This is a very-nearly-accurate version of the essay that appears in *The Palgrave Centenary Companion to Principia Mathematica*, B. Linsky and N. Griffin (eds), Palgrave Macmillan 2013, pp 59-78..

From Logicism to Metatheory[1]

Patricia Blanchette
University of Notre Dame

From 1914, when Behmann first lectured on *Principia* in Göttingen, to 1930, when Gödel proved the incompleteness of its system, *Principia Mathematica* played a large role in the development of modern metatheory.[2] The *Principia* system, with its explicit axiomatic approach to the fundamental principles of logic, was just what was needed to make possible the precise formulation and treatment of meta-logical questions. One might have thought, then, that at least by the time of finishing his work on *Principia*, Russell would have been in just the right position to appreciate such straightforward metatheoretical issues as those of the completeness and soundness of a logical system, of the independence of its axioms, and so on. But, notoriously, he seems curiously far removed from anything like modern metatheory. Russell never formulates a completeness theorem, or even raises anything like a modern completeness question about his system. He even seems strangely confused about what we now take to be an entirely straightforward method of proving the independence of logical axioms. In *Principles of Mathematics*, Russell remarks that

> [W]e require certain indemonstrable propositions, which hitherto I have not succeeded in reducing to less than ten. Some indemonstrables there must be; and some propositions, such as the syllogism, must be of the number, since no demonstration is possible without them. But concerning others, it may be doubted whether they are indemonstrable or merely undemonstrated; and it should be observed that the method of supposing an axiom false, and deducing the consequences of this assumption, which has been found admirable in such cases as the axiom of parallels, is here not universally available. For all our axioms are principles of deduction; and if they are true, the consequences which appear to follow from the employment of an opposite principle will not really follow, so that arguments from the supposition of the falsity of an axiom are here subject to special fallacies. Thus the number of indemonstrable propositions may be capable of further reduction, and in regard to some of them I know of no grounds for regarding them as indemonstrable except that they have hitherto remained undemonstrated. [*Principles of Mathematics* §17]

This view, that we can't use standard methods to demonstrate the independence of logical axioms, is one that Russell maintains up to and including the period of writing *Principia*.[3]

[1] A version of this essay was presented at the "PM @ 100" conference at McMaster University in May, 2010. Thanks to the organizers, Bernie Linsky and Nick Griffin, for an enjoyable and enlightening conference. And many thanks to various participants for their helpful comments; in addition to Linsky and Griffin, thanks are especially due to Sebastien Gandon, Warren Goldfarb, Greg Landini, Chris Pincock, Alasdair Urquhart, Jan Wolenski and Richard Zach. Thanks also to Paolo Mancosu for helpful comments on an earlier version of this essay.

[2] For helpful discussions of this history, see Mancosu 1999, Mancosu 2003, Mancosu/Zach/Badesa 2004, Zach 1999, Sieg 1999, Moore 1997, Awodey & Reck 2002.

[3] See *Principia Mathematica* *1, which cites this *Principia* passage approvingly. Also see Russell 1906 p. 160 footnote, and the letter to Jourdain from April 1909 (as quoted in Grattan-Guinness p. 117).

Why doesn't Russell, apparently well placed to appreciate modern metatheoretical questions and techniques, ever raise, employ, or even appear to understand them?

One answer to this question has been proposed by a group of scholars including Burt Dreben and Jean van Heijenoort, Warren Goldfarb, Tom Ricketts, and Joan Weiner. To quote the first pair:

> [N]either in the tradition in logic that stemmed from Frege through Russell and Whitehead, that is, logicism, nor in the tradition that stemmed from Boole through Peirce and Schröder, that is, algebra of logic, could the question of the completeness of a formal system arise.
>
> For Frege, and then for Russell and Whitehead, logic was universal: within each explicit formulation of logic all deductive reasoning, including all of classical analysis and much of Cantorian set theory, was to be formalized. Hence not only was pure quantification theory never at the center of their attention, but metasystematic questions as such, for example the question of completeness, could not be meaningfully raised. ... we have no vantage point from which we can survey a given formalism as a whole, let alone look at logic whole. [Dreben and van Heijenoort [1986] p. 44]

As Goldfarb puts it,

> If the system constitutes the universal logical language, then there can be no external standpoint from which one may view and discuss the system. Metasystematic considerations are illegitimate rather than simply undesirable. [Goldfarb [1979] p. 353]

In short, the early logicist position is one from which it makes no sense, because of the "universal" scope of logic as understood in this tradition, to stand back and raise evaluative questions about a given system of logic.

I would like to suggest that this cannot be the right answer to our question.[4] The subsidiary question with which we'll begin is that of whether "universalism" about logic does in fact rule out metatheory.

I. Universalism and Metatheory

(a) *What's Metatheory?*

In what follows, we'll take questions of "metatheory" to include questions about the adequacy, in various senses, of formal systems of logic. Included here are questions falling under the following three broad categories:

Reliability: Given a formal system S, one can ask whether proofs in S are *reliable* indicators of whatever S was designed to provide. The purpose of Frege's *Begriffsschrift*, for example, is the demonstration that the thought expressed by the concluding sentence of a proof is indeed a logical consequence of the thoughts expressed by its premise-sentences. Judged against this standard, the system of *Begriffsschrift* is reliable, while that of *Grundgesetze* isn't.

[4] For critical discussion of the anti-metatheory interpretation of the early logicist tradition, particularly with respect to Frege's work, see Stanley 1996, Sullivan 2005, Tappenden 1997, 2000.

