Did Gentzen Prove the Consistency of Arithmetic?

Daniel Waxman

Draft version: please don't cite

Abstract

In 1936, Gerhard Gentzen famously gave a proof of the consistency of Peano arithmetic. There is no disputing that Gentzen provided us with a mathematically valid argument. This paper addresses the distinct question of whether Gentzen's result is properly viewed as a proof in the epistemic sense: an argument that can be used to obtain or enhance justification in its conclusion. Although Gentzen himself believed that he had provided a "real vindication" of Peano arithmetic, many subsequent mathematicians and philosophers have disagreed, on the basis that the proof is epistemically circular or otherwise inert. After gently sketching the outlines of Gentzen's proof, I investigate whether there is any epistemically stable foundational framework on which the proof is informative. In light of this discussion, I argue that the truth lies somewhere in between the claims of Gentzen and his critics: although the proof is indeed epistemically non-trivial, it falls short of constituting a real vindication of the consistency of Peano arithmetic.

1. Motivation and Preliminaries

The aim of this paper is to examine a famous and very special case of a mathematical consistency proof: Gentzen's proof of the consistency of Peano Arithmetic (PA), the most widely accepted formal axiomatization of arithmetic.[1] There is no doubt that Gentzen provided a proof in the mathematical sense of the term: a rigorous, formalizable deduction, in which the conclusion is shown to be derivable by recognizably valid steps from a set of axioms (usually, albeit informally, taken to be set theory as codified by Zermelo Fraenkel Set Theory with the Axiom of Choice (ZFC) or some other generally accepted foundational theory whose status is, for the purposes of the derivation, not in question). The question this paper seeks to answer, by contrast, is whether Gentzen's result is a proof in a distinct *epistemic* sense: something like a procedure or demonstration capable of cogently providing epistemic support for its conclusion. (This will be made more precise as we proceed). An easy way to see that the notions

[1] See Gentzen [1969], Chapters 4 and 8.

come apart is that for anyone working within a given set of axioms, the trivial deduction of one of the axioms themselves will always be possible; yet nobody, I take it, would argue that a deduction of this sort yields any epistemic gain. The obvious fact that a proposition can be derived in any theory which contains it as an axiom provides it with no epistemic support whatsoever. Naturally Gentzen's proof is very far from being a trivial one-line proof; nevertheless, one of the questions that will occupy us is whether it can be shown to exhibit some analogous kind of epistemic defect.

Before discussing the proof, some brief motivation is in order. The status of Gentzen's proof relates to the more general question of whether and how we are justified in believing in the consistency of our best mathematical theories – a question which I believe is central to the epistemology of mathematics. Many of us are are strongly inclined to believe that our best theories – paradigm cases being theories like PA and ZFC – are in fact consistent; but on reflection, it is not at all obvious that there are any rational grounds for possessing this conviction, let alone for possessing it as strongly as it is held. We know, from Gödel's second incompleteness theorem, that if T is a consistent theory containing a minimal theory of arithmetic, then T fails to prove its own canonical consistency statement ConT.[2] So if we do indeed possess justification in the consistency of a given theory, there are basically only two possibilities as to its source: either from proving the theory's consistency within some *other* theory or from a method distinct from mathematical proof altogether. Whichever of these disjuncts obtains, the resulting account promises to illuminate the architecture of mathematical justification.

Another reason for interest in the justification of consistency is that it (and closely related notions like coherence or conservativeness) play a key role in many live views in the philosophy of mathematics. A nice illustration is provided by structuralism, whose motivating slogan is that mathematics is the study of abstract structures. This slogan can be interpreted in many ways, but the most prominent involve either reifying structures ("*ante rem* structuralism") or interpreting away talk of structures by speaking of all possible systems of a certain kind ("*in re* structuralism"). On both kinds of view, consistency or closely related notions play a fundamental role. A pressing question facing the *ante rem* structuralist concerns the ontology of structures: which structures exist? The most influential answer is Stewart Shapiro's *Coherence Principle*: every coherent theory characterizes a structure (or class of structures). And although "coherence" here

[2]See for instance Smith [2012]. I assume throughout that we are speaking of recursively axiomatized theories. The "canonical" consistency statement is one constructed using a provability predicate that perspicuously represents the definition of the informal notion of provability and satisfies the Bernays-Hilbert-Lob derivability conditions. I will assume without further comment that the canonical consistency statement for T expresses the intuitive property of consistency, and that we are justified in believing this; thus I slide between the informal claim that some theory is consistent, and the mathematical claim ConT.

is not exactly the same as consistency, it is a kind of second-order analogue of it – indeed, for first-order theories, coherence and consistency are co-extensional.[3] Similarly, the *in re* structuralist faces a pressing question: how are mathematical statements to be interpreted? The most influential answer is due to Geoffrey Hellman. A statement of arithmetic ϕ is roughly to be interpreted as involving two components: (i) a claim that in any possible ω-sequence – that is, about any possible system of objects satisfying the Peano axioms – the analogue of ϕ holds; and (ii) that some ω-sequence is possible.[4] The relevant notion of possibility is not, according to Hellman, metaphysical or physical; rather it is a "mathematico-logical" notion. Hellman's notion of mathematico-logical possibility is, like Shapiro's coherence, a kind of second-order analogue of consistency; and again, when first-order theories are in question, it is co-extensional with consistency. So structuralism assigns a central role to consistency, or something essentially similar to it, as far as first-order mathematics is concerned.

Structuralism is not alone here; there is a strong case to be made (although I will not attempt to make it here) that consistency or similar is central to many other views too: to neo-Fregeanism, to certain ("plenitudinous") brands of platonism, to formalism, and perhaps even to fictionalism.[5] And if consistency is central to these views, then it is reasonable to suppose that its epistemology is central to the epistemology of mathematics more generally.

