact that the singular terms in Hume's

Principle are the complex semantic units '$Nx : Fx$' and '$Nx : Gx$' which are made up of a term-forming operator 'Nx...' and the concepts 'F' and 'G'. The concepts 'F' and 'G' are sortal concepts which must be taken as grasped prior to the definition, since we cannot understand Hume's Principle unless we recognise that the same 'F' and 'G' which instantiate the equivalence relation occur also on the left-hand side. As such, the 'F' and 'G' must be part of the *definiens* of the definition and this is not represented if we take the *definiendum* to be the singular terms '$Nx : Fx$' and '$Nx : Gx$'. This goes back to Trueman's point about 'Socrates': we must define '$Nx : Fx$' and '$Nx : Gx$' not as block singular terms but as having semantic complexity such that 'Nx...', 'F' and 'G' operate in the singular term as relevant units, suitable to be replaced.

The same reasoning holds for the suggestion that the *whole sentence* on the left-hand side of Hume's Principle is the *definiendum*. Indeed, in addition to involving 'F' and 'G' it also involves identity which we have seen to be grasped prior to abstraction and so also forms part of the *definiens* and not the *definiendum* (Hale & Wright 2001, 148-150).

Once we have set out the objection of this chapter, Hale and Wright may have a new reason to revise their view and claim that the singular terms are the true *definiendum* of Hume's Principle. I predict that this would be a natural point from which they might try to avert the damage of the dilemma. This will be addressed in a later section where I argue that the objection will stand regardless of how the *definiendum* is identified.

For now, let us agree with Hale and Wright's present identification of the *definiens* to be 'Nx...'. If Hume's Principle is a successful definition, then what it defines is *the number function*. Which is to say that it settles the meaning of the term-forming operator which takes predicates F and G to singular terms. In doing so, it secures reference to a function which takes one-one concepts to the same object and so is apt to associate numbers with concepts. Note that the genuine occurrence of singular terms is thus achieved by establishing the meaning of the 'Nx...' as operating on 'F' and 'G' (which we already recognise as concepts) to form singular terms. What is needed in this story is for 'Nx...' to be recognised *as a singular term-forming operator*. This is done by employing the surface syntax of the output expression '$Nx : Fx$'. In other words, it is the syntax which determines that '$Nx...$' maps predicates to singular terms rather than to predicates or sentences as would be the case with '$Nx : Fx$' with something like the following (nonsensical) abstraction principle:

$$\forall F \forall G(Nx : Fx \leftrightarrow F \approx G),$$

here '$Nx : Fx$' would be read as something like, 'the number is F'.

Having identified our *definiendum*, let us again apply the dilemma: either '$Nx...$' has a prior meaning and the definition is circular or we have an explicit definition of a concept one level higher than '$Nx...$'.

If '$Nx...$' were already meaningful this would require an antecedently given number function. This is to give up providing a reconstruction of the numbers and to instead lay down an axiom – in which case we may just as well lay down the Dedekind-Peano axioms. Hale and Wright would obviously not be happy to accept this horn of the dilemma since it undermines the theoretical motivations of abstraction as preferable alternatives to axioms.

This pushes us onto the second horn, that Hume's Principle gives an explicit definition of a concept one level above '$Nx...$'. As I think this is the horn of the dilemma that Hale and Wright land on, let us spend some time unpacking what this diagnosis amounts to.

First, let us establish the order of the higher-order concept which is defined. To do this we must establish the order of '$Nx...$'. In 'Function and Concept' Frege states that the order of a function (and similarly of a concept) is determined by the order of its argument. In particular, a function or concept is one level higher than whatever features in its argument place (1980*g*, 146-8). Therefore, if '$Nx...$' referred to a function it would refer to a function which is at least second-order since this function would have concepts as its arguments (F, G) and concepts – according to the Fregean hierarchy – must be at least first-level simply because they are unsaturated. This makes the explicit definition a definition of a concept which is at least third-order. Henceforth let us assume that concepts F and G are first-order and so speak of Hume's Principle as defining a third-level concept.

I think there is an inherent tension with Frege's general picture here which deserves explicit notice before we proceed. Simply put, if the type of a function and the type of a concept is determined by the type of their argument, and if we are to be able to count the things falling under concepts of different levels, e.g. the number of moons of Mars; the number of primitive geometric concepts, etc., then the very same function must take concepts of any level and map them to the same object (same number) when and only when those concepts are equinumerous. However, this means that the function itself is of indeterminable logical type since its argument can be filled by any logical type.

In order for Frege to consistently maintain that functions are typed by their arguments and that one can count things falling under any level of sortal concept, it seems that he will be forced to fracture the number function into infinitely many distinct logically typed functions, all of which map to numbers. Each function will be of level $n+1$, where its argument is level n. This concession leaves Frege with an unsatisfying conceptual analysis of the concept of number. It is difficult to see how Frege can go on to say that these distinctly typed functions in some sense constitute the number function. Thus Frege seems forced to abandon any attempt to provide a unified conceptual analysis of number. At the very least, this is a drastic revision of Frege's theoretical aims.

For the sake of presenting the current argument, we will set this issue aside and proceed by noting that in the case of Hume's Principle the candidate number function being defined is at least third level. For the ease of exposition we will mostly speak of Hume's Principle as straightforwardly defining a third level function. This does not mean we will take it for granted that Frege must define infinitely many number functions. All we will assume is that the level of the function defined by Hume's Principle is dependent – by uncontroversial Fregean principles – on the level of the concepts in its argument places, F and G.

We can now ask what the third-order concept is. It is the third-order concept in which the function expression 'Nx...' in Hume's Principle features as a *variable*. We can formally represent the concept more clearly by replacing '$Nx...$' with another variable term:

Hume's Principle* $\forall F \forall G(\mathbf{X}_x : Fx = \mathbf{X}_x : Gx \leftrightarrow F \approx G)$.

For this reason it is a mistake to infer from the surface syntax of Hume's Principle that '$Nx...$' is contributing to the definition as a candidate *name* for the number function which binds the 'x'. Adapting the surface syntax as above shows that 'X_x' is not apt to name a number function, but is instead apt to be *replaced* with a term referring to a number function. In this way, Frege's diagnosis gives us alternative syntactic evidence for the logical syntax of the abstraction principle. It shows that there is no genuine presence of a term-forming operator in Hume's Principle but only a place holder for one. The above formalisation also shows that it is misleading to think of Hume's Principle as partitioned into a left and right side since this makes it look as if the *definiens* is restricted to the right-hand side when in fact all of Hume's Principle is involved in defining a higher-order concept, i.e., the concept of *a number-function*.

Let us go over this important point again. To infer from the surface syntax of Hume's Principle that it must contain an expression whose semantic contribution is to name a function, is akin to the following mistake which the King makes to Alice in *Through the Looking-Glass*:

> "I see nobody on the road," said Alice. "I only wish I had such eyes" the King remarked in a fretful tone. "To be able to see Nobody! And at that distance too!" (Carroll 1871, 53).

Here, the King has understood "Nobody" as a name, when in fact it is a quantifier expression. The reason that the King has made this mistake is precisely because he was mislead by the surface syntax of Alice's remark. To think that Alice referred to a particular person is akin to understanding Hume's Principle as defining a second-level function expressed by '$Nx...$' when in fact it defines a third-order concept in which '$Nx...$' functions as a variable *ranging over* such second-level functions.

As a further illustration we can adapt Frege's actual objection to Hilbert in the following way:

> The characteristic marks you [Hale and Wright] give in [Hume's Principle] are apparently all higher than first-level; i.e., they do not answer to the question "What properties must an object have in order to be a [number]?", [the characteristic marks are also all higher than second-level, i.e., they do not answer to the question "What properties must a function have in order to be a number function?"], but they contain, e.g., [third]-level relations, e.g., between the [second-level number function] and [some first-level] concepts [F and G]. It seems to me that you really want to define [third]-level concepts but do not clearly distinguish them from first [or second level] ones (Frege, 1900/1980*b*, 46, adaptation mine).

This adaptation neatly summarises the diagnosis of the second horn of the dilemma.

The consequences of this diagnosis for Hale and Wright are immediate and damaging. If 'Nx...' is functioning as a variable in Hume's Principle then Hale and Wright lose their syntactic evidence for the presence of singular terms. This is because the singular terms were given to us by way of the term-forming operator '$Nx : Fx$' . Applying Frege's diagnosis we see that the term-forming operator is not itself furnished with meaning and because of this it cannot secure well-defined singular terms as its values. Instead, the expression in question operates like a variable "$X_x : Fx$" which is certainly not a singular term but an incomplete expression forming part of a predicate – in particular, the predicate expressed by the entirety of Hume's Principle .

This produces a domino effect: if there are no singular terms, then there is no reference to objects and if there is no reference to objects then there is no logical reconstruction of the numbers.

Thus, merely from the fact that abstraction principles are implicit definitions, we can use Frege's dilemma to provide us with alternative syntactic evidence regarding the expressions in Hume's Principle which debunks the appeal to surface syntax. This puts Hale and Wright in a difficult position, for they can neither avoid their need for the genuine presence of singular terms nor revise their claim that Hume's Principle is an implicit definition.

4.3.4 Anticipating a counter-objection

In later chapters I will deal with ways in which Hale and Wright can defuse or incorporate my objection and I will argue that there is no way for them to do so which is not ontologically revisionary. In this section, I identify what is, as far as I can see, the only potential route of *avoiding* my objection entirely. This route would be tempting for Hale and Wright because it would involve making only a minimum amendment to their current account. I then show why the tempting strategy does not avoid the objection.

At the start of applying Frege's dilemma we took the *definiendum* of Hume's Principle to be 'Nx...', the term-forming operator, in line with Hale and Wright. However, it seems that if Hale and Wright were to ignore the worry of semantic idleness, they could take the *definiendum* to be the purported singular terms themselves '$Nx : Fx$' and '$Nx : Gx$'.

At first glance, this seemingly small revision is able to avoid my objection in the following way: if the *definienda* of the definition where the singular terms '$Nx : Fx$' and '$Nx : Gx$' then the *definienda* would be of level 0, since the expressions would refer to objects. This would mean that what was defined by Hume's Principle was of level 1, by Frege's dilemma for implicit definition. Thus, Hume's Principle would define a concept of a lower-order than I identified in my objection, one which could be suitable for use as a number concept. It seems that making a small change in the *definienda* could thus avoid the damage of Frege's dilemma altogether.

This potential dodge is intuitively appealing but we can block off this escape route in the following way. We can concede that the definition would now define a lower-order concept but when we look at the proposal in detail it is clear that *this lower-order concept is not suitable for use as the concept of number*. This is brought out clearly if we again make the surface syntax of Hume's Principle match the new logical syntax by replacing the identified *definienda* with a variable:

$$\forall F \forall G(\mathbf{Y} = \mathbf{Z} \leftrightarrow F \approx G).$$

If the purported singular terms are the *definienda* of the definition then what we have is a two-place first-order predicate expression. This predicate picks out the concept of two objects being identical when any two concepts stand in an equivalence relation. This does not provide us with a number function. A number function is a function which maps two concepts to the same object only when *those concepts* are one-one. This concept, by contrast, does not preserve the relation between the objects and the concepts F and G. It does not associate the objects with the concepts standing in an equivalence relation and for this reason it is unsuitable as a candidate *definiens*.

Another way to emphasis the same point is to note that the concept is not of a high enough level to even be related to a number function. The alternative *definiendum* 'Nx...' had the disadvantage of yielding a concept which was one level higher than the desired number function. That being said, the concept defined was the concept *of* a number function and so at least it preserved a link between the concept defined and the function which we needed. In this case, what is defined is a concept one level *lower* than the desired second order number function and which preserves no such link. That is just to say that the concept of two things being identical whenever there are two concepts in an equivalence relation is irrelevant to the mapping of equinumerous concepts to the same object.

In fact, to identify '$Nx : Fx$' and '$Nx : Gx$' as the *definienda* is actually a more direct way to lose the functioning of singular terms than our original application of Frege's diagnosis. Before, the singular terms were lost as an indirect result of losing the term-forming operator. But with this change to the *definienda*, we have that the singular terms themselves function as variables in such a way that there is no way to recover the complexity of the number function or the genuine occurrence of F and G.

For these reasons it is clear that, although initially tempting, this strategy is a false hope, because it would define a *different* first-order concept and not the concept of a number function.

An appeal to the surface syntax of the purported singular terms will be of no help here. We have already accepted Wright's conception of the surface syntax working as part of the *definiens* to fix the meaning of the expressions. Thus the surface syntax of the other expressions in Hume's Principle establishes that they have the function they seem to; in particular, the identity refers to the identity relation with which we are familiar, and the same F and G occur on both sides of the bi-conditional. Furthermore, the surface syntax of the *definiendum* fixes the level of the concept defined by the implicit definition. It enables us to recognise that if we did make the term 'Nx...' directly meaningful then we would have to define a second-level function. But what Frege's diagnosis reveals is that in the case of the *definiendum* the logical syntax, i.e. the actual functioning of the defined primitive is not to be read off the surface syntax. The numerical primitives contribute as variables, marking the argument place in the concept defined by abstraction. The surface syntax does play a role in establishing the meaning of the variable and it does this by representing the order of the concepts which are replaceable for the variable when the variable is understood as an argument place. This exploits the potential gap we identified in Wright's account, namely allowing that an appeal to the surface syntax may coherently replace an appeal to the reference of a term in establishing the term's meaning, but this does not imply that the established meaning can be directly read from the surface syntax.

Therefore, we are able to accept both Wright's conception of implicit definition and that the *definiendum* of Hume's Principle is the term-forming operator, the singular terms, or the left-hand side of the bi-conditional, and still show – using alternative semantic evidence from Frege – that there are no expressions which are functioning as genuine singular terms in abstraction principles. Since it is from the presence of genuine singular terms that Hale and Wright recapture the numbers, this is a loss which directly undermines their theoretical ambitions to provide a logicist result.

Conclusion

In this chapter, we have introduced another foundational account in mathematics and novelly connected neo-Fregean logicism with the Frege-Hilbert controversy. This was done by using Frege's dilemma for implicit definition, which we identified during our exposition in such a way that anticipated this connection and ensured it could be made cleanly and straightforwardly. Although the details of Frege's diagnosis had to be adapted, we exploited a startlingly simple similarity between the accounts to make the connection: namely, both employ implicit definitions as foundational principles in order to give a reconstruction of some domain of mathematics – in the one case geometry, and in the other case arithmetic.

Showing how Frege's diagnosis can be adapted to understand Hale and Wright's account was only half of the work. What enabled us to produce a new objection was the fact that the diagnosis undermined two important commitments of the neo-Fregean logicists: that abstraction principles are implicit definitions and that there is the genuine presence of singular terms in Hume's Principle. On account of the centrality of these commitments, the objection we gave was very problematic and (thanks to Frege) also diagnostic. In the next chapter I will consider whether Hale and Wright can revise their commitment to singular terms and in this way avoid the brunt of this new objection.

Chapter 5

An objection to Hale and Wright's Plan B

Introduction

In this chapter we will consider the alternative routes that Hale and Wright can take – or have taken – to recover numbers as logical objects without the aid of singular terms.

As we might expect given the impressive neo-Fregean logicist cannon, Hale and Wright have set out a proposal that has the potential to let them absorb the objection that I have raised while safeguarding their logicist program.

The first section will expound Hale and Wright's plan B. The next section will undermine the proposal by again drawing on objections which Frege raises to Hilbert. The last section will consider the ontological cost Hale and Wright have to pay if they are to avoid this second application of Frege's objections to Hilbert. The subsequent chapter will document the way in which Hale and Wright have already forfeited their Fregean ontology.

This strategy will have the following effect: since we are varying the commitment to singular terms, it holds constant the ontological consequences of the commitment to implicit definitions. This is what is of interest to us: that the nature of the foundational sentences has an influence on the ontology they mean to characterise.

5.1 Hale and Wright's Plan B: conditional forms of Hume's Principle

In "Implicit Definition and the A Priori", Hale and Wright consider some conditional forms of Hume's Principle. They conclude that a direct stipulation of an abstraction principle is possible and preferable to a less direct approach. However, if it turned out that for some reason a direct stipulation of the principle was problematic – perhaps because it was shown to be "arrogant", for example – then we could still introduce numbers using a more indirect approach. They compare this alternative to the way in which scientific entities are implicitly defined by their theories. Here we will survey their proposals and identify their broad strategy.

5.1.1 Conditional introduction of numbers

Hale and Wright begin their survey of different versions of Hume's Principle with the conditional given by Harty Field in his "Platonism for the Cheap" (1989, 167-70).[1]

If numbers exist, then $\forall F \forall G(Nx : Fx = Nx : Gx \leftrightarrow F \approx G)$. [2]

Hale and Wright classify this as a *Carnap conditional*, using Horwich's terminology (1997, 425).[3]

Let us carefully unpack how this conditional is to be understood. The interesting part of the conditional is its antecedent. However, the semi-formal presentation of the

[1] It is interesting that Kemp – who, as we have seen, is sensitive to the connection between the Frege-Hilbert controversy and neo-Fregean logicism – also suggests in a footnote that Hume's Principle be understood as conditional, if it is a definition (Kemp 2005, 186 ft.7).

