Number theory and elementary arithmetic*

Jeremy Avigad

June 12, 2002

Abstract

Elementary arithmetic (also known as "elementary function arithmetic") is a fragment of first-order arithmetic so weak that it cannot prove the totality of an iterated exponential function. Surprisingly, however, the theory turns out to be remarkably robust. I will discuss formal results that show that many theorems of number theory and combinatorics are derivable in elementary arithmetic, and try to place these results in a broader philosophical context.

1 Introduction

The relationship between mathematical logic and the philosophy of mathematics has long been a rocky one. To some, the precision of a formal logical analysis represents the philosophical ideal, the paradigm of clarity and rigor; for others, it is just the point at which philosophy becomes uninteresting and sterile. But, of course, both formal and more broadly philosophical approaches can yield insight, and each is enriched by a continuing interaction: philosophical reflection can inspire mathematical questions and research programs, which, in turn, inform and illuminate philosophical discussion.

My goal here is to encourage this type of interaction. I will start by describing an informal attitude that is commonly held by metamathematical proof theorists, and survey some of the formal results that support this point of view. Then I will try to place these formal developments in a more general philosophical context, and clarify some related issues.

In the philosophy of mathematics, a good deal of attention is directed towards the axioms of Zermelo-Fraenkel set theory. Over the course of the twentieth century, we have come to learn that *ZFC* offers an extraordinarily robust foundation for mathematics, providing a uniform language for defining mathematical notions and conceptually clear rules of inference. One would therefore

*This is a DRAFT. In October of 2001, parts of an early version were presented to an in-house workshop at the University of Pittsburgh's Center for Philosophy of Science. Since then, I have received a draft version of Arana [1], which addresses similar issues from a complementary perspective. I am grateful to Andrew Arana, Steve Awodey, Teddy Seidenfeld, Wilfried Sieg, and Neil Tennant for comments, suggestions, and corrections.

like to justify this choice of framework on broader ontological or epistemological grounds. On the other hand, incompleteness phenomena and set theoretic independences show that *ZFC* is insufficient to settle *all* mathematical questions, so it would also be nice to have robust philosophical principles that can help guide the search for stronger axioms.

In the proof theory community, however, there is a long tradition, from Weyl's *Das Kontinuum* [81] and Hilbert and Bernays' *Grundlagen der Mathematik* [32] through the work of Takeuti [77] and the school of Reverse Mathematics today [71], of studying theories that are significantly weaker. The general feeling is that most "ordinary" mathematics can be carried out, formally, without using the full strength of *ZFC*.

I should qualify these remarks, and, in fact, I will do so in many ways throughout the course of this essay. One issue that arises has to do with the choice of formal representation of the informal bodies of mathematics under investigation. When dealing with weaker theories, linguistic and axiomatic restrictions force one to pay careful attention to the way in which the relevant mathematical notions are defined; and there is the unsettling fact that the outcome of the analysis can depend, unnaturally, on one's choice of definitions. So, for example, the question as to whether the mean value theorem of undergraduate calculus can be derived in a certain restricted theory may well depend on one's definition of the real numbers, or of a continuous function. When it comes to finitary objects like numbers, finite sets, and finite sequences, however, issues of representation seem less problematic, because it usually turns out that various natural definitions are easily shown to be equivalent, or at least have equivalent properties, on the basis of minimal systems of axioms. With this in mind, it makes sense to limit the playing field by asking which finitary theorems of number theory and combinatorics can be derived in a particular formal theory; restricting one's attention in this way thereby provides a useful touchstone for the analysis.

Now, in the hierarchy of formal foundations for mathematics, Zermelo-Fraenkel set theory is stronger than Zermelo set theory, which is in turn stronger than higher-order arithmetic, second-order arithmetic, first-order arithmetic, and primitive recursive arithmetic, in order of descreasing strength. In the next section, I will describe a first-order theory known as *elementary arithmetic*, *EA*, which is even weaker than all of these. *EA* is so weak that it cannot prove the totality of an iterated exponential function, so, from the point of view of any set theorist, *EA* is almost laughably weak when considered as a foundation for mathematical reasoning.

But the proof theorist begs us to consider whether there is more to *EA* than meets the eye. From the point of view of finitary number theory and combinatorics, *EA* turns out to be surprisingly robust. So much so that Harvey Friedman has made the following

Grand conjecture. *Every theorem published in the* Annals of Mathematics *whose statement involves only finitary mathematical objects (i.e. what logicians call an arithmetical statement) can be proved in elementary arithmetic.*

Friedman's conjecture is a clear and pointed manifestation of the proof-theoretic attitude alluded to above. Unlike most mathematical conjectures, this one may be spectactularly true, spectacularly false, or somewhere in between. Since the conjecture was posted to the Foundations of Mathematics discussion group [84] on April 16, 1999, it covers the following special case:

Specific conjecture. *Fermat's last theorem is derivable in elementary arithmetic.*

We are a long way from settling even the more restricted conjecture; making real progress towards that end will require combining a deep understanding of some of the most advanced methods of modern number theory with the proof theorist's fetish for developing mathematics in restricted theories. But the conjectures are interesting because many proof theorists consider them plausible, whereas, I suspect, most mathematicians would lay strong odds against them.

In the next two sections I will discuss some formal developments that support the proof theorist's intuition. In Section 2, I will describe elementary arithmetic and some conservative extensions thereof. In Section 3, I will consider Dirichlet's theorem on primes in an arithmetic program and the prime number theorem as case studies, and cite work by Patrick Cegielski and Oliver Sudac that shows that these can be derived in restricted theories of arithmetic. Of course, assessments as to the weight of such evidence in favor of Friedman's conjecture will vary, and since there is, at present, no clear mathematical or philosophical sense one can ascribe to notions of "evidence" and "likelihood" with respect to a mathematical conjecture, an evaluation of the formal results in these terms would be, at best, of heuristic or sociological value. Rather than pursue this course, I will instead use the discussion of Friedman's conjecture to clarify and explore some broader philosophical issues. Section 4 therefore addresses the question as to what we, as philosophers, are to make of the formal results.

I do not wish to convey the impression that contemporary proof-theoretic research is rallied around a crusade to establish Friedman's conjecture. To be sure, a number of branches of proof theory involve formalizing mathematics in weak or restricted theories; the fields of reverse mathematics, constructive mathematics, weak theories of arithmetic, and Kohlenbach's "proof mining" are some examples.[1] But even in these fields, a greater emphasis is usually placed on studying metamathematical properties of the theories under investigation, such as their models, measures of their strength, and interpretations between them; independence results; and mathematical applications, such as extracting additional information from classical proofs. And, as far as formalization is concerned, one is always interested in particular cases; the assumption that large portions of mathematics can be treated in this way is often left implicit. So one should not view the present work as a survey of any particular research program but, rather, as an exploration of a tacit *Weltanschauung* that guides contemporary research in proof theory.

[1] See the appendix to this paper for references to these subjects.

2 Elementary arithmetic

I will take the language of elementary arithmetic to be the first-order language with a constant symbol, 0, function symbols $S, +, \times$, and $\uparrow$, and a binary relation symbol, $<$. In the intended interpretation these denote zero, the successor function, addition, multiplication, exponentiation, and the usual ordering of the natural numbers, respectively. Bounded quantification is defined by taking $\forall x < t\ \varphi$ to denote $\forall x\ (x < t \to \varphi)$, and taking $\exists x < t\ \varphi$ to denote $\exists x\ (x < t \wedge \varphi)$, where t is any term that does not involve x. A formula is said to be Δ_0, or *bounded*, if each of its quantifiers is bounded, i.e. occurs in one of the contexts above. Since, in such a formula, quantified variables range over only finite segments of the natural numbers, the properties they define (in the standard model) can be decided algorithmically, simply by testing all the relevant values.

Elementary arithmetic is the theory axiomatized by the following set of axioms:[2]

- $S(x) \neq 0$
- $S(x) = S(y) \to x = y$
- $x + 0 = x$
- $x + S(y) = S(x + y)$
- $x \times 0 = 0$
- $x \times S(y) = (x \times y) + x$
- $x \uparrow 0 = S(0)$
- $x \uparrow S(y) = (x \uparrow y) \times x$
- $x < y \leftrightarrow \exists z\ (x + S(z) = y)$
- The schema of induction for bounded formulae:

$$\varphi(0) \wedge \forall x\ (\varphi(x) \to \varphi(S(x))) \to \forall x\ \varphi(x),$$

  where $\varphi(x)$ is Δ_0, possibly with free variables other than x.