So much for motivation. The plan of the paper is as follows. Section 2 provides a brief overview of Gentzen's proof of the consistency of PA. Here my aim is not to provide the full technical details or a historically focused exposition of the proof.[6] Rather, I'll give an overview of the structure of the proof (emphasizing the sense in which it is a kind of cut-elimination theorem) and discuss the resources needed for its formalization. The remaining sections address our main question: the epistemic status of the proof. Section 3 considers the objection, evident in some of the early reactions to Gentzen by his critics, that the proof exhibits a problematic kind of epistemic circularity. I attempt to make this complaint more precise by reference to the recent epistemological literature on transmission failure. After reconstructing the objection to the best of my ability, I argue that it fails: Gentzen's proof is not straightforwardly circular, for reasons related to the formulation of mathematical induction required for the proof to go through. But I do not think that the matter ends there. In Section 4 I take up the

[3]Shapiro [1997, 133]. Shapiro's primary reason for not identifying coherence and consistency is that consistency is at least partly a mathematical notion, which would lead to problematic circularity for his purposes.

[4]Hellman [1989, Ch1].

[5]See for instance Hale and Wright [2001], Balaguer [1998], Weir [2010], and Field [1989].

[6]More detailed accounts of the proof can be found in standard texts on proof theory; see for instance Takeuti [1987]. The original proofs can be found in Gentzen [1969]. For further information about the historical and mathematical context of the proof, see the essays in Kahle and Rathjen [2015].

distinct question of whether Gentzen's result represents, as he himself believed, a "real vindication" of the consistency of Peano Arithmetic. I approach that issue by considering a number of 'foundational equivalence theses' seeking to equate various formal mathematical systems with intuitive conceptions of mathematical subject-matters. In light of this discussion, I conclude that the truth lies somewhere in between the claims of Gentzen and his critics: although the proof is epistemically non-trivial, it falls short of constituting a real vindication of the consistency of arithmetic.

2. Gentzen's Consistency Proof

2.1 Background

The story of the foundations of mathematics in the early 20th century has been frequently told, so I will be brief in setting the scene. In the late 1920s and early 1930s, metamathematics was preoccupied by Hilbert's Program, which (motivated in part by the discovery of Russell's paradox for naive set theory and the desire to demonstrate that mathematics could be adequately formalized) sought to place mathematics on a secure footing.[7] Hilbert's Program had two main aspects: (i) a conviction in the absolute epistemic security of *finitary* mathematics (a notion left unformalized by Hilbert and his immediate followers, but which can be crudely characterized as reasoning which presupposes no completed infinite totalities); and (ii) a view that *infinitary* mathematics could be shown secure by providing consistency proofs using only finitary resources. However, the program was fatally wounded by Gödel, who in effect showed that part (ii) is impossible. For if infinitary mathematics is a proper extension of finitary mathematics, and if finitary mathematics contains minimal arithmetical resources, then there can be no finitary proof of the consistency of infinitary mathematics on pain of the inconsistency of finitary mathematics.

Working in light of this background, Gentzen undertook the project of proving the consistency of arithmetic in a way that was not ruled out by Gödel's results. This project came to fruition in 1936, when he published "The Consistency of Elementary Number Theory", the paper which arguably founded contemporary proof theory. Although the main elements of the result were essentially already present in 1936, they were refined and clarified by Gentzen in a number of subsequent papers.

In this section I will give a brief and schematic exposition of Gentzen's proof. In fact I'll sketch it in two different ways: first a (somewhat anachronistic) version of the proof due to Schütte, and then one closer to Gentzen's own.[8] The first version is

[7] For nice discussions of Hilbert's Program, see Giaquinto [2002] and Zach [2007].
[8] Schütte's proof can be found in Schütte [1977].

included for mainly expositional reasons: it shows the sense in which the proof can be seen as a cut-elimination result. The second version is valuable since it allows us to see precisely what resources are required to formalize the proof. As we will see, the details of the formalization matter a great deal when it comes to assessing the proof's epistemic worth.

2.2 Consistency as Cut-Elimination

In earlier work, Gentzen introduced the now-familiar deductive system of natural deduction. In the course of studying it, he formulated the sequent calculus, whose basic notion is a *sequent*: a string of the form $A_1, ..., A_n \Rightarrow B_1, ..., B_n$, where the double-arrow "$\Rightarrow$" is intended to formalize the notion of natural deduction derivability. Roughly speaking a sequent expresses that the conjunction of the formulas in the antecedent entails the disjunction of the formulas in the succedent. The sequent calculus has both axioms and rules of inference. The axioms are "basic" sequents of the form $A \Rightarrow A$. The rules of inference indicate ways in which one sequent can be validly obtained from another, and are typically categorized in two ways: (i) rules governing the behaviour of the connectives and (ii) structural rules. As an example of rules for connectives, here are the (left and right) rules for negation:

$$[\neg\text{L}] \quad \frac{\Gamma \Rightarrow A, \Delta}{\Gamma, \neg A \Rightarrow \Delta}$$

$$[\neg\text{R}] \quad \frac{\Gamma, A \Rightarrow \Delta}{\Gamma \Rightarrow \neg A, \Delta}$$

and as an example of a structural rule, here is weakening:

$$[\text{Weakening L}] \quad \frac{\Gamma \Rightarrow \Delta}{\Gamma, A \Rightarrow \Delta}$$

$$[\text{Weakening R}] \quad \frac{\Gamma \Rightarrow \Delta}{\Gamma \Rightarrow A, \Delta}$$