[2] This particular formulation is from Hale and Wright (2001, 143). It is not original to Field.

[3] Lewis canonically refers to it as a Carnap *sentence* (1970, 427-46).

conditional presented by Hale and Wright obscures how this antecedent is supposed to work. To make it clearer what is going on in the antecedent, we can use the sentence which makes the logical syntax of Hume's Principle more perspicacious:

Hume's Principle* $\forall F \forall G(\mathbf{X}_x : Fx = \mathbf{X}_x : Gx \leftrightarrow F \approx G)$.

Notice that if we bind the variable which stands in for the term-forming operator then we get the following Ramsey sentence.[4]

Ramseyfied Hume's Principle* $\exists \mathbf{X} \forall F \forall G(\mathbf{X}_x : Fx = \mathbf{X}_x : Gx \leftrightarrow F \approx G)$.

We can then use this sentence as the antecedent of the conditional. Call this the Carnap conditional, in line with Hale and Wright's use of the term,

Carnap Conditional of Hume's Principle*

$$\exists X \forall F \forall G(X_x : Fx = X_x : Gx \leftrightarrow F \approx G) \rightarrow \forall F \forall G(Nx : Fx = Nx : Gx \leftrightarrow F \approx G).$$

Hale and Wright are the first to admit that the Carnap conditional is too weak to deliver the numbers by itself. Wright points out that it requires supplementation in order to affirm its antecedent (Wright 1983, 148-52). Indeed, precisely what Field does is *deny* the antecedent while maintaining the truth of the conditional. According to Hale and Wright, the Carnap conditional is akin to a theorist stipulating that, "if there are any things satisfying such-and-such laws then there are electrons" (2001, 141). What are we to make of this explanation? It is clear they want to imply that the numbers can be thought of as theoretical entities in some similar respect to how we can think of electrons. However, to understand what this similarity amounts to we must set out the broader proposal, in the context of which Hale and Wright introduce the Carnap conditional. Hale and Wright only suggest the conditional after they survey a way of understanding the content of a scientific theory where it is bifurcated into two components:

[4]Thanks to Michael Potter who first suggested to me the idea of Ramsifying Hume's Principle. This is also noticed by Trueman (2014, 365) and Hale & Wright (2001, 139, 148).

> ... one encapsulating the distinctive empirical content of the theory without deployment of the novel theoretical vocabulary, the other serving to fix the meaning(s) of the theoretical term(s) we seek to introduce. The theory's total empirically falsifiable content is, roughly, that there exist entities of a certain kind, viz. entities satisfying (a schematic formulation of) the (basic) claims of the theory (Hale & Wright 2001, 139).

Hale and Wright go on to give an example of what they have in mind which is in tandem with our reflections here.

> Thus if, focusing for simplicity on the case where a single new theoretical term, 'f', is introduced, the undifferentiated formulation of the theory is '$\#f$', then its empirical content is exhaustively captured by its Ramsey sentence, '$\exists x(\#x)$', where the new variable 'x' replaces 'f' throughout '$\#f$'. The new term 'f' can then be introduced, by means of what is sometimes called the *Carnap conditional*: '$\exists x(\#x) \rightarrow \#f$', as denoting whatever (if anything) satisfies '#_' (on the intended interpretation of the old vocabulary from which it is constructed). This conditional expresses, in effect, a convention for the use of the new term 'f'. Being wholly void of empirical content, it *can* be stipulated, or held true a priori, without prejudice to the empirical disconfirmability of the theory proper (Hale & Wright 2001, 140).

From this context, the Carnap conditional should be understood as introducing a *convention* for how to understand the denotation of a new theoretical term. Further, the stipulation of this convention is not undermined by the empirical content of the theory turning out to be false, nor apparently even by the complete absence or overabundance of referential candidates.

Before we unpack this approach any further, I wish to emphasise an important disanalogy between a theory of electrons and a theory of number. While in the former case what is supposedly being defined is a theoretical *object*, in the latter case it is a *function*.

That is to say that the Carnap conditional provides not a theoretical treatment of numbers *per se*, but a treatment of the number function. As such, the opaque explanation we began with is better rendered as follows: if there are any *number functions* then they satisfy such-and-such-laws. This point is noticed by Kit Fine where he notes that Field's denial of the antecedent of a conditional form of Hume's Principle:

> ... just amounts to the claim that there is *no operator* that conforms to the Law. If we regard Hume's Law as part of a "scientific" theory, then this response is equivalent to a Ramsey-style treatment of theoretical terms (Fine 2002, 524 fn.10, emphasis mine).

Here Fine anticipates the motivations behind Hale and Wright's conditional introduction strategy but notices that the proper *definiendum* is the number function. This is clear if we carefully adapt Hale and Wright's simplified example to the case of Hume's Principle. Then we have that the Carnap Conditional of Hume's Principle* expresses a convention for understanding the denotation of the new term 'Nx...'. In particular, 'Nx...' refers to whatever function satisfies Hume's Principle* with the intended interpretation of its logical vocabulary. Since Hume's Principle* expresses the higher-order concept which we have been concerned with, this proposal is tantamount to introducing a function by stipulating the following: it is the function which falls under the concept of being a number function, if anything does. Thus, the conditional introduction strategy that Hale and Wright consider provides another way to put the point that Hume's Principle is a definition of a higher-order concept rather than a definition of a term-forming operator referring to a number function.

Hale and Wright go on to say that their favoured use of conditional introduction is the inverse of such a procedure. They tell us that the theorist stipulates "if there are electrons, they satisfy such and such laws" rather than "if there are any things satisfying such-and-such laws then there are electrons" (Hale & Wright 2001, 141). This they call the *inverse* Carnap conditional:

$\forall F \forall G \forall u \forall v((u = Nx : Fx \wedge v = Nx : Gx) \rightarrow (u = v \leftrightarrow F \approx G)$.[5]

It is unclear what we are to make of this conditional. One clue is that earlier in the discussion Hale and Wright say that '$\exists F \exists x(x = Ny : Fy)$' can be understood as the hypothesis 'numbers exist'. Perhaps, then, we should read it as follows: if some objects in the domain turn out to be numbers, these numbers will satisfy this law. However, Hale and Wright explain it in a slightly different way. This principle, they say,

> ... tells us what numbers are in just the same way that the inverse Carnap conditional for any (other) scientific theory tells us what the theoretical entities it distinctively postulates are – by saying what (fundamental) law(s) they must satisfy, if they exist. That there are numbers is itself no conceptual or definitional truth – it is rather the content of a theory (Hale & Wright 2001, 144).

However, it doesn't seem to be the content of the inverse Carnap conditional *that there are numbers* – rather – it seems to say that if the term-forming operator refers to objects then these objects satisfy a certain condition. There is nothing in the content of the conditional that requires there to be numbers, unless there are already numbers in the domain. Consider what Field would likely say about the inverse Carnap conditional. He would want to deny the antecedent, but how would he go about this? He would have to ensure that the universal quantifiers binding u and v do not range over numbers. In that case, that there *are* numbers depends on whether or not we have quantified over numbers. The principle says something more akin to the following: if we have quantified over the numbers, then the numbers satisfy these laws. If this is the case then it seems that Field and Hale and Wright would just understand the inverse Carnap conditional as ranging over a different domain. In this case, the interesting dispute seems to be with respect to how the domain is fixed. This brings us to the well trodden issue of the impredicativity of abstraction principles.

[5] Hale & Wright (2001, 143)

5.1.2 Impredicativity

Dummett, Potter and Sullivan have pressed onto Hale and Wright a circularity worry regarding the impredicativity of the quantifiers in abstraction principles. They worry that if the objects given in abstraction are in the domain of the quantifiers on the right-hand side of abstraction principles – i.e. the principles are impredicative – then it seems that we are required to have some grasp of these objects prior to being given them by abstraction.[6] After all, the quantifiers are grasped in advance of the abstraction. And presumably, for their meaning to be made determinate, the domain must be independently circumscribed in a way that must involve at least some kind of indirect reference to the desired objects.

Hale and Wright counter Potter and Sullivan's version of the impredicativity worry by appealing to a universal domain. They argue that if such a notion turns out to be coherent then they are absolved of the need to demarcate a domain – such that even if abstraction principles are to be understood impredicatively, we need not presuppose an independent means of reference to the members of the domain. Thus, they claim, there is no worrying circularity with respect to this kind of quantification,

> ...so long as we can legitimately take recourse to absolutely unrestricted (what [Potter and Sullivan] term 'genuinely universal') quantification, there will be no further difficulty in the idea that the first-order quantifiers in an impredicative abstraction range over independently existing objects of which we may have no antecedent conception – and indeed, we would wish to add, of which there may be no possible conception – independent of the abstraction itself (Hale & Wright 2008, 203).

[6]The relevant quantifiers on the right-hand side only appear in the unabbreviated formulation of Hume's Principle. I have emphasised them below:

$$\forall F \forall G(Nx : Fx = Nx : Gx \leftrightarrow \exists R(\forall \boldsymbol{x}(Fx \to \exists! y(Gy \land Rxy)) \land \forall \boldsymbol{y}(Gy \to \exists! x(Fx \land Rxy))).$$

Hale and Wright use a different strategy to counter Dummett's version of the impredicativity worry. They argue that it is indisputable that there is a "range inspecific" grasp of quantifiers such that one can grasp a sentence which quantifies over a domain without grasping all of the members of that domain. The neo-Fregean logicist's most powerful argument for this understanding of quantification is that it is unclear how an infinitely large range could ever be specified without use of a generalisation (Hale & Wright 2001, 242). But if generalisation presupposes an articulation of its domain then it seems we could never get our foot on the ladder. Furthermore, their claim that:

> ... it is not and cannot in general be a prerequisite for a quantified statement to be determinate or determinate enough in content that the range of its quantifiers should have been specified in advance (Hale & Wright 2001, 242).

Has its precedent in Frege, where he insists that:

> If I utter a sentence with the grammatical subject 'all men', I do not wish to make an assertion about some Central African chief wholly unknown to me. It is thus utterly false that I am in any way designating this chief when I use the word 'man', or that this chief belongs in any way whatsoever to the reference of the word 'man' (Frege 1895/1980*a*, 227).

In this way, the neo-Fregean logicists maintain that numbers are in the domain of the quantifiers used in abstraction but that we need not demand a grasp of them prior to a grasp of the *definiendum* of abstraction. In order to grasp some kinds of general statements about them, we require an acquaintance with numbers (for example) no more than we require an acquaintance with an unknown man in Africa to grasp general statements about men.

We should note that the neo-Fregean logicists prefer an impredicative reading of abstraction principles but refuse to contrast this with a free reading. Instead they repeatedly claim the following:

> Both points are central to our conception of abstraction, on which abstraction principles are, in crucial cases, both *free and impredicative*. We do *not* assume the existence of referents for the associated abstract terms. Since the instances of an abstraction principle are material biconditionals, all that is required for their truth is that their left- and right-hand sides have the same truth-value, and this condition can be met by taking left-hand sides whose terms lack reference to be false – provided, of course, that their right-hand sides are likewise false. If the abstraction is conceived as a stipulation, this coincidence in truth-value will be precisely what is stipulated, in effect. Since we do not assume that all well-formed singular terms in the language have reference, the underlying logic must indeed be free. This does not, however, mean that the inference to any instance of an abstraction can *only* be safely made when the abstraction is supplemented with an *additional* premise asserting the existence of referents for its left-hand-side terms (Hale & Wright 2008, 197-8).

The singular terms and variables in Hume's Principles are thus construed freely in the sense that they can be laid down without assuming a denotation. What Hale and Wright are at pains to stress in the above passage is that, on their picture, reference is secured by means other than explicit existential assumptions. They go on to say that reference is secured by establishing singular terms in true atomic contexts (Hale & Wright 2001, 199). We may bracket this manoeuvre for now since it involves exploiting their theory of reference, which will be covered in the next chapter. More to the point, even on their free interpretation of abstraction principles Hale and Wright maintain that numbers already feature in the universal domain and as such there remains an explanatory burden. Hale and Wright are aware of this and do not pass over their responsibility to address the issues surrounding abstraction; they argue that if abstraction principles are to be construed impredicativity then this feature of them can be rendered harmless. Indeed, they argue further that we should expect numbers to fall within the range of the abstraction principle, citing Frege's observation in §14 of *Grundlagen* that it should be a feature of any adequate definition of number that it be applicable to its own proper objects (Hale & Wright 2001, 245).

The strongest case for the harmlessness of impredicativity comes from the combination of the arguments we have surveyed which Hale and Wright forward to Potter and Sullivan and to Dummett respectively. For, technically, it is not enough merely to argue for a universal domain without the supplementary assumption that this domain can be grasped and that it can be grasped in a domain-unspecific way. And – in the other direction – technically it is not enough merely to argue that there is such a thing as a “range-unspecific" grasp of a quantifier without also making a case for why abstraction principles admit such a grasp. Hale and Wright are themselves well aware of this and in fact invoke the idea of unrestricted generality in support of there being a range-unspecific grasp of a quantifier (Hale & Wright 2001, 242-243). Putting both arguments together we get a more complete exposition of their view: there is a universal domain and there must be a “range-unspecific" grasp of this domain so that to grasp an unrestricted quantifier I need not grasp everything that there is.

5.1.3 The strategy for a conditional introduction of numbers

Looking at the options for conditional introduction of the numbers it seems, from a first glance, that the inverse Carnap conditional can only deliver the numbers if they are already in the domain. Hale and Wright defend this impredicativity with the following picture. In abstraction, there is a genuinely universal domain and a grasp of that domain which is “range-unspecific"; which is to say that to grasp a quantifier whose range is entirely general we need not be acquainted with everything. The truth conditions of the quantifiers over a domain can be determinate without requiring that we are in a position to independently specify the domain, i.e. without requiring that we have independent means of reference to each of its members. Hale and Wright take these ideas together to avoid the lingering worry of epistemic circularity and thus render impredicativity harmless.

The Carnap conditional is modest in restricting its ontological commitments to its consequent. The problem with this conditional, however, is that there seems to be no straightforwardly un-arrogant way to assert its antecedent given that this would involve commitment to infinitely many objects. Nevertheless, what is important to note about the general conditional strategy is that – as Hale and Wright are keen to stress – in both cases, if these conditionals are considered to be implicit definitions, then even without ontological commitment to numbers someone like Field has to admit that some kind of concept is defined.

In this way, the neo-Fregean logicists intend to move the stress away from an ontological stipulation of the existence of numbers to an introduction or explanation of the concept of a number. They tell us that,

> When an abstraction is put forward as an implicitly definitional stipulation, there is no attempt to create *objects* or stipulate their existence – what is created, if all goes well, is not objects, but *grasp of a concept* (Hale & Wright 2008, 202).

This brings their position closer to the diagnosis we have made of Hume's Principle with the aid of Frege. However, if this is all that is supposed to be produced by the process of abstraction, it remains for us to explain how the neo-Fregean logicist envisages that this concept can carry the burden of a logicist result.

There is an important point to bear in mind here: in their later writing, Hale and Wright stress that abstraction principles provide us with numbers in the sense that they provide us with a means of referring to objects which we could have no other means of referring to. In the following passage they explain this conception in detail.

> Our approach to abstraction, then, must be fashioned by the idea that what abstraction does is to provide a referential route to objects which are already in the domain. This constitutes a reconstruction of said objects because the objects were not grasped prior to the abstraction, not because they where not

> there but because we had no means of picking them out of the domain other than by the route which is afforded to us by abstraction (Hale & Wright 2009, 210).

Of course, for this proposal to be viable we must grant Hale and Wright that there are objects which plausibly admit of no independent means of specification despite falling in the range of our quantifiers, and we must grant that numbers are among these objects. These assumptions will prove to be highly relevant and we will return to them in detail in the next chapter.

Hale and Wright also claim here that the only means of conceiving of these abstractionist objects is to present them, "as the values of the relevant function". From this, we might venture that it is perhaps the concept which is explained or introduced by conditional abstraction principles which *constitutes* the novel referential route to the numbers. In particular, because this concept – the concept of a number function – can be said to represent the objects in the domain *as the values of the functions which fall under it*. This is another point which we will later return to. For now, let us step back and suggest a general strategy on the basis of these considerations that might offer a suitable escape route for Hale and Wright in the face of Frege's diagnosis.

5.1.4 Generalising the strategy

Here is a very basic point: on a Fregean language-first conception of reality there are two potential routes to an object. The most natural one is via the reference of a singular term. However, we could also get to objects via the reference of a predicate which was such that only the desired objects fell under the concept the predicate refers to.

In the last chapter, our concern was to take away the appeal to singular terms from Hale and Wright. This left open the possibility that they take the second, more indirect, route. Thus, Hale and Wright could accept the diagnosis that Hume's Principle defines a higher-order concept with open arms if they could find a way to recover the numbers *from this higher-order concept*.

The point here is that Frege's dilemma only works as an objection against Hale and Wright if we hold fixed their commitment to recovering numbers using singular terms. However, we have just seen that they have a proposal which amounts to taking the indirect route of recovering numbers with a *predicate*. Hale and Wright themselves admit that their suggestion is not a fully developed proposal, but what they do discuss provides enough for us to attribute to them a plan B.