Numerals $1, 2, 3, \ldots$ can be defined as $S(0), S(S(0)), S(S(S(0))), \ldots$, and since we can easily derive $S(x) = x + 1$, we can use the latter, more common, mathematical expression in place of $S(x)$. For readability, I will adopt a number of

[2] Variations of this theory go by different names in the literature. Harvey Friedman called an analogous theory *elementary function arithmetic*, or *EFA*. Jeff Paris, Alex Wilkie, and others studied a closely related theory they called $I\Delta_0(exp)$ (see [29]). Alternatively, one can restrict primitive recursive arithmetic to bounded recursion, as described below, and obtain a theory called *elementary recursive arithmetic*, or *ERA*. Note that the name "elementary arithmetic" is also sometimes used to denote classical first-order (Peano) arithmetic, though this usage is becoming less common.

common notational conventions and shorthand, for example, dropping parentheses, writing xy instead of $x \times y$, and writing x^y instead of $x \uparrow y$.[3]

Classical first-order arithmetic is obtained by extending the schema of induction to all formulae in the language. Since, arguably, full induction is justified under our intuitive understanding of the natural numbers, elementary arithmetic may come across as a severely (and perhaps unnaturally) restricted fragment. The fact that it is proof-theoretically weak can be expressed in various ways. For example:

Theorem 2.1 *The consistency of EA can be proved in primitive recursive arithmetic.*

Many (e.g. Tait [76]) take primitive recursive arithmetic, *PRA*, to be a reasonable formalization of Hilbert's informal notion of "finitary" reasoning, and those that dispute this identification typically maintain that finitary reasoning is properly stronger. As a result, this first theorem is usually understood to say that elementary arithmetic has a finitary consistency proof. Here is another sense in which it is weak:

Theorem 2.2 *If EA proves $\forall x\, \exists y\, \varphi(x, y)$, where φ is bounded, then there is a term t, not involving y, such that EA also proves $\forall x\, \exists y < t\, \varphi(x, y)$.*

This second theorem is a special case of a more general theorem due to Rohit Parikh [54] (see also [14]), and implies that each Δ_0-definable function of *EA* is bounded by a finite iterate of the exponential function.

A first objection to the claim that a good deal of mathematics can be carried out in *EA* is that the language is not even rich enough to *express* basic mathematical notions. The language *is* barely expressive enough to state Fermat's last theorem:

$$\forall x, y, z, w\, (x \neq 0 \wedge y \neq 0 \wedge z \neq 0 \wedge w > 2 \rightarrow x^w + y^w \neq z^w).$$

Similarly, one can define notions like divisibility

$$x|y \equiv \exists z\, (xz = y)$$

and primality[4]

$$\mathit{Prime}(x) \equiv (x > 1 \wedge \forall y, z\, (x = yz \rightarrow y = 1 \vee z = 1)).$$

But, at least on the surface, there is no way to refer to ordered pairs of numbers, or finite sets and sequences; and surely these are staples of any introductory text on number theory.

[3] One can show that, in fact, *EA* is finitely axiomatizable; see [29, Theorem V.5.6]. I am grateful to Robert Solovay for correcting my claim to the contrary in an earlier draft.

[4] In a more general algebraic context, this formula expresses that x is *irreducible*, and the primality of a nonzero element x is better represented by the formula $\forall yz\, (x|yz \rightarrow x|y \vee x|z)$. For the natural numbers, however, these two notions coincide, and this fact is derivable in *EA*.

The objection is easily met by noting that one can introduce such notions in definitional extensions of EA. For example, we can freely introduce relation symbols $R(\vec{x})$ to abbreviate Δ_0 formulae $\varphi(\vec{x})$, with defining axiom $R(\vec{x}) \leftrightarrow \varphi(\vec{x})$. The obvious translation, replacing R by φ everywhere, allows us to eliminate the use of the new symbol from proofs, even if the induction schema is extended to bounded formulae involving R. Similarly, whenever we can prove $\forall \vec{x}\ \exists y\ \psi(\vec{x}, y)$ for some Δ_0 formula $\psi(\vec{x}, y)$, we can introduce a new function symbol f with defining axiom $\psi(\vec{x}, f(\vec{x}))$. An only slightly more involved translation allows us to eliminate these new function symbols as well.[5]

We can use such definitions to augment the language of EA. For example, fixing an appropriate means of coding pairs of natural numbers as a single number, we can introduce a function symbol $\langle \cdot, \cdot \rangle$ for pairing and functions symbols $(\cdot)_0, (\cdot)_1$ for projections, and prove

- $(\langle x, y \rangle)_0 = x$
- $(\langle x, y \rangle)_1 = y$

More importantly, we can define functions *length*, *element*, and *append* to handle sequences of numbers, and, writing $(s)_i$ instead of $\mathit{element}(s, i)$ to denote the ith element of s, we can prove

- $\mathit{length}(0) = 0$
- $\mathit{length}(\mathit{append}(s, x)) = \mathit{length}(s) + 1$
- $\forall i < \mathit{length}(s)\ ((s)_i = (\mathit{append}(s, x))_i)$
- $(\mathit{append}(s, x))_{\mathit{length}(s)} = x$

The idea is that 0 denotes the empty sequence, and, more generally, each number s represents a sequence

$$\langle (s)_0, (s)_1, \ldots, (s)_{\mathit{length}(s)-1} \rangle.$$

With a theory of sequences in hand, we can justify the definition of additional functions by bounded primitive recursion: given g, h, and k of appropriate arities, we can define a new function f by the equations

$$\begin{aligned} f(0, \vec{y}) &= g(\vec{y}) \\ f(x+1, \vec{y}) &= \begin{cases} h(x, f(x, \vec{y}), \vec{y}) & \text{if this is less than } k(x, \vec{y}) \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

This is just the familiar schema of primitive recursion, with the added stipulation that the function under definition cannot grow faster than another function, k,

[5] The first step is to replace $\psi(\vec{x}, y)$ by a formula $\psi'(\vec{x}, y)$ that determines y uniquely by requiring it to be the least value satisfying ψ. In other words, if we define $\psi'(\vec{x}, y)$ to be $\psi(\vec{x}, y) \wedge \forall z < y\ \neg\psi(\vec{x}, z)$, then ψ' is still Δ_0, and it satisfies $\forall \vec{x}\ \exists! y\ \psi'(\vec{x}, y)$ as well as $\forall \vec{x}, y\ (\psi'(\vec{x}, y) \rightarrow \psi(\vec{x}, y))$.

previously defined. The class of functions obtained by starting with 0, sucessor, addition, multiplication, and exponentiation, and closing under composition and the schema above is known as the class of *Kalmar elementary functions*, from which elementary arithmetic gets its name; roughly, these are the functions that can be computed in time and/or space bounded by a fixed iterate of the exponential function.[6] The schema of bounded recursion above can, in turn, be used to justify fancier forms of recursion, providing us with additional flexibility in defining new functions and relations.

We can then proceed to code finite sets of numbers, for example, intepreting $x \in y$ to mean "the xth least-significant bit in the binary representation of y is 1." With this definition in hand we find, all of a sudden, that we have access to the apparatus of modern set theory, as long as we restrict our attention to finitary objects. For example, a function from a (finite) set A to a (finite) set B is just a (finite) set of ordered pairs F, such that for every x in A there is a unique y in B such that $\langle x, y \rangle$ is in F; a (finite) group consists of a (finite) set, paired with a binary operation satisfying the usual group axioms; and so on. Bounded quantification over sets, $\forall x \in y$ and $\exists x \in y$, reduces to ordinary bounded quantification in the language of *EA*.

In fact, with appropriate definitions of the union and power set operations, the following are all provable in elementary arithmetic:

- Δ_0 Separation: $\exists z\, \forall x\, (x \in z \leftrightarrow x \in y \wedge \varphi(x))$, where φ is Δ_0 and z is not free in φ
- Union: $\forall x\, (x \in \bigcup y \leftrightarrow \exists z \in y\, (x \in z))$
- Power set: $\forall x\, (x \in \mathcal{P}(y) \leftrightarrow x \subseteq y)$
- Δ_0 Foundation: $\exists x\, \varphi(x) \rightarrow \exists x\, (\varphi(x) \wedge \forall y \in x\, \neg\varphi(y))$, where φ is Δ_0.

The reader may recognize these as key axioms of Zermelo set theory, with restrictions on the formulae that can appear in the separation and foundation axioms. It is a familiar fact that much of ordinary mathematics takes place in Zermelo set theory; so much so that Quine judged objects unavailable in that theory to be "without ontological rights" (Quine [58], quoted in Feferman [20]). Without an axiom of infinity, we have to relinquish the set of natural numbers, and we have restricted separation and foundation in the axioms above; but otherwise, Zermelo set theory remains intact. This goes a long way towards explaining why elementary arithmetic turns out to be a robust foundation for *finitary* mathematics.[7]

More dramatic interpretations and translations can be used to extend the reach of *EA* even further. In the context of a theory, a formula is said to be Σ_1 (respectively, Π_1) if it is equivalent to the result of adding a single block

[6] There are a number of equivalent characterizations of the Kalmar elementary functions, many of which are presented in Rose [61]. Rose credits the definition of the elementary functions to Kalmar (1943) and Csillag (1947).

