Logique & Analyse **237** (2017), 107-128

HILBERT ON CONSISTENCY AS A GUIDE TO MATHEMATICAL REALITY

Fiona T. Doherty

Abstract

In his early work Hilbert puts forward the principle that in mathematics consistency is enough for existence. Moriconi (2003) claims that the standard understanding of Hilbert's contention is that he is assuming the completeness of his system. I look at the evidence for this interpretation and conclude that at the time he made this claim Hilbert had not yet developed a sophisticated conception of meta-mathematical concepts like consistency and completeness to allow him to formulate the completeness theorem. I then consider how we should understand Hilbert's contention in light of this and suggest that, for Hilbert, consistency is *conceptually prior* to existence. On the basis of this I present a new reading of Hilbert's Principle which recovers Hilbert's true contention, and along with it the philosophical significance of Hilbert's early work which – in particular – provides a new approach to questions of ontology in mathematics.

Keywords: Philosophy of Mathematics, David Hilbert, The Frege-Hilbert Controversy, Consistency, Mathematical existence, Conceptual priority.

1. Hilbert's Principle

David Hilbert is the best known proponent of the striking thesis that, in mathematics, 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*a*); in his lecture "Über den Zahlbegriff" (1900*b*) and in a letter he writes to Frege in 1899. In the letter we find the first and best known 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*a*, 39-40, emphasis mine)

doi: 10.2143/LEA.237.0.3197361.

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.

Note that Hilbert's Principle does not provide an interpretion of what Hilbert actually says in the quote. Rather, it attempts to extract a thesis from Hilbert on the basis of his remark. A lesser known proponent of the same view is Poincaré who – in his paper "Mathematics and Logic" – asserts that … in mathematics the word exist can have only one meaning, it means free from contradiction. (Poincaré 1912*b*, 454)[1]

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. This would be unproductive because Hilbert's Principle is not – by itself – a fully fleshed out thesis. It tells us nothing, for example, 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 the contention behind it. As such, the concern of this paper will be neither to attack *nor* to defend Hilbert's view but to discover it. Our guiding question will be:

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

According to Moriconi, the near unanimous answer to this question is that Hilbert's Principle is an anticipation of the completeness theorem. I will refer to this as *the standard reading* and lay it out in more detail in §2. In §3 I will argue that the best defence of the standard reading is problematic because it attributes to Hilbert a modern conception of consistency and completeness which he did not then have. In §4 I will give a new answer to (Qu) and suggest that Hilbert's Principle is a conceptual priority claim. Finally in §5 I will consider the elements of Hilbert's understanding of mathematical ontology which the new reading brings to light.

[1] 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.

2. The standard reading of Hilbert's Principle

Moriconi (2003, 130-131) claims that the "standard view" of Hilbert's Principle is that it is an anticipation of the completeness theorem and evidence that Hilbert assumed the completeness of his system.[2]

Most recently, Shapiro makes the suggestion,

> …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)

Resnik, the seminal expositor of the Frege-Hilbert controversy, claims,

> …[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)

In *Frege on the Consistency of Mathematical Theories* Dummett claims that in the Frege-Hilbert controversy, part of Frege's objection to Hilbert's view was

> …a rejection of that variety of formalism according to which the existence of mathematical entities is tantamount to the consistency of the theory relating to them. It is in fact incorrect for a first-order theory, in view of the completeness theorem, but true for a higher-order theory. Frege, however, did not know the completeness theorem for first-order logic … (Dummett 1976, 1)

This interpretation of Frege's position assumes that Hilbert's contention regarding consistency and existence was in line with what is established by the completeness theorem. All three come dangerously close to this interpretation of Hilbert's Principle:

> **Standard 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 *specified deductive system*. The second feature is that existence is understood as the existence of a *model* for those sentences, which consists

[2] With the exception of Moriconi (2003) who argues 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.

of a domain and a valuation function which maps from the domain to truth values. What we should notice is that in order for the reading to provide a plausible answer to (Qu.) it must be that in Hilbert's Principle Hilbert meant existence to be the existence of a model and consistency to be proof theoretic. To this point we will return in §3.