For Hilbert, Frege's diagnosis that concepts defined by implicit definition are one level higher than their *definiendum* is not an objection. It is a sophisticated way of understanding how his axioms worked. In the same way, Frege's dilemma should not be considered an objection but an insight into the proper *definiendum* of Hume's Principle. This is viable if Hale and Wright can make a case for the suggestion which we have just raised: that Hume's Principle explains a concept which constitutes a novel referential route to objects in the domain. In particular; the objects are values of the function which falls under the concept. Furthermore, if this concept was explained only using logical laws and definitions, this could be the sense in which those objects are presented as *logical* objects.

5.2 Frege's critique of Plan B

This new proposal seems inherently promising. However, by another appeal to an important point of the Frege-Hilbert controversy, we are able to pull the promising rug from under Hale and Wright's feet.

Recall that Frege's actual objection to Hilbert only came later, when he makes the comparison that:

> Your system of definitions is like a system of equations with several unknowns, where there remains a doubt whether the equations are soluble and, especially, whether the unknown quantities are uniquely determined (Frege, 1900/1980*b*, 45).

Here Frege raises two doubts with respect to how Hilbert has secured the geometric primitives: he has not established that they are unique, nor that they exist. We considered these worries with respect to Hilbert's ontology; now we bring the same two-fold objection to the proposed conditional introduction of numbers. In fact, we will apply Frege's objection twice: once with respect to the numerical objects and once with respect to the numerical function. However, it should be kept in mind that, due to the intimate relation between object and function, these applications are not unrelated but merely separated for the sake of clarity.

5.2.1 The first application

First we have a question more directly analogous to Frege's point to Hilbert: can Hume's Principle establish the existence and uniqueness of the numbers? The existence point is not new, we have already considered the charge that abstraction principles do not secure the existence of the objects they introduce. Having now outlined the proposed escape plan, we can make this charge more specific.

The conditional versions of Hume's Principle define a concept which provides a new referential route to the numbers. Independently, we have shown the concept defined by Hume's Principle to be the concept of being a number function. This is (at least) a third-order concept. Consequently, it does not have objects falling under it; it has functions falling under it and the objects appear as the values of these functions. Frege's first charge then, is that defining such a second-order concept does not ensure that there are any objects falling in the value range of the functions which satisfy the concept. In other words, it does not ensure that the referential route leads us to any objects in the domain. It could be that the concept defined by the abstraction principle is inconsistent and thus has nothing falling under it, as with basic law (V). Herein enters the bad-company objection.

The second charge is that even if there are objects that are the values of the function, the concept that is explained by Hume's Principle can be satisfied by more than one function (any function that satisfies Hume's Principle*). In the next section we will show that the concept is satisfied by infinitely many functions if it is satisfied by any functions at all. This gives us different sets of objects which are the values of different functions, all of which satisfy the number concept. Julius Caesar could act as the value of a function just as well as the number 2 could. The referential route will be entirely indeterminate with respect to these competing collections of objects.

Therefore, using a higher-order concept to refer to objects does not pick those objects out uniquely, nor does it ensure they even exist. These objections are worrying but they are made worse when we consider that the existence and the uniqueness of the number function is also called into question.

5.2.2 The second application

Now we apply Frege's objection to the claim that Hume's Principle has determined the number function. The first doubt is whether a number function *exists* and the second doubt is whether a number function is picked out *uniquely* by the concept defined by Hume's Principle.

I think the first objection – that conditional forms of Hume's Principle do not resolve whether the conditional is 'soluble' – can most clearly be understood as the following simple point: Hale and Wright must establish the truth of the antecedent. We saw this to be the Ramsey sentence:

Ramseyfied Hume's Principle $\exists \mathbf{X} \forall F \forall G (\mathbf{X}_x : Fx = \mathbf{X}_x : Gx \leftrightarrow F \approx G).$

However, as we have noted, the sentence is outspokenly arrogant, i.e. it requires us to unparsimoniously posit the existence of infinitely many entities. It seems hard to know how the above Ramsey sentence could be established as true without illegitimate appeal to a number function which satisfies it.

The second worry is that conditional forms of Hume's Principle do not secure whether the "unknown quantities are uniquely determined". In order to substantiate this objection I will first use a permutation argument to show that there are infinitely many number functions.[7] I will then explain why Hume's Principle does not determinately refer to any of these functions.

Consider a number function, '$Nx...$' mapping concepts to objects. Now consider the objects which fall in the range of this function, i.e. all of the cardinal numbers. Take the first two of these numbers and permute them so that all two-membered concepts get mapped to the number one and all one-membered concepts get mapped to the number two. Since the relevant conception of functions is extensional, the function's nature is entirely exhausted by their arguments and values. Thus, the permuted function is distinct from the original function purely on the basis that it takes one-membered things to the number two while the original function maps them to the number one. Furthermore, it is obvious that we could permute these objects infinitely many times to produce infinitely many distinct functions. This gives us the result that if there is one number function, there are infinitely many number functions.

What all of these functions have in common is that they are *number* functions. They all map concepts to the same object when and only when those concepts stand in an equivalence relation. Notice, however, that this is all Hume's Principle tells us about the function. This is just another way of expressing the higher-order concept defined by Hume's Principle. Hume's Principle is unable to discriminate between the functions since it picks them out using a higher-order property they all have in common. Hence Hume's Principle does not refer uniquely to any one number function. This would be the case even if we were to establish the Ramsey sentence as true.

This enhances the worry we had regarding the existence and uniqueness of the objects, for the number functions are the mediators between the higher-order concept and the objects. If there is no number function picked out uniquely and if there is doubt that Hume's Principle by itself can secure that there are any functions satisfying the higher-order concept (without illegitimate appeal to an antecedent domain of functions), then this make it even more difficult to even suggest that reference to the objects can be se-

[7]Thanks to Michael Potter for this permutation argument and for pointing out where it might be useful. A closely related permutation argument is suggested by (Rosen 2003, 230-32).

cured determinately by way of the concept; the supposed 'referential route' given by abstraction.

Therefore, not only can Frege's dilemma be transplanted to undermine Hale and Wright's recovery of numbers using singular terms, but Frege's other objections to Hilbert can be adapted to undermine Hale and Wright's only alternative proposal; to recover numbers via a predicate.

5.3 Way out for Hale and Wright, ontological revision

We will now explore what I take to be the only potential route for Hale and Wright to avoid Frege's two-fold objections and still recover the numbers as logical objects.

As the first half of this thesis has made us well aware, Frege originally raises his worries regarding existence and uniqueness with Hilbert. Towards the end of chapter 3, we saw that Hilbert avoided these worries because of his structuralist conception. A natural suggestion, therefore, is that in order to avoid Frege's pressing objections Hale and Wright could adopt a similar conception. However, I want to stress right away that whether or not this approach proves successful, it amounts to *abandoning a Fregean conception of objects in favour of a conception of objects as Hilbertian basic elements*.

Let us briefly remind ourselves of Hilbert's conception before considering how it could help Hale and Wright avoid the objection at hand. We articulated Hilbert's conception of a basic element on the basis of passages such as the following:

> It is surely obvious that every theory is only a scaffolding or schema of concepts together with their necessary relations to one another, and that the basic elements can be thought of in any way one likes. If in speaking of my points I think of some system of things, e.g. the system: love, law, chimney-sweep... and then assume all my axioms as relations between these things, then my propositions, e.g. Pythagoras' theorem, are also valid for these things (Hilbert, 1899/1980*d*, 40-41).

These Hilbertian basic elements were distinguished from Fregean object in the following way:

df. A *Hilbertian basic element* is an object which – with respect to some axioms – has all and only those properties that can be derived as logical consequences of its satisfaction of those axioms.

df. A *Fregean object* is an object which – with respect to some axioms – can have more properties than those that can be derived as logical consequences of its satisfaction of those axioms.

Hilbertian basic elements contrast with Fregean objects by the fact that they depend for their existence on the concept – or structure – under which they fall. Whereas Fregean objects exist independently of the concepts that they fall under. Furthermore, Hilbertian basic elements are characterised exhaustively by the structure they depend upon, whereas Fregean objects are capable of having different modes of presentation. We employed Linnebo's characterisation of non-eliminativist structuralism to categorise these features of basic elements:

i. **The Incompleteness Claim:** Mathematical objects are incomplete in the sense that they have no "internal nature" and no non-structural properties.

ii. **The Dependence Claim:** Mathematical objects from one structure are dependent on each other and on the structure to which they belong.

We saw where Hilbert most explicitly denies that geometric objects are Fregean objects:

> I do not want to assume anything as known in advance; I regard my explanations in sect. 1 as the definition of the concepts point, line, plane – if one adds again all the axioms of group I to V as characteristic marks. If one is looking

> for other definitions of a 'point', e.g., through paraphrase in terms of extensionless, etc., then I must indeed oppose such attempts in the most decisive way; *one is looking for something one can never find because there is nothing there*; and everything gets lost and becomes vague and tangled and degenerates into a games of hide-and-seek (Hilbert, 1899/1980*e*, 39, emphasis added).

Finally, we saw that Hilbert's conception of basic elements avoids Frege's powerful two-fold objections concerning existence and uniqueness. Towards the latter, we first recognised that Hilbert is providing a reconstruction of the geometric objects point, line, etc., *as basic elements*. With respect to the basic elements, if we attribute (i) to Hilbert then he can avoid Frege's uniqueness objection in virtue of the fact that the basic elements have no-non structural properties. It follows from this that the axiomatisation uniquely defines a concept and as a result uniquely picks out a system of basic elements falling under that concept. In short, the permutation argument cannot get off the ground. We then recognised that if we attribute (ii) to Hilbert then he avoids Frege's existence objection in virtue of the fact that the existence of the basic elements is given by the existence of the structure of concepts to which it belongs. This structure, is in turn, is secured by the consistency of the axioms which define it.

All in all, we characterised Hilbert as an eliminativist with respect to Fregean objects and a non-eliminativist with respect to basic elements. Hilbert's structuralist position is nicely captured by Bernay's following remark:

> Indeed, one might say that there is no clear difference at all between a direct entity and a system of laws to which it is subject, since a number of laws present themselves by means of formal developments which, on their part, possess the character of the direct entity (Bernays 1950/2002, 23).

There is a good argument to be made for interpreting Hilbert as an eliminativist about both Fregean objects and Hilbertian basic elements. This would avoid Frege's objections in an even more direct way. In short, this view would hold that the basic elements *do not exist* and from this it follows that they do not have to be characterised uniquely (or at all). The reason we were interested in exploring Hilbert's non-eliminativist interpretation is that it concerns a species of structuralism which avoids Frege's worries but at the same time *attempts to recover a system of objects*. It is now apparent why this species of structuralism is of interest to our general purpose. Hale and Wright must provide a reconstruction of objects of some sort – lest they deliver an ontology with only Fregean concepts and functions. Such an ontology would be at odds with the logicist conjecture that numbers are logical *objects*.

Let us now consider how Hilbert's non-eliminativist structuralism could aid Hale and Wright in avoiding Frege's two-fold objection. Hale and Wright would first have to adopt (i) the incompleteness claim and (ii) the dependence claim, with respect to mathematical objects. Perhaps less obviously, they would also have to adopt (ii) with respect to functions, as we shall see.

This amendment would block the permutation argument, because there would only be one number function. In virtue of the incompleteness of mathematical objects, the objects in the range of the number function could not be permuted to produce distinct functions. Functions would still be extensional but if we attempted to permute the first basic element with the second basic element then the first basic element would become the second basic element and the second would become the first. This would happen because you cannot swap around positions in a structure in the same way you can swap around Fregean objects. We are only able carry out a permutation of some objects if we have an independent means of keeping track of those objects. This requires there to be a means of specifying those objects that is independent of abstraction, which is precisely what is denied on the Hilbertian conception. In virtue of (i) the incompleteness claim, the nature of the basic elements are exhausted by their being in the position they are in. As a result, the infinitely many number functions which we produced using Frege's conception of an object all collapse into one function which has in its range a position for all the numbers as basic elements. In this way, Hume's Principle is able to pick out a unique function. Or, to be more precise, it can pick out a unique function if functions are conceived of extensionally in the sense that the are exhausted by their arguments and values.

If Hume's Principle can pick out a unique function in this way and if the values of the function are the nodes of the structure, then it seems hopeful that defining a higher-order concept could enable us to uniquely identify a function that has numbers as its values – so long as these numbers were Hilbertian basic elements. The question that remains is whether we can secure the existence of these basic elements. That is, even with the function secured, how are we to be sure that there is anything which falls under the structure?

In virtue of (ii), the existence of the basic elements depends upon the existence of the number function which ranges over them. In turn, the existence of this function depends upon the existence of the higher-order concept defined by Hume's Principle. To prevent an upward regress of dependence we need a means of establishing existence at *some* level. To achieve this, we can again turn to Hilbert for whom the existence worry is resolved by appeal to the consistency of the axioms which define the structure. In the same way, Hale and Wright could employ Hilbert's Principle (with respect to Hume's Principle) and prove as an immediate consequence of the consistency of Hume's Principle, the existence of the higher-order concept it defines. From the existence of the higher-order we could obtain the existence of the number function and the numbers by invoking (ii).

Hale and Wright are already alive to the importance of establishing the consistency of Hume's Principle. Towards this end, impressive progress has been made; it has already been proven that Hume's Principle is equi-consistent with second order arithmetic (Wright 1983, 273 ; Boolos 1998, 296).Given our credence in the consistency of second-order arithmetic, this certainly seems to put the consistency of Hume's Principle beyond any reasonable doubt. However, Hale and Wright settle for the weaker conditional thesis that Hume's Principle can recover the numbers as logical objects under the highly probable assumption that it is consistent – along with a host of other criteria like harmony, generality and conservativeness (Hale & Wright 2001, 137-8).

All in all, this would give the following picture: the likely consistency of Hume's Principle, would put beyond doubt the existence of the higher-order concept it defines by application of Hilbert's Principle. Then, from (ii) with respect to functions, we could infer the existence of a number function since *its* existence depends entirely on the existence of the higher-order concept it satisfies. Similarly, from (ii) with respect to basic elements, we could infer the existence of numbers as basic elements, since their existence depends entirely on the structure of concepts which they satisfy. In this picture, the existence of the concept becomes more important than the existence of the objects. If we have secured the higher-order concept by means of the consistency of Hume's Principle, then we can extract from it the structuralist objects we require. The objects are extracted as *those objects which fall in the range of the function which falls under the concept*.

This only constitutes the bare bones of a proposal. There are many things which still need clarified. For instance, it is unclear whether the relative consistency result would be enough to employ Hilbert's Principle. Furthermore, it is likely Hale and Wright would be reluctant to accept Hilbert's Principle in the first place. However, this provides us with a sketch of a way out for Hale and Wright whereby both of Frege's charges could potentially be avoided. For if they adopted (i) and (ii) and if their new logicist proposal could be made good, then this would mean that the prospect that the higher-order concept defined by abstraction provided a novel referential route to some abstractionist objects would be considerably strengthened.

At this point we can reflect on what has been conceded in order to avoid Frege's objections. It certainly seems extremely revisionary and unexpected that the objections force Hale and Wright to adopt a structuralist ontology in order to maintain any semblance of a viable logicism. After all, this kind of Hilbertian logicism is entirely distinct from the logicism which Hale and Wright proclaim to defend. The Fregean conception of a number as an object is a central Fregean tenet; in particular the conception of a number as a kind of object that admits of different modes of presentation which are independent of the object's characterisation or use in a theory. For Hale and Wright to abandon this commitment is for them to abandon a project that is recognisable as any kind of neo-*Fregean* logicism. In the revised project numbers are not determinate objects – *the* numbers in platonic heaven – but basic elements falling under (the function that satisfies) a concept defined by the process of abstraction.

The drastic revision of this proposal seems reason enough to steer clear of any suggestion that Hale and Wright could adopt a Hilbertian non-eliminativist structuralism as a way of avoiding Frege's objections. However, in the next chapter what I will suggest is that Hale and Wright have implicitly adopted this starkly unFregean ontology already. That is to say that they have already forfeited the central element of their project that characterises their logicism as distinctively Fregean rather than Hilbertian or Russellian, etc. This ontological concession has not been noticed or admitted but it has been adopted in order for Hale and Wright to strengthen their account against the many and difficult objections which they have faced. The aim of the last chapter will not be to question the viability of the breed of logicism they have adopted; I leave it as an open question whether some such proposal can be made to work. Rather, the last chapter will merely aim to expose the extent of Hale and Wright's heretical abandonment of a Fregean ontology in favour of a non-eliminativist structuralist logicism.

Conclusion

Having charitably articulated a strategy on behalf of Hale and Wright from the basis of their suggestions about a conditional introduction of numbers, we can see that this Plan B will be of no help to the neo-Fregean logicists. This is because we only have to apply Frege's other objections to Hilbert to see that the problematic aspect of the account remains. Hale and Wright's abstraction principles do not recapture the numbers as objects. Since we have given up one of the two commitments which problematized the account in the last chapter, and the problem still remains, it seems likely that it is being generated from the commitment we have not given up: that Hume's Principle is an *implicit definition*. We saw that the best way to avoid the problem seems to be to follow Hilbert in adopting a kind of non-eliminativist mathematical structuralism, but that this means giving up on a reconstruction of Fregean objects. This substantiates what it means to say that our conception of foundational sentences seem to constrict, or indeed dictate, our conception of the ontology which we use them to capture.