Of the structural rules, the most interesting is the cut rule. A here is the *cut-formula*, so-called because it is "cut" from the derivation:

$$[\text{Cut}] \quad \frac{\Gamma \Rightarrow A, \Delta \quad A, \Sigma \Rightarrow \Pi}{\Gamma, \Sigma \Rightarrow \Delta, \Pi}$$

Gentzen's study of the sequent calculus culminated in his *Hauptsatz* or Cut Elimination Theorem: any sequent that admits of a derivation admits of a *cut-free* derivation, i.e. one that does not use the cut rule. One of the interesting corollaries of this theorem – the crucial fact that connects it with the question of consistency – is that sequent cal-

culus deductions have the *subformula property*: whenever a sequent $\Gamma \Rightarrow \Delta$ is derivable, it has a derivation all of whose formulae are subformulae of the formulae of Γ and Δ. The consistency of the deductive system follows almost immediately: note that there is a derivation of an inconsistency, i.e. $\emptyset \Rightarrow A \wedge \neg A$, if and only if there is a derivation of the empty sequent $\emptyset \Rightarrow \emptyset$.[9] So if the system is inconsistent, there must be a derivation of the empty sequent featuring only its subformulas. But there are no such subformulas, and it is easy to demonstrate that the empty sequent could be derived only by an application of cut.[10] So the sequent calculus must be consistent.

So far all that has been shown is the consistency of the background logic: recall that the axioms of the system under consideration are simply logical axioms of the form $A \Rightarrow A$. One way to think about the Gentzen consistency proof for arithmetic is as a cut-elimination result for a system expanded to include not just logical but also mathematical axioms. The mathematical theory in question is Peano Arithmetic (PA): the familiar theory over the language $\mathcal{L}_{\mathsf{PA}} = (0, S, +, \cdot)$ with the usual axioms for successor, addition, and multiplication:

PA1 $(Sx = Sy \rightarrow x = y)$

PA2 $\neg(Sx = 0)$

PA3 $x + 0 = x$

PA4 $x + Sy = S(x + y)$

PA5 $x \times 0 = 0$

PA6 $x \times Sy = (x \times y) + x$

PA7 $[\phi(0) \wedge \forall x(\phi(x) \rightarrow \phi(Sx))] \rightarrow \forall x \phi(x)$ where $\phi(x)$ is a formula in the language of arithmetic.

The connection with cut-elimination results is brought out most clearly in a version of the proof given by Schütte.[11] The consistency of PA is approached by considering an alternative theory PA_ω which differs in two key respects. First, the axioms of PA_ω are:

$\mathsf{PA}_\omega 1$ $\emptyset \Rightarrow A$ where A is a true atomic sentence

$\mathsf{PA}_\omega 2$ $B \Rightarrow \emptyset$ where B is a false atomic sentence

[9] Proof: from $\emptyset \Rightarrow A \wedge \neg A$ we can derive $A \Rightarrow \emptyset$ and $\emptyset \Rightarrow \neg A$ and and hence, by cut, the empty sequent; conversely, any sequent can be obtained from the empty sequent by weakening.

[10] Each of the other rules either preserves or increases the number of formulas in the sequent.

[11] My exposition here is indebted to Rathjen [1999], which constains an extremely approachable introduction to ordinal analysis.

$\mathsf{PA}_\omega 3$ $\quad F(s_1, s_2, ..., s_n) \Rightarrow F(t_1, t_2, ..., t_n)$ whenever s_i and t_i are terms evaluating to the same numeral.

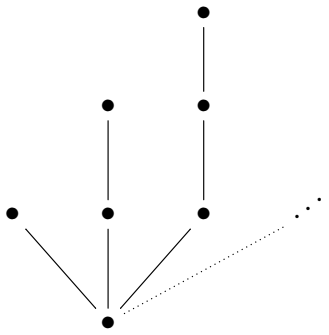
Second, the rules of inference of PA_ω involve replacing the usual quantifier rules with left and right versions of the ω-rule:

$$\omega R \quad \frac{\Gamma \Rightarrow \Delta, F(0) \quad \Gamma \Rightarrow \Delta, F(1) \quad ... \quad \Gamma \Rightarrow \Delta, F(n) \quad ...}{\Gamma \Rightarrow \Delta, \forall x F(x)}$$

$$\omega L \quad \frac{F(0), \Gamma \Rightarrow \Delta \quad F(1), \Gamma \Rightarrow \Delta \quad ... \quad F(n), \Gamma \Rightarrow \Delta \quad ...}{\exists x F(x), \Gamma \Rightarrow \Delta}$$

Notice that the ω-rule is infinitary, since there are infinitely many formulas above the inference line. Thus, derivations within PA_ω can be thought of as infinite well-founded trees, with their conclusion as a single root.

A *branch* is a maximal chain in the tree (/derivation); its *length* is the ordinal to which it is isomorphic. The *height* of a tree is the supremum of the length of the branches within it. Derivations give rise to well-founded trees, i.e. whose branches are each of finite length. But since we are considering trees with infinitely many branches (due to the addition of the ω-rule), it is possible for the height of a tree to be infinite.



A tree of height ω

The *cut-rank* of a derivation is the length of the longest cut-formula within it. Introducing some notation, say that $\mathsf{PA}_\omega \left|\frac{\alpha}{k}\right. \Gamma \Rightarrow \Delta$ if there is a derivation within PA_ω of the sequent $\Gamma \Rightarrow \Delta$ whose cut-rank is k and whose height is α. The first important fact to note is that derivations in PA correspond to derivations in PA_ω in the following sense:

Fact 1. *If* $\mathsf{PA} \vdash \Gamma \Rightarrow \Delta$, *then* $\mathsf{PA}_\omega \left|\frac{\alpha}{k}\right. \Gamma \Rightarrow \Delta$ *for some ordinal* $\alpha < \omega + \omega$ *and finite* k.

