PROOF THEORY IN PHILOSOPHY OF MATHEMATICS

ANDREW ARANA

Abstract. A variety of projects in proof theory of relevance to the philosophy of mathematics are surveyed, including Gödel's incompleteness theorems, conservation results, independence results, ordinal analysis, predicativity, reverse mathematics, speed-up results, and provability logics.

Proof theory is the branch of mathematical logic in which proofs are studied as formal objects in their own right. It originated in David Hilbert's aspiration to prove the consistency of arithmetic using what he took to be epistemically secure mathematical methods. In a series of papers (most notably [Hil25]), Hilbert divided mathematics into a "real" part that is finitary and contentual, and a non-contentual "ideal" part that concerns, for instance, imaginary numbers and infinitary propositions in analysis and set theory. The objects of real mathematics were said to be finite strings of symbols, and hence visualizable. This, Hilbert seemed to think, ensured the reliability of real methods for establishing results. In addition, he believed that the objects of real mathematics were recognizable by all human investigators, and hence that real methods provided for the intersubjectivity of mathematics. However, he acknowledged that ideal methods were indispensable in practice, on account of their efficiency in the discovery and proof of new theorems. He thus set forth to license the use of ideal methods with respect to his epistemic standards by seeking a real proof that real theorems provable by ideal methods are also provable by real methods. This would ensure that the epistemic security of real methods is "conserved" by ideal methods (cf. [Det86], pp. 16–19, 62–73, for a discussion of whether Hilbert was right to think this). Hilbert envisioned his proof theory as a precise means of proving such a "conservation result". More generally, a theory T_2 is said to be *conservative* over a theory T_1 if and only if every sentence φ in the language of T_1 that is provable in T_2 is also provable in T_1.

Date: October 31, 2009.

Forthcoming in *Philosophy Compass*. Thanks to Jeremy Avigad, Curtis Franks, Paolo Mancosu, Doug Patterson and Sean Walsh for comments on drafts.

Hilbert's nascent program received a jolt when Kurt Gödel [Göd31] revealed his two incompleteness theorems, asserting that for formal theories containing a basic arithmetic core, there are sentences that are true in those theories but unprovable (the "first incompleteness theorem"), and that, provided such a theory is consistent, there are sentences that can be said to express the consistency of these theories yet are unprovable in those theories (the "second incompleteness theorem"). In both cases the sentences are expressed in the language of the theories they're purportedly 'about', and yet are unprovable in those theories (cf. [Fra05]). The importance of Gödel's work for Hilbert's program is that, if finitary mathematics is formalizable in a theory to which Gödel's work applies, his second incompleteness theorem implies that the sentence expressing the consistency of finitary mathematics cannot be provable by finitary methods. Since this consistency sentence can be proved in a variety of non-finitary theories (for instance, set theory; cf. [Lin97]), Gödel's second incompleteness theorem seems to show that the conservation result Hilbert sought cannot be obtained.

Gödel's results apply in particular to first-order Peano Arithmetic (PA), a theory widely believed to implicitly define the elementary arithmetic of the natural numbers. The language of PA consists of constants 0 and 1; function symbols $+$ and $\cdot$; and relation symbol $<$. Its axioms consist of the ordinary arithmetic laws of addition, multiplication, and less-than, plus an induction scheme for all formulas in the language.

Gödel's basic ideas in proving the incompleteness theorems were as follows. His first insight was to see how to "code" symbols and strings of symbols in a formal language into natural numbers; and how to express syntactic relations like "is a proof in PA" by relations between natural numbers, that is, by relations on the codes of strings. The hard part was to ensure that when p is a proof in PA of a sentence φ, then PA *proves* that the code $\ulcorner p \urcorner$ of p has the analogous natural number relation to the code $\ulcorner \varphi \urcorner$ of φ—that is, that $\mathrm{PA} \vdash \mathrm{Proof}_{\mathrm{PA}}(\ulcorner p \urcorner, \ulcorner \varphi \urcorner)$—and that when p is not a proof in PA of φ, that $\mathrm{PA} \nvdash \mathrm{Proof}_{\mathrm{PA}}(\ulcorner p \urcorner, \ulcorner \varphi \urcorner)$. Gödel showed that any syntactic relation whose obtaining can be confirmed by a computer (to put it somewhat anachronistically) can be "represented" in this way in PA. Gödel's next insight yielded a way of obtaining sentences in the language of PA that express facts about themselves, known as 'diagonalization'. In particular, Gödel showed that for every formula $P(x)$ in the language of PA, there is a sentence S such that $\mathrm{PA} \vdash S \leftrightarrow P(\ulcorner S \urcorner)$. He then applied diagonalization to the syntactic relation "is not provable from the axioms of PA"—that is, to the formula $\neg\mathrm{Prov}_{\mathrm{PA}}(x)$, where $\mathrm{Prov}_{\mathrm{PA}}(x) = \exists y \mathrm{Proof}_{\mathrm{PA}}(y, x)$—and obtained a sentence G such that $\mathrm{PA} \vdash G \leftrightarrow \neg\mathrm{Prov}_{\mathrm{PA}}(\ulcorner G \urcorner)$. G thus expresses its own unprovability in PA. Gödel then argued that G is true but not provable in PA, giving his first incompleteness theorem. For his second incompleteness theorem, Gödel observed that, supposing that PA is consistent, PA cannot prove sentences expressing its own consistency, for instance Con(PA) defined as $\forall p \neg \mathrm{Proof}_{\mathrm{PA}}(p, \ulcorner 0 = 1 \urcorner)$. Its proof involves showing that the reasoning of the first incompleteness theorem can be carried out within PA.