[7] The interpretation of set theory in even weaker fragments of arithmetic has been considered by Sazonov, e.g. in [62].

of existential (respectively universal) quantifiers to a Δ_0 formula. A formula is said to be Δ_1 if it is provably equivalent to both Σ_1 and Π_1 formulae. Bearing in mind that Δ_0 properties are finitely checkable, it should not seem surprising that there is a precise sense in which the Σ_1-definable properties are exactly the *computationally verifiable* ones; the Π_1-definable properties are the *computationally refutable* ones; and the Δ_1-definable properties are the ones that are *computationally decidable*. This hierarchy can be extended; for example, a Π_2 formula is obtained by prepending a block of universal quantifiers to a Σ_1 formula. Σ_1 *collection* is the axiom schema,

$$\forall x < a\ \exists y\ \varphi(x, y) \rightarrow \exists w\ \forall x < a\ \exists y < w\ \varphi(x, y),$$

for Σ_1 formulae φ. In set-theoretic terms, this is a restricted version of the replacement principle that distinguishes Zermelo-Fraenkel set theory from Zermelo's original version. Harvey Friedman and Jeffrey Paris have shown, independently and using different model-theoretic proofs, that one can add this principle to EA, without changing the Π_2 consequences of the theory (see [29, 35]).[8] Wilfried Sieg [66] has used Gerhard Genzten's method of cut elmination to obtain an effective translation from one theory to the other (see also [10], or [2] for a model-theoretic version of Sieg's proof). More recently, Petr Hájek [28] has shown that one may even obtain this conservation result by a direct intepretation.

The principle of Σ_1 collection can be used to justify the principle of induction for Δ_1 formulae, i.e. induction for decidable properties.[9] One consequence of this is that one can develop, in a conservative extension of EA, a workable theory of recursive sets and functions. To make this precise, let us take the language of second-order arithmetic to have both number variables $x, y, z, \ldots$, and set variables $X, Y, Z, \ldots$ with a binary relation $x \in Y$ relating the two. (Functions can be interpreted in the usual way, as sets of ordered pairs.) The theory RCA_0^* of [71, 72] is obtained by extending the schema of induction in EA by allowing free set variables to occur in the induction formulae, and adding the schema

[8] A Π_2 sentence $\forall x\ \exists y\ \psi(x, y)$ can be understood as making the computational assertion that for every input x, a program that searches for a y satisfying $\psi(x, y)$ is bound to find one. Conversely, statements of the form "the function computed by algorithm e (or Turing machine e) is total" are Π_2. Thus, the Π_2 consequences of a theory can be said to characterize its computational content.

Many, but of course not all, important theorems of number theory and combinatorics can be expressed naturally as Π_2 statements. Andrew Arana has, however, suggested the Hilbert-Waring theorem as an example of a statement that is not of this form. Proved by Hilbert in 1909, this theorem asserts that for every k there is a bound N such that every natural number can be written as the sum at most N kth-powers, a Π_3 assertion. But in the 1920's Hardy and Littlewood provided an explicit (and computable) bound on N in terms of k, and incorporating any such strengthening in fact renders the assertion Π_1. (See [30, Notes on Chapter XXI] and [52, Section 11.5] for notes and references.) On the other hand, Roth's theorem is, famously, a Π_3 assertion for which it is still open as to whether there is a computable bound; see e.g. [49, 50]. See also the discussion of the prime number theorem in footnote 12 below.

[9] It is still open as to whether Σ_1 collection is strictly stronger than Δ_1 induction; see, for example, [29, 11].

RCA of *recursive comprehension axioms*:

$$\forall x\ (\varphi(x) \leftrightarrow \psi(x)) \rightarrow \exists Y\ \forall x\ (x \in Y \leftrightarrow \varphi(x)),$$

where φ and ψ are Σ_1 and Π_1, respectively, possibly with free set and number variables. In words, *RCA* asserts that every computably decidable property determines a set of numbers. So, in RCA_0^* one can reason about infinite sets of natural numbers, and functions from numbers to numbers. While the axioms of RCA_0^* are true when set variables are interpreted as ranging over computable sets, none of the axioms *commit* one to this intepretation; they are satisfied equally well by the collection of, say, *arithmetic* sets, or the full power set of $\mathbb{N}$. By interpreting sets in terms of indices for computable functions, we can interpret RCA_0^* in *EA* plus collection, so the former theory is conservative over the latter for formulae in the common language.

In fact, one can do even better. In the 1980's Harvey Friedman drew attention to an axiom he called *weak König's lemma*, or *WKL*, which expresses the second-order assertion that every infinite binary tree has an infinite path. This is a form of compactness; in the context of RCA_0^* it can be used to prove the compactness of any closed bounded interval of real numbers, in the sense that every covering by open sets has a finite subcover, as well as the completeness and compactness of first-order logic. In fact, in the spirit of reverse mathematics, we can show that over RCA_0^* weak König's lemma is even *equivalent* to these principles. Sieg [67] has shown that nonetheless WKL_0^* is still conservative over elementary arithmetic for Π_2 sentences. Stephen Simpson and his student, Rick Smith, have shown [71, 72] that, moreover, adding weak König's lemma does not change the Π_1^1 theorems of RCA_0^*. Simpson and Smith used a model-theoretic argument to prove the conservation theorem for RCA_0^*, but using the methods of Avigad [4] a direct interpretation of WKL_0^* in RCA_0^* can also be obtained.[10]

[10]The unrestricted version of König's lemma [44] asserts that every finitely branching tree with arbitrarily long paths has an infinite path; weak König's lemma is the restriction of this principle to binary trees (this is, trees on $\{0, 1\}$). This restriction has an interesting history. Its (constructively inequivalent) contrapositive, i.e. the assertion that every binary tree with no infinite path is finite, is essentially the *fan theorem* of L. E. J. Brouwer (see [80, Volume 1]). Stephen Kleene showed that the theorem is false when interpreted as ranging over *computable* sets and functions: there is a computable infinite binary tree with no computable path. It was Friedman who first observed that the principle is logically weak in the context of certain fragments of second-order arithmetic; in particular, he used a model-theoretic argument to show that the theory WKL_0, which adds Σ_1 induction to WKL_0^*, is conservative over primitive recursive arithmetic for Π_2 sentences. Proof-theoretic proofs were later obtained by Sieg [68], who used cut elimination, and Kohlenbach [39, 41], who used the Dialectica interpretation and extended the result to weaker theories (see also [6]). Leo Harrington was able to strengthen Friedman's theorem by showing that WKL_0 is conservative over RCA_0 (i.e. RCA_0^* plus Σ_1 induction) for Π_1^1 sentences. (RCA_0 is directly interpretable in the the restriction $I\Sigma_1$ of Peano arithmetic in which induction is only allowed on Σ_1 formulae; the latter had been shown to be conservative over *PRA* for Π_2 sentences by Grigori Mints, Charles Parsons, and Gaisi Takeuti, independently). Harrington's proof used a forcing argument inspired by the "low basis theorem" of Carl Jockusch and Robert Soare [34]; both Friedman's and Harrington's proofs can be found in see [71]. Syntactic proofs of Harrington's theorem, yielding explicit translations, are due to Hájek [28] and Avigad [4]. The work of Simpson and Smith described above is the analogue of Harrington's theorem for RCA_0^*. Fernando Ferreira proved

One need not stop with sets and functions; one can just as well design conservative extensions of elementary arithmetic with higher-order functionals, i.e. operations mapping functions to functions, and so on. One way to proceed is simply to interpret such a theory in RCA_0^* by developing, in RCA_0^*, a theory of computable functionals, such as Kleene's hereditarily recursive operations (see, for example, [6, 79]). Ulrich Kohlenbach has shown that various forms of Gödel's *Dialectica* intepretation can also be used to obtain conservation results, yielding additional information. Indeed, as part of his proof mining program Kohlenbach [37] has developed many analytic notions—continuous functions, integration, transcendental functions like e^x, sine and cosine—in theories that are even weaker, and has explored ways of eliminating various types of analytic principles, weak König's lemma among them [39, 41]. Much of the analysis carries over, a fortiori, to the types of systems we are interested in here.

To consider one last dimension, one can even extend higher-order conservative extensions of elementary arithmetic with "nonstandard" reasoning, i.e. reasoning involving nonstandard natural numbers as well as infinitesimal rationals (see [3, 7, 17, 73]). This provides yet another weak framework in which one can develop analytic notions in a natural way.