The second fact is that the cut-rank of a derivation can be reduced, at a 'cost' of increasing the height of the derivation by exponentiation by ω:

Fact 2. *If* $\mathsf{PA}_\omega \left|\frac{\alpha}{k+1}\right. \Gamma \Rightarrow \Delta$, *then* $\mathsf{PA}_\omega \left|\frac{\omega^\alpha}{k}\right. \Gamma \Rightarrow \Delta$.

Thus, if $\mathsf{PA}_\omega \vert\!\frac{\alpha}{k}\ \Gamma \Rightarrow \Delta$, then by appealing to Fact 2 k-many-times, we have that there is a cut-free derivation of the same sequent with height $\omega^{\omega^{\cdots^{\alpha}}}$, i.e. α exponentiated by ω k-many times. This is the reason for the relevance of the ordinal ϵ_0, defined as

$$\epsilon_0 = lim(\omega, \omega^\omega, \omega^{\omega^\omega}, ...) = \text{the least } \alpha \text{ such that } \omega^\alpha = \alpha$$

For combining Facts 1 and 2, we have that

$$\text{if } \mathsf{PA} \vdash \emptyset \Rightarrow \emptyset, \text{ then } \mathsf{PA}_\omega \vert\!\frac{\beta}{0}\ \emptyset \Rightarrow \emptyset, \text{ for some } \beta < \epsilon_0$$

In other words: if PA is inconsistent, then the empty sequent $\emptyset \Rightarrow \emptyset$ can be derived in PA_ω and, what is more, this derivation will be cut-free. But just as in the case for the background logic, it is easy to see that PA_ω does not allow for a cut-free derivation of the empty sequent.

2.3 Formalizing the Proof

I hope the foregoing gives a heuristic sense of how the proof runs. Now, the sketch just provided was given in informal terms, without any explicit concern for the background theory in which we were operating and whose principles we were presupposing. But in order to make the proof rigorous and to evaluate its epistemic credentials, we need to know in more detail which resources are required for its formalization. Here it is helpful to return to a version of the proof closer to Gentzen's own original presentation. It proceeds via the following steps:

(I) Each derivation D of PA is assigned an ordinal $Ord(D) < \epsilon_0$.

(II) A reduction procedure $\mathcal{R}$ is introduced such that whenever $D_\perp$ is a derivation of a contradiction, then so too is $\mathcal{R}(D_\perp)$. Furthermore, for any derivation $D_\perp$ of a contradiction, $Ord(\mathcal{R}(D_\perp)) < Ord(D_\perp)$. Note the strict inequality here.

Given (I) and (II), suppose that PA is inconsistent. Then there exists a derivation $D_\perp$ of a contradiction. The reduction procedure $\mathcal{R}$ can be applied to $D_\perp$ to yield a derivation $\mathcal{R}(D_\perp)$ whose ordinal is strictly less than that of $D_\perp$. Furthermore, this procedure can be iterated, yielding an infinitely descending chain of ordinals:

$$\epsilon_0 > \alpha_0 = Ord(D_\perp) > \alpha_1 = Ord(\mathcal{R}(D_\perp)) > \alpha_2 = Ord(\mathcal{R}(\mathcal{R}(D_\perp))) > \cdots$$

Contrapositively, from the assumption that no such infinite descending chain of ordinals below ϵ_0 exists – the principle also known as transfinite induction up to ϵ_0 –

the consistency of PA follows.

Turn now to the question of how the foregoing reasoning might be formulated in a suitable metatheory. There are a number of requirements:

1. A means of representing derivations in the sequent calculus in order to even talk about them in the first place.

2. A means of representing the ordinals (below ϵ_0) and the standard ordering relation $<$ on them.

3. The definability of functions corresponding to the ordinal assignment function *Ord* (which assigns ordinals to derivations) and the reduction procedure $\mathcal{R}$.

4. The provability of the fact that, if $D_\perp$ is a derivation of the empty sequent, then $Ord(\mathcal{R}(D_\perp)) < Ord(D_\perp)$.

5. Transfinite induction[12] up to ϵ_0: the principle that

$$\forall x, y < \epsilon_0[(x < y \to Px) \to Py] \to (\forall y < \epsilon_0)Py$$

It turns out that conditions (1-4) are all satisfied in Primitive Recursive Arithmetic (PRA), a relatively weak theory that captures reasoning about primitive recursive functions.[13] Since we will in due course take an interest in the epistemic status of this theory, it is worth setting it out in full.

PRA is a quantifier-free theory with symbols for 0 and the successor relation, together with a function-symbol for each primitive recursive function.[14] Its axioms are:

PRA1 $\quad Sx \neq 0$

PRA2 $\quad Sx = Sy \to x = y$

PRA3 for each primitive recursive function f, the recursion equations for f

PRA4 the induction rule (schema): from $\phi(0)$ and $\phi(x) \to \phi(Sx)$ infer $\phi(x)$

[12]The reason for calling this a form of induction is clear by comparing with an ordinal-theoretic formulation of standard arithmetical induction: $\forall x, y < \omega[(x < y \to Px) \to Py] \to (\forall y < \omega)Py$

[13]See for instance Troelstra and Schwichtenberg [2000].