The standard reading renders Hilbert's Principle less ontologically exuberant and less controversial than a surface reading, depending – of course – on the ontological import of a model. However, such a principle has well known limitations as Curtis Franks points out:

> 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)

Franks – upholding the standard reading – makes the point that while Gödel's completeness theorem shows that, in the first-order case, the standard reading of Hilbert's Principle is true, Gödel's incompleteness theorems show that in the second-order case it is false. Since Hilbert was writing before Gödel's results, he was simply not aware that his principle is only true if the background logic is complete. Thus the standard reading goes hand-in-hand with the standard assessment of Hilbert's Principle as weak and philosophically uninteresting, because it was shown to be restricted only to first-order cases and because it was superseded by the completeness theorem. It is because of the standard reading that Hilbert's Principle is seen to deserve little philosophical attention. This I hope to remedy by the new reading presented in §4.

3. The problem with the standard reading

Having introduced the standard reading, this section will show why it is problematic. To do this I will offer a defence of the standard reading against its most damaging problem and conclude that even in light of the defence the standard reading is untenable.

3.1. *Early Hilbert Completeness and Consistency*

The standard reading attributes to Hilbert both a modern understanding of syntactic consistency and a modern understanding of semantic completeness.

The standard reading 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. In making such a claim Hilbert is implicitly appealing to the *semantic completeness* of his system. We have seen this to be the standard answer to (Qu).

However, Moriconi points out that this was not the conception of completeness that Hilbert had at the time. Around 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, for example of 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).

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

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. However, for the standard reading to be plausible its proponents must argue that early Hilbert already had the proof theoretic understanding of consistency and conception of semantic completeness which they attribute to him.

The reason that the standard reading has been almost unanimously accepted is that it is so tempting to read 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 to us. 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. 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).[3]

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

3.2. *In defence of the standard reading*

It is difficult to see how the standard reading could relieve this burden of proof without 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 shall do just that and seek to construct a defence of the standard reading by employing all the textual sources which I believe the proponent of the standard reading might appeal to in support of the argument that Hilbert had a proof theoretic understanding of consistency in this early period.

One substantive piece of evidence for the standard 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" (1900*a*). 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 1899, 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 qualify a conception of consistency as proof theoretic. The proof theoretic approach, of course, is the idea of investigating the properties of 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 that 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 proof theoretic we require corroborating evidence that he was in possession of these two characterising elements. Let us label these; (A) consistency as a relation holding between mere *sentences* and (B) those sentences being part of a *closed system of deductive rules* which are *formally specified*.

The evidence for (A) comes from an important passage in Hilbert's *Festschrift*, the introduction to §9 (1899*b*). 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 his consistency proofs aim to establish. 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 indeed that Hilbert speaks of sentences as the relata of the consistency relation. Strictly speaking he calls the *consequent* of the axioms sentences, but we can assume that the consequence relation holds between the same relata. This gives good contemporary textual support Hilbert had (A), i.e. that he already thought of consistency as a relation between mere sentences.

There is also evidence that Hilbert had (B), and so that his conception of consistency was proof theoretic because he understood consistency as holding between sentences which were manipulated *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).

Hilbert there offers 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*a*, 50, ft. 4)

Very significantly, 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 specified deductive system in the fact that 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 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 Hilbert makes more explicit 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. Indeed, in a sketch, this is the proof theoretic method.

This section has presented the contemporaneous textual evidence that defenders of the standard reading could appeal to in order to defend the view that Hilbert had a modern proof theoretic understanding of consistency around 1900. In the next section I will argue that a careful re-examination of the best evidence shows the evidence to be inconclusive.