The purpose of an axiomatization of geometry, on the other hand, might be the proof of the truths of Euclidean geometry; a system is reliable with respect to this standard iff nothing incompatible with a truth of Euclidean geometry is a theorem of the system. For (virtually) any formal system, consistency is a necessary condition of reliability, so that consistency itself, of various kinds, comes under the rubric of reliability results.

Modern results included under the heading of Reliability include what we now know as the soundness and consistency of formal systems.

Comprehensiveness. Questions of comprehensiveness are questions having to do with the extent to which the system in question provides *all* of the proofs it is designed to produce. One might ask whether the system includes proofs corresponding to all of the instances of logical entailment expressible in its language, whether it includes as theorems all of the truths of Euclidean geometry, and so on. Modern results falling under this heading include completeness of various kinds, including Gödel-completeness, i.e. the claim that semantic consequence in S implies provability in S.

Expressive Richness. Questions of expressive richness have to do with the extent to which the system's language and its semantic resources are up to the task of representing the logical structure of the arguments it treats. Included here are e.g. the truth-functional completeness of propositional systems, the categorical representation of important theories, and so on.

(b) *What's Universalism?*

The question we're interested in is that of whether a "universalist" view about logic or about logical systems is one from which metatheoretical questions turn out to be incoherent. The answer here will turn on what one means by "universalism." We can sketch the relevant possibilities as follows.

One sense in which one might be a "universalist" about logic is simply to hold that logic as a whole, i.e. that collection of principles underlying all correct inference, is universal in the sense that it applies everywhere, and (hence) that it serves as the grounds of all justification and explanation. Here the universality in question has to do not with a particular proposal for codifying the principles of valid inference, but with those very principles themselves, the principles in virtue of which a conclusion does or does not follow logically from a collection of premises. As Ricketts puts it, specifically with respect to Frege, the universalism in question is the view that

> Any explanation will draw on the principles of logic. In this way, logic, the maximally general science, provides a framework that embraces every science… Indeed, because of logic's maximal generality, as Frege understands justification and explanation, no other science can have justificatory or explanatory relevance to logic.[5]

[5] Ricketts [1998] p. 141 of Beaney/Reck

Frege is, to be sure, a universalist in this sense. From the Fregean perspective, logic applies everywhere, and the relation of logical entailment is essential to the justificatory force of all explanations and lines of argument. So too for Russell and Whitehead. From this form of universalism, it follows immediately that one can never engage in non-circular justifications of the fundamental principles of logic, or of, in van Heijenoort and Dreben's term, "logic as a whole." Any such justification would presuppose the very principles it seeks to justify.

But this restriction has of course nothing to do with metatheory. Metatheory is never an attempt to justify logic as a whole; it is always an attempt to evaluate particular formal systems, i.e. particular codifications of those universal logical principles. At issue in metatheoretic investigations are the virtues of the codification, not of the underlying principles being codified.[6]

A second sense of "universalism," more in line with the passage from Dreben and van Heijenoort quoted above, is that in which a formal system S is taken to be universal in virtue of being applicable everywhere, to all areas of inquiry. In this sense too, both Frege and Russell-Whitehead were certainly universalists. Frege held that his formal systems, suitably modified by adding vocabulary as required, could serve as frameworks for presenting proofs not just about arithmetic, but about any area of discourse in which rigor of proof was at a premium. Similarly for Russell and Whitehead: the theory of types and its associated formalism were to have been applicable everywhere. But this form of universalism, again, brings with it no difficulties for metatheory: that a system is (intended to be) universally applicable in this sense is no barrier to our asking meaningful questions about it, e.g. about whether it does in fact have the expressive and deductive resources required for universal applicability, about whether it is reliable, and so on.

The barrier to metatheory arises not when one takes it simply that logic in general is universal, or when one takes it that one's formal system is universally applicable, but when one holds the considerably stronger thesis about that formal system that its derivations offer the *only* way of presenting compelling or scientifically-acceptable arguments. Where S is the formal system in question, call this thesis "exclusivism about S."

Exclusivism about S causes problems for metatheory in two ways.

The first difficulty is one of circularity: If the only way to present legitimate or scientifically-compelling arguments is via derivations in S, then all attempts to give such arguments presuppose the reliability of S (in either the strong form that derivability in S guarantees logical validity, and hence that S's theorems are truths of logic, or in the weaker sense that derivations in S are truth-preserving and hence that its theorems are true). Hence to try to argue in a legitimate or scientifically-compelling way for S's reliability would be, by the universalist's lights, to reason in a very small and vicious circle. I take it that this is what Ricketts has in mind when making the following claim about Frege:

[6] For a nice discussion of this issue, see Peter Sullivan 2005.

> A formalized proof of the soundness of Frege's codification of logic would thus presuppose the truth of the axioms and the soundness of the inference rules of that codification. Given Frege's view of justification as explanation within unified science, such a proof would be scientifically pointless. ([1998] p. 177.)

That is, because Frege is an exclusivist about the completed, universal formal system that grows out of *Grundgesetze,* it's impossible to give a scientifically-acceptable, or a justification-conferring, demonstration of the reliability (or the soundness) of that system or any fragment thereof.