Chapter 6

Hale and Wright's ontological concession

Introduction

In the previous chapter we saw that whether Hale and Wright attempt to recover numbers via a singular term, or a predicate, their deep rooted commitment that abstraction principles are implicit definitions seems to affect the ontology they secure by means of abstraction. Indeed we suggested that the only way they seem able to escape from Frege's objections to Hilbert regarding implicit definitions was to follow Hilbert in his rejection of Fregean ontology. Such a move, we noticed, would be extremely revisionary.

In fact, Hale and Wright have made this move already. We can see this in particular from the way in which they defend their account against their opponents. To show this the first section will extract Hale and Wright's commitment to an unFregean ontology in what I take to be the most simple way: their assumption that there are no available independent means of specification of abstractionist objects. The remaining sections will then consider three important objections: the Julius Caesar problem, the MP problem, and the need for a abstractionist meta-ontology. What we will see is that to defend themselves against these objections Hale and Wright essentially take the same approach in all three cases: they deny that abstraction is even intended to provide a logical reconstruction of Fregean objects.

6.1 An independent means of specification

We have already noted that in their later work Hale and Wright insist they must be granted the assumption that abstractionist objects have "no independent means of specification" (2008, 204). I understand them to mean that the only mode of presentation of these objects – the only way of securing reference to them – is to posit them as *those objects which fall under the concept defined by the process of abstraction*. This is to say that under this assumption abstractionist objects can only be grasped by first grasping a concept and then grasping what kind of objects must fall within the range of that concept.

Of course, this does not mean that we have no means of referring to numbers apart from abstraction principles, for the neo-Fregean Logicist's account is not concerned with whether we do *in fact* become acquainted with numbers by means of their abstraction principle or whether we *should* learn about numbers in this way. Rather, it is concerned with providing a *logical reconstruction* of how numbers might be given to us so as to establish their logicality (MacBride 2003). For the numbers to count as logical objects – in other words – Hale and Wright do not need to show that every way of being acquainted with them is logical but only that there *is* at least one potential logical route to doing so even if no one has taken this route, or ever will. And in order to showcase this logical route we must suspend all the other routes we are familiar with for accessing the numbers. This is similar to how we may suspend other kinds of archaeological information like art and artefacts to try and show that the physical features of prehistoric people can be known via the route of their skeleton alone.

I think that we can identify the loss of Fregean objects solely from the nature of this single assumption of Hale and Wright. To examine the assumption in more detail let us look again – this time in full – at one of the passages in which they describe this assumption in most detail:

> Abstraction principles purport to introduce fundamental means of reference to a range of objects, to which there is accordingly no presumption that we have any prior or independent means of reference. Our conception of the epistemological issues such principles raise, and our approach to those issues, need to be fashioned *by the assumption that we may have – indeed there may be possible – no prior, independent way of conceiving of the objects in question other than as the values of the relevant function.* So when Boolos asks, what reason do we have to think that there is any function of the kind an abstraction principle calls for, it is to skew the issues to think of the question as requiring to be addressed by the adduction of some kind of evidence for the existence of a function with the right properties that takes elements from the field of the abstractive relation as arguments *and objects of some independently available and conceptualisable kind as values*. If the best we can do, in order to assure ourselves of the existence of a relevant function or, relatedly, of the existence of a suitable range of objects to constitute its values, is to appeal to our independent ontological preconceptions – our ideas about the kinds of things we take to exist in any case – then our answer provides a kind of assurance which is both insufficient and unnecessary to address the germane concerns: insufficient, since independent ontological assurance precisely sheds no light on the real issue – viz. how we can have reason to believe in the existence of the function purportedly defined by an abstraction principle, and accordingly of the objects that constitute its range of values, when proper room is left for the abstraction to be fundamental and innovative; unnecessary since, *if an abstraction can succeed when taken as fundamental and innovative, it doesn't need corroboration by an independent ontology* (Hale & Wright 2009, 92, emphasis mine).

By now the picture is familiar: what the neo-Fregean logicists are aiming to provide via abstraction is a referential route to objects which are in the domain. We saw them argue that the domain can be set out and the quantifiers ranging over it can be grasped, without a grasp of these objects (therefore, they argue, impredicativity is harmless). However, even if we assume in their favour that quantifying over some objects does not presuppose that we can referentially individuate some objects, and also that abstractionism can provide a referential route to some objects (although the last chapter has argued that for this to be plausible they require a structuralist ontology), this still leaves the question of why the referential route given by abstraction cannot be counted as one amongst many. It makes sense for them to argue that abstraction affords one such route, but is this the only possible route?

Before we answer this, let us ask another more basic question: what exactly is a referential route? Most simply, it is a way of referring to an object, so that we can go from a grasp of the quantifier in 'all men' to being equipped with a way to refer to the central African chief. What way is this? On Hale and Wright's picture we present the chief as the value of a function (or concept). The important thing to note is that this amounts to *a mode of presentation* of the chief. Asking why there is only one route to an object is to ask – in part – why the *same* object could not be the value of *different* concepts. For, if the object could be the value of different functions then this would show up in the fact that it could be presented or grasped via a different referential route. Thus, this question is equivalent to asking why the objects recovered by abstraction *do not admit of different modes of presentation*.

When we understand the assumption in this light we see that precisely what Hale and Wright are asserting in the passage above (especially in the first two emphasised parts) is that the objects they are recovering have only one possible mode of presentation. This amounts to a denial that they are even aiming to recovering Fregean objects. Hale and Wright can surely be granted this request and may go on to offer an account which reconstructs these other kinds of objects, but they must be challenged on the claim that their conception and approach is fashioned by the aim to recover a nonFregean ontology of things.

Furthermore, to deny that there is a mode of presentation of that object *independent* of the abstraction principle is to make the properties and nature of that object dependent on the abstraction, which links back to the incompleteness claim (ii). Indeed, the last emphasised remark sounds very much like Hilbert's denunciation of Fregean objects and the search for independent definitions or characterisations of geometrical objects as a misguided game of hide and seek.

In this way we can uncover Hale and Wright's commitment to a Hilbertian (or at least unFregean) ontology merely from their request to be granted the assumption that *the objects recovered by abstraction admit of no means of specification independent of the abstraction principle*.

Hale and Wright are well known for being happy to revise their positions in light of criticism. Perhaps they can revise what they have said here in light of this objection. However, I very much doubt that they can avoid their commitment to the single referential route of abstractionist objects. As we will see, it is an assumption which appears in many different parts of the debate. Furthermore, I think the implicit commitment to a Hilbertian ontology runs deep and is not one Hale and Wright can easily avoid. I will show this by laying out in detail three disparate parts of the neo-Fregean logicists' account which have been subjected to criticism. Our concern will be to set out how Hale and Wright have defended their account from these attacks. In particular, we shall look at their solution to the Julius Caesar problem; their reply to the MP objection; and their insistence that minimalism (their theory of reference) replaces any need for a meta-ontology.

I take these to be three of the most potentially damaging systemic objections to Hale and Wright.[1] I will show that Hale and Wright rely on a Hilbertian ontology to defend themselves against these objections. As a result, a revision of this ontology is not an option for them lest they incur the fatal consequences of each objection. We will also see that when assume a Fregean ontology, Hale and Wright's account is once again vulnerable to the original force of the three objections.

[1] I think that the embarrassment of riches objection, most forcefully presented in Weir (2003) is another systemic objection of this sort of importance; but its connection to the background ontological commitments of neo-Fregean logicism is more subtle and difficult to characterise and anyway the three mentioned objections will be more than enough to make the point.

6.2 Sortal inclusion principle

In this section, we will present Hale and Wright's complicated solution to the Julius Caesar problem. We will then consider its ontological costs in the way we have just outlined.

6.2.1 Sortals and categories: Exposition of the distinctions appealed to by Hale and Wright

Sortal concepts are those expressed by count nouns, i.e. those Fs for which the question 'How many Fs are there?' has a determinate answer. Hale and Wright delineate four distinct subgroups from this broad category: pure sortals, impure sortals, phase sortals and functional concepts. A pure sortal is such that "it belongs to the essence of the object to be a thing of that kind" such as with person, tree or city (Hale & Wright 2001, 387). Impure sortals are concepts which restrict a sortal concept by detailing an inessential characteristic, such as with tiger with a thorn in its paw, or boy who cried wolf. Phase sortals are concepts which characterise essential features of an object but for which an object can survive transition from one phase to another such as with caterpillar, concubine, child. Functional concepts are concepts which occupy a certain role where it is not essential to the intrinsic nature of the object that it occupy this role: nonetheless, the nature of the object indirectly constrains the object's suitability to inhabit this role such as with lunch, doorstop, examiner.

As we will see, the most important of these sub-categories for the neo-Fregean logicists are pure sortals. This is true especially with respect to contrasting sortals with functional concepts by their claim that pure sortals are more 'epistemically autonomous' in the sense that, for a pure sortal F 'x is the same F as y' will sometimes have a determinate truth-value which can be known purely by grasp of the sortal F and without further knowledge of what kind of thing x and y are (Hale & Wright 2001, 386-88).

A pure sortal concept is associated with both a criterion of application and a criterion of identity: to understand what it is to be an F – for a pure sortal F – one must grasp whether a given thing is an F and whether any given Fs are the same or distinct. Two distinct pure sortals can have the same criterion of identity, for example tiger and cat. The most maximally general pure sortal which includes all the objects associated under a particular identity criterion (perhaps animal in this case) is a *category*; this the neo-logicist takes from Dummett (1981, 76).

Later, much will turn on how the application and identity criteria relate to one another. Dummett argued that a criterion of identity cannot be derived from the criterion of application (1981, 73–5) but the opposite direction is of more significant to the neo-Fregean logicists, i.e. extracting a criterion of application from an identity criterion (given by an abstraction principle). As Kim (2014) has noted, this seems most plausible in the case of categories, since an identity criterion will be uniquely associated with a category F so that it seems natural to characterise what it is for a given thing to be an F as something whose identity can be determined by the particular identity criterion (Kim 2014, 408).

The neo-logicists identify the background ontology that is assumed by this stratification most explicitly and lucidly here:

> It is the outline of a world in which all objects belong to one or another of a smallish range of very general categories, each of these subdividing into its own respective more or less general pure sorts; and in which all objects have an essential nature given by the most general pure sort to which they belong. Within a category, all distinctions between objects are accountable by reference to the criterion of identity distinctive of it, while across categories, objects are distinguishable by just that – the fact that they belong to different categories (Hale & Wright 2001, 389).

Bearing much weight in this picture is the appeal to the sameness and distinctness of criteria of identity. The membership of various sub-sortals to a single category depends on their association with the same identity criterion; the most general non-overlapping categories are distinguished from each other by being associated with distinct identity criteria. Further – as is brought out clearly above – the specification of the identity criteria is needed to do the work of providing the essential nature of the objects which fall under it, which is no easy task. To this issue we shall return shortly.

6.2.2 What is the Sortal Inclusion Principle and how does it circumvent the Caesar problem?

Since Hume's Principle introduces numbers by making identity statements of the form 'the number of Fs = the number of Gs' meaningful, it does not seem apt to provide a criterion of application. Indeed we can even doubt that it provides us with a criterion of identity for numbers. This is called into question by Frege's own objection to using Hume's Principle to introduce numbers; the Julius Caesar problem. Frege objects that the statements which are made meaningful by Hume's Principle do not tell us what sorts of objects numbers are: for all it tells us numbers might be in charge of Roman legions. Frege's objection is made in passing and his contention is disputed but we can say generally that Frege seems to consider there to be some kind of explanatory inadequacy in the fact that Hume's Principle does not provide the means of determining whether the number of Fs = q, except in the single case where q is in the form 'the number of Gs'.

The neo-Fregean logicist's most extended attempt to circumvent this charge has been to supplement Hume's Principle with a further principle which they call the sortal inclusion principle and which they take to do the work of distinguishing numbers from Roman generals:

> **Sortal Inclusion Principle** (SI#): Some Fs are G only if F and G are each sub-sortals of one and the same category.[2]

(SI#) is intended to give us a way of deciding the truth-value of some identity statements (namely, what Hale and Wright call mixed identity statements). It follows from

[2] Hale & Wright (2001, 396)

the principle that if we know two objects fall under distinct categories then this is enough to know that an identity statement involving them will be false. Here appears the epistemic autonomy which we noted was used to contrast sortal with functional concepts: we need know nothing further regarding the nature of these objects to know an identity statement involving them to be determinately false. Importantly we see that the contrast of epistemic autonomy is properly between functional concepts and categories rather than between functional concepts and all sortal concepts. This is because if objects a and b belong to distinct sortals A and B respectively, this does not preclude the objects from being identical, as with keychain and bottle-opener. In this case the distinctness or identity of the relevant objects will be a question further to and undecided by (SI#). This situation cannot arise with categories which are the most general sortals associated with a particular criterion of identity; we can conclude from the fact that objects fall into different categories that they do not fall under any sortals in common and, most importantly, that they have no identity criterion in common. Thus the statement that they are the same thing must always be false.

How, then, is (SI#) used to give a solution to the Caesar problem? The Caesar problem is that Hume's Principle does not provide a means of determining whether Julius Caesar is a number; for instance, whether he is the number of moons of Mars. The identity condition in question is therefore one between objects which fall into two distinct sortals; Julius Caesar uncontroversially under *people* and the number of moons under *number*. The principle tells us that the objects falling under one sortal can only be identical to the objects falling under another sortal if both sortals are subordinate to the same category. As we saw, sortals are subordinate to a general sortal or category when (necessarily but not sufficiently) the sortals have the same identity conditions as the category. Here we see how the principle depends quite essentially on the identification and distinction of criteria of identity.

Hale and Wright claim that the sortals *number* and *people* must be associated with different identity conditions. This is because Hume's Principle derives the identity criterion for numbers by way of an appeal to equivalence relations between concepts and a criterion for personal identity can never be determined by appeal to equivalence relations between concepts. This claim is made even more intuitive if we recall that we are dealing with pure sortals which are those which characterise the essence of an object. The thought is that the intrinsic properties of personhood can never be established by an appeal to concepts and relations between them. Thus, they must belong to distinct categories and cross-categorical identity statements involving them must be false, or as problematically indeterminate as cross-categorical statements in general (Hale & Wright 2001, 394-96).

6.2.3 The need for a criterion for criteria of identity

The crux of the maneuver past Caesar, then, is Hale and Wright's claim that the sortal *number* and the sortal *person* have distinct identity conditions and thus are not subordinate to the same category. Once they have established this, they can appeal to the sortal inclusion principle to establish a mixed identity statement involving numbers to be false. Let us take some time to unpack the substance of this claim.

Returning to the issue of a lack of criterion of identity for criteria of identity, it is not obviously the case that the neo-Fregean logicists need an explicit principle to do this work. To consider this let us reconstruct the substantive parts of their view regarding distinguishing identity criteria: identity criteria are established in general by constructing a biconditional sentence from an identity together with a sentence which is apt to make the identity sentence meaningful. If the identity is successfully imbued with meaning then the resulting sentence involving the biconditional will express the criterion of identity for the relevant objects: that is, it will express a way to tell when two given objects are the same. This is an abstract characterisation of the process which the neo-Fregean logicists present as *definition by abstraction*. More concretely, in the case of giving a criterion of identity for numbers, those most interesting of logical objects, Frege uses exactly this method considering whether the bi-conditional 'the number of Fs is equal to the number of Gs if and only if the Fs and the Gs are equinumerous' can adequately provide the concept of number. Here the sentence on the right-hand side is intended to establish a meaning for the identity sentence on the left-hand side and as such introduce us to a criterion of identity for numbers. For the neo-Fregean logicists, identity criteria will therefore be distinct when they are established by way of distinct abstraction principles.

This comes some way, but so far only shifts the explanatory gap onto abstraction principles; for how are we now to distinguish different abstraction principles? The neo-Fregean logicists, as we will see, do not need an explicit criterion for the distinction of abstraction principles but rather the mere claim that, at the very least, we can be assured that abstraction principles are distinct when the sentences on their right-hand side refer to different things; or – more precisely – when the content of their *definiens* is distinct. Here again we can grant Hale and Wright that they need not provide a whole semantic theory to do this work but instead allow their general appeal to non-borderline cases. In the relevant case, for instance, they make appeal to the fact that sentences involving reference to concepts have distinct content from sentences involving reference to people, or indeed any objects. More particularly, this claim is based on the background premise that abstract objects are individuated differently from concrete objects such that reference to concepts can never serve to provide a means of distinguishing concrete objects but can individuate certain abstract objects.

This assumption seems innocuous enough when isolated, but it cannot by itself do the work of identifying all mixed identity statement. That is to say that along with the sortal inclusion principle it will deliver a means of determining whether the number of Fs $= q$, but only in the cases where q is a sentence which refers to a concrete object. If q is a sentence which refers to an abstract object then the sortal inclusion principle cannot be of any help for deciding whether appeal to equivalence relations between concepts could establish its identity criterion. Without supplementation of their account, therefore, the neo-Fregean logicists will re-encounter the Caesar problem in the abstract realm: is the number of authors of *Principia* identical to the letter A? This depends upon whether abstract objects falling under the sortal *letter* could be distinguished by appeal to equivalence relations between concepts. Since we have antecedent access to the concept of a letter this seems unlikely; but we have no determinate way of settling this question. Where this becomes problematic is of course in cases of objects that are not numbers but which plausibly might be distinguished by appeal to equivalence relations between concepts: is the number of authors of *Principia* identical to the set which has Russell and Whitehead as its members?