[14]The class of primitive recursive functions is defined as follows. The basic primitive recursive functions are: (i) the constant function 0; (ii) the successor function S; (iii) the projection functions P^i_n, which, when applied to n arguments, returns the i^{th}. In addition, the class is closed under two operations. First, composition: if f and g are primitve recursive, then the composite function $f \circ g$ is primitive recursive. Second, primitive recursion: if f and g are primitive recursive, and h is such that $h(0, \bar{x}) = f(\bar{x})$ and $h(Sy, \bar{x}) = g(\bar{x}, y, h(y, \bar{x}))$, then h is primitive recursive. If h can be written in this way, these two equations are called its recursion equations.

It is worth noting that since PRA lacks quantifiers, induction cannot be formulated with a universal quantifier; this is why it appears in rule form. Despite its lack of quantifiers, PRA is nevertheless able to simulate some simple universally quantified claims by means of formulas with free variables. [15]

Despite the relative weakness of PRA, it provides a natural setting for formalizing Gentzen's argument. For returning to our list of requirements above:

1. PRA names all natural numbers. Therefore it can represent formulas of the language of arithmetic via a Gödel-numbering. Furthermore, arithmetical sequents (and deductions) can also be represented by coding them up as various sequences of formulas.

2. By judicious use of coding, PRA is also able to represent the ordinals below ϵ_0. One way to see this involves Cantor's Normal Form theorem, which implies that every ordinal $\alpha < \epsilon_0$ can be written in the form $\alpha = \omega^{\alpha_1} \cdot k_1 + \omega^{\alpha_2} \cdot k_2 + ... + \omega^{\alpha_n} \cdot k_n$ where $\alpha > \alpha_1 > \alpha_2 > \cdots > \alpha_n$ and $k_i \in \mathbb{N}$. Since the ordinals α_i can themselves be written in Cantor Normal Form with exponents that are smaller still, each ordinal $\alpha < \epsilon_0$ can be represented uniqely by a term in the alphabet $\omega, \cdot, +, 0, 1, 2, ...$ – and this representation can itself be encoded in the language of arithmetic. In such a context, we operate not with ordinals directly but with their codes. Furthermore the relation $\prec$, which holds between codes of ordinals just when they themselves stand in the usual ordering relation $<$, is definable within PRA.

3. An analysis of the ordinal-assignment function *Ord* and the reduction procedure $\mathcal{R}$ shows that are both primitive recursive, and so are represented by functions in PRA.

4. PRA proves the fact that $Ord(\mathcal{R}(D_\perp)) \prec Ord(D_\perp)$ where $D_\perp$ is a derivation of the empty sequent.

5. Transfinite induction up to ϵ_0 is not provable within PRA. (If it were, then PRA would be able to prove the consistency of PA, and thus its own consistency, and hence would be inconsistent.) But it can be *expressed* in rule form: from $y \prec x \rightarrow (\phi(y) \rightarrow \phi(x))$ infer $\phi(x)$.

   A further fact about the required formulation of transfinite induction is worth emphasizing: because PRA is stated in a quantifier-free language with symbols for primitive recursive functions, transfinite induction restricted to formulas of

[15] In particular those that are Π_1, i.e. of the form $\forall x \phi(x)$ where ϕ contains only bounded quantifiers.

that language is all that is needed for the proof to go through. The fact that induction is restricted in this way will prove to be of significance when evaluating the proof.

Summing it all up: writing $\mathsf{TI}(\epsilon_0)$ for transfinite induction up to ϵ_0 restricted to quantifier-free formulas, we have:

$$\mathsf{PRA} + \mathsf{TI}^{QF}_{PR}(\epsilon_0) \vdash \mathrm{Con}(\mathsf{PA})$$

This concludes the overview of the proof. We turn now to its evaluation.

3. Circularity?

There is no question that Gentzen's argument is valid, in that it meets the usual logical and mathematical standards for an acceptable proof. Nor can it be doubted that it is of great mathematical significance, constituting one of the founding results of proof theory. It provides us with a means of validly deducing $\mathrm{Con}(\mathsf{PA})$, and thus – waiving any doubts about our formalization of consistency claims – the consistency of PA itself. But earlier we distinguished between a mathematical proof – of which Gentzen's argument is a paradigm case – and an epistemic proof – a procedure or demonstration capable of cogently providing epistemic support for its conclusion. Our question is whether Gentzen's argument is a proof in this epistemic sense. In other words: does it provide us with a means of obtaining justification in the consistency of PA?

Gentzen himself appeared to believe so. In the preamble to his paper he claims to have proven the consistency of arithmetic "by means of forms of inference that are *completely unimpeachable*".[16] Later, he writes that

> I am inclined to believe that in terms of the fundamental distinction between disputable and indisputable methods of proof, the proof of the finiteness of the reduction procedure [i.e. the part of the proof involving transfinite induction] can still be considered indisputable, so that the consistency proof represents a *real vindication* of the disputable parts of elementary number theory.[17]

However, many have disagreed. The most usual grounds for doubt involve the suspicion that, although the Gentzen proof is fully rigorous, it is nevertheless trivial or problematically circular in a way that makes it epistemically impotent. Unfortunately,

[16] Gentzen [1969], p. 135. Italics added.
[17] Gentzen [1969], p. 197.

in this domain, quips are more common than arguments. Hermann Weyl, for instance, reportedly remarked that

> Gentzen proved the consistency of arithmetic, i.e. induction up to the ordinal ω, by means of induction up to ϵ_0[18]

– the joke here being that ϵ_0 is an ordinal much larger than ω. Tarski is similarly reported to have commented that knowledge of the Gentzen proof increased his confidence in the consistency of arithmetic "by an epsilon" – the joke *here* being that "epsilon" is used both within the name of ϵ_0 and also in analysis to denote an arbitrarily small positive real number (as in the common first line of proofs "let $\epsilon > 0$").[19] Kleene prefers to reserve judgement, writing that:

> to what extent the Gentzen proof can be accepted as securing classical number theory in the sense of that problem formulation is in the present state of affairs a matter for individual judgement, depending on how ready one is to accept induction up to ϵ_0 as a finitary method.[20]

As I'll argue (in §5), Kleene is close to the truth here, but the situation is slightly more subtle than his remark suggests. In particular I'll argue that the real question is not whether transfinite induction up to is *finitary*, but whether it can be motivated (in a sense which I'll try to make more precise) in a way that's independent of the considerations motivating PA.