There is good reason to wonder whether Gödel's incompleteness theorems warrant the conclusion that Hilbert's sought conservation result cannot be obtained. Firstly, it is not clear that the conditions of Gödel's theorem apply to finitary mathematics (cf. [Det86] and [Det90]). In particular, it is not clear that finitary mathematics can be formalized; for instance, Gödel resisted this initially (cf. [Göd31], p. 195; also [Fef08]). This is, in part, because it is unclear from Hilbert's writings exactly what counts as finitary (cf. [HB70], p. 290). Secondly, even if the sought conservation result were unobtainable, it might be obtainable for every statement of mathematical interest. Thirdly, it may be that results closely related to the conservation result Hilbert sought are obtainable, even if Hilbert's original program is itself unrealizable. We shall address each of these points in what follows.

With respect to the first point, Tait (cf. [Tai81], [Tai05]) has argued that finitary mathematics should be identified with the first-order theory Primitive Recursive Arithmetic (PRA) obtained from PA by adding axioms defining every primitive recursive function, adding symbols to the language of PA for each such defined function, and restricting the induction scheme to quantifier-free formulas. Tait argues for this identification firstly by stressing Hilbert's identification of finitary mathematics with those methods "presupposed by all nontrivial mathematical reasoning about numbers" (p. 525), and showing that PRA uniquely meets this criterion.

With respect to the second point, the question is whether a conservation result like the one Hilbert sought can be obtained for all real theorems of antecedent mathematical interest, if not for all real theorems simpliciter. One way this question has been approached is by investigating whether results long sought by mathematicians but presently unproved are instances of arithmetic incompleteness. For example, it may turn out that number theorists have failed to prove either the Goldbach conjecture (that every even integer greater than two is the sum of two primes) or its negation because they are unprovable in PA. However, at present there is no good reason to think that the Goldbach conjecture is such an example, nor is any other such example known in present-day ordinary mathematics. Hence this approach does not resolve the question of whether every real theorem of antecedent mathematical interest has a real proof. However, proof theorists have found examples of mathematically "natural" truths that are unprovable in arithmetic and even set theory. Of particular note is the work of Paris, Kirby and Harrington on a finite form of Ramsey's theorem (a general case of the "pigeonhole principle" that if there are more objects to be put into boxes than there are boxes, one box will have more than one object) that is shown to be unprovable in PA (cf. [PH77]); and work of Friedman on truths in finite mathematics that are unprovable in Zermelo-Fraenkel set theory (ZFC) (cf. [Fri98], [Fria] and [Wei09]). In both cases the truths are shown to be unprovable by noting that the truths are equivalent to a consistency sentence for the appropriate theory; their unprovability then follows from Gödel's second incompleteness theorem. Both types of examples are said to be "natural" in the sense that they're "close" to truths studied in present-day mathematics, meaning that they *could* have arisen in present-day practice, though they did not (or so it is argued). So while this work does not answer whether every real theorem of *antecedent* mathematical interest has a real proof, it answers (negatively) the related question of whether every mathematically natural real theorem has a real proof.

Relatedly, Gödel observed in his 1931 paper that his true but unprovable sentences become provable when the incomplete theory in question is augmented by new axioms of higher 'type'; for instance, Gödel sentences for theories of arithmetic type become provable when appropriate set-theoretic axioms are added (cf. [Göd31], footnote 48a; also [Göd95] and [Fef87]). In the case of Friedman's results, the truths of finite mathematics he identifies are not provable in ZFC but are provable in ZFC augmented by new axioms asserting the existence of infinite cardinal numbers that otherwise cannot be proved to exist in ZFC (i.e., "large cardinals"). These results have accordingly been explored as components of a "regressive" argument for accepting new axioms for mathematics, though Feferman has urged caution on the grounds that the sentences in question have been "cooked up" using metamathematics (since these sentences are equivalent to consistency sentences), rather than arising naturally from mathematical practice (cf. [Fef87] and [FFMS00]).

The third point, that results closely related to the one Hilbert's program sought may be fruitfully pursued, has been approached in several different bodies of work. Here three will be considered in further detail. The first focuses on proving the consistency of arithmetic and theories of analysis by means that are as epistemically secure as finitary reasoning, despite containing non-finitary elements. As articulated by

Gerhard Gentzen (cf. [Gen36]), Takeuti (cf. [Tak87], Chapter 2, §11) and Schütte (cf. [Sch77], p. 3), the goal is "constructive", if not strictly finitary, consistency proofs for theories of both arithmetic and analytic type (cf. [Fef00]). In [Gen36] Gentzen provided what he took to be such a proof for PA. He proceeded by showing that every proof in PA can be reduced by finitely many steps to a 'minimal form', and that this cannot be done for derivations of false arithmetic assertions such as $1 = 2$. It follows that false arithmetic assertions like $1 = 2$ are not provable in PA, and hence that PA is consistent. Non-finitary methods enter in proving the finite reducibility of proofs of PA. Gentzen associated with each proof in PA an ordinal number measuring a proof's complexity, and showed that the reduction of a proof has a smaller ordinal than the initial proof. To measure the complexity of inductive proofs, in which the universally quantified conclusion of an inductive inference is more complex than the infinitely many particular instances surveyed by the inductive step, Gentzen used "transfinite" ordinals, that is, ordinals that permit counting after the totality of natural numbers has been counted. For PA this measurement requires ordinals no greater than ϵ_0, the least ordinal α such that $\omega^\alpha = \alpha$ (as can be shown by inspecting the possible forms of proofs in PA). Then, to show the finite reducibility of proofs in PA, Gentzen used "transfinite induction" through ϵ_0—which is like ordinary mathematical induction except that it ranges over all ordinals through ϵ_0 instead of just the natural numbers. In [Gen43] Gentzen showed that the consistency of PA requires for its proof *exactly* transfinite induction through ϵ_0; no lesser ordinal suffices because, he showed, transfinite induction through every ordinal beneath ϵ_0 is provable in PA. This suggests using ϵ_0 as a measure of PA's strength, and so ϵ_0 is called the "proof-theoretic ordinal" of PA.