3 Case studies

An *arithmetic progression* is a sequence of the form

$$a, a + d, a + 2d, \ldots$$

where a and d are natural numbers. If a and d have a factor in common, then every term in the sequence will share that factor. For example, in the sequence

$$6, 15, 24, 33, \ldots$$

every term is divisible by 3, and so none of the terms are prime. When a and d have no factor in common, the situation is entirely different:

Theorem 3.1 *If a and d are relatively prime, the arithmetic progression $a, a + d, a + 2d, \ldots$ has infinitely many primes.*

Adrien-Marie Legendre made some attempts towards proving this in his *Théorie des Nombres* of 1798, but a proof was not obtained until 1837, when Peter Gustav Lejeune Dirichlet succeeded in doing so via a dramatic use of analytic methods. In modern terms, to each value of d one assigns a group of *Dirichlet*

a version of the conservation result for a theory of polynomial time computable arithmetic (see [25, 24]). There are also connections between weak König's lemma and nonstandard arithmetic (see Avigad [3] and Tanaka [78]). See also Kohlenbach [38] for a discussion of uniform versions of weak König's lemma in a higher-order setting.

An ω-model of WKL_0, i.e. a collection of sets closed under recursive definability and containing a path through every infinite binary tree, is called a *Scott set*. Their importance to the study of models of arithmetic was first demonstrated by Dana Scott [63], and was rediscovered and put to dramatic use by Friedman (see Kaye [35]).

characters, that is, functions χ that map the group of natural numbers relatively prime to d homomorphically to complex roots of unity (and map the rest of the natural numbers to 0). For each such character χ one has the *Dirichlet series*

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s},$$

where s ranges over the complex numbers. The proof of Theorem 3.1 then uses methods of complex analysis—limits, differentiation, integration, and analytic continuation—to characterize the behavior of $L(s, \chi)$ when s is close to 1.

It would be hard to overstate the importance of Dirichlet's theorem to contemporary mathematics. This theorem, together with Dirichlet's paper *Sur l'usage des séries infinies dans la théorie des nombres*, launched the field of analytic number theory. Characters and L-series are still central to that subject, and to algebraic number theory as well.

The theorem is also interesting from a foundational point of view. In the early nineteenth century, analysis was thought to be grounded in geometric intuitions about the continuum of real numbers, and relying on these intuitions to establish a number-theoretic assertion must have seemed unsettling, or methodologically impure, to some. Howard Stein [74] surmises that this was one of the motivations behind the program, often attributed to Dirichlet, of interpreting analytic notions in terms of the natural numbers. Dirichlet's influence on Kronecker and Dedekind is well known, as are the radically different ways the two interpreted this challenge: for Kronecker, the reduction was to take place via explicit symbolic representations and algorithmic procedures, whereas Dedekind's conceptual reduction involved abstract set-theoretic notions. This tension is clearly discernible in foundational writings by Cantor, Kronecker, and Dedekind in the 1880's and 1890's, which, in turn, influenced foundational developments in the early twentieth century. So, by forcing the mathematical community to come to terms with the use of analytic methods in number theory, Dirichlet's theorem can be seen as a point at which views as to the appropriate goals and methods of mathematics began to diverge, with important foundational effects later on. (See Avigad and Reck [8] for a survey of such mathematical developments, and their influence on logic via the Hilbert school.)

Modern presentations of Dirichlet's proof are accessible to beginning graduate students. Ireland and Rosen [33] provides a nice account, as does Serre [65]. Note that Dirichlet's theorem can be expressed, directly, in the language of elementary arithmetic, using "arbitrarily large" as the appropriate reading of "infinitely many":

$$\forall a > 0, d > 1 \; (\mathit{RelPrim}(a, d) \rightarrow \forall x \; \exists y \; (y > x \wedge \mathit{Prime}(a + yd)))$$

where

$$\mathit{RelPrim}(a, d) \equiv (\forall z \; (z|a \wedge z|d \rightarrow z = 1))$$

expresses that a and d are relatively prime, and $|$ and *Prime* express divisibility and primality, as discussed above.[11]

Despite the use of analytic methods in the original proof, it turns out that Dirichlet's theorem can be derived in weak fragments of arithmetic. Patrick Cegielski [16] has shown that primitive recursive arithmetic, or PRA, suffices; the proof amounts to formalizing the usual analytic methods in RCA_0, and then invoking Π_2 conservativity over PRA. There seems to be no bar to adapting the definitions to RCA_0^*, and, with some extra care, getting the argument to go through there; this would yield derivability in EA.

Indeed, there are alternative proofs of Dirichlet's theorem that seem to be amenable to direct formalization in EA. In 1949, the number theorist Atle Selberg published an elementary proof of Dirichlet's theorem [64], avoiding the use of complex characters and infinite sums. A discussion of Selberg's result, as well as a slightly less elementary proof that uses infinite limits but avoids complex analysis, can be found in Melvyn Nathanson's book, *Elementary methods in number theory* [52]. Georg Kreisel mentioned Dirichlet's theorem in [47, Section 3.3] and [48, Section 3.2] (see also [50]) as an example of a proof amenable to his "unwinding" program, which was designed to extract useful information from mathematical proofs.

Keep in mind that the word "elementary" is being used in many different ways in this discussion. In informal mathematical usage, "elementary" is often used to mean "simple" or "easy to understand." In the Selberg/Nathanson sense it is used to mean "avoiding the use of infinitary (or analytic) methods." The fact that Selberg's proof of Dirichlet's theorem is harder to understand than the standard analytic one shows that these two senses can be at odds! I have already noted that the use of the word "elementary" in the phrase "elementary arithmetic" is due to the fact that the axiom system is closely related to the class of "elementary functions" in the sense of Kalmar. To complicate matters even further, among logicians the word "elementary" is often used to mean "first-order."

Returning to the issue of formalizability in weak theories, the prime number theorem provides another interesting case study. This asserts that the quotient $\pi(x) \log x/x$ approaches 1 as x approaches infinity, where $\pi(x)$ denotes the number of primes less than or equal to x. The prime number theorem was first proved by Hadamard and de la Vallée Poussin independently in 1896, again using the methods of complex analysis. But Gaisi Takeuti [77] developed enough complex analysis to carry out the proof of the prime number theorem in a conservative extension of first-order arithmetic, and with additional care Oliver Sudac has shown [75] that $I\Sigma_1$ suffices.[12] Once again, elementary proofs have been found,

[11] Bringing the quantifiers over x and y to the front of the formula and bounding the quantifier over z shows that this formulation of Dirichlet's theorem is (equivalent to a sentence that is) Π_2.

[12] Here, however, more work needs to be done to show that the prime number theorem is provable in EA, or even PRA. Sudac [75] incorrectly claims (pages 186 and 235) that $I\Sigma_1$ ($RI\Sigma_1$ in his notation) is a conservative extension of PRA. This is of course false, since, for example, PRA does not prove Σ_1 induction itself. Although $I\Sigma_1$ *is* conservative over PRA

by Selberg and Paul Erdös. Selberg's proof is described in both Nathanson's book as well as Hardy and Wright [30], and both sources provide historical notes and references.

Incidentally, the method of working with conservative extensions can be used to clarify slightly different senses in which one can respond to the challenge of showing that a particular theorem can be derived in elementary arithmetic. One can ask for:

- an informal proof that there is a formal derivation of the theorem in some conservative extension of EA;
- an informal description of a formal derivation in some conservative extension of EA, and an explicit procedure for translating this to a derivation in EA;
- an informal description of a derivation of the theorem in EA; or
- a formal derivation of the theorem in EA.

The first option is the most liberal: one can use *any* mathematical argument at all, constructive or not, to establish the existence of the required derivation. For example, model-theoretic methods are commonly used to obtain conservation results, and often have such a non-constructive character. This is enough to establish the existence of the purported derivation by general mathematical standards. But, from a foundational point of view, one prefers to have an explicit (and finitary) proof that there is a derivation; or, better, an explicit description of the derivation itself. In fact, the conservation results and case studies discussed here all meet the second, more stringent, requirement above; and, with some metamathematical trickery (e.g. formalizing proofs of conservativity and partial soundness) can be construed as yielding proofs of the third type as well. The latter is akin to the standards that we use when we claim, for example, that an ordinary mathematical proof can be justified on the basis of the axioms of set theory: without presenting a formal derivation, we rest content with a proof that uses only definitions and methods whose formal representations are taken to be straightforward. But one might worry that our intuitions with respect to provability in weak theories are not as good as our intuitions with respect to provability in ZFC. In that case, the fourth type of evidence might be the most convincing, if not the most illuminating: one demands an explicit symbolic derivation, say, in a format amenable to mechanical verification. One strategy towards obtaining such a derivation is to use a semi-automated computerized proof assistant to derive the theorem in a conservative extension of EA, and then implement the appropriate translation. This has never been

for Π_2 sentences, natural arithmetic formalizations of the prime number theorem (including the one used by Sudac on page 234) are at best Π_3. So to obtain provability in PRA, one needs to either strengthen the conclusion (say, by incorporating an explicit, computable rate of convergence), or work in PRA itself. In this case, formalizing Selberg's elementary proof may be more straightforward.