It is worth noting that this circularity-argument does not rule out all metatheory. The metatheoretic claims one might make about a formal system can be divided into two camps as follows. The first contains just those claims whose truth is presupposed when we treat derivations within the system as expressing justification-conferring proofs. Included here are most obviously the reliability of the system, its consistency, the truth-preservation of its derivations, and various forms of soundness. Into the second camp fall those claims about formal systems that are not presupposed in making such justificatory appeal to its derivations, i.e. those metatheoretic claims about a system whose falsehood regarding that system does not undermine its reliability. Included here for example are various claims of comprehensiveness for the system, and completeness in the modern sense. That a system fails to include *all* of the derivations of a certain class does not mean that the derivations it does include are in any way faulty. The exclusivist position regarding a formal system S immediately and obviously entails that those metatheoretic claims about S that fall into the first camp – the "reliability" camp – cannot be non-circularly demonstrated. But there is no such quick route from exclusivism about S to the indemonstrability of those second-camp claims (those from the "comprehensiveness" camp) about S. That we would have to presuppose the *reliability* of S in order to demonstrate its comprehensiveness or its completeness – in the same way that S's reliability would be presupposed in order to demonstrate anything at all, on this line of argument - does not make such demonstrations circular.[7]

The second difficulty posed for metatheory by this form of universality is that of semantic paradox. In order to obtain from the system of *Grundgesetze* or of *Principia* a formal system U sufficient not just for e.g. physics and astronomy, but also for semantics, one would need to add primitive terms for, amongst other things, the fundamental properties of interest in metatheory. And we know that, when the subject of the metatheoretic investigations is U itself, this can't be done. If U is consistent, then it must lack some of the resources (e.g. a well-behaved negation

[7] That is: the quick argument just outlined which gets one from exclusivism about S to the circularity of demonstrations of S's reliability is not available in the case of demonstrations of S's comprehensiveness etc. This does not mean that there is no argument from this premise to the conclusion in question: one might well wonder whether the exclusivist position makes all metatheoretic claims or demonstrations problematic on the basis of difficulties about self-reference, since from the exclusivist position it follows that arguments about S will be carried out in S. We turn to this question below.

operator, referring terms for its own formulas, generous rules of sentence-formation, a well-behaved truth-predicate, and so on) that one might reasonably take to be essential for significant metatheory. If the exclusivist holds that the only scientifically-acceptable investigations are those that can be carried out in a comprehensive deductive system, then he holds a view from which very little metatheory about that system or fragments thereof can count as scientifically acceptable.

In short: if Frege, or Russell and Whitehead, had been exclusivists about their (or about any) formal systems, then they would have occupied a position from which the pursuit of metatheoretical questions about those systems would have been clearly incoherent.

<u>II. Russell and Metatheory</u>

The first thing to note about exclusivism is that it carries the immediate consequence that no arguments, and hence no mathematics, can have provided justification for their conclusions prior to the advent of the formal system in question. Neither Frege nor Russell holds this. Despite recognizing in the usual way that some mathematical arguments have been flawed, neither Frege nor Russell advocates the wholesale indictment of the history of mathematics prescribed by exclusivism. With regard e.g. to Dedekind's proof procedure, one which is considerably less rigorous than is his own, Frege remarks that

> [H]is procedure may have been the most appropriate for his purpose; …. The length of a proof ought not to be measured by the yard. It is easy to make a proof look short on paper by skipping over many intermediate links in the chain of inference and merely indicating large parts of it. Generally people are satisfied if every step in the proof is evidently correct, and this is permissible if one merely wishes to be persuaded that the proposition to be proved is true.[8]

His own extraordinarily-careful procedure, as Frege sees it, goes well beyond what's required simply to establish the truth of the propositions proven. His logicist purposes, which include not just the demonstration of the truth of arithmetical claims, but a clear exhibition of the fundamental grounds of each of the (typically already-known) premises, requires that "considerably higher demands must be placed on the conduct of proof than is customary in arithmetic."[9] The rigor imposed by expressing one's proofs as derivations in a Frege-style formal system is, in short, significantly more demanding than is required of arguments sufficient for establishing the truth of their conclusions, even within the demanding field of mathematics.

Frege does not claim that the work of previous mathematicians must be re-cast as formal *Grundgesetze*-style derivations in order to be persuasive, and does not claim that careful arguments couched in ordinary language or in ordinary-cum-mathematical language are unscientific or otherwise illegitimate. This is as one should expect of an author whose careful

[8] Gg I Introduction p. viii.
[9] Gg I §0, p.1

philosophical arguments in the non-formal *Grundlagen*, e.g. to the effect that numbers are objects, that statements of number are assertions about concepts, that arithmetic is not about ideas, and so on, are intended to be taken seriously.

Frege's careful arguments in *Grundgesetze* also involve, it's worth noting, a good deal of metatheory. In Vol I §10, prior to providing a stipulation governing courses-of-value, Frege provides a careful and deliberate argument that the system as defined to this point can be consistently extended by the addition of just such arbitrary stipulations. In §11, he argues that the definite-description operator is well-defined, i.e. that terms formed via the use of this operator are neither empty nor ambiguous. A similar argument at Vol I §35 establishes that the $\cap$ symbol is well-defined, which, as Frege notes, is essential for "the correctness of … proofs." In §§29-31 he provides a careful (if flawed) proof that function-names and singular terms of *Grundgesetze* are all uniquely denoting. And so on. Frege is no stranger to the practice of standing back and viewing his formal system as a tool, to raising and answering questions of adequacy about that tool, or to providing detailed and persuasive arguments for those answers.

Turning to Russell: It is worth recalling that having noticed the inconsistency in Frege's formal system, Russell was by 1902 in a position to appreciate vividly the importance of the consistency-question for formal systems of logic. It would have been very strange for him to have failed to recognize the importance of just that question for his own system.