After Gentzen proof theorists have attempted to give similar consistency proofs for stronger theories (that is, with larger proof-theoretic ordinals), including theories of analysis and set theory (cf. [Rat06]). Besides being of pure mathematical interest, this work is intended to determine for these theories the degree to which their consistency can be proved "constructively" if not finitarily, and thus the extent to which results closely related to Hilbert's program can be obtained. Relatedly, proof theorists seek information about the constructive or finitary content of classical, nonconstructive proofs. An important example of reduction of the non-constructive to the constructive was given by Gödel in his translations of PA into the intuitionistic theory of elementary arithmetic known as Heyting Arithmetic (cf. [Göd33b] and [Göd58]; and [AF98]). More recently, Kohlenbach's "proof mining" program has aimed at gathering constructive information from non-constructive proofs (cf. [Koh08] and [Avi09]).

The philosophical interest of these projects rests largely on the extent to which "constructive" methods are judged more valuable than non-constructive methods. On the one hand, it seems to be accepted by many that constructive methods provide for more secure or reliable knowledge than non-constructive methods (even though the *reasons* for thinking that vary quite a lot, for instance from intuitionists to strict finitists). On the other hand, if it is granted that non-constructive methods can provide proofs of theorems, Kreisel's question remains: "What more do we know if we have proved a theorem using restricted means than if we merely know that it is true?" (cf. [Kre58]).

A second body of work aimed at results close to Hilbert's original program is concerned with predicativity and the reduction of non-predicative systems to systems with predicative strength (cf. [Fef64] and [Sch77], Chapter 8; and [BFPS81], [Poh09], and [Fef05]). A set is "predicatively definable" if it is defined in terms of natural numbers, or in terms of predicatively definable sets that have already been defined. Sets defined by way of a collection of sets that includes the set to be defined are thereby excluded, such as the 'set' of all sets that do not contain themselves giving rise to Russell's paradox. Hermann Weyl, a student of Hilbert, helped focus attention on predicative reasoning (cf. [Wey18]). Feferman has developed Weyl's vision in a series of works, arguing that classical and modern analysis can for the most part be carried out in predicative theories that are conservative over PA. He argues that as a result, the predicativist need not admit any transfinite sets beyond the countably infinite, since, he maintains, commitment to PA only entails commitment to at most the countably infinite. This is valuable, Feferman says, because the uncountably infinite is more problematic than the countable infinite. As he puts it:

> [T]he continuum itself, or equivalently the power set of the natural numbers, is not a definite mathematical object. Rather, its a conception we have of the totality of "arbitrary" subsets of the set of natural numbers, a conception that is clear enough for us to ascribe many evident properties to that supposed object (such as the impredicative comprehension axiom scheme) but which cannot be sharpened in any way to determine or fix that object itself. On my view, it follows that the conception of the whole of the cumulative hierarchy, i.e., the transfinitely cumulatively iterated power set operation,

> is even more so inherently vague, and that one cannot in general speak of what is a fact of the matter under that conception. (cf. [FFMS00], p. 405)

In addition, Feferman has argued that predicative analysis suffices for proving all "scientifically applicable analysis", and hence, by the Quine-Putnam indispensability argument (cf. [Put79]), the ontological commitments of science should include the objects of predicative analysis but not necessarily the uncountable infinite (cf. [Fef93b]; also cf. Hellman's [Hel04] and Feferman's response [Fef04]).

A third body of work aimed at partially realizing Hilbert's program is concerned with subsystems of second-order arithmetic, that is, arithmetic theories in which some of the quantifiers range over sets of natural numbers. A great deal of mathematics can be developed in such theories, including basic and advanced results in analysis, algebra, and logic. A handful of such theories have been singled out for their mathematical fruitfulness; two will be mentioned here. Their characterization uses the following syntactic notion: a sentence in the language of arithmetic is Π^0_1 (resp., Σ^0_1) if it consists of finitely many first-order universal (resp., existential) quantifiers followed by a formula with only bounded first-order formulas. A key example of a Π^0_1 sentence is the sentence Con(PA) expressing the consistency of PA defined earlier as $\forall p \neg \mathrm{Proof}_{\mathrm{PA}}(p, \ulcorner 0 = 1 \urcorner)$. We may now turn to the theories. Firstly, RCA_0 is the subsystem of second-order arithmetic obtained by adding to PRA a comprehension scheme for recursively definable sets (hence the "R" in RCA_0) , and replacing PRA's induction scheme with an induction schema for Σ^0_1 formulas, possibly with set parameters. Secondly, WKL_0 is the theory RCA_0 augmented by weak König's lemma, which yields the existence of paths through infinite binary trees (more precisely, $\{0, 1\}$-trees). Friedman observed that WKL_0 is conservative over PRA for Π^0_1 sentences.[1] Simpson has argued (cf. [Sim88], [Sim99], pp. 381–382) that this conservation result permits the identification of WKL_0 with finitary reasoning, at least for Π^0_1 sentences such as Con(PA) that were of central importance to Hilbert's goal. He thus claims that this result provides a partial realization of Hilbert's program. More generally, to determine what part of Hilbert's program

[1] In fact, the conservation result holds for Π^0_2 sentences. The result is a consequence of Harrington's result that WKL_0 is conservative over RCA_0 for Π^1_1 sentences, and the result of Mints, Parsons and Takeuti (independently) that $\mathrm{I}\Sigma_1$ (which is PA with induction restricted to Σ^0_1 formulas) is conservative over PRA for Π^0_2 sentences. This is sufficient since WKL_0 and RCA_0 have the same first-order part, $\mathrm{I}\Sigma_1$. For proofs, cf. [Sim99], pp. 369–372.

can be carried out, Simpson asks which subsystems of second-order arithmetic are conservative over PRA for Π^0_1 sentences.