done for a theorem as complex as Dirichlet's, but it seems to be within striking distance of today's technology.[13]

At this stage, we can sum up the proof-theorist's cause for optimism with respect to Friedman's conjecture. Using only straightforward definitional extensions, EA is strong enough for a direct formalization of the methods of finitary number theory, say along the lines of Hardy and Wright's classic textbook [30]. Furthermore, paradigmatic uses of central ideas and methods from analytic number theory can be formalized in appropriate conservative extensions. Of course, this is not to say that *every* analytic method can be formalized in EA. For example, analytic principles like the least upper bound theorem and the Bolzano-Weierstrass theorem are equivalent, over RCA_0, to an arithmetic comprehension principle that yields a conservative extension of full first-order arithmetic, but has consequences that go beyond the strength of PRA. Infinitary combinatorial principles like the infinitary version of Ramsey's theorem have a similar character, and the methods of descriptive set theory go well beyond that.[14] But concerted efforts from the proof theory community have failed to turn up even a single example of a "standard" number theoretic or combinatorial result that uses such infinitary principles in an essential way. One *can* cook up examples of finitary combinatorial statements that require strong axioms: in addition to consistency statements, there are, for example, Goodstein's theorem, the Paris-Harrington theorem, and Friedman's finitary version of Kruskal's theorem (see, for example, [55, 26] and Friedman's contribution to [23]). But all of these were designed by logicians who were explicitly looking for independent statements of this sort. Restricting our attention to ordinary mathematical literature, then, there is some positive evidence for Friedman's conjecture, and a striking absence of counterexamples. Barring the corrupting influence of logicians on the *Annals of Mathematics*, the conjecture may seem to be in good stead.

At this point the skeptic can object that the proof theorist's purported evidence is hopelessly meager. Dirichlet's theorem and the prime number theorem may have been state of the art more than a century ago, but mathematics has grown considerably since then; Wiles' proof of Fermat's last theorem [82], for example, represents a degree of difficulty that is greater by leaps and bounds. The optimistic proof theorist can respond that *difficulty* is not the central issue here: if there is one lesson to be learned from the case studies above, it is that it is a mistake to conflate mathematical difficulty with the need for strong axioms. So, to say that the proof of Fermat's last theorem uses strong axioms in an essentially ineliminable way means not just that it is more difficult than the proof of, say, Dirichlet's theorem; but that the basic concepts and methods involved are of an entirely different *character*. To this, our skeptic can reply that, indeed, the methods involved in the proof of Fermat's last theorem may well have such a different character: it is foolhardy to assume that the modern

[13] Interactive proof systems like Isabelle [86], Coq [83], and HOL [85] provide support for the formalization of elaborate proofs. Isabelle is especially well adapted to this type of research, since it allows the user to specify the underlying deductive system.

[14] See Simpson [71] for a thorough investigation of these issues.

machinery associated with elliptic curves, modular forms, and Galois representations is fundamentally "like" the ancestral technology. Whatever one's bias, it should be clear that more information is needed to mount a convincing case either way.

One can imagine that a fruitful collaboration between a proof theorist and a number theorist could yield progress towards resolving the issue. The number theorist could explain the central concepts and types of inferences involved in the proof of Fermat's last theorem; the proof theorist could explain the various tricks and representations that are used to keep axiomatic requirements down; together they might be able to home in on the potential problems and difficulties. Attempts to formalize and adapt a few key lemmata could serve to bolster optimism or justify a more skeptical stance. As indicated in the introduction, however, I think it is fruitless here to speculate as to the outcome. Instead, I would like to address the issue of *why* we might want to embark on such a pursuit. In other words, why should we care?

Without drawing a fine line between mathematical and philosophical motives, let me quickly dispense with some of the more mathematical reasons one might be interested in such a program. As the work of Selberg, Erdös, and Nathanson suggests, the use of restricted methods can be of mathematical interest in its own right. In particular, the act of "mathematizing with one's hands tied" can often yield new proofs and a better understanding of old results. It can, moreover, lead to interesting questions and fundamentally new results, such as algorithms or explicit bounds that are absent from nonconstructive proofs. By the completeness theorem, the question as to the derivability of Fermat's last theorem in EA is equivalent to asking whether FLT is true of a class of structures bearing sufficient similarity to the natural numbers. The question as to whether FLT is true of the natural numbers occupied some of the greatest minds of mathematics for more than three and a half centuries; the question as to whether it is true of the more general class of structures might well be worth a second look.

In the next section, however, I want to consider why Friedman's conjecture might be of interest to the working philosopher. I will not be so much intersted in traditional questions related to the metaphysics and epistemology of mathematics, though, as I will indicate below, I believe that the discussion can be brought to bear on such issues. More broadly, I take it that at least one aspect of philosophy of mathematics involves the attempt to characterize both the methods and goals of mathematics, and to understand the extent to which the methods are suited to the goals. With this in mind, I can formulate the question I wish to address more pointedly: what do formal results related to Friedman's conjecture have to tell us about the methods of contemporary mathematics?

4 Philosophical issues

I have described a general program of exploring the extent to which theorems of number theory and combinatorics can be derived in elementary arithmetic. To

avoid any misunderstandings, I think it wise to enumerate some of the claims that are *not* implicit in such a pursuit.

First of all, of course, one should not claim that *all* theorems of combinatorics and number theory can be derived in weak theories, since Gödel's incompleteness theorem shows that in any reasonable theory of arithmetic there will be true arithmetic statements that are underivable. Indeed, a formalization of the theorem of Matjasevic, Robinson, Davis, and Putnam shows that in any consistent extension of elementary arithmetic there are true universally quantified Diophantine equations that cannot be derived (see the discussion of C. Dimitracopoulos' formalization in [29, Section I.3.d]). Beyond the independent combinatorial statements mentioned in the last section, in recent years Harvey Friedman has made a good deal of progress towards producing "natural" examples of combinatorial statements that require strong axioms (see [23]). The question as to the extent to which Friedman's examples are, or may become, part of mainstream mathematics is an interesting and important one. But this opens an entirely new avenue for discussion, one that I do not wish to pursue here.

Even restricting one's attention to ordinary theorems of number theory and combinatorics (that is, those of the sort covered by Friedman's conjecture), it is important to keep in mind that I am not making the claim that the usual proofs go through in weak theories, unchanged. As noted above, adapting proofs to weak systems usually involves restricting the generality of intermediate claims and paying close attention to the linguistic complexity of relevant notions, tasks which are cumbersome and annoying (and generally unnecessary) for the working mathematician.

For these reasons, it is important to note that research towards establishing Friedman's conjecture does not inherently involve normative or descriptive claims. In other words, one need not claim that mathematicians *should* restrict their methods to those that can be formalized in elementary arithmetic, nor that elementary arithmetic supplies faithful representations of what ordinary mathematicians do, in any reasonable sense. Similarly, logicians studying weak formal theories need not maintain that much stronger theories are not also worthy of study. Set theory aims to provide a rich and broad foundational framework, and to explore concepts that are either mathematically interesting in their own right or have bearing on other branches of mathematics. The question as to what extent large cardinal and determinacy axioms contribute in this regard is also a subject worthy of debate, but, in my view, the question as to whether Fermat's last theorem can be derived in elementary arithmetic is, for the most part, irrelevant to the discussion. (The panel discussion [23] provides an in-depth and thoughtful presentation of some of the major schools of thought with respect to these issues.)

Given these qualifications, what are we to make of the evidence in favor of Friedman's conjecture? There are, I think, two clear morals one can extract.

The first is that it is a mistake to confuse mathematical difficulty with logical strength; in other words, there is a difference between saying that a proof is hard, and saying that it requires strong axioms. It is trivial to show that

Zermelo-Fraenkel set theory is consistent, assuming the existence of an inaccessible cardinal. In contrast, Fermat's last theorem may well be provable in elementary arithmetic; but it is unlikely that there is an easy proof.

The second moral is that insofar as the methods of contemporary mathematics are worthy of philosophical attention, and insofar as contemporary mathematics can be formalized in weak theories, weak theories are worthy of philosophical attention as well. In other words, there is a lot more going on "low down" than is commonly assumed. Much of the excitement of the nineteenth century transition to modern mathematics traces to the possibility of using "abstract" methods—operations on infinitary objects and structures, with no explicit computational interpretation—to derive "concrete" results about numbers and finitary objects. The reduction of analytic number theory to elementary arithmetic can be seen as just one way of trying to understand the infinite in concrete terms; surely this can tell us something about the mathematical concepts involved.