The aim of the rest of the paper is to get clear on the sorts of issues raised by these commentators on Gentzen's proof and to arrive at a stable evaluation of its epistemic status. The remainder of this section considers the complaint that the proof is objectionably circular; I will argue that it is not. The first task, though, is to make the objection more precise. To do that, we turn to the notion of *transmission of warrant* in the sense much discussed within recent epistemology.

3.1 Transmission, Transmission Failure, and Circularity

What is the point of engaging in logical deduction? A natural answer is that it is a way of improving one's epistemic position: if the premises have a positive epistemic status (I'll focus here on justification, though I think nothing essential would be lost if reformulated in terms of knowledge), then, via deduction, that positive status attaches

[18] Girard [2011], p. 9.

[19] Kahle [2015] cites Kreisel [1979] as reporting "familiar jokes (for example, by Tarski whose confidence [in the consistency of arithmetic] was increased by ϵ", or by Weyl who was astonished that one should use "ϵ_0 induction to prove the consistency of ordinary, that is ω-induction)." See also Kleene [1967], p. 257.

[20] Kleene [1952], p. 479.

to the conclusion. But recent work in epistemology has made a convincing case that not all deductions can work in this way. Here are three famous examples from the literature:

- Dretske's zebras. On the basis of a perceptual experience as of a zebra, we believe P – that the animal in front of me is a zebra – and go on to deduce Q – that the animal in front of me is not a cleverly disguised mule. [21]

- Bootstrapping. On the basis of the gas gauge on my car reading 'FULL', we believe P – that my car's gas tank is full – and go on to deduce Q – that the gas gauge is working reliably on this occasion.[22]

- Moore's 'proof' of an external world. On the basis of a perceptual experience as of a hand, we believe P – that there is a hand in front of me – and go on to deduce Q – that the external world exists.[23]

The intuitive judgement in at least the first two cases is that there is something *epistemically* problematic about deducing Q from P (at least when P is justified in the manner described).[24] But each deduction is valid (or can be made so with the addition of true and presumably justified enthymematic premises, e.g. that donkeys aren't mules). To get slightly clearer on the issue, distinguish *arguments* from *deductions*. An argument consists of premises and a conclusion, which is (or is supposed to be) logically supported by the premises. A deduction occurs when an argument is put to use: it is a particular, dateable mental event in which an agent infers the conclusion of an argument on the basis of its premises in a logically competent way. Now let us say that a deduction *fails to transmit justification* if the agent in question obtains no new or enhanced justification in the conclusion as a result of carrying it out. The moral suggested by the examples above is that, sometimes, perfectly valid deductions fail to transmit justification.[25]

Can we say anything substantive about when exactly a deduction fails to transmit justification? A natural first thought is that what is going wrong in the examples

[21] See Dretske [1970].

[22] See for instance Vogel [2000].

[23] Moore [1939].

[24] Moore's 'Proof' is far more controversial than the other two. It is defended by a number of contemporary epistemologists, most prominently Pryor [2004]. Many however are sceptical of its efficacy; see for instance Wright [2002].

[25] Transmission must be distinguished from closure. Roughly, closure principles claim that if the premises of a competent deduction are justified, then so is the conclusion; transmission principles claim that this justification arises *as a result* of the deduction. It should thus be clear that a denial that arguments always transmit does not require a rejection of general closure principles. As far as I know, the distinction between transmission and closure was first recognized in Wright [1986]. See also Zalabardo [2012] for an externalist account of transmission failure.

above is that they involve *problematically circular* inferences. Here is one way to flesh that worry out. Sometimes, when we are justified in believing a claim, that justification *depends* on other epistemically relevant facts. For example, take a simple case of perceptual justification: I am justified in believing that I am sitting at a desk in front of a laptop; that justification presumably depends, at least in part, on the visual experience I am currently undergoing, as of being seated at a desk in front of a laptop. That is not necessarily to make the psychological claim that my belief is inferentially based, i.e. that I *infer* it from some other beliefs or mental states. But whatever the correct psychological description, there is overwhelming pressure at the level of epistemic analysis to say that my *justification* that I'm sitting at a desk depends (at least in part) on the perceptual experiences I'm having.

So we have a notion of *epistemic dependence*: a relation that holds between, on the one hand, items of justification, and on the other, the sorts of things that can give rise to this justification. Let's leave it fairly loose what the relata of the second kind are: there's at least a *prima facie* case for thinking that justification might depend on, among other things, having certain experiences, receiving certain testimony, carrying out certain (inductive or deductive) reasoning, or being justified to believe certain other propositions.

All this suggests a hypothesis concerning transmission failure: that it arises when a deduction involves a *viciously circular pattern of epistemic dependence between its premises and its conclusion*. More precisely, in the form of a sufficient condition:

(Dependence Circularity) If S carries out a competent deduction from premises to conclusion C, and S's justification to believe one or more of the premises epistemically depends on justification to believe C, then the deduction fails to transmit justification to C.