3.3. *Undermining the defence of the standard reading*

In this section I will argue that although the evidence I have presented to support the standard reading appears to be very strong, it is not enough to establish that Hilbert had a 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 from Hilbert that Hilbert had a method for proving consistency, which was distinct from the model-theoretic method in *Festschrift*, and of which Frege was sceptical. 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 (A) the idea of consistency as a relation holding between sentences, came from an appeal to the introduction of §9 of Hilbert's *Festschrift*. However, the important quote used by Ajdukiewicz is actually very 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 anzugenbén, in der sämtliche Axiome der fünf Gruppen erfüllt sind. (Hilbert 1899*b*, §9)

Ajdukiewicz has mistranslated "Tatsache" as "sentence", when it is the ordinary word for *fact*.[4] This striking mistranslation is explained by the more general

[4] This is not a question of a difference in translation, any translation of *Tatsache* will render it as more than a syntactic notion.

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*b*, §9, translation mine)

Indeed, nowhere at this time does Hilbert speak of his axioms or their consequences as "sentences". Most of the time he refers to his axioms simply as "Axiome"; in §9 he refers to them as facts; and in the correspondence he calls them concepts, telling Frege:

> It is surely obvious that every theory is only a scaffolding (schema) of concepts together with their necessary connections… (Hilbert 1899*a*, 42)

In short, there is no direct textual evidence, that around 1900 Hilbert had already made the leap to (A) understanding *sentences* as the relata of the consistency relation.

We also considered evidence that Hilbert had (B) the idea of a closed system of deductive formally specified rules. This came from Hilbert's two addresses "Mathematical Problems" (1900*a*) 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 direct proof of the consistency of the axioms of arithmetic. 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 had the idea of specifying a deductive system *formally*. We have seen that what Hilbert says here is *compatible* with the proof theoretic method – and in particular (B) – but, by itself, it is too meagre to *constitute* (B). In other words, what Hilbert delivers in this address is a manifesto, not a formally specified deductive system.

In the 1904 address "Über die Grundlagen der Logik und der Arithmetik" Hilbert gives much more of a substantive account of how a direct consistency proof is to be carried out. What he presents can certainly be regarded as a sketch of the proof-theoretic method. However, taking into account some of the other aspects of Hilbert's view at the time we see that Hilbert straightforwardly lacked the tools to realise this sketch. Most significantly, around 1900 Hilbert did not yet have a rigorous logical formalism.[5] Held up by this lacking

[5] 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.

and by Poncairé'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 1920's.[6] In consequence, Hilbert would not have been able in 1904 to specify a deductive system formally, which is to say that he lacked (B).

Furthermore, it is clear from several places in the very same address that Hilbert lacked (A);

> Let an object of our thought be called a *thought-object* (*Gedankending*) or, briefly, an *object* (*Ding*) and let it be denoted by a sign… (Hilbert 1904, 131)
>
> Having thus established a certain property for the axioms adopted here, we recognize that they never lead to any contradiction at all, and therefore we speak of the thought-objects defined by means of them, *u*, *f*, and *f'*, as consistent notions or operations, or as consistently existing. (Hilbert 1904, 134)

Hilbert speaks of "thought-objects" as a technical term, not for signs themselves, but for those things *denoted* by a sign. It is again clear that at this stage Hilbert went beyond a syntactic appeal.

It is not contentious to point out that early Hilbert lacked a logical formalism and that his conception of the relata of a negative consequence relation like consistency was not purely formal. Blanchette, for instance, notes these 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)

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

Therefore, there is no conclusive textual 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 (A) and (B).

This is not to deny that Hilbert had already begun to move towards the proof theoretic approach. There is no question that the textual evidence shows that even in 1900 Hilbert thought there was *some* way of directly proving consistency without making appeal to existential assumptions. Further, he 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

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

to check whether the axioms yield any inconsistency. The most important conceptual advance that Hilbert had made around this time was to investigate 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, metamathematics was born.

However, Hilbert still thought of the axioms as loosely semantic and had no way to model the rules that governed them because he had no formal system of deduction. Therefore, at this early stage Hilbert had indeed begun to make key movements towards the proof theoretic understanding of consistency, but he was not yet in line with the now standard proof theoretic conception.