Of course, from a foundational point of view, such results are of central importance. At least on the surface, what was at stake in the heated debates of the 1920's and 1930's was the status of infinitary methods in mathematics. A verification of Friedman's conjecture would show that there is a precise sense in which the kind of infinitary methods found in the *Annals of Mathematics* can ultimately be justified relative to finitary ones. Indeed, the principles of elementary arithmetic are acceptable to even the most radical constructivist, and, in particular, fall clearly within the range of methods sanctioned by Kronecker; only the rare ultra-finitist may object. Even if one is skeptical of Friedman's conjecture, at this stage it is at least safe to say that large and interesting portions of mathematics have a finitary justification. As a result, issues that have, historically, shaped philosophical and foundational discussion over the course of the last century now seem rather benign. This, in and of itself, is no small metamathematical accomplishment.

But there is, I think, more to it than that. Many researchers in proof theory are attracted to the subject by its attention to the *methods* of contemporary mathematics, independent of metaphysical issues. The formal results discussed in the last two sections certainly tell us something about these methods. But they do not tell the whole story, and what is equally interesting, perhaps, is what is left out of the analysis.

The picture that emerges from the proof-theoretic inquiry is roughly this. In doing mathematics, mathematicians are drawn to abstractions and general principles that unify diverse branches of their subject, fix terminology, and simplify proofs. The introduction of infinitary, nonconstructive, set-theoretic reasoning towards the end of the nineteenth century is a case in point. In general, the adoption of new mathematical methods need *not* be conservative, which is to say, new principles can yield new consequences; and logicians are good at finding extreme examples that illustrate how this comes about. But, to a large extent, these logicians are trucking in outliers and pathologies, and, in particular, careful proof-theoretic analysis shows that in ordinary mathematical

practice, the full strength of infinitary set-theoretic reasoning is rarely needed.[15]

Thus, the proof theorist seeks both ontological and epistemological reduction, trying to see how little we can get away with. In the end, we learn that, in a sense, we do not need a very rich universe of mathematical objects, nor do we need strong principles of reasoning, as long we are willing to restrict the generality of our theories to the bare minimum needed to obtain concrete (read: finitary, number theoretic, combinatorial, or computational) results. Infinitary objects like topological spaces, manifolds, and measures can be coded as suitable sets of numbers, finitary objects can be coded as numbers, and some basic axioms of arithmetic are then enough to justify the desired mathematical inferences. So, in a sense, mathematics does not need analysis, algebra, or geometry; all it needs is a weak theory of arithmetic.

But certainly there is a sense in which this is false: *of course* mathematics needs topological spaces, manifolds, and measures, and certainly number theory would be crippled by any ban on group characters and complex integrals. Put simply, the success of the reductionist program poses the philosophical challenge of explaining this use of the word "need": if logical strength isn't everything, what else is there? Granted, the proof-theoretic reductions seem to wreak havoc on the underlying mathematical ideas and concepts. But can we say, in precise terms, what is lost in the translation? (Perhaps something to do with the "meaning" of the original proofs?)[16] Looking at things the other way around, if we can say what it is that is lost in the translation, we will have a better sense of what it is that the more abstract methods add to the mathematical enterprise.

Coming to terms with the nature and utility of mathematical concepts will

[15] It is worth emphasizing yet again that one can subscribe to this characterization of contemporary mathematics while denying the stronger claim that mathematics *has* to be this way. For example, it seems likely that Harvey Friedman would assent to the weaker claim, but, if he has his way, the situation will change, and large cardinal assumptions will, in the future, play an ineliminable role in ordinary combinatorics.

[16] One might think that considerations as to the lengths of the formal derivations are relevant here. But although there *are* provably dramatic speedups between ZFC and EA, the difference does not seem to be as large for the translations that come up in practice. For example, most of the conservation results described in Section 2 can be obtained by direct interpretation, with a polynomial bound on the increase in length of proof. Avigad [5], for example, provides an efficient way of eliminating symbols that are introduced in definitional extensions.

Although translations that use cut elimination and normalization *can* lead to superexponential increase in proof length, one can even use tricks to simulate even these translations efficiently. For example, EA can prove the cut-elimination theorem for proofs any fixed depth, so given a proof d of an arithmetic formula φ in the source theory, one can typically construct a closed term t and a proof, in EA, of the assertion $\exists d' < t$ ("d' is a proof of φ in EA"). Then, using a partial truth predicate and Solovay's method of shortening cuts (see [29, 57]) one can construct a short proof of the soundness of EA up to t (for formulae of complexity less than or equal to that of φ); and hence a short proof of φ.

To be sure, derivations will get longer when one uses tricks like these. But the increase is not dramatic, and it seems unlikely that it alone can account for the loss of intelligibility. This is not to deny that length has something to do with explaining how infinitary methods can make a proof simpler and more comprehensible. But the advantages of working in a conservative extension seem to have as much to do with the perspecuity and naturality of the notions involved, and using the number of symbols in an uninterpreted derivation as the sole measure of complexity is unlikely to provide useful insight.

necessarily involve addressing broader issues like the nature of mathematical discovery, explanation, intelligibility, and fruitfulness. To be sure, there *has* been progress along these lines. For example, both the work of Georg Polya and Imre Lakatos' *Proofs and Refutations* try to make sense of the nature of mathematical discovery, albeit in very different ways. Ken Manders has used a number of historical case studies to provide of deeper understanding of the role that conceptual advances play in the development of mathematics; Michael Resnik, Mark Steiner, Paolo Mancosu, and others have made initial attempts at characterizing mathematical explanation (see [51] and the references there); and a number of essays in a recent volume edited by Emily Grosholz and Herbert Breger [27] try to come to terms with the nature of progress in mathematics. These initial starts are promising. But, in large part, the problem is that the philosophy of mathematics still lacks an appropriate language and analytic methodology for discussing mathematical concepts, let alone the utility of abstract or infinitary methods in mathematical practice.

Furthermore, although the use of infinitary methods can simplify a proof, many mathematicians feels that their use comes at a cost. The discussion in the last section shows that even when infinitary proofs are available, more elementary ones are often deemed desirable. This, too, poses philosophical challenges: what is it that such an elementary proof brings to mathematics, over and above what is provided by a more abstract one? Does elementary arithmetic, or first-order arithmetic, provide a useful explication of mathematicians' informal notion of elementarity? Or are there better ways of understanding the notion of an "elementary" method or proof?[17]

Other mathematicians may object that much of the preceeding discussion is based on the implicit assumption that the primary use and justification of infinitary methods is in their application towards obtaining concrete results. Many consider the study of infinitary mathematical structures a worthwhile activity in and of itself; obtaining (or reobtaining) concrete results helps ground and situate the research, but may be no more important than, say, providing conceptual insight and forging connections between diverse branches of mathematics. The fact that attitudes vary in this respect only makes the philosophical task more difficult. What we would really like is a framework that can help us make sense of the range of viewpoints.

I am perhaps optimistic in supposing that the prospects are good for obtaining an interesting and informative theory of mathematical concepts, one that enriches and extends our current low-level accounts of mathematical practice. What is hoped for is a robust theory that can help us understand how mathematical knowledge is structured and applied, how mathematical theories develop, how conjectures are formed, what types of theorems are judged to be desirable, and what kinds of developments are understood to constitute mathematical

[17] One might hope to obtain a better formal notion of elementarity by somehow restricting elementary arithmetic in such a way as to avoid coding of sequences, and so on. For example, one might disallow bounded quantification in induction formulae, but augment the language with more specifically arithmetic and algebraic resources. Kreisel, for example, hints at such an approach towards characterizing the notion of a "direct proof" in [46, page 248].

progress; in much the same way that symbolic logic, axiomatic foundations, and philosophical reflection on deductive and semantic notions currently help us understand everyday mathematical language, constructions, and inference. If such a theory is possible it will likely involve the convergence of ideas from a number of disciplines, representing a diversity of approaches to understanding the nature of mathematics. It should combine a sensitivity to mathematical and computational issues with philosophical reflection, historical insight, and formal, logical analysis. Such a theory may even (contra Frege) benefit from a deeper understanding of human cognitive capabilities.

Inspiration as to how to obtain more robust accounts of mathematical practice may well from sources that have received relatively little philosophical attention, such as the fields of automated deduction, semi-automated deduction, and the computer-assisted formalization and verification of mathematics. After all, complete formalization is tedious, and tedium can inspire a good deal of creativity. Researchers in these fields have already begun to develop systems to facilitate the process;[18] the pragmatic tricks and tools that are designed towards formalizing, storing, accessing, and exchanging mathematical knowledge efficiently will, I believe, help us better understand how that knowledge is structured.

Appendix: further reading

In this essay I have touched on a number of topics in metamathematics and proof theory. As a survey, it is not comprehensive in any respect, or even a balanced overview. Taken together, the following references provide a fuller picture, and provide a starting point for further exploration.