Each of the three bodies of work just surveyed, in addition to bearing on partial realizations of Hilbert's program, bear on a less ideological project; as Feferman has put it, they help answer the question "what rests on what?" (cf. [Fef93a]). To answer this, one seeks to determine the weakest assumptions sufficient for proving given theorems of ordinary mathematics (e.g. number theory, analysis, algebra, topology). "Weakest" here must be understood with respect to some ordering of strength, for instance with respect to how much induction the assumptions include, or what types of sets the assumptions permit. Ordinal analysis of theories, as discussed earlier, is one means of determining this. Another means focuses on particular theorems rather than entire theories. For example, consider the prime number theorem (PNT), which states that the number of primes less than x is approximately $\frac{x}{\log x}$ in the limit. It was proved by Jacques Hadamard [Had96] and Charles de la Vallée Poussin [dlVP96] using complex analysis in 1896. Later Atle Selberg [Sel49] and Paul Erdős [Erd49] were able to prove it without using complex analysis (but still using real analysis). Both of these proofs have been found to carry over to weak arithmetic settings. Takeuti [Tak78] found a conservative extension of PA in which elementary complex analysis can be carried out and which is sufficient for formalizing the complex-analytic proof of the PNT; and Cornaros and Dimitracopoulos [CD94] formalized Selberg's proof in a theory with even less induction known as elementary arithmetic (cf. [HP98] and [Bus98] for more on weak arithmetics, and cf. [Avi03] for a philosophical discussion of this work).

This shows that the PNT is provable in quite weak theories of arithmetic. By contrast, it can turn out that the weakest theory in which a given result is provable is quite strong. An example of this is the work of Friedman mentioned earlier exhibiting truths of finite mathematics that are not provable in finite mathematics or even set theory as currently canonized (ZFC), but are provable in ZFC augmented by "large cardinal" axioms.

Another approach to determining what rests on what is to seek instead an answer to the question of *exactly* how strong a given theorem is. This is what is done in "reverse mathematics", pioneered by Friedman (cf. [Fri76], [Fri75]) and described in a definitive monograph [Sim99] by Simpson. Starting with a mathematical theorem and an interesting collection of set-theoretic theories, the goal of reverse mathematics is to determine which of these theories is logically equivalent to that theorem (using as background a theory that is logically weaker than the theorem and theories under consideration). It is called "reverse mathematics" because in showing this logical equivalence, the axioms of a theory are proved from a theorem, a reversal of usual mathematical methodology. When successfully carried out, this program locates both necessary and sufficient conditions for a given theorem, and thus locates *the* logically weakest theory (among a given set of set-theoretic candidate theories) for proving a given theorem. The search for reversals is made easier by the striking fact that a very small number of set-theoretic theories have turned out, in practice, to provide for reversals for a wide variety of theorems in ordinary mathematics (over the background theory RCA_0).[2] The theory WKL_0 mentioned earlier is one of these; among the others is an extension of RCA_0's comprehension axiom to all arithmetically definable sets called ACA_0; and a theory known as ATR_0 whose first-order part is the same as Feferman's predicative analysis (cf. [Sim99], p. 41). Two examples of theorems reversing to WKL_0 are the completeness theorem for first-order logic, and the theorem that every continuous function on $0 \leq x \leq 1$ is bounded.

It is reasonable to wonder about the value of determining the weakest theory in which a given result is provable. One answer is that these results help identify more precisely the minimal ontological commitments of mathematics (as Feferman emphasizes in his work on predicativity). Another is that, if weaker theories are thought to be more secure than stronger theories, they help determine how much mathematics can be proved securely. The work surveyed in this essay so far has been aimed in these directions. Yet another answer is that this work helps engage the mathematician in "the sport of seeing how little one can get away with", as Avigad has put it (cf. [Avi05]). Hilbert offered an answer of a still different type. He seems to have understood this kind of project as revealing a "grounding" of mathematical theorems:

> By the axiomatic study of mathematical truth I understand an investigation which does not aim to discover

[2]One reservation about reverse mathematics as presently carried out is that RCA_0 is too strong a base theory, because in RCA_0 significant amounts of real analysis can be coded. This can distort the meaning of reversals of theorems of analysis over RCA_0, because it is hard to tell how much of the reversal is a result of the coding apparatus of RCA_0 and how much is a result of the logical complexity of the theorem being reversed. As a result, the complexity of some theorems of analysis may be measured lower than it intuitively ought to be (on this issue, cf. [Avi03]). For this reason, Friedman has been developing a "strict reverse mathematics" in which the base theory is weakened to avoid these issues (cf. [Fric], [Frib]).

> new or more general theorems with the help of given truths, but rather the position of a theorem within the system of known truths and their logical connections in a way that indicates clearly which conditions are necessary and sufficient for the grounding [*Begründung*] of that truth. (cf. [Hil03], p. 50)

How "grounding" ought to be understood here is unclear (cf. [Ara08]), though Hilbert may have been drawing on views on grounding of Bernard Bolzano (cf. [Bol72] and [Bol99]; and [Tat02], [Tatng], and [Man99]) and Gottlob Frege (cf. [Fre80], §2, 3; and [Det88]). Along these lines, Simpson has noted a connection between reversals in reverse mathematics and "scientific demonstrations" in Aristotle's sense (cf. [Sim99], pp. 31–2, and [Sim00]).

As noted earlier, Hilbert thought ideal methods worth defending because of their efficiency in the discovery and proof of new theorems, compared with real methods generally. For instance, he stressed the "fecund" use of ideal elements in geometry, noting that they often resulted in "simple and perspicuous" work (cf. [Hil25], p. 379). Proof theorists have accordingly sought to determine what gains in efficiency are purchased by moving from a theory to a conservative extension of that theory, measuring the efficiency of proofs in terms of their syntactic length. To measure efficiency a basic distinction is made between "polynomial" and non-polynomial speed-up, with the former being regarded as relatively insignificant and the latter as relatively significant. Somewhat more precisely, for T_2 a conservative extension of T_1, T_1 is at most a *polynomial speed-up* of T_2 when for every φ provable in T_2, the length of the shortest proof (measured in terms of total number of symbol occurrences) of φ in T_2 is less than some fixed polynomial multiple of the length of the shortest proof of φ in T_1 (cf. [CI05]). A variety of theories have been studied for speed-up. Solovay showed that set theory with classes (Gödel-Bernays) has non-polynomial speed-up over set theory without classes (ZF) (cf. [Pud98]); and Avigad and Hájek showed, independently, that WKL_0 has at most polynomial speed-up over RCA_0 (cf. [Avi96], [Háj93]).

