This is an updated version of the paper which appears in *Logique et Analyse* **237** (2017), 107-128

Hilbert on Consistency as a Guide to Mathematical Reality*

Fiona T. Doherty†

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" 1996. He also sets it out in his lecture "Über den Zahlbegriff" 1900 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*b*, 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

*Many thanks are due to the audience at the Mathematical Aims Beyond Justification workshop in Brussels, the Serious Metaphysics Group and the Graduate Seminar in Cambridge. I must also thank an anonymous referee for some very helpful comments. Special thanks in particular to Robin Cunnah, Ali Boyle, Adam Bales, Shyane Siriwardena, Georgie Stratham, Lukas Skiba, Alexander Greenberg, Fredrik Nyseth, Gary Kemp, Lucy Campbell, Chris Cowie, Felipe Contatto, Vera Flocke, James Hutton, Alex Grzankowski, Wolfgang Schwarz, Bryan Pickel, and Kyle Mitchell. Extra thanks to Chloé de Canson. Research funded by AHRC.

†Orcid.org/0000-0002-8493-5668. E-mail: fiona.t.doherty@gmail.com

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

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 paper will be neither to attack *nor* to defend Hilbert's view, but to discover it. As such, the guiding question of the paper 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.

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).

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 note that the focus in this paper will be on Hilbert's remark in the context of his contemporaneous writings around 1900.

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:

[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).

> **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 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:

[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.

> ...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, 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.

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 an offprint of his famous lecture "Mathematische Probleme" 1996. 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.

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

> 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, 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*a*). 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 1996, 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 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, 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.

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*a*, §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*a*, §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*b*, 42) and in his lectures even talks of "thought-objects" which are not themselves syntactic but are "denoted by a sign" (Hilbert 1904, 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" 1996 and "Über die Grundlagen der Logik und der Arithmetik" 1904. 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 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

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

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). 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.

[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). For more on the chronology of proof theory see Zach (2016) and von Plato (2016).

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.

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).

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, 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*b*, 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.

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.

[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.

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 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

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

primitive geometric terms such as "point", "line", "congruence" etc.[11] 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). Let us now 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 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.

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. In the remainder of this paper we will carefully articulate the difference between the priority reading and the misguided reading, and then make some suggestive

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

remarks about what our discussion has so far brought to light about Hilbert's early ontological position.

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 2017, 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] Elsewhere I suggest a better way of understanding Hilbert's metaphysics of mathematics is that he supports an intuitive kind of mathematical structuralism. Regardless, it is clear that 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, 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*b*, 39-40, emphasis mine).

[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).

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 paper 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.

References

Ajdukiewicz, K. (1996), 'The Logical Concept of Proof: A Methodological Essay', *Studia Logica* (19), 12–45.

Blanchette, P. A. (1996), 'Frege and Hilbert on Consistency', *Journal of Philosophy* **93**(7), 317–336.

Brown, J. R. (2005), *Philosophy of Mathematics: An Introduction to a World of Proofs and Pictures*, London: Routledge.

Ferreirós, J. (2009), 'Hilbert, Logicism, and Mathematical Existence', *Synthese* **170**(1), 33–70.

Franks, C. (2017), David Hilbert's Contribution to Logical Theory, *in* A. P. Malpass & M. Antonutti-Marfori, eds, 'The History of Philosophical and Formal Logic', Bloomsbury, pp. 1–16.

Frege, G. (1899), 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 1980', Blackwell, pp. 34–38.

Frege, G. (1900), 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 1980', Blackwell, pp. 43–48.

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.

Hilbert, D. (1899*a*), *Grundlagen der Geometrie*, Leipzig: Teubner. English translation of 10th edition by L. Unger, 1971. Chicago: Open Court.

Hilbert, D. (1899*b*), 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 1980', Blackwell, pp. 38–43.

Hilbert, D. (1900), *Über den Zahlbegriff*, *in* W. B. Ewald, ed., 'From Kant to Hilbert: A Source Book in the Foundations of Mathematics 1996', Oxford University Press, pp. 12–28.

Hilbert, D. (1904), *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. 1967', Harvard University Press, pp. 129–139.

Hilbert, D. (1918), '*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', 1967, Oxford University Press, pp. 464–480.

Hilbert, D. (1922), *Grundlagen der Mathematik*, Lecture Notes by Bernays, *in* W. B. Ewald, M. Hallett, M. Ulrich & W. Sieg, eds, 'David Hilbert's Lectures on the Foundations of Arithmetic and Logic 1917-1933. 2013', Berlin and Heidelberg: Springer, pp. 431–527.

Hilbert, D. (1996), 'Mathematical Problems: Lecture Delivered Before the International Congress of Mathematicians at Paris in 1900', *Journal of Symbolic Logic* **44**(1), 116–119.

Moriconi, E. (2003), 'On the Meaning of Hilbert's Consistency Problem', *Synthese* **137**(1–2), 129–139.

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.

Pudlák, P. (2013), *Logical Foundations of Mathematics and Computational Complexity: A Gentle Introduction*, Switzerland: Springer.

Resnik, M. D. (1974), 'On the Philosophical Significance of Consistency Proofs', *Journal of Philosophical Logic* **3**(1), 133–147.

Shapiro, S. (2005), 'Categories, Structures, and the Frege-Hilbert Controversy: The Status of Meta-Mathematics', *Philosophia Mathematica* **13**(1), 61–77.

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.

von Plato, J. (2016), The Development of Proof Theory, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', Metaphysics Research Lab, Stanford University.

Zach, R. (2016), Hilbert's Program, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', Metaphysics Research Lab, Stanford University.