For the model theory and proof theory of weak theories of arithmetic, see Hájek and Pudlák [29], Kaye [35], and Buss [14]. Together these cover most of the theorems mentioned in Section 2. A series of conferences known as Journées sur les Arithmétiques Faibles (JAF) is devoted to the study of weak fragments of arithmetic; see surveys by Denis Richard [60] and J.-P. Ressayre [59], as well as the conference web page [87].

For elementary arithmetic in particular, [29] provides the best overview. When it comes to formalizing metamathematics in elementary arithmetic, there is a big difference between provability with and without cut; [29] presents many of the seminal results that clarify this relationship, including a striking theorem due to Alex Wilkie that implies that elementary arithmetic cannot even prove the consistency of Robinson's Q.

In a sense, mathematical questions having to do with formalizing mathematics in weak theories become more interesting when one passes to fragments of arithmetic that are even weaker than EA, since in such theories many of the usual finitary techniques of number theory and combinatorics can not be carried out. See [56] and [18] for examples of what can be done in such theories. For

[18] See, for example, Harrison [31].

an overview of the subject, as well as some of the interesting connections to the subject of computational complexity, see Krajícek [45], as well as [29, 59].

As far as the formalization of mathematics in theories of second-order arithmetic, Simpson [71] is the standard reference, combining such formalization with the study of metamathematical and model-theoretic properties of the theories involved. The book is also a significant contribution to Friedman's program of Reverse Mathematics, where one aims to show that the theories used in the formalization are minimally sufficient, in the sense that over a weak base theory the theorems under investgation are in fact equivalent to the axioms used to prove them.[19] (See also [70].)

Simpson's book draws heavily on the tradition of constructive mathematics, which places different restrictions on the methods involved. The goal of the latter is to develop portions of mathematics in a computationally informative (or intuitionistically justified) way. Despite L. E. J. Brouwer's antagonism towards formalism, there is currently a good deal of interest in specifying formal deductive systems that represent the informal practice, and understanding their metamathematical properties. See [9, 12, 80], and [13] for a recent survey.

From the 1950's onwards, Georg Kreisel urged the application of proof theoretic methods towards obtaining additional mathematical information from nonconstructive proofs in various branches of mathematics. His "unwinding program" met with mixed results; see the discussion in [21, 50] and other essays in [53]. But, under the rubric of "proof mining," Ulrich Kohlenbach has made important progress in the use of metamathematical methods to extract useful information and sharper results from nonconstructive proofs in numerical analysis and approximation theory. See [41, 42, 37] for an overview of the methods, and [40, 43] for some applications.

Contemporary proof theory can be characterized broadly as the general study of deductive systems, mathematical or otherwise. By now, a sprawling array of loosely affiliated disciplines can be grouped under this general heading. In this essay, I have focused on (one aspect of) the *metamathematical* branch of the subject, where, in the tradition of Hilbert and Bernays' *Grundlagen der Mathematik* [32] and Kleene's *Introduction to Metamathematics* [36], the goal is to study theories which minimally capture local features of mathematical practice, with an eye towards understanding these theories in relatively concrete, constructive, or computational terms. For a historical overview, see [8].

Finally, for positive examples of a healthy interaction between formal and more broadly philosophical approaches to the philosophy of mathematics, see, for example, [22, 53, 69].

[19] Wilfried Sieg points out that the spirit of such reversals can be found as early as 1872 in Dedekind's famous *Stetigkeit und die irrationale Zahlen* [19], at the end of which Dedekind shows that his "continuity principle" is in fact equivalent to fundamental theorems of analysis like the least upper-bound principle. The same spirit can be found in the well-known fact that many interesting consequences of the axiom of choice are known to be *equivalent* to the axiom of choice over Zermelo-Fraenkel set theory.

References

[1] Andrew Arana. Methodological purity as mathematical ideal. Preprint, November 2001.

[2] Jeremy Avigad. Saturated models of universal theories. To appear in the *Annals of Pure and Applied Logic*.

[3] Jeremy Avigad. Weak theories of nonstandard arithmetic and analysis. To appear in Stephen Simpson, editor, *Reverse Mathematics 2001*.

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

[5] Jeremy Avigad. Eliminating definitions and Skolem functions in first-order logic. In *Proceedings of the 16th annual IEEE symposium on logic in computer science*, pages 139–146, 2001. Final version to appear in *ACM Transactions on Computer Logic*.

[6] Jeremy Avigad and Solomon Feferman. Gödel's functional (Dialectica) interpretation. In Buss [15], pages 337–405.

[7] Jeremy Avigad and Jeffrey Helzner. Transfer principles for intuitionistic nonstandard arithmetic. To appear in the *Archive for Mathematical Logic*.

[8] Jeremy Avigad and Erich H. Reck. "Clarifying the nature of the infinite": the development of metamathematics and proof theory. Technical Report CMU-PHIL-120, Carnegie Mellon University, 2001.

[9] Michael Beeson. *Foundations of Constructive Mathematics*. Springer, Berlin, 1985.

[10] Lev Beklemishev. A proof-theoretic analysis of collection. *Archive for Mathematical Logic*, 37:275–296, 1998.

[11] Lev Beklemishev. On the induction schema for decidable predicates. Technical Report Logic Group Preprint Series 208, Department of Philosophy, Utrecht University, 2000.

[12] Errett Bishop and Douglas Bridges. *Constructive Mathematics*. Springer, Berlin, 1985.

[13] Douglas Bridges and Steve Reeves. Constructive mathematics in theory and programming practice. *Philosophia Mathematica*, 7:65–104, 1999.

[14] Samuel Buss. First-order proof theory of arithmetic. In Buss [15].

[15] Samuel Buss, editor. *The Handbook of Proof Theory*. North-Holland, Amsterdam, 1998.

[16] Patrick Cegielski. Le théorème de Dirichlet est finitiste. Technical Report 92.40, L.I.T.P., Paris VII, 1992.

[17] Rolando Chuaqui and Patrick Suppes. Free-variable axiomatic foundations of infinitesimal analysis: a fragment with finitary consistency proof. *Journal of Symbolic Logic*, 60:122–159, 1995.

[18] Paola D'Aquino and Angus Macintyre. Non-standard finite fields over $I\Delta_0 + \Omega_1$. *Israel Journal of Mathematics*, 117:311–333, 2000.

[19] Richard Dedekind. *Stetigkeit und irrationale Zahlen*. Vieweg, 1872. Translated by Wooster Beman as "Continuity and irrational numbers" in *Essays on the theory of numbers*, Open Court, Chicago, 1901; reprinted by Dover, New York, 1963. The Beman translation is reprinted, with corrections by William Ewald, in Ewald, W., editor, *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, Clarendon Press, Oxford, 1996, volume 2, pages 765-779.

[20] Solomon Feferman. Why a little bit goes a long way: logical foundations of scientifically applicable mathematics. In *PSA 1992*, volume 2, pages 422–455. Philosophy of Science Assocation, East Lansing, 1993. Reprinted with minor corrections in [22], pages 284–298.

[21] Solomon Feferman. Kreisel's "unwinding" program. In Odifreddi [53], pages 247–273.

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

[23] Solomon Feferman, Harvey Friedman, Penelope Maddy, and John Steel. Does mathematics need new axioms? *Bulletin of Symbolic Logic*, 6:404–446, December 2000.

[24] António Fernandes and Fernando Ferreira. Basic applications of weak König's lemma in feasible analysis. Preprint.

[25] Fernando Ferreira. A feasible theory for analysis. *Journal of Symbolic Logic*, 59:1001–1011, 1994.

[26] Harvey Friedman. Internal finite tree embeddings. To appear in [69].

[27] E. Grosholz and H. Breger, editors. *The growth of mathematical knowledge*. Kluwer Academic Publishers, The Netherlands, 2000.

[28] Petr Hájek. Interpretability and fragments of arithmetic. In Peter Clote and Jan Krajíček, editors, *Arithmetic, Proof Theory, and Computational Complexity*, pages 185–196. Oxford University Press, Oxford, 1993.

[29] Petr Hájek and Pavel Pudlák. *Metamathematics of first-order arithmetic*. Springer, Berlin, 1993.

[30] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers*. Oxford, fifth edition, 1979.

[31] John Harrison. Formalized mathematics. Technical Report 36, TUCS, 1996. http://www.cl.cam.ac.uk/users/jrh/papers/form-math3.html.

[32] David Hilbert and Paul Bernays. *Grundlagen der Mathematik*. Springer, Berlin, first volume, 1934, second volume, 1939.

[33] Kenneth Ireland and Michael Rosen. *A Classical Introduction to Modern Number Theory*. Springer, New York, second edition, 1990.