And indeed, Russell is quite clearly interested in straightforward metatheoretical questions, including that of consistency. In the early pages of *Principia*, we find an explicit concern with consistency and comprehensiveness:

> The proof of a logical system is its adequacy and its coherence. That is: (1) the system must embrace among its deductions all those propositions which we believe to be true and capable of deduction from logical premises alone … and (2) the system must lead to no contradictions, namely in pursuing our inferences we must never be led to assert both *p* and not-*p*, *i.e.* both "|-.p" and "|- .~p" cannot legitimately appear.[10] [*PM* pp 12-13]

In the *Principles of Mathematics*, having introduced the propositional calculus, Russell remarks that

> From this point we can prove the laws of contradiction and excluded middle and double negation, and establish all the formal properties of logical multiplication and addition – the associative, commutative and distributive laws. Thus the logic of propositions is now complete. [PoM §19]

And after the additional introduction of the calculus of classes and the calculus of relations:

> Mathematics requires, so far as I know, only two other primitive propositions, the one that material implication is a relation, and the other that $\in$ … is a relation. We can now develop the whole of mathematics without further assumptions or indefinables. [PoM §30]

[10] *PM* Ch. 1, pp 12-13 (pagination from the 1962 edition (to *56)).

In the Introduction to the second edition of *Principles*, Russell notes the difficulty of establishing comprehensiveness:

> [I]t is difficult to see any way of proving that the system resulting from a given set of premises is complete, in the sense of embracing everything that we should wish to include among logical propositions. (PoM, Intro to 2nd Ed'n; xii)

In the introduction to the second edition of *Principia*, the concern with the comprehensiveness of the propositional fragment comes out as follows, just after the introduction of Nicod's proposition:

> From the above proposition, together with the rule of inference, everything that logic can ascertain about elementary propositions can be proved, provided we add one other primitive proposition… [PM, Intro to 2nd Ed, p. xix]

The concern for reliability is often mentioned in conjunction with that of comprehensiveness, as in this passage from the Preface to *Principia*:

> In constructing a deductive system such as that contained in the present work, there are two opposite tasks which have to be concurrently performed. On the one hand, we have to analyse existing mathematics, with a view to discovering what premisses are employed, whether these premisses are mutually consistent, and whether they are capable of reduction to more fundamental premisses. On the other hand, when we have decided upon our premisses, we have to build up again as much as may seem necessary of the data previously analysed… It is not claimed that the analysis could not have been carried further: we have no reason to suppose that it is impossible to find simpler ideas and axioms by means of which those with which we start could be defined and demonstrated. All that is affirmed is that the ideas and axioms with which we start are sufficient, not that they are necessary. (PM, Preface (1910) pp v-vi.)

On the next page, we find a nice expression of Russell's picture of the role of the formalism and of the underlying type theory: in addition to pointing out the importance of consistency, Russell offers essentially the claim that, as restricted to mathematics, the theory of types is conservative over ordinary mathematics:

> The particular form of the doctrine of types advocated in the present work is not logically indispensable, and there are various other forms equally compatible with the truth of our deductions. … [H]ardly anything in our book would be changed by the adoption of a different form of the doctrine of types. In fact, we may go further, and say that, supposing some other way of avoiding the contradictions to exist, not very much of our book except what explicitly deals with types, is dependent upon the adoption of the doctrine of types in any form, so soon as it has been shown (as we claim that we have shown) that it is *possible* to construct a mathematical logic which does not lead to contradictions. It should be observed that the whole effect of the doctrine of types is negative: it forbids certain inferences which would otherwise be valid, but does not permit any which would otherwise be invalid. Hence we may reasonably expect that the inferences which the doctrine of types permits would remain valid even if the doctrine should be found to be invalid. [PM Preface (1910) p. vii.]

Reliability and comprehensiveness come up together again at the beginning of Part I, Section A ("The Theory of Deduction") of *Principia*:

> [E]very deductive system must contain among its premises as many of the properties of implication as are necessary to legitimate the ordinary procedure of deduction. In the present section, certain propositions will be stated as premises, and it will be shown that

> they are sufficient for all common forms of inference. It will not be shown that they are all *necessary*, and it is possible that the number of them might be diminished. All that is affirmed concerning the premises is (1) that they are true, (2) that they are sufficient for the theory of deduction, (3) that we do not know how to diminish their number. [*PM* Part I, Section A; *p. 90]*

As this small sample should make clear: though Russell's way of treating metatheoretical questions is not up to modern standards, and his terminology doesn't always match ours, he has no aversion either to adopting the metatheoretical "perspective" in general, or to raising specific metatheoretical questions about his own formal system. We can't attribute either Russell's failure to formulate a modern completeness question, or his puzzling remarks about independence proofs, to a general logicist-inspired inability to make sense of metatheory.