4. A new reading of Hilbert's principle

In this section I will develop the view that Hilbert intended far less by his principle than the standard reading claims. On my reading, Hilbert's Principle is best understood as a conceptual priority claim. I will show that this novel reading of Hilbert's Principle does not require attributing to Hilbert a modern conception of consistency or completeness. This will provide us with an understanding of Hilbert's Principle that is sensitive to the textual evidence that we have examined and – importantly – also brings out the philosophically interesting contention behind Hilbert's remark. It is important to bear in mind that Hilbert's Principle is not supposed to be a direct interpretation of Hilbert's quoted remarks; rather, what is at stake between different readings is the explanation of *why* Hilbert makes the remarks that he does, i.e. the answer to (Qu). Hilbert's Principle is the name I have given to any such explanation of what motivates Hilbert to take consistency as the criterion for existence in mathematics.

4.1. *The priority reading*

The contemporaneous texts that we have examined support the claim that Hilbert thought it was possible to directly prove consistency. However, the evidence is not substantial enough to support the claim that this was to be done using proof theory proper, nor that what Hilbert meant by his principle was that a syntactic consistency proof establishes the existence of a model *because of the completeness of his system*. What is needed is a reading of Hilbert's Principle that avoids attributing to early Hilbert an understanding of consistency and completeness that outstrips his methods of proof, but accommodates the fact that he had already made progress towards a proof theoretic understanding of consistency. This section will present such a reading.

The kernel of the new reading will be that – by his principle – Hilbert is advocating to Frege a particular conception of consistency that is conducive to producing a direct consistency proof. In what sense can an understanding of consistency be conducive to producing a proof? If consistency is understood as the kind of concept that can only be established by employing existential assumptions, then there can never be a *direct* proof of consistency because the proof will always be *relative* to the truth of those assumptions. In other words, it cannot be that the only way to establish that some axioms are *possibly* true together is to first establish that they are *actually* true together and have their consistency follow as a trivial consequence. On this conception (which is essentially Frege's) there is no way to avoid making existential assumptions in our consistency proof. The alternative conception of consistency, which Hilbert adopts, is one on which existence and truth are thought of as established *by consistency* rather than the other way around.

I think it becomes clear that this is Hilbert's contention when we bring in the full context of Hilbert's Principle, rather than analysing it as an isolated remark. Notice that Hilbert only offers his principle as a comment on a remark made by Frege. Hilbert identifies Frege's remark as what might prevent Frege from understanding Hilberts consistency proofs. What we have is Hilbert explaining to Frege the key to understanding his (and allegedly Cantor's) approach to consistency. But to do this Hilbert does not offer a proof sketch but something else; a particular conception of truth and existence in mathematics. Here is the complete remark made by Hilbert:

> 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*a*, 39-40, emphasis mine)

Considering Hilbert's Principle in context, I think, makes it clear his intention is to advocate a *conception* of consistency as the criterion for truth and existence, no less and no more. It also substantiates what this conception amounts to. We have said that, on the priority reading, the most important feature of Hilbert's conception is that by reversing the default understanding of the relationship between consistency and existence, Hilbert is also reversing the requirements of proof. Consistency – since it acts as its criterion – must be established *before* existence. Unpacking this important point will provide the detail of the new reading and so will be the concern of the remainder of this section.

Let us start by asking; why does Hilbert say that consistency is to be understood as the *criterion* for existence? As this forms part of an explanation Hilbert is offering to Frege, this is the same as the (Qu) *what does Hilbert mean by Hilbert's Principle?* I think we can make the contention of Hilbert's Principle entirely explicit by articulating the following conditions:

1. There is no non-circular way to establish existence which does not rely on consistency.
2. There *is* a non-circular way of establishing consistency which does not rely on existence.

We can make this 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; 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 sharing DNA and B be the concept of being siblings. Then A is *connected* to B because one can establish that two people are siblings by appeal to their DNA; and A can be established *directly* because one can establish that two people share DNA without appeal to them being siblings. Let us label these two further conditions as follows;

Connect. There is a way of establishing B using a substantive appeal to A.
Direct. There is a way of establishing A without making any appeal to B.[7]

Note that it is because we can gerrymander any proof to introduce and eliminate an appeal to any arbitrary concept, that the appeal to A must be doing *substantive* work in the proof.