[34] Carl Jockusch and Robert Soare. Π^0_1 classes and degrees of theories. *Transactions of the American Mathematics Society*, 173:33–56, 1972.

[35] Richard Kaye. *Models of Peano Arithmetic*. Clarendon, Oxford, 1991.

[36] Stephen Kleene. *Introduction to Metamathematics*. North-Holland, Amsterdam, 1952.

[37] Ulrich Kohlenbach. Proof theory and computational analysis. In *Third Workshop on Computation and Approximation (Comprox III) (Birmingham, 1997)*, page 34 pp. (electronic). Elsevier, Amsterdam, 1998.

[38] Ulrich Kohlenbach. On uniform weak König's lemma. *Annals of Pure and Applied Logic*, 114:103–116, 2002.

[39] Ulrich Kohlenbach. Effective bounds from ineffective proofs in analysis: an application of functional interpretation and majorization. *Journal of Symbolic Logic*, 57:1239–1273, 1992.

[40] Ulrich Kohlenbach. New effective moduli of uniqueness and uniform a-priori estimates for constants of strong unicity by logical analysis of known proofs in best approximation theory. *Numerical Functional Analysis and Optimization*, 14:581–606, 1993.

[41] Ulrich Kohlenbach. Analyzing proofs in analysis. In W. Hodges et al., editors, *Logic: From Foundations to Applications: European Logic Colloquium '93*, pages 225–260. Clarendon Press, Oxford, 1996.

[42] Ulrich Kohlenbach. Mathematically strong subsystems of analysis with low rate of growth of provably recursive functionals. *Archive for Mathematical Logic*, 36:31–71, 1996.

[43] Ulrich Kohlenbach. A quantitative version of a theorem due to Borwein-Reich-Shafrir. *Numerical Functional Analysis and Optimization*, 12:641–656, 2001.

[44] Dénes König. Über eine Schlussweise aus den Endlichen ins Unendliche: Punktmengen. Kartenfärben. Verwandschaftsbezeihungen. Schachspiel. *Acta Litterarum ac Sceintarum (Ser. Sci. Math) Szeged*, 3:121–130, 1927.

[45] Jan Krajíček. *Bounded Arithmetic, Propositional Logic, and Complexity Theory*. Cambridge University, 1995.

[46] Georg Kreisel. On the interpretation of non-finitist proofs, part I. *Journal of Symbolic Logic*, 16:241–267, 1951.

[47] Georg Kreisel. Neglected possibilities of processing assertions and proofs mechanically. In Patrick Suppes, editor, *University-level computer-assisted instruction at Stanford 1968–1980*, pages 131–147. Stanford University, 1981.

[48] Georg Kreisel. Logical aspects of computation. pages 205–278. Academic Press, London, 1990.

[49] Horst Luckhardt. Herbrand-analysen zweier Beweise des Satzes von Roth: Polynomial Anzahlschranken. *Journal of Symbolic Logic*, 54:234–263, 1989.

[50] Horst Luckhardt. Bounds extracted by Kreisel from ineffective proofs. In Odifreddi [53], pages 289–300.

[51] Paolo Mancosu. On mathematical explanation. In Grosholz and Breger [27], pages 103–119.

[52] Melvyn Nathanson. *Elementary Methods in Number Theory*. Springer, New York, 2000.

[53] Piergiorgio Odifreddi, editor. *Kreiseliana: About and Around Georg Kreisel*. A K Peters, Wellesley, Massachusetts, 1996.

[54] Rohit Parikh. Existence and feasibility in arithmetic. *Journal of Symbolic Logic*, 36:494–508, 1971.

[55] Jeff Paris and Leo Harrington. A mathematical incompleteness in Peano arithmetic. In Jon Barwise, editor, *The Handbook of Mathematical Logic*, pages 1133–1142. North-Holland, Amsterdam, 1977.

[56] Jeff Paris, Alex Wilkie, and Alan Woods. Provability of the pigeonhole principle and the existence of infinitely many primes. *Journal of Symbolic Logic*, 53:1235–1244, 1988.

[57] Pavel Pudlák. The lengths of proofs. In Buss [15], pages 547–637.

[58] W. V. O. Quine. Reply to Charles Parsons. In H. Hahn and P. A. Schilpp, editors, *The Philosophy of W. V. O. Quine*, volume I, pages 396–403. Open Court, La Salle, 1986.

[59] J.-P. Ressayre. Weak arithmetics. *Theoretical Computer Science*, 257:1–15, 2001.

[60] Denis Richard. What are weak arithmetics? *Theoretical Computer Science*, 257:17–29, 2001.

[61] H. E. Rose. *Subrecursion: Functions and Hierarchies*. Clarendon Press, Oxford, 1984.

[62] Vladimir Sazonov. On bounded set theory. In *Logic, Methodology, and Philosophy of Science X*, volume I, pages 85–103. Kluwer Academic, 1997.

[63] Dana Scott. Algebras of sets binumerable in complete extensions of arithmetic. In J. C. E. Dekker, editor, *Recursive Function Theory*, volume V of *Proceedings in Symposia in Pure Mathematics*, pages 117–122. American Mathematical Society, 1962.

[64] Atle Selberg. An elementary proof of Dirichlet's theorem about primes in an arithmetic progression. *Annals of Mathematics*, 50:297–304, 1949.

[65] Jean Pierre Serre. *A Course in Arithmetic*. Springer, New York, 1973. Translation of *Cours d'arithmetique*.

[66] Wilfried Sieg. Fragments of arithmetic. *Annals of Pure and Applied Logic*, 28:33–72, 1985.

[67] Wilfried Sieg. Reductions for theories of analysis. In G. Davis and P. Weingartner, editors, *Foundations of Logic and Linguistics*, pages 199–231, New York and London, 1985. Plenum Press.

[68] Wilfried Sieg. Herbrand analyses. *Archive for Mathematical Logic*, 30:409–441, 1991.

[69] Wilfried Sieg, Richard Sommer, and Carolyn Talcott, editors. *Reflections on the Foundations of Mathematics: Essays in honor of Solomon Feferman*, volume 15 of *Lecture Notes in Logic*. Association for Symbolic Logic, 2001.

[70] Stephen Simpson, editor. *Reverse Mathematics '01*. To appear, 2001.

[71] Stephen Simpson. *Subsystems of Second-Order Arithmetic*. Springer, Berlin, 1998.

[72] Stephen Simpson and Rick Smith. Factorization of polynomials and Σ^0_1 induction. *Annals of Pure and Applied Logic*, 31:289–306, 1986.

[73] Richard Sommer and Patrick Suppes. Finite models of elementary recursive nonstandard analysis. *Notas De la Sociedad de Matematica de Chile*, 15, 1996.

[74] Howard Stein. Logos, logic, and logistiké. In William Aspray and Phillip Kitcher, editors, *History and Philosophy of Modern Mathematics*, pages 238–259. University of Minnesota, 1988.

[75] Olivier Sudac. The prime number theorem is PRA-provable. *Theoretical Computer Science*, 257:185–239, 2001.

[76] William Tait. Finitism. *Journal of Philosophy*, 78:524–546, 1981.

[77] Gaisi Takeuti. *Two Applications of Logic to Mathematics*, volume 13 of *Publications of the Mathematical Society of Japan*. Iwanami Shoten and Princeton University Press, 1978.

[78] Kazuyuki Tanaka. Non-standard analysis in WKL_0. *Mathematical Logic Quarterly*, 43:401–412, 1997.

[79] A. S. Troelstra. Realizability. In Buss [15].

[80] A. S. Troelstra and Dirk van Dalen. *Constructivism in Mathematics: An Introduction*, volumes 1 and 2. North-Holland, Amsterdam, 1988.

[81] Hermann Weyl. *Das Kontinuum. Kritische Untersuchungen über die Grundlagen der Analysis.* Veit, Leipzig, 1918. Second edition (1932). Translated as *The Continuum. A Critical Examination of the Foundation of Analysis*, Dover, New York, 1994.

[82] Andrew Wiles. Modular elliptic curves and Fermat's last theorem. *Annals of Mathematics*, 142:443–551, 1995.

[83] The Coq proof assistant. Developed by the LogiCal project. http://pauillac.inria.fr/coq/coq-eng.html.

[84] Foundations of Mathematics online discussion forum. http://www.math.psu.edu/simpson/fom.

[85] The HOL system. Developed by the University of Cambridge Computer Laboratory. http://www.cl.cam.ac.uk/Research/HVG/HOL/.

[86] The Isabelle theorem proving environment. Developed by Larry Paulson at Cambridge University and Tobias Nipkow at TU Munich. http://www.cl.cam.ac.uk/Research/HVG/Isabelle/index.html.

[87] Journées sur les Arithmétiques Faibles (JAF). http://www.univ-paris12.fr/lacl/jaf/.