These results might be thought to bear on Hilbert's point that the addition of ideal elements to a theory improves its efficiency. However, Detlefsen (cf. [Det96], p. 87, and [Det90], pp. 370, 376), Avigad ([Avi03], p. 276) and Potter ([Pot04], pp. 234–236) have counseled care in interpreting these results, pointing out that length of proof is only one measure of proof complexity, and a flawed one at that. Proofs in conservative extensions may provide for simplicity of other types unmeasured by the syntactic standard surveyed here (for instance, simplicity of understanding).

Gödel's work on intuitionistic logic also initiated *provability logic*, which is closely connected with modal logic (for surveys, cf. [Boo93] and [Art07]) and has been extensively used in formal epistemology (cf. [FHMV95]). In [Göd33a] Gödel gave what he took to be basic laws of provability (for instance, the law that if p is provable, then p), noting that these laws were simply the axioms of the modal logic known as S4. Gödel thus recognized a parallel between provability and necessity. This parallel is natural in light of the fact that provability and consistency are modal duals to one another, in the same way that necessity and possibility are. Thus if we interpret $\Box\varphi$ as 'φ is provable' and $\Diamond\varphi$ as 'φ is consistent,' then the familiar equivalence $\Diamond\varphi \leftrightarrow \neg\Box\neg\varphi$ reads 'φ is consistent if and only if it is not provably false,' in line with our informal understanding of these notions.

Gödel's interest in this matter was to show how to interpret intuitionistic propositional logic in a classical propositional theory augmented by a provability operator. However, he noted that S4 are laws only of *informal* provability, rather than provability in a formal system such as PA, because in the latter case some theorems of S4 are false (they contradict the second incompleteness theorem). This raises the question of what exactly is the logic of formal provability in PA, i.e. of the formal provability predicate $\mathrm{Proof}_{\mathrm{PA}}(x)$ defined earlier. Kripke conjectured that a propositional modal logic known as GL answers this question (cf. [BS91], pp. 9–10). To understand this conjecture, it is helpful to return to Gödel's work on the incompleteness theorems. The "G" of GL comes from Gödel, and more specifically from the inclusion in GL of modal analogues of the Hilbert-Bernays derivability conditions giving sufficient conditions for a formal provability predicate of a formal theory to prove Gödel's second incompleteness theorem (and hence codifying what it means for a theory to have "enough" arithmetic to prove the second incompleteness theorem; cf. [Fef61]). The "L" comes from the inclusion of a modal analogue of "Löb's theorem" (cf. [Löb55]). Gödel proved his first incompleteness theorem by obtaining, by diagonalization, a sentence G in the language of PA that "expresses" its own unprovability in e.g. PA—that is, $\mathrm{PA} \vdash G \leftrightarrow \neg\mathrm{Prov}_{\mathrm{PA}}(\ulcorner G \urcorner)$—and then showing that G is true but unprovable in PA. Leon Henkin noted that diagonalization could be used to obtain a sentence L expressing its own *provability*—that is, $\mathrm{PA} \vdash L \leftrightarrow \mathrm{Prov}_{\mathrm{PA}}(\ulcorner L \urcorner)$—and asked whether L is provable in PA. Martin Löb showed that it is, demonstrating that if $\mathrm{PA} \vdash \mathrm{Prov}_{\mathrm{PA}}(\ulcorner L \urcorner) \to L$, then $\mathrm{PA} \vdash L$. GL, then, consists of modal versions of the Hilbert-Bernays derivability conditions and of Löb's theorem, in addition to basic axioms for classical propositional logic.

Solovay verified Kripke's conjecture that GL is the logic of formal provability in PA by using "translations" of sentences in the language of propositional modal logic to sentences in the language of PA that respect the Boolean connectives while sending sentences $\Box A$ to $\mathrm{Prov}_{\mathrm{PA}}(\ulcorner A \urcorner)$. There are many such translations. Solovay [Sol76] showed that the sentences of propositional modal logic for which *each* such translation is a theorem of PA are exactly the theorems of GL. Hence, Solovay concluded, the logic of formal provability in PA is GL.

This essay has touched on several projects in proof theory relevant to the philosophy of mathematics, belonging mostly to what is known as *reductive proof theory*. It is fitting to mention briefly other proof-theoretical projects of philosophical interest. Chief among these are proof-theoretic semantics and structural proof theory, both of which originate in Gentzen's [Gen35]. Gentzen showed how to formulate a system of logical deduction in which the semantics of the logical connectives are given by rules governing their introduction into and elimination from proofs, in contrast to "model-theoretic semantics" in which the meanings of the logical connectives are given in terms of their truth conditions. Gentzen's idea has been developed by Prawitz [Pra65] and Martin-Löf [ML96], among many others, and incorporated into a defense of anti-realism by Dummett in [Dum91] (also cf. [Ten87] and [Ten97]). In the same paper Gentzen also developed a formal system that kept the natural deduction approach of using introduction and elimination rules to describe the semantics of the logical connectives and to give the rules of inference, while eliminating the use of non-logical axioms in proofs. Instead, non-logical axioms were pushed into the antecedents of conditionals that Gentzen called "sequents". This "sequent calculus" turned out to be mathematically elegant and continues to yield insights into the logical structure of proofs (cf. [NvP01], [TS00]; and [Ara09] for a discussion of some philosophical issues arising therein).