If both of these conditions 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 share DNA which does not rely on them being siblings. Combining (Connect) and (Direct) with

[7] Note that although these simplified conditions will be the ones most useful to us, they should be considered as shorthand for the fully articulated conditions:

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

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

In the example we have used x would stand for a *pair* of people.

the relevant concepts *consistency* and *existence* we can now identify four conditions;

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

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

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

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

On the new reading of Hilbert's Principle, Hilbert's contention can be exhaustively characterised by his acceptance or rejection of these four conditions. I will now present evidence to show that Hilbert accepts all of the conditions except Direct (1).

The most immediate fit with what Hilbert explicitly writes is Connect (2), existence is connected to consistency. Indeed, it is hard to understand any sense in which consistency acts as the criterion for existence which does not entail this condition. We saw that in the full context of Hilbert's Principle, Hilbert gives an illustrative example in which he says that the existence of an equation's root is "proven, as soon as" the axiom 'Every equation has a root' is shown to be added without contradiction. Thus appealing to the consistency of the axiom establishes the existence of the root as proven, and does so immediately (Hilbert 1899*a*, 40). This is entirely in line with Connect (2).

Concerning Direct (2), we have seen from the textual evidence that around 1899 Hilbert thought there was *some* way of establishing consistency directly, even if he had not yet invented the proof theoretic method. Furthermore, it seems implausible that Hilbert is making the claim that consistency is the criterion for existence *and* that the criterion for consistency is existence. If this were the case Hilbert would have made no progress towards avoiding an appeal to existence to achieve a direct proof. Consistency and existence would exist in a primitive and inaccessible circle of interconnectedness. To avoid attributing Hilbert with such a weak position it must be that, for Hilbert, consistency can be established directly, i.e. Direct (2).

Connect (1) – that consistency can be established using an appeal to existence and truth – is entailed by the remark of Frege's that Hilbert quotes. It is implausible that Hilbert would deny such an innocuous principle as Connect (1). Indeed, he does not tell Frege that what he says is wrong, but instead that it was of interest to him because his own conception was "the exact reverse". Furthermore, the topic of debate between Hilbert and Frege is Hilbert's *Festschrift* in which Hilbert offers a proof which exploits the *existence* of a model to establish the *consistency* of his axioms.

I think we should characterise Hilbert as accepting Connect (1). However, this requires a hasty qualification. Attributing Connect (1) to Hilbert is not to claim that he is in agreement with Frege's view. It is to say that Hilbert's disagreement with Frege is not about whether one can infer consistency from truth but – rather – which conception of consistency is needed. For, although Hilbert agrees that consistency can be inferred from existence and truth, this is not to say that Hilbert thinks it easy to secure truth and existence in the first place.

This brings us to Direct (1), whether existence can be established directly without invoking consistency. This is the only condition that Hilbert denies. Certainly at the time of writing, Hilbert had no actual means of establishing existence directly. We must bear in mind that Hilbert was first and foremost a mathematician and as such the only legitimate means of 'establishing' existence, for him, is *proof*. Intuitively, however, existence is not the kind of concept which admits of formal proof. Even Frege has no way of establishing existence mathematically,

> 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, 43)

Rather than admit a logically unhygienic appeal to pointing at objects, Hilbert has a different conception of existence whereby existence is proven once consistency is proven. In this way he hoped to make questions of existence in mathematics tractable and rigorous. This is shown when Hilbert realises his consistency proof of his geometric axioms in his *Festschrift* builds a model from another theory – specifically a fragment of the real numbers – and so that the proof only establishes the consistency of his axioms *relative* to the theory of the Reals. As we have said, Hilbert wants a stronger result than this and the second of his 23 problems asks for a remedy to this shortfall. Most tellingly, his second problem does not posit the task of proving the *existence* of his models – but rather – the task of proving the *consistency* of the axioms of arithmetic. (Hilbert 1900*a*, §2)