The real substance of Hale and Wright's solution – then – is that we can tell enough about numbers from the fact that their identity conditions are established by appeal to equivalence relations between concepts to be able to distinguish them from certain other things. In particular, and in increasing generality, we can distinguish them from: Caesar, people, concrete objects and any objects whose identity criterion– whatever they turn out to be – could not successfully be established by appeal to these means. Importantly, we can tell that numbers are distinct from such other types of objects without further appeal to what sort of thing a number is. This satisfies the epistemic autonomy of numbers. However, without a substantiated account of what qualifies as a successful appeal to these means, Caesar looks in danger of raising his head, albeit as an *abstracta*.

6.2.4 The sortal ontology and Frege's ontology

Here we move to considering the status of the sortal inclusion principle itself and more particularly, the ontological assumptions that it needs in place in order to function as a solution to the Caesar problem by the neo-Fregean logicists' own lights.

The best way to bring out the worrying assumption which is presupposed by (SI#) is that, as formulated, it has only sortal concepts in its domain. This is to say that in order for two concepts F and G to have the sortal inclusion principle applied to them they must first be recognised as sortal concepts. Recall in §6.2.1 we explained that sortal concepts are those which have a criterion of identity and a criterion of application associated with them. Further, pure sortal concepts require that it is of the essence of a thing that it falls under the relevant concept. The question we shall carefully consider is whether the neo-Fregean logicists are entitled to the assumption that there is a sortal concept for the sortal inclusion principle to operate on (in the hope of producing a solution to the Julius Caesar problem).

The sortal inclusion principle must be envisaged by Hale and Wright to work together with Hume's Principle in the following way: Hume's Principle works to introduce a sortal concept by implicit definition and the sortal inclusion principle supplements the definition by exploiting the given sortal concept. Thus we can isolate the assumption to be that Hume's Principle provides a sortal concept. There are two questions about the general picture here: first, that of a circularity depending on how much we must assume regarding the concept supposedly given by Hume's Principle, secondly what grounds acceptance of (SI#) as a supplementary principle. Let us consider now each in turn.

Certainly we cannot – without circularity – assume that Hume's Principle introduces us to the concept of number. But can we assume that it introduces us to a sortal concept in order to solve the Julius Caesar objection when the Julius Caesar objection seemed to challenge the adequacy of Hume's Principle for delivering the concept of number? This will depend, of course, on the nature of the charge of inadequacy here. The circularity worry raises its head only if it is part of what is in thrown into doubt by the Caesar problem that the concept given by Hume's Principle is sortal.

In what sense might we doubt that Hume's Principle does not provide a sortal concept? Perhaps in the straightforward sense that Hume's Principle fails to give a concept at all. Perhaps because it fails to achieve what it is formulated to do; establish the identity criterion for numbers. Perhaps because it does not provide the application criterion for numbers. Perhaps because it establishes a concept for which it is not essential to the object's nature that they fall under it, thus Hume's Principle might provide a phase or functional concept. Relatedly, we might have reservations that the concept of number which we set out to recapture by way of logicism can be antecedently assumed to be a pure sortal concept and not an impure sortal or functional concept.

The circularity, however, will only pertain to these doubts if they are raised by the Caesar problem. The worry that Hume's Principle does not successfully establish the identity criterion for numbers by way of the identity sentence on the left-hand side can plausibly be reconstructed as part of Frege's concern that definitions provide total functions; but it is not one that can be much pressed on the neo-Fregean logicists pending a more sophisticated criterion of identity for identity criteria to hold them to; or even a fuller account of the success conditions for establishing identity criterion, which they leave pending. The general understanding of the Caesar problem fits best with the accusation that Hume's Principle does not successfully establish the application conditions for numbers since it does not provide a generalised understanding of what it is for a given thing to be a number, except in the case where the object is presented to us in the form 'the number of *G*s'. There is a further interpretation of the Caesar problem by which we might understand it as challenging the view that Hume's Principle provides a concept which gives the essential nature of the numbers: for, if by the lights of Hume's Principle Julius Caesar could be a number, then we must have failed to give a characterisation of number which captures the essential properties of numbers – which amongst others, plausibly include the fact that numbers are abstract objects. The point might be put as follows: a functional concept will also be associated with a criterion of identity and application but unlike a pure sortal these criteria will not characterise the object by its essential features – which is to say that it will not be of the nature of the object that it must fall under that concept. Many different objects can be doorstops or lunches. In this sense, there is nothing to guarantee that Hume's Principle delivers a sortal concept rather than a functional one.

In fact, Hale and Wright explicitly distance themselves from this particular circularity worry in a reply they make to Rosen. In doing so they bring out more clearly the assumption which the sortal inclusion principle is premised upon. Rosen (2003) avoids engaging with the later stronger arguments in Hale and Wright's paper for the reason that each "assumes as a premise that 'Number' picks out a pure sortal" whereas he sees this as the claim that the neo-Fregean logicists must establish to defend their account (Rosen 2003, 235). Hale and Wright concede that if they had made the straightforward assumption that 'Number' succeeds in picking out a sortal concept then their reply would be impotent not only against Rosen's contention but against its very purpose, to rebut the Caesar problem. However, they deny that they rely on such a circular assumption and distinguish between the assumption that number is a sortal concept and the assumption that Hume's Principle is to be understood as offering an explanation of a sortal concept:

> What is certainly true is that the sortal exclusion argument presupposes that Hume's Principle is put forward as explaining or implicitly defining a (pure) sortal concept, rather than a 'role' or 'office' sortal comparable with *doorstop*, *bookmark* and the like (Hale & Wright 2001, 255).

They go on to claim they are able to make this assumption because whether the principle succeeds in providing this explanation is a further question. Thus, in order to avoid circularity Hale and Wright build it into the very success conditions of concept acquisition that Hume's Principle not only satisfies a host of constraints such as harmony, generality, conservativeness, etc. but also delivers a pure sortal concept – as distinct from a phase, impure or functional concept. Therefore, the only assumption that (SI#) is premised on is that Hume's Principle is plausibly a candidate for successfully defining a sortal concept. To this point we shall return shortly.

The second species of worry is about what can ground our acceptance of the sortal inclusion principle, as a supplementation of Hume's Principle. Hale and Wright have in places addressed the delicate issue of the status of the principle. Hale is the more explicit and insists that (SI#) is not merely the stipulation that *objects which have not been introduced as numbers are not numbers*; but rather that this result follows from the laying down (SI#) along with the certain features inherent in Hume's Principle (Hale 2001, 198).

If (SI#) is not to be understood as being laid down by mere stipulation, the principle certainly needs justification. Hale and Wright's most direct provision of this justification is their claim that (SI#) is grounded in "the idea that a criterion of application for the sortal concept number is extractable from its associated criterion of identity" (Hale &

Wright 2001, 369). The thesis that a criterion of application can be extracted in general from the criterion of identity as we said before has plausibility only when the concept in question is a category and not a mere sub-sortal. If the concept was a sub-sortal then merely from knowing its criterion of identity we would not be able to tell which sub-sortal it was. This is because sortals can share criterion of identity (as do all the sortals subsumed in one category) and are distinct if they differ in their criterion of application – as, for example, with cat and tiger.

The thesis underlying the sortal inclusion principle – then – is that since the essential nature of the objects is given by the category to which they belong, then from the identity conditions for the objects in the category we can extract the application conditions for the objects. That is, we can tell whether a given category F applies to a concept by asking whether its identity can be determined in a certain way. In the relevant case, Hume's Principle gives us the identity criterion for numbers by way of an equivalence relation holding between concepts. Thus, if number is a category, then what it is to be a number is extractable from this criterion in the sense that numbers are essentially those things whose identity can be determined by appeal to equivalence relations between concepts.

This brings out that the neo-Fregean logicists not only need the assumption that Hume's Principle is a plausible candidate for providing a sortal concept, but for defining a special species of sortal concept, namely a category.

Let us take a step back. It should be clear by now that to say that abstraction principles must define a category is just to say that they must define a concept which is such that the objects falling under that concept are exhaustively characterised by this concept. This is what it really means to say that we can extract its criterion of application from its criterion of identity: that what these objects are, is given entirely by the fact that they fall under this concept. As we have seen, the concept most plausibly defined by Hume's Principle is a higher-order concept which has *functions* and not *objects* falling under it. In this case, we can say that the nature of the objects is exhaustively characterised by their being presented as the value of a function which falls under the concept defined by the abstraction principle – otherwise the abstraction principle is not successful. That is the sense in which it is part of the success conditions of abstraction that an abstraction principle define a category.

What does this tell us about the nature of the objects if the objects are such that they are exhaustively characterised in this way? It is clear how this picture is compatible with Hilbertian basic elements since these objects are always exhausted by the concepts or structures under which they fall, so that we can count any well-defined concept as a category. Categories will be many and easy to find; any consistent well-defined set of axioms will give us a category. Could this success condition provide Fregean objects which were exhaustively characterised by their falling under some concept not because they were dependent on the concept but because of an ontological coincidence? The answer must be no because of the fact that there can be no legitimate antecedent appeal to the objects external to the category via an independent mode of specification, once we understand what a category is. This coincidence makes sense for Fregean objects because they can always admit of different modes of presentation. But it makes no sense to say that the objects falling under categories could be verified as having all and only those properties in virtue of which they satisfy the category.

This does not show exactly that the category ontology is not Fregean; it only shows that the independent modes of presentation of an object must take place *within* a category. It is when we consider how these categories are delineated that we meet with a distinctively unFregean ontology. For, even if we grant that (SI#) is not a mere stipulation, we must acknowledge that it is not an innocent principle in the sense that it requires the import of a whole ontology of stratified concepts – as articulated in the quote at the end of the last section.

Frege, himself, would have admitted only one category of objects. As we have seen, Hale and Wright have divided up this category. How can we say that their ontology is Fregean until we ensure that they admit all and only the same categories that Frege would? Unfortunately, if they were to make this concession, then they would no longer have any solution to the Caesar problem. The Julius Caesar problem occurs precisely because, for Frege, *Julius Caesar and the number one are both objects which fall under sub-*

sortals of the same category, i.e. the category *object*. If we are to split these categories and replace them with more non-overlapping categories (for example the categories; abstract object and concrete object) then we are again moving away from a Fregean ontology and multiplying categories as the Hilbert conception would have us do.

This shows us the sense in which Hale and Wright's purported solution to the Caesar problem is dependent on an unFregean ontology, in this case an ontology of categories.

6.3 MP Problem

We now come to a more straightforward, but nevertheless systemic objection, for which it is more straightforward to see the sense in which Hale and Wright defend themselves by employing an ontology which is unFregean.

The MP problem was raised (twice) by Sullivan and Potter (1997 and 2005) who suggest introducing MPs by the following abstraction principle:

> **(MP)** $a's$ Member of Parliament is the same as $b's$ Member of Parliament if and only if a lives in the same constituency as b.[3]

On the face of it, such a principle is as suitable as any other abstraction principle for introducing the concept of an MP. In this sense, this is a species of the bad company objection. What is particularly bad about the company of this abstraction principle is that, while it is *true* of MPs it does not exclude political representatives having further properties outside their job – from having birthdays and bank accounts and so on. Such an abstraction is able to introduce us to Fregean objects since MPs are not merely the objects distinguished by co-constituency relations between people but have different modes of presentations beyond those given by (MP).

Hale and Wright defend themselves by claiming that this objection 'founders' on a dilemma which they explain as the following:

[3] Potter & Sullivan (1997, 139)

> If (MP) is intended as a genuine abstraction, purportedly introductory of a sortal concept and fully comparable in all relevant respect to the Direction Equivalence, *et al.*, then it is no paradox to deny that MPs are people and that no objection to the *Frege's Conception* approach if it carries that consequence. On the other hand, if the principle is offered as analytic of the concept, Member of Parliament, as ordinarily understood, then, first it is arguably not even a truth as it stands – it needs an existential proviso to cover the case of temporarily unrepresented constituencies, etc.; and second, it in any case remains that the concept it governs is not *sortal* but – in the sense illustrated at the beginning of this section – *functional*: to be an MP is not to be a certain basic sort of object but it is rather to occupy a certain kind of functional role (and hence to be some other kind of object first and foremost). But if (MP) does not introduce such a sortal concept (in the relevant sense), then it is not properly subjected to principled constraining overlap and inclusion amongst such concepts (Hale & Wright 2001, 380).

We can extract a simplified version of this dilemma:

(MP) Dilemma: Either MPs are *abstracta* or MPs are *unsuitable for abstractionist recovery*.[4]

This dilemma is extremely telling. Furthermore, Hale and Wright conclude their reply to Potter and Sullivan with the unambiguous statement that:

> Our position remains, unrepentantly, that even in the impredicative case, the newly introduced terms may provide our sole underived means of reference to the objects in question, and that warrant to regard those terms as referring need be sought no further afield than in the abstraction principle itself, along with our ordinary and antecedently available means of appraising the truth of instances of its right-hand side (Hale & Wright 2008, 206).

[4] Hale & Wright (2008, 206)

We again find the request that the objects suitable for abstraction admit of no independent means of specification – which we have already shown to imply a Hilbertian ontology.

If the MPs are understood as *abstracta* then they have no exotic properties, such as bank accounts or holidays, because all of the properties they have follow as a logical consequence of how they have been introduced. Our intuitions are meant to push us towards the second horn of the dilemma: that people are not the kind of objects which abstraction principles are fit to introduce.

We should ask the following question here: what is to decide the issue of which objects are suitable for abstractionist recovery? It seems to be the *kind* of objects that we want to recover. In particular, it seems like any objects which have properties independent of abstraction – not just concrete objects – will land on the second horn of the dilemma. As such, it seems that all Fregean objects will land on the second horn. Indeed, the only things suitable to land on the first horn are what we have characterised as Hilbertian basic elements, since these are exhausted by their satisfaction of the principle and have no additional nature. Thus, even the formulation of this dilemma amounts to the very surprising result that *Fregean objects are not the kind of objects which are suitable to be recovered by means of abstraction principles.*

6.4 Minimalism

In this last section we look the aspect of Hale and Wright's account which is most directly relevant to the issue of their background ontology; that is, their theory of reference and the ontological conception they build into that theory.

6.4.1 The need for a meta-ontology

As we have seen, abstraction is a process in which a bi-conditional is stipulated between a statement known in advance and a statement with distinct existential import in such a way that an epistemic route is carved from one to the other. The claim that an implicit definition can have a novel ontological export in this way is what Hale and Wright have called the abstractionist's 'central *ontological* idea' (2009, 181).

Various parties have put pressure on Hale and Wright to admit appeal to a 'meta-ontology' in order to validate this central claim. The thought is that to legitimate their move we need some variety of additional metaphysical commitment to avoid merely defining numbers into existence, as Boolos (1998) worried might be the case. One such meta-ontology is maximalism. On a maximalist view, everything that can exist does exist. On a quantifier variance account, there are distinct and equally valid quantifier meanings and so the selection of different meanings for the quantifiers in abstraction principles is the best way to make sense of reconceptulisation, or, the metaphor of content re-carving.[5]

Putting in place a meta-ontological thesis – such as maximalism or quantifier variance – would independently secure the existence of the objects being introduced in such a way as to legitimate the move to the novel half of the bi-conditional. Despite the left-hand side's – not inconsequential – ontological commitments, i.e. the existence of infinitely many objects. In this way, a meta-ontology would allow for a kind of ontological innocence which would support the plausibility of the substantive epistemic export of definition by abstraction.

Hale and Wright consider maximalism and quantifier variance and conclude that neither are up to task of securing the abstractionists' central ontological idea. They argue that the maximalist account is not compatible with a logicist reconstruction of numbers since the ontology is generated from a metaphysical commitment rather than by abstraction. With respect to a quantifier variance account the alternative meanings which are posited are so obscure that the account no further develops the way in which implicit definitions can have non-circular novel ontological import.

[5]It is Katherine Hawley (2007) who argues that maximalism is the best supplementation for neo-Fregean logicism and Ted Sider (2007) who argues they must accept a quantifier variance account.

Rather than offer an alternative meta-ontology Hale and Wright then deny that their account requires any such supplementation. In order to resist adopting any meta-ontology the neo-Fregean logicists must do two things. To begin with they must argue that none of the meta-ontological views put forward are already entailed by the account they have developed. Indeed, they must show more generally that no substantive metaphysical thesis is already implicit in their account. This general aim will partly be achieved by their second task: to make explicit and develop the features of their account which fill the theoretic gap the meta-ontologist has identified and so offer an alternative understanding of the ontological picture underlying abstraction.

Hale and Wright claim that this theoretical gap can be filled by an ontologically innocent theory of reference. They call this the *preferred conception of abstraction* which they take to be able to support the central idea without suspending the logicist hope. In order to introduce their preferred theory they first of all go to great lengths to distinguish their view from another theory of reference which they call Lockeanism. They claim that they have been wrongly interpreted by their objectors as having a Lockean theory of reference.