III. But what then?

That Russell doesn't formulate what we now know as the "completeness" question for formal systems is a straightforward result both of timing and of his conception of logical entailment, i.e. of that relation which it's the purpose of a formal system to formalize. His concern with the comprehensiveness of his formal system is the concern that it be rich enough to "legitimate the ordinary procedure of deduction,"[11] that it "embrace among its deductions all those propositions which we believe to be true and capable of deduction from logical premises alone,"[12] that it "embrac[e] everything we should wish to include among logical principles,"[13] and so on. That is to say, while the purpose of the formal system is to formalize all legitimate logical inferences, Russell doesn't provide a reductionist account of this class of inferences. There is no counterpart in Russell's work to the idea of logical entailment as reducible to truth-preservation across structures or interpretations. One straightforward reason for this is simply chronological: the fruitfulness of this now-canonical semantic approach was still several years in the future. Perhaps more importantly: the idea of the formal language as reinterpretable, while not foreign to Russell (see below) was never a focus of his concern. In Russell's view, the important logical properties of his formalism were never those having to do with characteristics preserved under reinterpretation, but were instead those having to do with the language under its single canonical interpretation. As Goldfarb has emphasized, logic for Russell had to do with the kinds of relations that linked fully-interpreted sentences or the things expressed by them, and was not a matter of relations between reinterpretable pieces of syntax. This is of course not a point about the coherence of metatheory, but about how from a Russellian point of view the metatheoretical question of comprehensiveness might be framed. Because Russell doesn't understand entailment in terms of the behavior of formulas across interpretations, the question of the Gödel-

[11] PM p. 90
[12] PM p 12.
[13] PoM, Intro to 2nd Edn, xii.

completeness of a system, while not an unintelligible or uninteresting one from the Russellian point of view, cannot from this viewpoint serve as a version of the all-important question of comprehensiveness.[14]

Let's return to the 1903 passage quoted above regarding independence-proofs, a passage taken to indicate Russell's inability to view the axioms of his system as objects of investigation. The heart of the passage is as follows:

> [T]he method of supposing an axiom false, and deducing the consequences of this assumption, which has been found admirable in such cases as the axiom of parallels, is here not universally available. For all our axioms are principles of deduction; and if they are true, the consequences which appear to follow from the employment of an opposite principle will not really follow, so that arguments from the supposition of the falsity of an axiom are here subject to special fallacies.

At first glance, Russell seems to be claiming that it's impossible to do just what Bernays did in 1918, namely to establish the independence of various of the logical axioms of *Principia* from one another by constructing models.

But the method of constructing models in the modern sense is pretty clearly not what Russell is talking about here.

An independence proof of what we'll call the "modern" kind is a proof that turns on an assignment of values to the sentences in question, usually via a reinterpretation of important parts of the vocabulary of those sentences. Where the goal is to prove the independence of a sentence I from a set S of sentences, the fundamental idea is the assignment of values to I and to the members of S in such a way that (a) there's a particular value V assigned to all of the members of S, and known to be preserved under the relation of deductive consequence; and (b) I is not assigned V. It follows immediately that I is not a deductive consequence of S. The most familiar method of assigning values is to assign an interpretation to the non-logical terms of the sentences in such a way that each member of $S \cup \{\sim I\}$, as reinterpreted, expresses a theorem of a favored background theory B. Because the theorems of B are closed under deductive consequence, we know immediately that the value "theorem of B" is assigned not just to each member of $S \cup \{\sim I\}$, but also to all of the deductive consequences of S. And if B is consistent, then the value "theorem of B" is not assigned to I, with the immediate result that I is not a deductive consequence of S.

Compare the method Russell refers to. When Russell talks here about the method used to prove the independence of the parallels postulate, he talks about a strategy of "supposing an axiom false, and deducing the consequences of this assumption." Notice that the modern model-theoretic strategy, in which the axiom-sentence in question is interpreted in such a way as to

[14] Russell's lack of interest in what we now think of as completeness questions comes out nicely in Peter Milne's [2008]. As Milne shows here, Russell demonstrated in [1906] essentially the crucial elements of a kind of completeness proof for propositional logic, but never understood what he had done in this light.

express the negation of a theorem of the background theory, involves no such deduction of consequences. In proving via the modern procedure the independence of the parallels postulate from the other axioms of Euclidean geometry, we don't suppose the axiom false and then deduce consequences from this assumption; we instead re-interpret the sentence expressing this axiom, and conclude immediately from the reinterpretation, as above, that the axiom is independent if the background theory is consistent. No deductions from the negated axiom, or from the assumption of its falsehood, enter into the demonstration at any point.

But this modern procedure requires the relatively late understanding of the language in question, and of the axioms whose independence is at issue, as reinterpretable frameworks, as syntactic "scaffoldings" (to use Hilbert's term) whose terms stand as place-holders for assigned content. It's only under this understanding of the formulas in question as freely-reinterpretable forms that the modern strategy of independence demonstrations can get off the ground. The idea that the parallels axiom can be shown to be independent of the other Euclidean axioms via an interpretation under which these axiom-sentences express claims having no geometric content whatsoever can only make sense after the logical structure of the geometric theory has become understood as entirely divorced from its geometric content.

The strategy Russell refers to is an older one, one familiar from the kinds of independence proofs for the parallels postulate provided by figures like Gauss, Bolyai, Lobachevsky, and Beltrami. Here the idea is that we prove the parallels postulate to be independent of the other axioms of Euclidean geometry by constructing or describing a "manifold," a kind of space, or a surface on which the other Euclidean axioms hold but on which the parallels postulate is false. The subject-matter of this described or constructed space or surface is still geometric, and while it's often appropriate to understand what's going on as involving some re-interpretation of terms (so that e.g. the term "line" now refers to a circumference of a sphere), the ordinary contents of most terms are retained, and the described surface (for example) stands as a representative of how a plane could in principle be configured. The "assumption" of the negation of the parallels postulate is the assumption that the space or surface as described will in fact falsify the parallels postulate. And the reason it's important to deduce the consequences of this assumption is that it's only by deducing the consequences of this assumption (taken together with the fundamental assumptions about the space or surface given in its original description) that one can come to a reasonable conviction that the space described is in fact coherent, i.e. that in supposing the space to satisfy the descriptions proposed (including that of the negation of the parallels postulate), one has not engaged in self-contradiction.