Proof theory remains a vibrant area of research with important philosophical connections, and there is good reason to think this will continue to be the case. One reason for confidence is that there has been a move in recent years toward a practice-based philosophy of mathematics, in which philosophical reflection is brought to bear on the work actually done by mathematicians, historically through the present, not just in logic but in algebra, analysis, and geometry as well (cf. [Man08]). This essay has tried to shed light on how proof theory has been concerned with ordinary mathematics since its inception. These recent developments in philosophy of mathematics suggest that proof theory will continue to be philosophically critical.

References

[AF98] Jeremy Avigad and Solomon Feferman. Gödel's functional ("Dialectica") interpretation. In *Handbook of proof theory*, pages 337–405. North-Holland, Amsterdam, 1998.

[Ara08] Andrew Arana. Logical and semantic purity. *Protosociology*, 25:36–48, 2008. Reprinted in *Philosophy of Mathematics: Set Theory, Measuring Theories, and Nominalism*, Gerhard Preyer and Georg Peter (eds.), Ontos, 2008.

[Ara09] Andrew Arana. On formally measuring and eliminating extraneous notions in proofs. *Philosophia Mathematica*, 17:208–219, 2009.

[Art07] Sergei Artemov. Modal Logic in Mathematics. In Patrick Blackburn, Johan van Bentham, and Frank Wolter, editors, *Handbook of Modal Logic*, pages 927–969. Elsevier, 2007.

[Avi96] Jeremy Avigad. Formalizing forcing arguments in subsystems of second-order arithmetic. *Annals of Pure and Applied Logic*, 82:165–191, 1996.

[Avi03] Jeremy Avigad. Number theory and elementary arithmetic. *Philosophia Mathematica*, 11:257–284, 2003.

[Avi05] Jeremy Avigad. Weak theories of nonstandard arithmetic and analysis. In *Reverse mathematics 2001*, volume 21 of *Lecture Notes In Logic*, pages 19–46. Association for Symbolic Logic, La Jolla, CA, 2005.

[Avi09] Jeremy Avigad. The metamathematics of ergodic theory. *Annals of Pure and Applied Logic*, 157:64–76, 2009.

[BFPS81] Wilfried Buchholz, Solomon Feferman, Wolfram Pohlers, and Wilfried Sieg. *Iterated inductive definitions and subsystems of analysis: recent proof-theoretical studies*, volume 897 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1981.

[Bol72] Bernard Bolzano. *Theory of Science*. University of California Press, Berkeley, 1972. Originally published in 1837. Translated by A. George.

[Bol99] Bernard Bolzano. Purely analytic proof of the theorem that between any two values which give results of opposite sign there lies at least one real root of the equation. In William Ewald, editor, *From Kant to Hilbert*, volume 1, pages 227–248. Oxford University Press, 1999. Originally published in 1817; translated by S. Russ.

[Boo93] George Boolos. *The logic of provability*. Cambridge University Press, Cambridge, 1993.

[BS91] George Boolos and Giovanni Sambin. Provability: the emergence of a mathematical modality. *Studia Logica*, 50(1):1–23, 1991.

[Bus98] Samuel R. Buss. First-order proof theory of arithmetic. In *Handbook of proof theory*, pages 79–147. North-Holland, Amsterdam, 1998.

[CD94] Charalampos Cornaros and Costas Dimitracopoulos. The prime number theorem and fragments of PA. *Arch. Math. Logic*, 33(4):265–281, 1994.

[CI05] Patrick Caldon and Aleksandar Ignjatović. On mathematical instrumentalism. *Journal of Symbolic Logic*, 70(3):778–794, 2005.

[Det86] Michael Detlefsen. *Hilbert's Program*. Reidel, Dordrecht, 1986.

[Det88] Michael Detlefsen. Fregean hierarchies and mathematical explanation. *International Studies in the Philosophy of Science*, (3):97–116, 1988.

[Det90] Michael Detlefsen. On an alleged refutation of Hilbert's program using Gödel's first incompleteness theorem. *Journal of Philosophical Logic*, 19(4):343–377, November 1990.

[Det96] Michael Detlefsen. Philosophy of mathematics in the twentieth century. In *Philosophy of Science, Logic, and Mathematics*, volume 9 of *Routledge History of Philosophy*, pages 50–123. Routledge, London and New York, 1996. Edited by Stuart G. Shanker.

[dlVP96] Charles de la Vallée Poussin. Recherches analytiques sur la théorie des nombres premiers. *Ann. Soc. Sci. Bruxelles*, 20:183–256, 1896. Reprinted in Collected Works Volume 1, edited by P. Butzer, J. Mawhin, and P. Vetro, Académie Royale de Belgique and Circolo Matematico di Palermo, 2000.

[Dum91] Michael Dummett. *The Logical Basis of Metaphysics*. Harvard University Press, 1991.

[Erd49] Paul Erdős. On a new method in elementary number theory which leads to an elementary proof of the prime number theorem. *Proceedings of the National Academy of Sciences, USA*, 35:374–384, 1949.

[Fef61] Solomon Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, 49:35–92, 1960/1961.

[Fef64] Solomon Feferman. Systems of predicative analysis. *Journal of Symbolic Logic*, 29:1–30, 1964.

[Fef87] Solomon Feferman. Infinity in Mathematics: Is Cantor Necessary? In Giuliano Toraldo di Francia, editor, *L'infinito nella scienza*, pages 151–190. Istituto della Enciclopedia Italiana,, 1987. Reprinted with minor changes in [Fef98], pages 28–73, 229–248.

[Fef93a] Solomon Feferman. What rests on what? The proof-theoretic analysis of mathematics. In *Philosophy of mathematics (Kirchberg am Wechsel, 1992)*, volume 20 of *Schriftenreihe Wittgenstein-Ges.*, pages 147–171. Hölder-Pichler-Tempsky, Vienna, 1993. Reprinted with minor changes in [Fef98].