As we will see, this is extremely telling because it is the Lockean theory of reference which is compatible with reference to Fregean objects and not Hale and Wright's preferred minimalist theory of reference.

6.4.2 Lockeanism

First, let us remind ourselves why Hale and Wright do much work to demarcate a theory of reference. This is to secure what we have called their central ontological idea; the idea that an implicit definition can establish a route from a known statement to an unfamiliar statement despite the fact that the unfamiliar sentence has new and impressive existential consequences which are not a feature of the original. It was the need to justify this idea, without assuming definitions have creative powers, which motivated the attribution of a meta-ontology which Hale and Wright want to resist.

Any general theory of reference will have to provide some kind of account of three key ingredients: first, we need a vehicle of reference; this is the role of expressions – of the syntactic categories of a language. Secondly, we need a means to make the vehicle work so as to aim it at the referents; this is the role of sense or meaning, of the semantic categories of language. Thirdly, we need something for the words to point *at*; we need there to exist something in the world which can constitute the reference; this is the role of the ontology.

The theory of reference which Hale and Wright claim to be in conflict with the abstractionists' project, they refer to as *the Lockean conception*. They attribute this conception of abstraction to Potter and Sullivan with regard to the following passage:

> What did Locke realise about 'gold'? Effectively, that there is an element of blind pointing in our use of such a term, so that our aim outstrips our vision. Our conception fixes what (if anything) we are pointing at but cannot settle its nature: that is a matter of what's out there. One image of the way [Hume's Principle] is to secure a reference for its terms shares a great deal with this picture (Potter & Sullivan 1997, 145-46).

On this view, although our expressions are perfectly adequate vehicles for picking out referents, this process of aiming our words at the world is inherently limited. Limited, in the sense that when we refer to things we do not determine or change what they are. We might say that referring has *no ontological consequences*. On some understanding of ontological consequence, however, the neo-Fregean logicists will agree with this; the act of referring does not bring things into existence. The kind of consequence Potter and Sullivan are concerned with is different: referring to an object under some aspect does not prevent it from having other aspects as well, which is to say our means of referring to an object does not settle its nature. As such, the ontological question of what that object is like cannot be given the same answer as the question of how we are able to refer to it. On this picture an object could always have properties which are not derivable from those properties which we use to pick it out. If an object is a fully paid member of objective reality then there could be no guarantee that the object has been exhaustively characterised when we refer to it – or, rather – no such guarantee could come from the mere success of our conception of that object achieving reference.

Let us extract two features from this view, both of which Hale and Wright claim to be at odds with the abstractionists' view:

$\mathcal{A}$. If it is to be maintained that the conception of objects we use for reference does not exhaust them then it must be that the objects, at least potentially, have different aspects to them. As such we must, at least potentially, be able to be given those objects by their different aspects, that is, to refer to them via different modes of presentation.

$\mathcal{B}$. It is possible for reference to be unsuccessful solely for the reason that the referent does not exist. That is to say; because of the gap between pointing and the world, we may offer a vehicle which is entirely suitable for reference but which fails to refer because reality fails to provide any suitable denizen.

These commitments which are extracted directly from Hale and Wright's characterisation of their opposing theory of reference amount to a denial of the characterisation we have given of the non-eliminativist structuralist's objects:

i. **The Incompleteness Claim**: Mathematical objects are incomplete in the sense that they have no "internal nature" and no non-structural properties.

ii. **The Dependence Claim**: Mathematical objects from one structure are dependent on each other and on the structure to which they belong.

In other words, this conception of reference is suitable for Fregean objects since it follows immediately from our characterisation ($\mathcal{A}$) and ($\mathcal{B}$). Let us now consider whether Hale and Wright's alternative theory of reference is just as suitable for Fregean objects. This will reveal what it is about the Lockean view that Hale and Wright go to such pains to disavow.

6.4.3 Hale and Wright's preferred theory of reference

Hale and Wright explain their theory of reference partly with reference to an abundant theory of properties and partly with reference to Aristotle. Here is their most detailed expression of the theory:

> On the abundant view of properties, predicate sense suffices for reference. But it is not the abstractionists' view of singular terms that sense suffices for reference – the view is that the truth of atomic contexts suffices for reference. However everyone agrees with that. The controversial point is what it takes to be in a position reasonably to take such contexts to be true. The point of analogy with the abundant view is that this is not, by minimalism, conceived as a matter of hitting off, Locke-style, some 'further' range of objects. We can perfect the analogy if we consider not simple abundance but the view that results from a marriage of abundance with Aristotelianism. Now the possession of sense by a predicate no longer suffices, more or less, for reference. There is the additional requirement that the predicate be true of something, and hence that some atomic statement in which it occurs predictively is true. That is a precise analogue of the requirement on singular terms that some atomic statement in which they occur referentially be true Hale & Wright (2009, 208).

With respect to the abundance theory of properties, Hale and Wright say here that predicate sense suffices for reference. By this they do not mean that sense *establishes* reference in the way of creative definition. Rather, they mean that because the metaphysical thesis of abundant properties is in play, all that is needed to ensure the success of reference is that the vehicle is functional, since the metaphysical backdrop will do the work of assuring the denizen.

With respect to Aristotelianism, Hale and Wright say that if a predicate successfully refers then it must meet the requirement formulated by Aristotle, that the predicate *be* *true of something*. They then draw a quick and close connection between the requirement that the predicate be true of something and the requirement that there is a true atomic statement in which it occurs as a predicate. This they claim to be (precisely) analogous to a requirement for singular terms. Let us spell out the translation of the analogy more slowly than is done above: first we can articulate a corresponding Aristotelian requirement on singular terms, namely, that there is something that is true *of them*. Next, we can make a quick and close connection between this and the requirement that there is a true atomic statement in which the expression occurs as a singular term. Finally we note that – on a general Fregean picture – the genuine occurrence of a singular term, in any sentence (including atomic ones) qualifies that expression as a *referential candidate*. Altogether, on this view to establish the reference of singular terms what we need is the truth of a relevant atomic context.

This view rejects the central Lockean idea that our aim could outstrip our vision. It does so by rejecting in particular the two theses of the Lockean conception which we articulated above. It rejects ($\mathcal{A}$) because it does not accept that objects must potentially have different aspects to them if this is to mean that there is the availability of some different means of reference to those objects. To deny ($\mathcal{A}$), however, is to accept (i) that the objects have no properties not already given by the abstraction.

Minimalism rejects ($\mathcal{B}$) because it denies the possibility that reference could fail because reality did not provide any candidate. Indeed, Hale and Wright characterise this denial as the essence of their preferred conception of abstraction, insisting that, "there *is* no metaphysical hostage in the transition, no need for an 'assist' from the World" (2009, 205). Instead, they say, if all goes well with the abstractionist recovery then there can be no significant risk of reference failure. This comes very close to but does not quite amount to an acceptance of the dependence of the abstractionist objects on the property which defines them (ii).

Although this appears quite straightforward, the theory of minimalism is in fact much more subtle than the rough characterisation we have given. In order to truly get to the heart of its ontological commitment we will require a much more detailed examination of its two appeals. This will also serve to prevent any accusation that Hale and Wright's position has not been understood, as they often complain to Potter and Sullivan.

The Substance of the Appeal to the Abundance Theory

The appeal to the abundance theory of properties gives rise to some potential confusion, in particular because it seems that Hale and Wright are making use of an abundance theory despite the fact that such a theory is surely a species of maximalism. This is worrying because, in this case, Hale and Wright appear to be motivating their theory of reference by an implicit appeal to the very problematic meta-ontology which they sought to avoid. As a further point of potential confusion, Hale and Wright call their view *minimalism*. We will here examine these sources of confusion in more detail and show that they can be understood unproblematically.

Let us carefully consider, first, how the abundance theory is related to maximalism. We have until now employed the same understanding of the maximalist thesis as Hale and Wright: that whatever could exist does exist. Of course, one can subscribe to such a principle and maintain a restricted understanding of the modal claim, so that what exists would still be maximised from all the possible things by the principle of plenitude, but the principle would not commit us to a plentiful ontological inventory.

How are we to understand the abundance theory itself? Lewis formulates abundant properties as the following:

> Any class of things, be it ever so gerrymandered and miscellaneous and indescribable in thought and language, and be it ever so superfluous in characterising the world, is nevertheless a property. So there are properties in immense abundance ... There are so many properties that those specifiable in English, or in the brain's language of synaptic interconnections and neural spikes, could only be an infinitesimal minority (Lewis 1983, 346).

The first thing to observe is that the theory only regards properties and as such can only be properly compared with a sub-species of maximalism: property maximalism. We can then formulate the related principle of plenitude as follows: whatever property can exist does exist. Indeed, given that Lewis (1983, 344) tells us earlier that properties are classes of things the first sentence of the quote can be understood as a direct endorsement of a principle of plenitude for properties. Further, the threshold for being a possible property is very low. We are told there are many more properties than there are predicates; more than those possible to describe in language and even in thought. This brings out clearly that the abundant theory is foremostly a metaphysical thesis and not a semantic one. For these reasons it seems that the abundant theory, at least as it is originally formulated by Lewis, should be understood as a subset of maximalism, at least as maximalism is formulated by Hale and Wright.

Hale and Wright contrast the abundance theory with a sparse theory, to which we will now divert. For the sparse theorist there is a limit on the ontological inventory and this has a knock-on effect for reference: the vehicle for reference might be entirely capable but fail to pick anything out simply because referents are harder to come by in a sparse universe. As such, endowing a term with sense is not enough to ensure that it refers. On this picture, we require a further assurance of the existence of the referent in order to declare the reference relation is successful.

Could the truth of atomic contexts give the further 'assist' from the world that is needed to secure reference? That is to say, does this view cohere with Hale and Wright's claim that all parties will agree that the truth of atomic contexts suffices for reference? When Lewis characterises a sparse theory (of universals) he does so in the following way:

> A satisfactory inventory of universals is a non-linguistic counterpart of a primitive vocabulary for a language capable of describing the world exhaustively (Lewis 1983, 346).

Although Lewis does not claim that universals are to be directly read off the primitive vocabulary of a language it seems fair to say that he draws a link between them such that the atomic language might, at the very least, give us some guide as to what universals there are. Indeed, nothing further is said by Lewis regarding how the sparse universals are otherwise to be specified. For these reasons, it seems that the sparse view sufficiently coheres with Hale and Wright's assumption that there is no disagreement that the truth of atomic contexts suffices for reference, though there is disagreement regarding how this truth is established.

We have then, on the one hand, the view that properties are in abundance so that reference to them is secure once we have fitted some appropriate expressions with a sense. On the other hand, we have the view that properties are sparse and that to secure reference we need to ensure both that the predicate has a sense and that its purported referent exists. Although Hale and Wright contrast these views in developing their theory of reference in fact neither is compatible with their view. This is because they are metaphysical positions which presuppose a common understanding of reference which cannot be reconciled with the neo-Fregean logicist project. On both views the existence of the referent is determined by something independent of language; in particular, a metaphysical thesis about properties. As we have seen, however, the neo-Fregean logicists cannot help themselves to any first-order metaphysical resources on pain of calling off a logicist expedition to the numbers. Further, since the abundance theory is a species of maximalism, it offers no alternative to the view supposedly being avoided by Hale and Wright in developing their own account.

For these reasons it cannot be the case that the neo-Fregean logicists intend to adopt the abundance theory in a straightforward sense. But if this is so, then we must establish in what capacity they employ the abundance theory in their account of reference. The precise claim made by Hale and Wright is that *their conception of reference stands to the Lockean conception of reference as the abundance theory of properties stands to a sparse theory of properties* (Hale & Wright 2001, 23). As such, the abundance theory is being employed as part of a comparison which is used to elucidate Hale and Wright's theory of reference. The potential confusion arises in mistaking which precise feature of the contrasting pairs merits the comparison.

One understanding of the comparison is the following. The sparse theorist and the Lockean conception both require a guarantee from the world to ensure certain expressions refer independently of their having a determinate sense. The abundance theorist and the neo-Fregean logicists deny that there is any such guarantee needed to plug a gap between language and the world. They both deny this for the same reason: that if we believe the world is abundantly populated then there will never be a case where the world lets us down by failing to assist us with a referent for our expression. Let us call this understanding of Hale and Wright's claim, the *Strong Comparison*.

There is another understanding of the comparison available. The abundant theorist is, as before, akin to the neo-Fregean logicists in denying that we need an independent guarantee for reference to work, but they deny this for *different* reasons. The abundant theorist denies the need for a worldly assist because the universe is so abundantly populated and the neo-Fregean logicists deny the need for a worldly assist for reasons which appeal to no such metaphysical thesis. Rather, they employ the requirement that there is a true atomic statement which the term occurs in referentially; elucidating this requirement by appeal to Aristotelianism. Thus the comparison is merited by the common denial of the need to establish the existence of the referents independently of their being referred to, and not by any overlap in their background metaphysical view. Call this understanding of Hale and Wright's claim, the *Weak Comparison*.

Confusion threatens because Hale and Wright can be read as making the Strong Comparison claim when they are in fact making the Weak Comparison. It is only the Strong Comparison claim that would involve the neo-Fregean logicists adopting a maximalist metaphysical view, which they neither need nor can rely on. That the neo-Fregean logicists are in fact making the Weak Comparison claim is supported by the consideration that had the neo-Fregean logicists adopted an abundant metaphysics, there would be no need for them to supplement their thesis with the requirement that the term must feature referentially in a true atomic sentence. If every meaningful expression was secured a reference simply because in the abundant universe there is plenty to go around, then this condition could always be trivially satisfied and so would be entirely superfluous. This is shown by the following: On an abundant view, every predicate expression which has a sense, will also have a reference (this contrasts with a sparse view, in which only some predicate expressions will have a reference). On a maximalist view, everything which can exist does exist. By an abundant-maximalist view, then, we could ensure the reference, not just of predicates, but of every expression which has a mode of presentation of something possible. Take, for example, a singular term A which has a sense and purports to refer to something possible (and thus actual). Any well-formed sentence that A features in, it will therefore feature in referentially. It might seem that we cannot deduce from this that it will feature in an *atomic* sentence, but since the neo-Fregean logicists count identity contexts as atomic, we can always construct a tautologous identity claim; $A = A$. Thus any expression akin to A can be trivially said to feature referentially in a true atomic sentence on the basis of the abundance-maximalist metaphysics alone. This shows that the neo-Fregean logicists must be understood as making the Weak Comparison which avoids adopting such an abundance theory since this would make the second half of their theory of reference an empty requirement. We can conclude that the neo-Fregean logicists are best understood as making an appeal to Aristotle in place of an appeal to a metaphysical view and not in addition to it.

We now have in view the substance of Hale and Wright's appeal to an abundant theory of properties. Their intention is not to employ the theory but to use it to elucidate a feature of their own account: namely, any view which requires independent assurance of the existence of properties or objects of abstraction in order to refer to them is to be denied.

The Substance of the Appeal to Aristotle

We are now well placed to consider the substance of the second element of Hale and Wright's theory of reference which they expound by appeal to Aristotle.

The above considerations make clear the role of the appeal: to avoid commitment to a metaphysical thesis while guaranteeing the referent of a singular term in a way analogous to the sense in which the abundance theory guarantees the existence of properties.

At the same time, the appeal is intended to narrow the class of singular terms for which reference is assured. Doing this gives the neo-Fregean logicists a better explanation of reference that hopes to avoid the charge of creative definition.

Let us remind ourselves how Hale and Wright present this appeal.

> We can perfect the analogy if we consider not simple abundance but the view that results from a marriage of abundance with Aristotelianism. Now the possession of sense by a predicate no longer suffices, more or less, for reference. There is the additional requirement that the predicate be true of something, and hence that some atomic statement in which it occurs predictively is true. That is a precise analogue of the requirement on singular terms that some atomic statement in which they occur referentially be true (Hale & Wright 2009, 208).

Hale and Wright begin, here, with the Aristotelian requirement that a predicate be true of something and draw from this their own requirement on reference: a singular term refers if there is a true atomic statement in which the singular term occurs in a referential position. Two steps are made to connect these different requirements. The first is the move between the requirement that the predicate be true of something and the requirement that the predicate occur referentially in a true atomic statement. The second is the move between a requirement for predicate reference and a requirement for singular term reference.

Before we examine these moves in turn, let us notice that the neo-Fregean logicists are not straightforwardly adopting a full-bloodied Aristotelian theory of reference and that they have no need to do so. As we saw was the case with the abundance theory, they are merely adopting certain features of the view which they adapt to their particular case and in this sense use the Aristotelian theory to elucidate their own theory of reference.