Even in the course of the correspondence with Frege, Hilbert makes clear that the feature that must be considered in order to investigate existence rigorously is consistency. He tells Frege,

> 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 1899*a*, 51-52, emphasis mine)

Thus, although we can infer consistency from existence, we have no rigorous means of *proving* existence directly. Instead, existence is proven "as soon as" consistency is proven (Hilbert 1899*a*, 40). Hilbert elucidates the concept of existence by aligning it with the metamathematical notion of consistency, adopting a conception on which existence is sometimes able to be proven, by means of its criterion. In other words, to posit consistency as the criterion of existence is to identify consistency as *the very means* by which existence can be proven. This is the sense in which Hilbert is denying Direct (1).

4.2. *A new answer to (Qu.)*

Let us now step back and consider what this detailed characterisation of Hilbert's contention amounts to. Most simply, it shows that by his principle Hilbert is making a kind of *priority* claim between two concepts. Not in the sense that one concept *grounds* the other, or is *reducible* to the other, or is *contained* within the other. It is a priority claim in the sense that there is an asymmetry between the way in which these (separate) concepts are related which is a result of Hilbert's denial of Direct (1). The asymmetry is given by the fact that consistency can be proven without appeal to existence, but a proof of existence needs an appeal to consistency.[8]

This gives us a new way of answering the question, (Qu.) *what does Hilbert mean by Hilbert's Principle?* The new reading – which I call the priority reading – will answer that,

> **Priority Reading of Hilbert's Principle:** Consistency is conceptually prior to existence in mathematics.

Understanding Hilbert as making a priority claim brings out the fact that when Hilbert talks to Frege of the conception that is needed to understand his proof, he is not merely speaking about the conception of consistency but also of existence, and of how these concepts are interrelated in mathematics. On Hilbert's view, each can be used to establish the other, they exist in a state of conceptual interconnection, but our entry point to the circle is consistency.

5. The priority reading and the standard reading

Now that the new reading has been presented let us ask what, in particular, makes it different to the standard reading and why it is important to recognise this difference.

[8] Although the relation of conceptual priority is variously taken to mean an epistemic, semantic or ontic relation, all that is meant by this reading is exhausted by (1) and (2) and what we have said regarding the conditions Connect and Direct.

5.1. *The difference between the two readings*

A concern which we might have about the priority reading is that it is entirely compatible with the standard reading. For the claim that Hilbert understood consistency as conceptually prior to existence is indeed compatible with the claim that Hilbert was anticipating the completeness theorem. If Hilbert already had the conception of completeness and of consistency required to formulate the completeness theorem then – of course – he would think of consistency as established directly and as being connected to existence (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 offer an alternative to.

This worry is worth mentioning because it brings into focus the fact that the actual contention of the priority reading is that Hilbert conceived of consistency as prior to existence and that *this is the full 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* think there was a direct method of proving consistency but, at this time, there is nothing to suggest that he already had the proof theoretic conception of consistency, as characterised by (A) and (B). Therefore, Hilbert's conception at the time is not sophisticated enough for him to make the assumption that his system is complete.

This brings us to the most significant difference between the new reading and the standard reading, which is that the priority reading does not attribute to Hilbert a modern understanding of consistency and completeness. Instead, Hilbert can be thought of as explaining to Frege how the intuitive concepts of consistency and existence are best understood in mathematics. He elucidates these concepts by relating them to each other in such a way that one serves as the *criterion* for the other and so can be understood as conceptually *prior* to the other in the way we have explained. Therefore, in order to grasp Hilbert's contention, Frege does not need to understand Hilbert as using a technical conception of consistency or to be indirectly speaking of the completeness of his system. Rather, Hilbert is introducing Frege to a conception of consistency and existence whereby *consistency is conceptually prior to existence*.