[Fef93b] Solomon Feferman. Why a little bit goes a long way: logistical foundations of scientifically applicable mathematics. In Micky Forbes David Hull and Kathleen Okruhlik, editors, *PSA 1992*, volume 2, pages 442–455. Philosophy of Science Association, 1993. Reprinted with minor changes in [Fef98].

[Fef98] Solomon Feferman. *In the Light of Logic*. Oxford University Press, New York, 1998.

[Fef00] Solomon Feferman. Does reductive proof theory have a viable rationale? *Erkenntnis*, 53(1-2):63–96, 2000.

[Fef04] Solomon Feferman. Comments on "Predicativity as a Philosophical Position" by G. Hellman. *Revue Internationale de Philosophie*, 229(3):313–323, 2004.

[Fef05] Solomon Feferman. Predicativity. In Stewart Shapiro, editor, *Handbook of the Philosophy of Mathematics and Logic*. Oxford University Press, 2005.

[Fef08] Solomon Feferman. Lieber Herr Bernays!, Lieber Herr Gödel! Gödel on finitism, constructivity and Hilbert's program. *Dialectica*, 62(2):179–203, 2008.

[FFMS00] Solomon Feferman, Harvey M. Friedman, Penelope Maddy, and John R. Steel. Does mathematics need new axioms? *Bull. Symbolic Logic*, 6(4):401–446, 2000.

[FHMV95] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning about knowledge*. MIT Press, Cambridge, MA, 1995.

[Fra05] Torkel Franzén. *Gödel's theorem*. A K Peters Ltd., Wellesley, MA, 2005. An incomplete guide to its use and abuse.

[Fre80] Gottlob Frege. *The Foundations of Arithmetic*. Northwestern Univ. Press, 1980. Translated by J.L. Austin.

[Fria] Harvey M. Friedman. Boolean relation theory and the incompleteness phenomena. Book draft, October 30, 2007, http://www.math.ohio-state.edu/~friedman/pdf/Whole103007.pdf.

[Frib] Harvey M. Friedman. The inevitability of logical strength: strict reverse mathematics. Draft, August 29, 2007, http://www.math.ohio-state.edu/~friedman/pdf/InevLogStr082907.pdf.

[Fric] Harvey M. Friedman. Strict reverse mathematics. Draft, January 31, 2005, http://www.math.ohio-state.edu/~friedman/pdf/StrictRM012305.pdf.

[Fri75] Harvey M. Friedman. Some Systems of Second Order Arithmetic and Their Use. In *Proceedings of the 1974 International Congress of Mathematicians*, pages 235–242. Canadian Mathematical Congress, 1975.

[Fri76] Harvey M. Friedman. Systems of second order arithmetic with restricted induction. I. *Journal of Symbolic Logic*, 41(2):557–8, June 1976.

[Fri98] Harvey M. Friedman. Finite functions and the necessary use of large cardinals. *Annals of Mathematics (2)*, 148(3):803–893, 1998.

[Gen35] Gerhard Gentzen. Untersuchungen über das logische schliessen. *Mathematische Zeitschrift*, 39:405–431, 1934-1935. Translated as "Investigations into logical deduction" in *The collected papers of Gerhard Gentzen*, M.E. Szabo (ed.), North-Holland, 1969.

[Gen36] Gerhard Gentzen. Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen*, 112:493–565, 1936. Translated as "The consistency of arithmetic" in *The collected papers of Gerhard Gentzen*, M.E. Szabo (ed.), North-Holland, 1969.

[Gen43] Gerhard Gentzen. Beweisbarkeit und Unbeweisbarkeit von Anfangsfallen der transfiniten Induktion in der reinen Zahlentheorie. *Mathematische Annalen*, 119(1):140–161, 1943. Translated as "Provability and Nonprovability of Restricted Transfinite Induction in Elementary Number Theory" in *The collected papers of Gerhard Gentzen*, M.E. Szabo (ed.), North-Holland, 1969.

[Göd31] Kurt Gödel. Uber formal unentscheidhare Sätze der *Principia Mathematica* und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931. Reprinted and translated in *Collected Works* Volume 1, Solomon Feferman et. al. (eds.), Oxford University Press, 1986.

[Göd33a] Kurt Gödel. Eine Interpretation des intuitionistischen Aussagenkalkuls. *Ergebnisse eines mathematischen Kolloquiums*, 4:39–40, 1933. Reprinted and translated in *Collected Works* Volume 1, Solomon Feferman et. al. (eds.), Oxford University Press, 1986.

[Göd33b] Kurt Gödel. Zur intuitionistischen Arithmetik und Zahlentheorie. *Ergebnisse eines mathematischen Kolloquiums*, 4:34–38, 1933. Reprinted and translated in *Collected Works* Volume 1, Solomon Feferman et. al. (eds.), Oxford University Press, 1986.

[Göd58] Kurt Gödel. Über eien bisher noch nicht benütze Erweiterung des finiten Standpunktes. *Dialectica*, pages 280–287, 1958. Reprinted and translated in Kurt Gödel, *Collected Works*, volume 2. Solomon Feferman et. al., editors. Oxford University Press, Oxford, 1995.

[Göd95] Kurt Gödel. Some basic theorems on the foundations of mathematics and their implications. In Solomon Feferman et. al., editor, *Collected Works*, volume 3, pages 304–323. Oxford Univ. Press, 1995.

[Had96] Jacques Hadamard. Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques. *Bull. Soc. Math. France*, 24:199–220, 1896.

[Háj93] Petr Hájek. Interpretability and fragments of arithmetic. In Peter Clote and Jan Krajícek, editors, *Arithmetic, proof theory, and computational complexity*, pages 185–196. Oxford University Press, 1993.

[HB70] David Hilbert and Paul Bernays. *Grundlagen der Mathematik. II.* Zweite Auflage. Die Grundlehren der mathematischen Wissenschaften, Band 50. Springer-Verlag, Berlin, 1970.