If this is kept in view then the second step from predicates to singular terms can be understood as an innocuous move, employed to translate a thesis which holds of one set of expressions to another. After all, Hale and Wright do not claim to derive one from the other but make the weaker claim that the Aristotelian thesis is a 'precise analogy' for the singular term thesis. This diffuses the objection that the step involves an obfuscatory slide between categories of expressions for which reference plausibly works very differently. Merely, instead of requiring for reference that predicates "be true of something" we require for singular terms that "something be true of them", to adapt Hale and Wright's turn of phrase. More precisely, this does not mean that something is true of the number of letters, or any other property of the typography of the singular term, but rather that *some referring expression is saturated by the singular term to form a true sentence.*

The first step remains to be unpacked. We can happily equivocate a predicate being true of something and its occurrence in a true sentence. The condition that the predicate occur referentially cannot be understood as requiring the predicates to possess determinate reference since in this case the condition would be circular. Therefore, to occur referentially in a sentence is not to refer, but to be *apt* to refer; to occupy a position in the sentence to which a semantic treatment would have to assign a reference, i.e. to occur with a determinate sense which contributes towards the truth conditions of the sentence. Hale and Wright themselves spell out the condition in a similar manner, but speak of the expression being reference-demanding rather than referentially apt. Either of these phrases indicates that what is intended by the neo-Fregean logicists is a weaker, non-circular condition which is distinguished from an expression's being referential in the sense of realising its referential potential. They introduce this terminology here:

> The only possible answer appears to be that such a feat of verification [that a singular term refers] must consist in verifying – if not an identity statement linking the term in question with another whose reference is assured – then *some* form or forms of statement embedding the term in question who truth *requires* that it refer: a statement, or range of statements, in which the term in question occupies a *reference-demanding* position. Such will be afforded

> by provision of the means to verify some form of *atomic* statement configuring such terms. Identity contexts are one kind of atomic statement. So abstraction itself – as a characterisation of putatively canonical grounds for the verification of such identity contexts – supplies a paradigm means, indeed an example it seems of the only foreseeable broad kind of means, for accomplishing the assurance required (Hale & Wright 2009, 204).

Here it is suggested that we can know a term is reference-demanding if it occurs in an atomic context. We will consider in a moment the difficulty with this appeal. We can say, however, that the condition that the expression be reference-demanding is an equivocal unpacking of the Aristotelian thesis if, like Frege, we grant the assumption that a well-formed predicate with a determinate sense will always have a referent, and indeed requires one for sentences in which they occur to have a truth-value. Granting this assumption means that to occur *as* a predicate in an expression is to occur *referentially*. This makes it equivalent to say that an expression occurs as a predicate in a true sentence and that an expression occurs referentially in a true sentence.

It should be noted here that precisely what the first objection in chapter 4 does is to demonstrate that there is no occurrence of a singular term in a referential position in Hume's Principle but that the principle as a whole is a predicate expression although it is an unsaturated one and so does not occur in any true atomic contexts.

The condition that the sentences are atomic seems intended to alter the requirement, but the restriction is obscure. The general idea is that there could be an expression which occurred in a true sentence but which failed to occur in a restricted and somehow privileged range of sentences and so did not qualify as fulfilling its referential potential. This could only be determined if we knew the relevant range of sentences and had a way of deciding whether an expression occurred in one. Neither of these things are given by the neo-Fregean logicists; instead we are merely assured that identity contexts are among them, for the reason that establishing statements of identity is the routine way in which we satisfy ourselves that singular terms refer (Hale and Wright, 2009:19). This provides us with a positive test for reference but not a negative one: if the expression occurs in a (true) identity context we can say it refers, but if it fails to occur in a true identity context (however this failure is supposed to manifest itself) then we cannot conclude from this that the expression does not refer, since it may be that it occurs in another of the unspecified atomic contexts. A negative test for reference and thus a full account of the relation is therefore pending an account of atomic contexts.

However the neo-Fregean logicists may demarcate the class of atomic sentences, it does not seem that the condition can be understood as a straightforward unpacking of an Aristotelian view in the same way as the other conditions. The two are best brought together by a restricted Aristotelian theory which holds only for reference of expressions in some special range of sentences, i.e. a theory that holds that an *atomic* predicate must be true of something, where an atomic sentence is a special class of sentence and an atomic expression is merely one that occurs in it. One consideration that should ease our anxiety that we have no general understanding of an atomic context is that we are assured of one example of such a context; identity contexts. Since it is the case that a principle qualifies as an abstraction principle in virtue of its form (though this cannot be sufficient if we are to keep out bad company) then all abstraction principles will exploit identity contexts on their left-hand side in the same way. In this sense, we only require the mysterious class of atomic sentences to have one well-motivated instance and this the neo-Fregean logicists provide. Nevertheless, failing a more general account of the subset of atomic sentences, we have no way of explaining why it is that such a class (identity contexts included) have the privilege that they do in a theory of reference. About this, the neo-Fregean logicists say nothing but merely exploit a vague background appeal to some form of atomism since this seems to be the only metaphysical conception which would support such an implicit assumption and is of course indicated by their choice of terminology. It seems, however, that the chance of a full explanatory account of reference is pending either the identification of a class of atomic sentences; or else an explicit appeal to atomism; or an explicit appeal to any thesis which invokes some fundamental level of language (reached by analysis or an analogous process) which gives us a privileged insight into reality.

Be that as it may, we can use the above considerations to provide a potential way in which Atomic Aristotelianism – i.e. the thesis that an atomic predicate must be true of something – can be unpacked to yield the thesis that an expression refers if it occurs referentially in an atomic true sentence, as the neo-Fregean logicists claim in the quotation. The questionable link is not the precise analogy drawn between singular term and predicate reference but rather between the claim that a predicate be true of something and that *hence* some atomic statement in which the expression occurs predictively is true. This derivation can only be drawn if atomism (whatever that spells out to be) is already assumed before Aristotelianism is employed to 'perfect the analogy' by elucidating Hale and Wright's theory of reference beyond its commonalities with the abundance theory.

Minimalism

We can now put together the detail of the theory of reference which Hale and Wright call *minimalism*.

As they point out in a footnote, this label is not to be mistaken as the counterpart of maximalism (Hale & Wright 2009, 207 ft.36). The *minimal* element of the view is not its ontological inventory but the required means of securing reference, which is to say that we can know a term refers with minimal assurance: we need not require that the world provide a suitable denizen as a referent but only that the world makes a range of privileged sentences in which the expression occurs as *true*.

There is another potential confusion regarding minimalism which Hale and Wright do not discuss: its similarity with the syntactic priority thesis, another thesis forwarded by Hale and Wright at various points. Let us lay this thesis out carefully so as to see the potential confusion and to establish that the two can in fact be kept distinct.

The syntactic priority thesis is attributed to Frege by both Wright and Dummett and holds that if an expression behaves syntactically like a singular term (in some range of true sentences) then it must refer. Of course, for Frege, if a singular term refers then it refers to an object and it is in this sense that the thesis is one of priority: the syntactic features of an expression are prior to its referent in determining the category of a singular term. This priority is taken by Wright and Dummett to issue from Frege's context principle: that only in the context of a proposition do words have meaning (Frege 1884, §62). Because of this, in places Hale and Wright attempt to establish the mere syntactic behaviour of an expression since they can then invoke the syntactic priority thesis to establish that this term must refer. Here is an example of a place where such an appeal is made:

> ...there is no question of our attaching a clear satisfaction condition to a predicate, for instance (or a clear condition on the identity of the value of the function denoted by an operator for a given argument) – yet somehow failing to supply such expression with *Bedeutung*... whatever *Bedeutung* is held to consist in for such expressions – one automatically confers a *Bedeutung* upon them by settling their meaning... (Hale & Wright 2001, 129).

The parallels between this appeal and minimalism as we have expounded it are immediate: there is no gap between the predicate and its referent and as a result we can guarantee that a term refers by establishing non-metaphysical features of it. Such as, that it has a determinate sense, that there is a genuine occurrence of the predicate, and that it occurs referentially. For the syntactic priority thesis, these are the only features relevant to reference. We saw, however, that the very reason Hale and Wright invoked an Aristotelian theory of reference was to distinguish their view from a theory of reference which held that settling the meaning of an expression was sufficient to settle its reference. Minimalism, instead, demands that further features of the expression are met before reference is secured; in particular its occurrence in a true atomic context. So while both theses will in common place a great weight on establishing the genuine syntactic occurrence of the expression and its determinate sense, this will ensure that the expression *refers* if syntactic priority is employed whereas on the minimalist picture all that this can ensure is that the expression is *reference-demanding*. The minimalist further requires that the sentence in which the expression demands reference is true and that it falls in a privileged range of sentences amongst which identity contexts prominently feature.

A truly minimal ontology

We have thus reconstructed Hale and Wright's preferred theory of abstraction in great detail and with great charity. Having done so, we may now ask what it amounts to and whether it is compatible with a Fregean ontology.

Let us consider Hale and Wright's dialectical position with respect to the Lockean view. The neo-Fregean logicists are not merely able to concede the coherence of alternative theories of reference; they are in fact not required to present any argument at all to establish the superiority of their theory of reference. How can their yoke be so light? Recall that the neo-Fregean logicists wanted to avoid commitment to a troublesome meta-ontology but to do so they needed to provide a theory that could justify their central ontological idea. To meet this aim, all that is required is that they produce one viable theory of reference which avoids the original troubling features of maximalism or quantifier variance, and which is able to support the central ontological idea. If there is more than one understanding of reference which meets these desiderata, the more the merrier. If there is an understanding of reference which is entirely legitimate but does not meet these desiderata, the neo-Fregean logicists may happily pay it no attention; it will not threaten the chance of providing their own theory which is fit for abstractionist purpose. The Lockean theory of reference falls into the latter of these two categories.

What the neo-Fregean logicists' argument against the Lockean conception brings out most importantly, therefore, is its *incompatibility with their desiderata for a theory of reference* and not the general theoretical superiority of minimalism. They say that on the Lockean conception whether reference is secured;

> ...is something that needs to be verified as a by-product of our, so to say, *finding* a range of objects 'out there' to which the conception embodied in the

> principle is (necessarily) faithful. And of course if that is to be possible, the objects in question must first be given to us under some *other* mode of presentation... But this [reference-fixing] spirit – necessary for the 'anxious metaphysical' stance – is simply in flat tension with the abstractionist conception of the matter; indeed, it is to view abstraction principles in a manner inconsistent with their capacity to serve the process of abstraction itself. Properly viewed, the very stipulate equivalence of the two sides of an instance of an abstraction principle is enough to ensure both that it is not proposed as part of a project of reference-fixing and that there *is no significant risk of reference failure* (Hale & Wright 2001, 22).

As we have seen Hale and Wright think that minimalism avoids this risk because it denies the need for substantive worldly assistance beyond ensuring the truth of an atomic context in which a meaningful expression features. This is not so on the Lockean view because although they are understood as agreeing that truth in an atomic context is enough for reference, their view of what is sufficient to establish the truth of such contexts is distinct. We are in a position to assert an atomic sentence as true if we have independent assurance that the referents of its predicates and singular terms exist. However, introducing the possibility of providing this independent assurance also introduces the possibility that the world *fails* to provide it and this is presented as the problem with any 'reference-fixing' theory of reference.

It is clear that this potential for referential failure makes such theories seem less suitable for Hale and Wright's purpose, but it is not clear why they make the stronger claim that this feature of the Lockean account actually incapacitates it as an account of abstraction. To see why they make this claim we must return to the two features of the Lockean view which we previously articulated:

$\mathcal{A}$. If it is to be maintained that the conception of objects we use for reference does not exhaust them then it must be that the objects, at least potentially, have different aspects to them. As such we must, at least potentially, be able to be given those objects by their different aspects, that is, to refer to them via different modes of presentation.

$\mathcal{B}$. It is possible for reference to be unsuccessful solely for the reason that the referent does not exist. That is to say; because of the gap between pointing and the world, we may offer a vehicle which is entirely suitable for reference but which fails to refer because reality fails to provide any suitable denizen.

So far we have discussed feature ($\mathcal{B}$) but Hale and Wright also invoke ($\mathcal{A}$) where they say that – for the Lockean – the objects which expressions aim to refer to must be given under some other mode of presentation. This is to say that the objects of abstraction are such that they can always have different aspects to them in addition to the ones by which they are given. In fact this second feature is the key for justifying Hale and Wright's stronger claim, as we will see.

First, recall that when we spelled out the contribution of abstraction we said that the neo-Fregean logicists understood abstraction principles as the only hope of rescuing objects from the universal domain they inhabit by providing a referential route to them. This contribution is substantive only on the assumption that, for the objects of abstractionist recovery, *there could be no means of referring to them independently of abstraction*. We again meet with the assumption that the objects of abstraction have a single aspect – that they are given to us only as the values of certain functions.

Thus, the Lockean view comes with an in-built feature which presupposes that a tenet of Hale and Wright's background conception of the abstractionist universe is false. And, in particular, that there is no way to be given the relevant objects except by abstraction and so that the recovered objects admit of no alternative mode of presentation. The Lockean view does not explicitly repudiate this thesis; technically it is compatible to have a Lockean account of reference and hold ($\mathcal{B}$). Such a view would maintain that we can never have assurance of reference since what would be needed is an independent specification that the objects are inert to provide. If the Lockean thesis is not to be understood as a sceptical one, however, we can safely say that it is not compatible with the central abstractionist tenet characterised in ($\mathcal{B}$). This is what merits the strong claim made by the neo-Fregean logicists that the Lockean understanding of abstraction principles forfeits their ability to 'serve the process of abstraction'. This process is only made possible, and indeed potentially explanatory, if the objects admit of exactly one mode of presentation, which happens to be the mode created by abstraction. For this reason the Lockean view does not meet Hale and Wright's desiderata for an account of reference and so can be ignored.

However, what is clear is that Hale and Wright's desiderata for an account of reference and so for abstraction is that it is only suitable for the recovery of Hilbertian basic elements, or some other kind of "thin" objects. More importantly, as we have just observed, the very reason that the Lockean theory of reference is rejected is not because as an account of abstraction it is inferior to minimalism, but precisely because it presupposes the recovery of Fregean objects. Their extended attacks on Lockeanism, therefore, shows the sheer extent to which Hale and Wright have abandoned the ontological ambitions of the Fregean project with which they set out.

As a final consideration, let us ask what the minimalist would consider sufficient to secure the truth of an atomic context. After all, this is the point which is represented by the neo-Fregean logicists as being the point of contention. We saw that on the Lockean view we require independent specification of the existence of the expression's referents to establish the truth of the atomic sentence and that this was denied by the Minimalist view. Assuming also that there is no disagreement that identity contexts are canonical instances of atomic contexts, how might we spell out the Minimalist's account of the establishment of the truth of atomic contexts, and how plausible is that account? On this view, the sentence on the left-hand side of an abstraction principle is given a determinate sense by the process of implicit definition; the truth of the sentence is not established immediately from its having a determinate sense but rather from the bi-conditional – since the world is such that the right-hand side of the bi-conditional is true then this establishes the truth of the left-hand side. Hence we have a very minimal worldly assist which nonetheless anchors the principle in reality. But, nonetheless, the truth of the identity sentence on the left-hand side needs only the success of the bi-conditional as an implicit definition to bestow sense upon it and to establish the truth of an atomic context as a result.

At the end of "The Metaontology of Abstraction" Hale and Wright concede that there are a number of questions which they acknowledge their opponents would need answered before considering abstraction to be a 'competitive option' (Hale & Wright 2009, 190). Amongst them they identify the metaphysical question:

> (M) What does the world have to be like in order for (the best examples of) abstraction to work?

To this they eventually provide the following answer:

> We have been rather neglecting question (M)... What, in the light of the foregoing discussion should now be said in answer? First, for each equivalence relation which is to underpin an abstraction – for all we have said, indeed, for *every* equivalence relation – there has to be an associated function... Second, the existence of such a function will of course require a properly behaved range of values... But if [minimalism] is accepted, the answer to question (M) could not be simpler: a world in which abstraction works is a world in which there are equivalence relations with non-empty fields (Hale & Wright 2009, 209).

This requirement sounds dazzlingly modest. However, what I have been suggesting is that in their rejection of the Lockean thesis Hale and Wright have preconditioned what the objects in this field are like. They have moved away from a universal domain and a Fregean conception of an object as a member of that domain to a collection of equivalence classes (or, categories) into which fall objects which are such that their nature is exhausted by their falling into an equivalence class. The abstractionists' preferred objects, then, are decidedly not Fregean and are more akin to the Hilbertian non-eliminativist objects which we have characterised. The *implicit* answer to (M) is that the world has to be made up of such 'thin' objects; truly a very minimal world indeed.

Conclusion

What we have been witness to is the disowning of a Fregean ontology in three central places of the neo-Fregean logicists' canon: in their ontology of categories with respect to their defence of the Julius Caesar problem; in their dilemma regarding which objects are suitable for abstractionist recovery with respect to their defence of the MP problem; and in their minimalist account of their central ontological idea. We found Hale and Wright's metaphysical picture to be minimal indeed, not because it made minimal ontological commitments as they claim, but because it presupposes a conception of objects as minimal, i.e. as exhausted by their being represented as falling in the value range of some function which falls under the concept defined by the process of abstraction (if successful). This conception is entirely at odds with a Fregean ontology of mathematics.

Conclusion

Hale and Wright have argued extensively that abstraction principles are a preferable alternative to mere axiomatic stipulation. A key element of their story is that abstraction principles – unlike axioms – are implicit definitions of the *abstracta* they recover. However, the substantive distinction between an axiom and an abstraction principle cannot be that one is a definition and the other is not. We saw that Hilbert very explicitly considered his axioms to be implicit definitions. What is the substantive difference, then, between an axiom and the purportedly privileged alternative of abstraction?