5.2. *Hilbert's ontology of mathematics*

Having isolated the difference between the standard reading and the priority reading, we may still ask why this difference is important to emphasise, aside from the motivation of historical accuracy. For the priority reading has been shown to be distinct from the standard reading, but it may still seem that the difference between the two readings is not significant. However, we will see that the difference between the readings is in fact relevant to contemporary philosophy of mathematics because Hilbert's contention – as reconstructed by the priority reading – has distinct philosophical implications from the standard reading.

Recall that the standard assessment of Hilbert's Principle – which goes hand in hand with the standard reading – is that Hilbert's Principle is philosophically uninteresting. This is because it is superseded by the completeness theorem itself and because Gödel showed it to be restricted to first-order systems. On the priority reading, however, Hilbert's Principle is not an anticipation of the completeness of his system but a much more basic view of the relationship between two fundamental concepts in mathematics. Thus, precisely because of the difference between the priority reading and the standard reading the philosophical significance of Hilbert's Principle is rendered in quite a different light.

On the priority reading, Hilbert is advocating a general position in which questions of existence in mathematics can be settled by investigating consistency. Since consistency is the very criterion for existence, it seems that on Hilbert's view all consistent things must exist – in the restricted domain of mathematics. To marry consistency and existence in this way means that it is incoherent to think of the existence of inconsistent things and to think of consistent things which do not exist. If we think of consistency as aligning with *possibility* in mathematics then we might understand Hilbert as something like a *modal actualist* with respect to mathematics.

Aside from offering an interesting local modal actualism, 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. Hilbert 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. Below I rerun the passage in which Hilbert makes this point, highlighting 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*a*, 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 consistency and existence is *part of his proof* in the sense that this conception must be understood for the proofs to be understood. Furthermore, Hilbert does not lay down a definition of existence as, for example, all and only those things that are consistent. If he did this he could deduce Hilbert's Principle from the definition as a trivial consequence. Instead, Hilbert is self-consciously adopting a legitimate way of *elucidating* the concepts of existence and consistency in mathematics.

What we can take from this is that – for Hilbert – the way in which we 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 the fruitfulness of that conception for proof. More generally speaking, the value of the conception is determined by progress in the relevant theoretic endeavour. We might think of Hilbert's general attitude towards adopting a conception as a methodology-first approach. In the same way that an erroneous proof might call the underlying conception into question; if the conception is fruitful in facilitating a proof then for Hilbert the success of the proof legitimates the conception. 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 (A) and (B) of its methodology; that it operates on sentences and that it exploits a formally specified deductive system. The methodology-first approach which I am attributing to Hilbert is typically used already to characterise proof theoretic consistency, and indeed model theoretic consistency. Quite simply, the former relation is *that kind of consistency that is established by proof theoretic means*, and the latter is *that kind of consistency that is established by model theoretic reasoning*.

It is therefore necessary that we observe that Hilbert did not have a sufficiently sophisticated conception of consistency to articulate the completeness theorem, in order to acheive the theoretical space to recover the *philosophical* contention of Hilbert's Principle.

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 textual context in which Hilbert states his famous principle, it is implausible that he was anticipating the completeness theorem. For, given the development of his methodology around 1900, this is simply too sophisticated a position for Hilbert to have held, even implicitly. Taking care not to read too much modern formal equipment into Hilbert's early writings allows us 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; one which is not undermined by Gödel's incompleteness theorems.

The 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 consistency and 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 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 contribution is to nevertheless make questions of existence tractable and rigorous by means of a conception of the relationship between the two meta-mathematical concepts whereby consistency is used to guide the fruitful investigation of mathematical reality.

Acknowledgements

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, Michael Potter, 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 special thanks to Chloé de Canson. Research funded by AHRC.

References

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

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

[3] Dummett, M. (1976), Frege on the Consistency of Mathematical Theories, *in* 'Frege and Other Philosophers', Oxford University Press UK, pp. 33–46.

[4] Ferreirós, J. (2009), 'Hilbert, Logicism, and Mathematical Existence', *Synthese* **170**(1), 33-70.