One way to put the central difference between the two kinds of proof, and the different independence-questions answered by them, is to put the issue in Frege's terms, in terms of the thoughts or nonlinguistic propositions expressed by the geometric sentences in question. As far as Frege is concerned, the independence of the parallels postulate is a matter of the non-provability of a given proposition, the parallels-postulate proposition, from the other Euclidean-axiom propositions. From his point of view, the *sentences* of Euclidean geometry are of merely secondary concern, and the derivability-relations between them are not the focus of interest in an independence-inquiry. What matters is the (typically richer) collection of logical relationships between propositions expressed. An independence-proof of a modern kind, one which involves a wholesale reinterpretation of the geometric vocabulary, is from this point of view a non-starter. For instead of dealing directly with the genuine geometric propositions, this kind of proof assigns to the geometric sentences an entirely new set of propositions, ones concerning e.g. constructions on the real numbers. Though of course this modern strategy is effective at demonstrating (relative) non-deducibility results between sentences, it fails badly from Frege's point of view, since it fails to demonstrate what he was primarily concerned with, namely non-provability results as these obtain not between bare sentences, but between geometric propositions. In short, while the old strategy, turning on the construction or description of a kind of space, delivers results about geometric propositions and possible combinations of truth-values thereon, the new (and considerably more tractable) strategy, using reinterpretations of the language, delivers results about partially-interpreted sentences and their deducibility-relations. It's for this reason that Frege rejects Hilbert's independence proofs, pointing out, correctly, that they don't address the question that Frege takes to be at issue, namely a rather old-fashioned (if nevertheless attractive) question not about theories construed as sets of sentences or multiply-instantiable scaffoldings, but about theories construed as sets of propositions, ones whose logical complexity does not always go hand in hand with the syntactic complexity of the sentences used to express them.[15]

Russell's attitude to modern-style independence proofs is not as uniformly negative as is Frege's, but he is also not entirely sanguine about the new methods. Russell shares, at least at some points, Frege's view of the propositions of mathematics as nonlinguistic entities, as opposed to the sentences to which modern techniques most immediately apply. As Russell puts it in *Principles* §13, commenting on McColl's treatment of variables:

> "[McColl] is led to speak of propositions as sometimes true and sometimes false, which of course is impossible with a genuine proposition.

At least here, propositions for Russell are things whose truth-value could have been different had things gone differently; but they are not the kinds of things that survive re-interpretation over different domains. In a similar vein, Russell's geometrical writings from the early monograph on the foundations of geometry to the much later Encyclopedia Britannica article regularly refer only

[15] See Frege's correspondence with Hilbert in Frege 1980; also Frege 1903, Frege 1906. For discussion of this issue, see my 1996, 2007.

to independence-proofs of the old kind, characterizing models of non-Euclidean geometries as representations of different kinds of space in which the parallels postulate is false.[16]

Russell would appear, additionally, to share Frege's view of Hilbert's work on the foundations of geometry. Recall that Frege criticizes Hilbert's method for its failure to show anything about the "real" axioms of geometry, i.e. the propositions expressed under the standard interpretation of the geometric sentences. As Frege sees it, Hilbert's re-interpretation technique results in demonstrations that leave the important questions (of consistency and independence for propositions) behind. In a 1904 letter to Couturat, Russell says:

> As for Frege, I have already seen his articles on geometry; I am completely in agreement with the opinions that he expresses there.[17]

But in some areas, Russell is quite happy with the new, interpretation-theoretic method for proving independence. In *Principles* *121, Russell cites with approval Peano's method of proving the independence of each of the Peano postulates from the others, a method which turns on re-interpretations of the arithmetical vocabulary.[18] More dramatically, Russell endorses at *22 of *Principia* Huntington's reinterpretation method of demonstrating the mutual independence of Huntington's own axioms for the theory of classes.[19] Here, though e.g. "∪" is used to indicate the union of two classes, Huntington assigns to (his version of) this symbol a variety of different relations (or rather their extensions) in order to demonstrate the various independence-claims.

The difference between contexts in which reinterpretations make sense, and those in which they are incoherent from Russell's point of view, comes out most vividly in his account of the two different ways in which we can view the language of formal logic. In a passage at the beginning of *4 of *Principia*, which is taken over almost verbatim from [1906], Russell writes:

> In this number, we shall be concerned with rules analogous, more or less, to those of ordinary algebra. It is from these rules that the usual "calculus of formal logic" starts. Treated as a "calculus," the rules of deduction are capable of many other interpretations. But all other interpretations depend on the one here considered, since in all of them we deduce consequences from our rules, and thus presuppose the theory of deduction. One very simple interpretation of the "calculus" is as follows: [Russell here gives an algebraic interpretation of the propositional connectives.] Symbolic logic considered as a calculus has undoubtedly much interest on its own account; but in our opinion this aspect has hitherto been too much emphasized, at the expense of the aspect

[16] See e.g. Russell 1897 Chapter I.
[17] Russell to Couturat of April 4, 1904. Translation by Jolen Galaguer. Presumably the "articles" Russell refers to are the series published in 1903. Thanks to Jolen Galaguer for pointing out to me this passage.
[18] Also see *Principles* section 377 for Russell's discussion of the independence of the axioms of descriptive geometry.
[19] Russell notes that the axioms are Huntington's. At *22.05, he notes that "The form of the above postulates is such that they are mutually independent, *i.e.* any nine of them are satisfied by interpretations of the symbols which do not satisfy the remaining one." One interesting question, which we won't pursue here, is whether there's any significance to Russell's having asserted independence of the *form* of the postulates, rather than of the postulates themselves.