On the basis of what we have discussed I think we can give an answer to that question. The difference between an axiom and an abstraction principle – which Hale and Wright are appealing to – is *the ontology they are capable of characterising*. While an axiom (in Frege's traditional sense) exploits a prior conception of the objects of some field of mathematics, an abstraction principle introduces us to objects which admit of no independent means of specification. Thus, axiomatic stipulation is capable of characterising objects which have a nature in addition to how they have been introduced, and definition by abstraction is capable of characterising objects which are exhaustive in that they have no properties further to those given by how they have been introduced.

Understanding the difference between abstraction principles and axioms in this way, we see that there is no longer an obvious sense in which definition by abstraction is superior to axiomatic stipulation, rather than an *alternative* to it. The claim of superiority only makes sense with respect to the corresponding ontological positions: Should we think of the objects of mathematics as thin in the sense that they are exhaustively characterised by our mathematical theories, or should we think of them as thick in the sense that they are always apt to surprise us with their other aspects?

For Hale and Wright to advocate a thin ontology of mathematical objects is entirely viable and actually in keeping with a logicist project. However, it is not in keeping with any kind of *Fregean* (or nearly Fregean) logicism. One consequence of this is that it becomes clear that the logicist project relies on some conception of mathematics and what its objects are like as well as some conception of logic. To put the point quite simply: whether mathematical objects are logical objects depends on what we think logical objects are and what we think mathematical objects are. Another consequence is that Hale and Wright should drop their 'neo-Fregean' prefix and label their project more honestly as *neo-Hilbertian* logicism (so long as this is clarified as *early* Hilbert) or perhaps *structuralist* logicism.

The foundational debates between Frege and Hilbert and between Hale and Wright and their opponents have provided us with a case-in-point of the complicated relationship between an axiomatisation and the reality it axiomatises. The controversy between Frege and Hilbert made this very clear since Frege resists Hilbert's employment of implicit definitions as axioms precisely because he is sensitive to the ontological consequences of this approach – as Hale and Wright have not been. In this way, this thesis has provided one sense in which one's conception of what kind of thing an axiom is turns out to be deeply interconnected to one's conception of what kind of thing a mathematical object is. It has also provided a result which can be explained as the stark observation that *Fregean objects are not suitable for abstractionist recovery*.

Bibliography

Ajdukiewicz, K. (1996), 'The Logical Concept of Proof: A Methodological Essay', *Studia Logica* (19), 12–45.

Bernays, P. (1950/2002), Mathematical Existence and Consistency, *in* S. G. Wilfried Sieg, Richard Zach, ed., 'Philosophie des sciences en hommage à Ferdinand Gonseth', Editions du Griffon, Neuchatel, pp. 11–25.

Blanchette, P. (2014), The Frege-Hilbert Controversy, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', Spring 2014 edn, Metaphysics Research Lab, Stanford University.

Blanchette, P. A. (1996), 'Frege and Hilbert on Consistency', *Journal of Philosophy* **93**(7), 317–336.

Boolos, G. (1998), *Logic, Logic, and Logic*, Harvard: University Press.

Brown, J. R. (2005), *Philosophy of Mathematics: An Introduction to a World of Proofs and Pictures*, London: Routledge.

Carroll, L. (1871), *Through the Looking-Glass, and What Alice Found There*, London: Harper Collins.

Dedekind, R. (1963), *Essays on the Theory of Numbers*, W. W. Behman, trans. New York: Dover Publications.

Dreben, B. & Kanamori, A. (1997), 'Hilbert and Set Theory', *Synthese* **110**(1), 77–125.

Dummett, M. (1976), Frege on the Consistency of Mathematical Theories, *in* 'Frege and Other Philosophers', Oxford: Oxford University Press, pp. 33–46.

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

Dummett, M. (1998), *Frege: Philosophy of Mathematics*, London: Duckworth.

Euclid (1956), *The Thirteen Books of Euclid's Elements*, Translation by Sir T. L. Heath. New York: Dover Publications.

Ferreirós, J. (2009), 'Hilbert, Logicism, and Mathematical Existence', *Synthese* **170**(1), 33–70.

Ferreirós, J. (2016), The Early Development of Set Theory, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', Fall 2016 edn, Metaphysics Research Lab, Stanford University.

Field, H. (1989), 'Platonism for Cheap? Crispin Wright on Frege's Context Principle', *Canadian Journal of Philosophy* **14**, 160–182.

Fine, K. (2002), *The Limits of Abstraction*, Oxford: Oxford University Press.

Franks, C. (forthcoming), David Hilbert's Contribution to Logical Theory, *in* A. P. Malpass, ed., 'The History of Logic', Continuum Press, pp. 1–16.

Frege, G. (1879/1964), *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*: Concept Script, a formal language of pure thought modelled upon that of arithmetic, *in* S. Bauer-Mengelberg & J. van Heijenoort, eds, 'From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931', Harvard: Harvard University Press, pp. 1–82.

Frege, G. (1882/1980*f*), Frege to Marty (1882), *in* G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, A. Veraart, B. McGuinness & H. Kaal, eds, 'Gottlob Frege: Philosophical and Mathematical Correspondence', Blackwell, pp. 99–102.

Frege, G. (1884), *Grundlagen der Arithmetik*, Breslau: Koebner.

Frege, G. (1895/1980*a*), A Critical Elucidation of Some Points in E. Schröder, Lectures on the Algebra of Logic, *in* P. Geach & M. Black, eds, 'Translations from the Philosophical Writings of Gottlob Frege', Blackwell, 3rd ed, pp. 86–106.

Frege, G. (1895/1980*c*), Frege to Hilbert (1.10.1895), *in* G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, A. Veraart, B. McGuinness & H. Kaal, eds, 'Gottlob Frege: Philosophical and Mathematical Correspondence', Blackwell, pp. 32–34.

Frege, G. (1897/1979), Logic, *in* K. F. Hermes, H. & F. e. Kaulbach, eds, 'Posthumous Writings: Gottlob Frege', Chicago: University of Chicago Press, pp. 126–151.

Frege, G. (1899/1980*e*), Frege to Hilbert (27.12.1899), *in* G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, A. Veraart, B. McGuinness & H. Kaal, eds, 'Gottlob Frege: Philosophical and Mathematical Correspondence', Blackwell, pp. 34–38.

Frege, G. (1900/1980*b*), Frege to Hilbert (06.1.1900), *in* G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, A. Veraart, B. McGuinness & H. Kaal, eds, 'Gottlob Frege: Philosophical and Mathematical Correspondence', Blackwell, pp. 43–48.

Frege, G. (1900/1980*d*), Frege to Hilbert (16.9.1900), *in* G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, A. Veraart, B. McGuinness & H. Kaal, eds, 'Gottlob Frege: Philosophical and Mathematical Correspondence', Blackwell, pp. 49– 50.

Frege, G. (1903/1971*a*), On the Foundations of Geometry: First Series, *in* E. H. W. Kluge, ed., 'On The Foundations of Geometry and Formal Theories of Arithmetic', New Haven and London: Yale University Press, pp. 22–37.

Frege, G. (1906/1971*b*), On the Foundations of Geometry: Second Series, *in* E. H. W. Kluge, ed., 'On The Foundations of Geometry and Formal Theories of Arithmetic', New Haven and London: Yale University Press, pp. 49–112.

Frege, G. (1906/1971*c*), Reply to Mr. Thomae's Holiday Chat, *in* E. H. W. Kluge, ed., 'On The Foundations of Geometry and Formal Theories of Arithmetic', New Haven and London: Yale University Press, pp. 121–127.

Frege, G. (1980*g*), Function and Concept, *in* P. Geach & M. Black, eds, 'Translations From the Philosophical Writings of Gottlob Frege', Philosophical Library, pp. 21–41.

Gabriel, G., Hermes, H., Kambartel, F., Thiel, C., Veraart, A., McGuinness, B. & Kaal, H. (1980), Gottlob Frege: Philosophical and Mathematical Correspondence, Blackwell, pp. 33–51.

Geach, P. (1955), 'Class and Concept', *Philosophical Review* **64**(4), 561–570.

Hale, B. & Wright, C. (2001), *Reason's Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*, Oxford: Oxford University Press.

Hale, B. & Wright, C. (2008), 'Abstraction and Additional Nature', *Philosophia Mathematica* **16**(2), 182–208.

Hale, B. & Wright, C. (2009), The Metaontology of Abstraction, *in* 'Metametaphysics: New Essays on the Foundations of Ontology', Oxford University Press, pp. 178–212.

Hallett, M. (1994), Hilbert's Axiomatic Method and the Laws of Thought, *in* A. George, ed., 'Mathematics and Mind', Oxford: Oxford University Press, pp. 158–200.

Hawley, K. (2007), 'Neo-Fregeanism and Quantifier Variance', *Aristotelian Society Supplementary Volume* **81**(1), 233–249.

Heck, R. (1993), 'The Development of Arithmetic in Frege's *Grundgesetze der Arithmetik*', *Journal of Symbolic Logic* **58**(2), 579–601.

Heck, R. (2005), 'Julius Caesar and Basic Law V', *Dialectica* **59**(2), 161–178.

Heck, R. (2011), *Frege's Theorem*, Oxford: Clarendon Press.

Hilbert, D. (1894/2004), *Die Grundlagen der Geometrie*, *in* M. Hallett & M. U. e. Majer, eds, 'David Hilbert's Lectures on the Foundations of Geometry, 1891-1902', Springer-Verlag, pp. 72–123.

Hilbert, D. (1895/1980*a*), Hilbert to Frege (04.10.1895), *in* G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, A. Veraart, B. McGuinness & H. Kaal, eds, 'Gottlob Frege: Philosophical and Mathematical Correspondence', Blackwell, p. 34.

Hilbert, D. (1899/1971), *Grundlagen der Geometrie*, Leipzig: Teubner. English translation of 10th edition by L. Unger, Chicago: Open Court.

Hilbert, D. (1899/1980*d*), Hilbert to Frege (29.12.1899), *in* G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, A. Veraart, B. McGuinness & H. Kaal, eds, 'Gottlob Frege: Philosophical and Mathematical Correspondence', Blackwell, pp. 38–43.

Hilbert, D. (1900/1980*c*), Hilbert to Frege (22.9.1900), *in* G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, A. Veraart, B. McGuinness & H. Kaal, eds, 'Gottlob Frege: Philosophical and Mathematical Correspondence', Blackwell, pp. 50–51.

Hilbert, D. (1900/1996*a*), *Über den Zahlbegriff*, *in* W. B. Ewald, ed., 'From Kant to Hilbert: A Source Book in the Foundations of Mathematics', Oxford University Press, pp. 12–28.

Hilbert, D. (1900/1996*b*), 'Mathematical Problems: Lecture Delivered Before the International Congress of Mathematicians at Paris in 1900', *Journal of Symbolic Logic* **44**(1), 116–119.

Hilbert, D. (1903/1980*b*), Hilbert to Frege (07.11.1903), *in* G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, A. Veraart, B. McGuinness & H. Kaal, eds, 'Gottlob Frege: Philosophical and Mathematical Correspondence', Blackwell, pp. 51–52.

Hilbert, D. (1904/1967*b*), *Über die Grundlagen der Logik und der Arithmetik*, *in* J. van Heijenoort, ed., 'From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931', Harvard University Press, pp. 129–139.

Hilbert, D. (1918/1967*a*), '*Axiomatisches Denken*', *Mathematische Annalen* **78**(1), 405–415. English translation in J. van Heijenoort, ed,' From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931', Oxford University Press, pp. 464–480.

Hilbert, D. (1922/2013), *Grundlagen der Mathematik*, Lecture Notes by Bernays, *in* W. B. Ewald & W. Sieg, eds, 'David Hilbert's Lectures on the Foundations of Arithmetic and Logic 1917-1933', Berlin and Heidelberg: Springer, pp. 431–527.

Horwich, P. (1997), 'Implicit Definition, Analytic Truth, and A Priori Knowledge', *Noûs* **31**(4), 423–440.

Kemp, G. (2005), 'Caesar from Frege's Perspective', *Dialectica* **59**(2), 179–199.

Kim, J. (2014), 'The Sortal Resemblance Problem', *Canadian Journal of Philosophy* **44**(3-4), 407–424.

Lewis, D. (1970), 'How to Define Theoretical Terms', *Journal of Philosophy* **67**(13), 427–446.

Lewis, D. (1983), 'New Work for a Theory of Universals', *Australasian Journal of Philosophy* **61**(4), 343–377.

Linnebo, O. (2008), 'Structuralism and the Notion of Dependence', *Philosophical Quarterly* **58**(230), 59–79.

MacBride, F. (2003), 'Speaking with Shadows: A Study of Neo-Logicism', *British Journal for the Philosophy of Science* **54**(1), 103–163.

Moriconi, E. (2003), 'On the Meaning of Hilbert's Consistency Problem', *Synthese* **137**(1-2), 129–139.

Oliver, A. (1994), 'Frege and Dummett are Two', *The Philosophical Quarterly* **44**(174), 74–82.

Parsons, C. (1965), Frege's Theory of Number, *in* M. Black, ed., 'Philosophy in America', Cornell University Press, pp. 180–203.

Peckhaus, V. (1991), '*Hilbertprogramm Und kritische Philosophie: Das Göttinger Modell Interdisziplinärer zwischen Mathematik Und Philosophie*', *Studia Logica* **50**(2), 351–354.

Poincaré, H. (1912*a*), 'The Latest Efforts of the Logisticians', *The Monist* **22**(4), 524–539.

Poincaré, H. (1912*b*), Mathematics and Logic, *in* H. Poincaré, ed., 'Mathematics and Science: Last Essays', New York: Dover Publications, pp. 65–74.

Poincaré, H. (1912*c*), 'The New Logics', *The Monist* **22**(2), 243–256.

Poincaré, H. (1952), *Science and Method*, New York: Dover Publications.

Potter, M. D. (2004), *Set Theory and its Philosophy: A Critical Introduction*, Oxford University Press.

Potter, M. & Sullivan, P. (1997), 'Hale on Caesar', *Philosophia Mathematica* **5**(2), 135–52.

Potter, M. & Sullivan, P. (2005), 'What is Wrong with Abstraction?', *Philosophia Mathematica* **13**(2), 187–193.

Pudlák, P. (2013), *Logical Foundations of Mathematics and Computational Complexity: A Gentle Introduction*, Switzerland: Springer.

Reck, E. H. (2003), 'Dedekind's Structuralism: An Interpretation and Partial Defense', *Synthese* **137**(3), 369–419.

Resnik, M. D. (1974), 'On the Philosophical Significance of Consistency Proofs', *Journal of Philosophical Logic* **3**(1), 133–147.

Resnik, M. D. (1980), *Frege and the Philosophy of Mathematics*, Cornell University Press.

Resnik, M. D. (1981), 'Mathematics as a Science of Patterns: Ontology and Reference', *Noûs* **15**(4), 529–550.

Rosen, G. (2003), 'Platonism, Semiplatonism and the Caesar Problem', *Philosophical Books* **44**(3), 229–244.

Shapiro, S. (1989), 'Structure and Ontology', *Philosophical Topics* **17**(2), 145–171.

Shapiro, S. (1997), *Philosophy of Mathematics: Structure and Ontology*, Oxford University Press.

Shapiro, S. (2000), 'Philosophy of Mathematics: Structure and Ontology', *Philosophical Quarterly* **50**(198), 120–123.

Shapiro, S. (2004), 'Foundations of Mathematics: Metaphysics, Epistemology, Structure', *Philosophical Quarterly* **54**(214), 16–37.

Shapiro, S. (2005), 'Categories, Structures, and the Frege-Hilbert Controversy: The Status of Meta-Mathematics', *Philosophia Mathematica* **13**(1), 61–77.

Sider, T. (2007), 'NeoFregeanism and Quantifier Variance', *Aristotelian Society Supplementary Volume* **81**(1), 201–232.

Sieg, W. (1998), Proof Theory, *in* 'The Routledge Encyclopedia of Philosophy', Metaphysics Research Lab, Stanford University.

Sieg, W. (1999), 'Hilbert's Programs: 1917–1922', *Bulletin of Symbolic Logic* **5**(01), 1–44.

Sieg, W. (2009), Hilbert's Proof Theory, *in* D. Gabbay, ed., 'The Handbook of the History of Logic', Elsevier, pp. 5–321.

Sullivan, P. M. (2004), Frege's Logic, *in* D. M. Gabbay, J. Woods & A. Kanamori, eds, 'Handbook of the History of Logic: The Rise of Modern Logic from Leibniz to Frege, vol. 3', Elsevier, pp. 658–769.

Trueman, R. (2014), 'A Dilemma for Neo-Fregeanism', *Philosophia Mathematica* **22**(3), 361–379.

von Plato, J. (2016), The Development of Proof Theory, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', Metaphysics Research Lab, Stanford University.

Weir, A. (2003), 'Neo-Fregeanism: An Embarrassment of Riches', *Notre Dame Journal of Formal Logic* **44**(1), 13–48.

Wright, C. (1983), *Frege's Conception of Numbers as Objects*, Aberdeen: Aberdeen University Press.

Zach, R. (2016), Hilbert's Program, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', Metaphysics Research Lab, Stanford University.