[5] Franks, C. (forthcoming), David Hilbert's Contribution to Logical Theory, *in* A. P. Malpass, ed., 'The History of Logic', Continuum Press, pp. 1–16.

[6] Frege, G. (1899), The Frege-Hilbert Correspondence, *in* G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, A. Veraart, B. McGuinness and H. Kaal, eds, 'Gottlob Frege: Philosophical and Mathematical Correspondence', Blackwell, pp. 33-51.

[7] Gabriel, G., Hermes, H., Kambartel, F., Thiel, C., Veraart, A., McGuinness, B. and Kaal, H. (1980), The Frege-Hilbert Correspondence, *in* 'Gottlob Frege: Philosophical and Mathematical Correspondence', Blackwell, pp. 33-51.

[8] Hilbert, D. (1899*a*), The Frege-Hilbert Correspondence, *in* G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, A. Veraart, B. McGuinness and H. Kaal, eds, 'Gottlob Frege: Philosophical and Mathematical Correspondence', Blackwell, pp. 33-51.

[9] Hilbert, D. (1899*b*), *Grundlagen der Geometrie*, Open Court Company (This Edition Published 1921).

[10] Hilbert, D. (1900*a*), 'Mathematical Problems. Lecture delivered before the International Congress of Mathematicians at Paris in 1900', *Journal of Symbolic Logic* **44**(1), 116-119.

[11] Hilbert, D. (1900*b*), Über den Zahlbegriff, *in* W. B. Ewald, ed., 'From Kant to Hilbert: A Source Book in the Foundations of Mathematics', Oxford University Press, pp. 2-1089.

[12] Hilbert, D. (1904), Über die Grundlagen der Logik und der Arithmetik, *in* A. Krazer, ed., 'Verhandlungen des dritten Internationalen Mathematiker-Kongresses in Heidelberg vom 8. bis 13. August', Teubne: Leipzig, pp. 174-185.

[13] Hilbert, D. (1918), 'Axiomatisches Denken', *Mathematische Annalen* **78**(1), 405-415.

[14] Hilbert, D. (1922), Grundlagen der Mathematik, lecture notes by Bernays, *in* W. B. Ewald and W. Sieg, eds, 'David Hilbert's Lectures on the Foundations of Arithmetic and Logic 1917-1933', Berlin and Heidelberg: Springer, pp. 431-527.

[15] Moriconi, E. (2003), 'On the Meaning of Hilbert's Consistency Problem (Paris, 1900)', *Synthese* **137**(1-2), 129-139.

[16] Peckhaus, V. (1991), 'Hilbertprogramm und kritische Philosophie: Das Göttinger Modell Interdisziplinärer zwischen Mathematik und Philosophie', *Studia Logica* **50**(2), 351-354.

[17] Poincaré, H. (1912*a*), 'The Latest Efforts of the Logisticians', *The Monist* **22**(4), 524-539.

[18] Poincaré, H. (1912*b*), Mathematics and Logic, *in* H. Poincaré, ed., 'Mathematics and Science: Last Essays', Dover Publications, pp. 65-74.

[19] Poincaré, H. (1912*c*), 'The New Logics', *The Monist* **22**(2), 243-256.

[20] Poincaré, H. (1952), *Science and Method*, New York: Dover Publications.

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

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

[23] Sieg, W. (1998), Proof Theory, *in* 'The Routledge Encyclopedia of Philosophy', pp. Accessed October 12, 2016.

[24] Sieg, W. (1999), 'Hilbert's Programs: 1917-1922', *Bulletin of Symbolic Logic* **5**(01), 1-44.

[25] Sieg, W. (2009), Hilbert's Proof Theory, *in* D. Gabbay, ed., 'The Handbook of the History of Logic', Elsevier, pp. 5-321.
[26] Von Plato, J. (2016), The Development of Proof Theory, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy'.
[27] Zach, R. (2016), Hilbert's Program, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy'.

Fiona T. Doherty
Trinity Hall
University of Cambridge
fiona.t.doherty@gmail.com