> in which symbolic logic is merely the most elementary part of mathematics, and the logical prerequisite of all the rest. [PM *4 (Summary); see [1906] p. 183.]

We can, in short, view the axioms of formal logic (in this case, roughly that of propositional logic) either as formulas of a mere calculus, susceptible to multiple interpretations, *or* as expressing the interpretation relied upon in *Principia*, on which they together express the fundamental principles of logic. Viewed in the former way, one can presumably give Huntington-style demonstrations of consistency and independence. But any such demonstrations, importantly, will be just about the bare and (from Russell's point of view) relatively uninteresting formalism; they won't be about the fundamental principles of logic themselves. If we want to reason about the fundamental principles of logic, we can't do so by examining the behavior of a "calculus" under varying interpretations. For once we reinterpret the calculus, we're not talking about the principles of logic at all.

This means that the only way to use the "geometric method" to demonstrate the independence of a given principles of logic from the others would be to proceed in the Gaussian way: i.e. to describe and then reason about an arena in which the target principle does not hold. (For to proceed in the modern way, i.e. via re-interpretations of vocabulary, would be, from Russell's point of view, to reason about something other than the principles of logic.) But – and this is Russell's point – such a Gaussian approach makes no sense as applied to principles of logic. The methodology of such a proof involves the requirement that the depicted state of affairs is a coherent one. It's essential to ask which propositions are true and which false in that state of affairs, and essential that the consideration of the depicted arena doesn't involve any alteration in the relations of logical entailment that hold between the propositions in question. We need to be able to reason about that arena in a way that begins with the propositions giving its fundamental properties – e.g. that the space has a particular kind of curvature – and that draws conclusions about which of the further propositions of interest are, in such a situation, true or false. The fact that the premise-propositions are true and the conclusion-proposition false in the depicted situation is the indication that the latter proposition isn't entailed by the former. This line of reasoning breaks down when the depiction of the purported state of affairs is incoherent. And this is the case when the depiction involves the denial of a logical principle. If we try to ask which propositions are true and which false in a situation in which both a proposition and its negation are true, or one in which a conjunction is true while its conjuncts are false, we've asked a question with no sensible answer.

Russell's description of the geometric method of proving independence in the 1903 passage should sound strange to a modern ear. He describes it as "the method of supposing an axiom false, and deducing the consequences of this assumption." When we proceed in the modern way via reinterpretations of vocabulary, no such deductions are necessary: we begin with an initial set of assignments to primitive vocabulary, which assignments induce a valuation on the whole language, governed by the semantic rules for that language. While we do (often) assign values in such a way that the sentence I is assigned the value "false," there is no point at which, on the reinterpretation technique, we reason about the consequences of assuming that the proposition ordinarily expressed by I is false. This kind of reasoning, which would indeed be incoherent were that proposition a fundamental, general law of logic, plays no role in the reinterpretive method of demonstrating independence. But that this is precisely what we do when demonstrating independence via reasoning about a described alternative space (or other arena) explains not only Russell's remark about such reasoning, but also his rejection of it in the case of logical principles.[20]

Russell's view of the effectiveness of the reinterpretation strategy stands at an interesting halfway-point between the entirely dismissive attitude of Frege and the modern, freewheeling attitude of early 20th-century Göttingen. The axioms of arithmetic and of the theory of classes are sufficiently closely identified with their canonical syntactic representations that the modern reinterpretation strategy as applied to these sentences establishes, as far as Russell is concerned, the independence of those axioms from one another. Russell is in this sense thoroughly modern. Frege would presumably not approve. But when it comes to the fundamental principles of logic, Russell is a Fregean: the principles themselves are not to be identified with formulas of the calculus, but are what's expressed under only one, i.e. the canonical, interpretation of the logical connectives. The reinterpretation strategy, accordingly, despite its intrinsic interest, can't tell us anything about the independence of the real principles of logic.

In 1909, Jourdain asks Russell about independence proofs as follows:

> When you enumerate the primitive propositions of *logic*, do you prove their independence by the usual method of giving certain interpretations to the primitive ideas, so that all but one (in turn) of the primitive propositions is verified? (Jourdain to Russell 28 April 1909, as reported in Grattan-Guinness p. 117)

Russell's reply is as follows:

---

[20] Here is Russell's description in his *Foundations of Geometry* monograph of early independence-arguments in geometry:

> A bolder method, suggested by Gauss, was carried out by Lobatchewsky and Bolyai. If the axiom of parallels is logically deducible from the others, we shall, by denying it and maintaining the rest, be led to contradictions. These three mathematicians, accordingly, attacked the problem indirectly: they denied the axiom of parallels, and yet obtained a logically consistent Geometry. They inferred that the axiom was logically independent of the others, and essential to the Euclidean system. [*FG* p. 8]

That independence is "proven" by denying the parallels postulate and following out the consequences of this denial without running into contradictions was a standard view at the time. See e.g. Hoüel 1867, p. 77, as quoted in Stump 2007:

> J. Bolyai and Lobachevskii drew consequences from this supposition, without ever finding themselves in contradiction with logic, but only with experience…

As Stump points out [Stump 2007], this less-than-conclusive strategy is shored up by the construction of non-Euclidean surfaces, but only becomes the modern and decisive proof-procedure familiar today after the adoption of a "formal" approach to the language of geometry, and the adoption of reinterpretive models around the turn of the century.