[Hel04] Geoffrey Hellman. Predicativity as a Philosophical Position. *Revue Internationale de Philosophie*, 229(3):295–312, 2004.

[Hil03] David Hilbert. Über den Satz von der Gleichheit der Basiswinkel im gleichschenkligen Dreieck. *Proceedings of the London Mathematical Society*, 35:50–68, 1902/03.

[Hil25] David Hilbert. On the infinite. In Jean van Heijenoort, editor, *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*, pages 369–392. Harvard University Press, 1925.

[HP98] Petr Hájek and Pavel Pudlák. *Metamathematics of first-order arithmetic*. Perspectives in Mathematical Logic. Springer-Verlag, Berlin, 1998. Second printing.

[Koh08] Ulrich Kohlenbach. *Applied proof theory: proof interpretations and their use in mathematics*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2008.

[Kre58] Georg Kreisel. Hilbert's programme. *Dialectica*, 12:346–372, 1958. Revised with postscript in *Philosophy of Mathematics*, second edition, P. Benacerraf and H. Putnam (eds.), Cambridge University Press, Cambridge, U.K., 1983, pages 207–238.

[Lin97] Per Lindström. *Aspects of Incompleteness*. Springer-Verlag, 1997.

[Löb55] M. H. Löb. Solution of a problem of Leon Henkin. *Journal of Symbolic Logic*, 20:115–118, 1955.

[Man99] Paolo Mancosu. Bolzano and Cournot on Mathematical Explanation. *Revue d'Histoire des Sciences*, 52:429–455, 1999.

[Man08] Paolo Mancosu, editor. *The Philosophy of Mathematical Practice*. Oxford University Press, 2008.

[ML96] Per Martin-Löf. On the meanings of the logical constants and the justifications of the logical laws. *Nordic Journal of Philosophical Logic*, 1(1):11–60, 1996.

[NvP01] Sara Negri and Jan von Plato. *Structural proof theory*. Cambridge University Press, Cambridge, 2001.

[PH77] Jeff Paris and Leo Harrington. A Mathematical Incompleteness in Peano Arithmetic. In Jon Barwise, editor, *Handbook of Mathematical Logic*, pages 1133–1142. North-Holland, Amsterdam, 1977.

[Poh09] Wolfram Pohlers. *Proof theory*. Springer-Verlag, Berlin, 2009.

[Pot04] Michael Potter. *Set theory and its philosophy*. Oxford University Press, New York, 2004.

[Pra65] Dag Prawitz. *Natural Deduction. A Proof-Theoretical Study*. Almquist and Wiksell, Stockholm, 1965. Reprinted in 2006 by Dover Publications.

[Pud98] Pavel Pudlák. The lengths of proofs. In Samuel R. Buss, editor, *Handbook of Proof Theory*, pages 547–637. North-Holland, 1998.

[Put79] Hilary Putnam. Philosophy of Logic. In *Mathematics, Matter and Method: Philosophical Papers, Volume 1*. Cambridge University Press, second edition, 1979.

[Rat06] Michael Rathjen. The art of ordinal analysis. In *International Congress of Mathematicians. Vol. II*, pages 45–69. Eur. Math. Soc., Zürich, 2006.

[Sch77] Kurt Schütte. *Proof theory*. Springer-Verlag, Berlin, 1977. Translated from the revised German edition by J. N. Crossley, Grundlehren der Mathematischen Wissenschaften, Band 225.

[Sel49] Atle Selberg. An elementary proof of the prime-number theorem. *Annals of Mathematics (2)*, 50:305–313, 1949.

[Sim88] Stephen G. Simpson. Partial realizations of Hilbert's Program. *J. Symbolic Logic*, 53(2):349–363, 1988.

[Sim99] Stephen G. Simpson. *Subsystems of Second Order Arithmetic*. Perspectives in Mathematical Logic. Springer, 1999.

[Sim00] Stephen G. Simpson. Logic and mathematics. In Stanley Rosen, editor, *The Examined Life*. Random House, 2000.

[Sol76] Robert M. Solovay. Provability interpretations of modal logic. *Israel Journal of Mathematics*, 25(3-4):287–304, 1976.

[Tai81] William W. Tait. Finitism. *The Journal of Philosophy*, 78(9):524–546, 1981.

[Tai05] William W. Tait. Remarks on finitism. In *The Provenance of Pure Reason*, pages 43–53. Oxford University Press, 2005.

[Tak78] Gaisi Takeuti. *Two Applications of Logic to Mathematics*. Princeton University Press, 1978.

[Tak87] Gaisi Takeuti. *Proof theory*. North-Holland, Amsterdam, second edition, 1987.

[Tat02] Armin Tatzel. Bolzano's theory of ground and consequence. *Notre Dame J. Formal Logic*, 43(1):1–25 (2003), 2002.

[Tatng] Armin Tatzel. Proving and grounding. Bolzano's theory of grounding and Gentzen's normal proofs. *History and Philosophy of Logic*, Forthcoming.

[Ten87] Neil Tennant. *Anti-Realism and Logic*. Clarendon Press, Oxford, 1987.

[Ten97] Neil Tennant. *The Taming of the True*. Clarendon Press, Oxford, 1997.

[TS00] A. S. Troelstra and H. Schwichtenberg. *Basic proof theory*. Cambridge University Press, Cambridge, second edition, 2000.

[Wei09] Andreas Weiermann. Phase transitions for Gödel incompleteness. *Annals of Pure and Applied Logic*, 157(2-3):281–296, 2009.

[Wey18] Hermann Weyl. *Das Kontinuum. Kritische Untersuchungen über die Grundlagen der Analysis*. Veit, Leipzig, 1918. Translated into English by S. Pollard and T. Bole as *The Continuum. A Critical Examination of the Foundations of Analysis*, 1994, Dover Publications.

Department of Philosophy, 201 Dickens Hall, Kansas State University, Manhattan, KS 66506-0803

E-mail address: `aarana@ksu.edu`