Although the literature on transmission failure is large and disputatious, Dependence Circularity seems to command widespread assent.[26][27] In what follows I'll take it as a working account of transmission failure; as we will see, it can be used to illuminate a main source of worry about the epistemic status of Gentzen's proof.

[26]Principles along these lines can be found in Wright [2002], Pryor [2012], and Neta [2013]. Although the principle is widely accepted, its application to cases is more controversial. That said, it provides a way of getting clear on (at least one aspect) of the recent dispute between those who endorse and those who reject the transmissiveness of Moore's "proof". If Dependence Circularity is correct, then that question reduces to whether or not perceptual justification about nearby hands depends on antecedent justification to believe that there is an external world.

[27]Strictly speaking, Gentzen's proof is an argument, not a deduction, so the question of whether it transmits or fails to transmit justification is not well posed; it is plausible that whether or not transmission occurs depends on the agent's evidence for the premises and perhaps other background aspects of the epistemic context in which the deduction is carried out. But our question is whether the proof can be put to use by a typical (mathematically informed) subject in typical epistemic circumstances to obtain or enhance their justification in the consistency of Peano Arithmetic; consequently no confusion is liable to result from talking this way.

3.2 Transmission Failure and the Gentzen Proof

The most explicit articulation of a circularity objection of which I'm aware can be found in a discussion by Crispin Wright:

> To accept the Gentzen proof is to be persuaded of a mapping between the proofs constructible in elementary number theory and the series of ordinals up to ϵ_0. And to understand the structure of the ordinals up to ϵ_0 is to grasp a concept which embeds and *builds on* an ordinary understanding of the series of natural numbers. So to trust the Gentzen proof is implicitly to forgo any doubt about the coherence of the concept of natural number.[28]

In light of the discussion of epistemic dependence and transmission failure, I suggest that this argument can be rendered as follows:

CA1 Justification in the premises used in the Gentzen proof depends on an understanding of the ordinals up to ϵ_0;

CA2 An understanding of the ordinals up to ϵ_0 depends on an understanding of the ordinals up to ω, i.e. the natural numbers;

CA3 An understanding of the natural numbers depends on justification in the consistency of PA;

CA4 Therefore, the Gentzen proof fails to transmit justification to the consistency of PA.[29]

Call this the Circularity Argument. Granted the transitivity of epistemic dependence and Dependence Circularity, the Circularity Argument is valid. In the remainder of this section, however, I will argue that it is unsound. Before I give my main reasons, a brief concern regarding CA3 should be mentioned. This claim plays a crucial role in the argument; for in order to apply Dependence Circularity to the Gentzen proof, the justification of least one of the premises must depend on justification for the conclusion, i.e. the consistency of PA. But CA3 is undermotivated. In fact it's open to substantial challenge: for all that has been said so far, the direction of dependence might be reversed; perhaps our justification in the consistency of PA depends, conversely, on our understanding of the natural numbers.

Nevertheless, even if this objection is waived, there is a more fundamental reason why the argument fails. The problem is that talk of "an understanding of the ordinals

[28] Wright [1994], p. 177.

[29] Again, read 'the Gentzen proof' as 'any deduction following the structure of the Gentzen proof'.

up to..." contains a crucial equivocation. It is true that the metatheory in which the Gentzen proof is carried out, $\mathsf{PRA} + \mathsf{TI}^{QF}_{PR}(\epsilon_0)$, does, in a sense, require an understanding of the ordinals up to ϵ_0, since one of its axioms is transfinite induction up to ϵ_0, and presumably such an axiom could be justified only via an appropriate understanding of the relevant ordinals. And it is true that an understanding of the ordinals up to ϵ_0 does, in a sense, require (and build on) an understanding of the natural numbers; after all, the natural numbers are the ordinals up to ω, and ϵ_0 is indeed an ordinal which extends ω. But there is a major difference in how these two theories – $\mathsf{PRA} + \mathsf{TI}^{QF}_{PR}(\epsilon_0)$ on the one hand and PA on the other – handle induction; a difference which allows us to see that although CA1, CA2, and CA3 have true readings, there is no reading of capable of making them true simultaneously.

Recall that the role of transfinite induction up to ϵ_0 in the Gentzen proof is to rule out the existence of an infinitely descending chain $\epsilon_0 > \alpha_0 = Ord(D) > \alpha_1 = Ord(\mathcal{R}(D)) > \alpha_2 = Ord(\mathcal{R}(\mathcal{R}(D))) > \cdots$ of the kind that would be generated by a derivation of an inconsistency. But as was emphasized in §2.4, transfinite induction is required in only a very limited form. Our analysis of the proof showed that any derivation of an inconsistency within PA would be witnessed by an infinitely descending chain of ordinals expressible by *quantifier-free formulas in the language of* PRA. So, although the proof requires transfinite induction up to ϵ_0, it is really only a highly restricted version of transfinite induction that is needed.

To see the full implications of this point, it helps to think of the inductive commitments of arithmetical theories as varying along two different dimensions. The first dimension, which we might call the *width* of induction, is the range of conditions or properties over which induction can be carried out. Here is how things play out for some relevant theories:

> PA is a first order theory containing an induction schema whose instances include all formulas in the language of arithmetic (i.e. in terms of zero, successor, addition, and multiplication). Its inductive commitments can thus be characterized as ranging over all and only arithmetical conditions, i.e. those expressed by a formula which is Π^0_n or Σ^0_n for some n.[30]
>
> Second order arithmetic Z_2 is obtained by replacing PA's induction schema with a second order induction axiom, quantifying into predicate position. The inductive commitments of Z_2 thus outstrip those of PA, since it is capable of formulating additional conditions via second order quantification. In these terms Z_2 allows induction over all and only analytic conditions, i.e. those which can be defined

[30] See e.g. Rogers [1987].

by a formula which is Π^1_n or Σ^1_n for some n.[31]

Primitive recursive arithmetic PRA allows induction in rule form, over a quantifier-free language with symbols for primitive recursive functions. Using free variables to simulate universal quantification, PRA in effect allows induction over the Π^0_1 conditions involving primitive recursive functions, a proper subset of the arithmetical conditions over which induction can be carried out in PA. There is thus a clear sense in which the inductive commitments of PRA are less extensive or 'narrower' than those of PA.