> I do not prove the independence of primitive propositions in logic by the recognized methods; this is impossible as regards principles of inference, because you can't tell what follows from supposing them false: if they are true, they must be used in deducing consequences from the hypothesis that they are false, and altogether they are too fundamental to be treated by the recognized methods. (Russell to Jourdain, April 1909, as reported in Grattan-Guinness p. 117)

What I hope to have clarified is that this attitude on Russell's part was not part of any hostility to metatheory in general, and that though it perhaps indicates a certain shortsightedness with respect to the then-emerging interpretation-theoretic use of models, nevertheless it is an entirely reasonable view regarding the applicability of an older geometric technique to questions of independence in logic.

References

Awodey and Reck 2002: "Completeness and Categoricity, Part I: Nineteenth-century Axiomatics to Twentieth-Century Metalogic," *History and Philosophy of Logic* 23 (2002) 1-30.

Beaney and Reck 2005: *Gottlob Frege: Critical Assessments of Leading Philosophers*, Routledge

Blanchette 1996: "Frege and Hilbert on Consistency," *The Journal of Philosophy* XCIII (July 1996), 317-336.

Blanchette 2007: "Frege on Consistency and Conceptual Analysis," *Philosophia Mathematica* 15, 3 (2007) 321-346;

Blanchette: "The Frege-Hilbert Controversy" in the *Stanford Encyclopedia of Philosophy*

Dreben and van Heijenoort - "Introductory Note to [Gödel] 1929, 1930 and 1930a" in Gödel's Collected Works Vol I, 44-59.

Frege [1893/1903] – *Grundgesetze der Arithmetik*. Partial English translation as *The Basic Laws of Arithmetic*, M. Furth (trans), University of California Press 1964.

Frege [1903] - "On the Foundations of Geometry" – First Series. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 12, pp 319-24, 368-75. English translation in Frege [1984] pp 273-284.

Frege [1906] – "On the Foundations of Geometry" – Second Series, *Jahresbericht der Deutschen Mathematiker-Vereinigung* 15, pp 293-309, 377-403, 432-30. English translation in Frege [1984] pp 293-340.

Frege [1980] - *Philosophical and Mathematical Correspondence*, Gabriel, et al eds. (Oxford: Blackwell, 1980)

Frege [1984] - *Collected Papers on Mathematics, Logic and Philosophy*, McGuinness ed. (Oxford: Blackwell, 1984)

Goldfarb 1979: "Logic in the Twenties: the Nature of the Quantifier" *The Journal of Symbolic Logic* Vol 44 #3 (Sep 1979) 351-368.

Grattan-Guinness 1977: *Dear Russell – Dear Jourdain*, Columbia University Press

Mancosu 1999: "Between Russell and Hilbert: Behmann on the Foundations of Mathematics," *The Bulletin of Symbolic Logic* Vol 5 #3 (Sep 1999), 303-330.

Mancosu 2003: "The Russellian Influence on Hilbert and His School," *Synthese* 137 (2003), 59 - 101.

Mancosu, Zach, Badesa 2004: "The Development of Mathematical Logic from Russell to Tarski: 1900-1935" in *The Development of Modern Logic*, Haaparanta (ed), New York and Oxford: Oxford University Press, 2004

Milne 2008: “Russell's Completeness Proof” *History and Philosophy of Logic* 29:31-62.
Moore 1997: “Hilbert and the Emergence of Modern Mathematical Logic,” *Theoria* vol 12 #1, pp 65-90.
Ricketts 1995: "Frege, the Tractatus, and the Logocentric Predicament," *Nous* Vol 19, #1, 3-15.
Ricketts 1998: “Frege’s 1906 Foray Into Metalogic,” *Philosophical Topics* 25, 169-188; reprinted in Beaney & Reck Vol II, 136-155.
Russell 1897: *An Essay on the Foundations of Geometry*, reprinted in 1956 by Dover.
Russell, *Principles of Mathematics* ([PoM])
Russell 1906: “The Theory of Implication,” *American Journal of Mathematics*, Vol 28 #2, pp 159-202.
Russell and Whitehead 1910-1913 ([PM]): *Principia Mathematica*, Cambridge University Press. Page references are to the abridged *Principia Mathematica to *56.*
Stanley 1996: “Truth and Metatheory in Frege,” *Pacific Philosophical Quarterly* vol 77, pp 45-70.
Sullivan 2005: “Metaperspectives and Internalism in Frege,” Beaney and Reck Vol II pp 85-105.
Tappenden 1997: “Metatheory and Mathematical Practice in Frege,” *Philosophical Topics* vol 25, pp 213-64.
Tappenden 2000: “Frege on Axioms, Indirect Proof, and Independence Arguments in Geometry: Did Frege Reject Independence Arguments?” *Notre Dame Journal of Formal Logic* 41 #2, 271-315.
Sieg 1999: "Hilbert's Programs: 1917-1922" *The Bulletin of Symbolic Logic* Vol 5 #1 (Mar 1999), 1-44.
Weiner 1990: *Frege in Perspective*, Ithaca: Cornell University Press
Zach 1999: "Completeness Before Post: Bernays, Hilbert, and the Development of Propositional Logic" *The Bulletin of Symbolic Logic* Vol 5 #3 (Sep 1999), 331-366