$\mathsf{PRA} + \mathsf{TI}^{QF}_{PR}(\epsilon_0)$ allows induction over the same conditions as PRA and thus has the same commitments regarding the width of induction.

The second dimension is the *height* of induction: how far into the ordinals can induction be carried out? One way to measure this is by the proof-theoretic ordinal of a theory: roughly the smallest ordinal whose well-ordering cannot be proven by the theory. [3234]Returning to the theories we have been considering:

Gentzen's result implies that PA cannot prove the well-ordering of ϵ_0 (on pain of being able to prove its own consistency). As Gentzen himself also showed, this result is best-possible, since the well-ordering of every ordinal below ϵ_0 is provable within PA.

The proof-theoretic ordinal of Z_2 is not currently known, although it is known to be far larger than ϵ_0.[35]

[31]There are many philosophical questions concerning the ontological and mathematical status of second (and higher) order quantification, but these do not need to be addressed here. (See for instance Shapiro [1991]. My point is simply that on certain – perfectly respectable – views of second order quantification, there are considerable differences in the inductive commitments of PA versus Z_2.

[32]Only 'roughly', because the present formulation elides difficult issues concerning ordinal notations. In the context of arithmetical theories, we deal not with ordinals directly, but with their codings or representatives in some ordinal notation system. But, as usual when codings are involved, pathological cases arise. Rathjen discusses the example (due to Kreisel) of a notation system of order type ω which is not provably well-ordered in PA because it in effect uses coding tricks to build in information about the consistency of PA. The usual response is to restrict attention to 'natural' ordinal notations. Rathjen expresses what I take to be the consensus view, that this notion cannot be mathematically precisely defined and is inherently informal, when he claims that "it is futile to look for a formal definition of 'natural well-ordering' that will exclude every pathological example".[33] While this means that the notion of "proof theoretic ordinal" is correspondingly informal, in practice, problems do not arise, since there appears to be considerable agreement among proof theorists about which ordinal notation systems count as natural.

[34]The relation between transfinite induction and well-ordering is this: T proves that λ is well-ordered (under a given ordinal notation $\prec$) if and only if T validates the inference rule: from $\forall x, y \prec \lambda((x \prec y \rightarrow \phi(x)) \rightarrow \phi(y))$ infer $\forall x \prec \lambda \phi(x)$.

[35]The ordinal analysis of significant fragments of Z_2, in particular $\Pi^1_0 - CA$ and closely related systems, has been carried out by Rathjen and Arai. See for instance Rathjen [1995].

PRA proves the well-ordering of ordinals up to, but not including, ω^ω. Since $\omega^\omega < \epsilon_0$ there is thus a clear sense in which PRA is committed to less extensive induction than PA.

$\mathsf{PRA} + \mathsf{TI}^{QF}_{PR}(\epsilon_0)$ clearly proves the well-ordering of ϵ_0 itself; thus its commitments regarding the height of induction go beyond those of PA.

Summarizing the above discussion:

Theory	**Mode of induction**	**'Width'**	**'Height'**
PA	axiom schema	arithmetical	ϵ_0
Z_2	second-order	analytical	unknown ($\gg \epsilon_0$)
PRA	rule form	q.f. over $\mathcal{L}_{\mathsf{PRA}}$	ω^ω
$\mathsf{PRA} + \mathsf{TI}^{QF}_{PR}(\epsilon_0)$	rule form	q.f. over $\mathcal{L}_{\mathsf{PRA}}$	$\epsilon_0 + 1$

The distinction between width and height allows us to see what exactly is wrong with the Circularity Argument. The Gentzen proof illustrates a trade-off between the two dimensions of inductive commitment: it shows us how the consistency of PA can be derived in a theory $\mathsf{PRA} + \mathsf{TI}^{QF}_{PR}(\epsilon_0)$ which entails the well-ordering of ϵ_0 and thus takes on *additional* commitments (relative to PA) concerning the height of induction, but which allows induction only over quantifier-free formulas involving primitive recursive functions and thus takes on *fewer* commitments (again, relative to PA) concerning the width of induction. The fundamental problem is thus equivocation in the 'understanding' mentioned in premises CA1, CA2, and CA3. For although it's true that justification in $\mathsf{PRA} + \mathsf{TI}^{QF}_{PR}(\epsilon_0)$ requires an understanding of the ordinals up to ϵ_0, our discussion of induction shows that the required understanding is in a perfectly precise sense weaker or less committal than an understanding of the natural numbers sufficient to motivate PA.

4. Triviality?

I've so far argued that the most promising argument for the circularity of the Gentzen proof fails. In this section, I'll outline a different way in which a consistency proof can be trivial and argue that the Genzten proof is not trivial in that sense.

Consider the "epistemic space" of views or doxastic attitudes that one might coherently adopt concerning mathematical theories, including claims about their consistency / inconsistency. Let the attitudes here include not only states of belief / disbelief / agnosticism but also gradable attitudes such as credences or degrees of belief. For instance, my own attitudes include the following:

all-or-nothing doxastic states concerning the consistency of various theories – e.g. belief that PA is consistent; agnosticism whether ZFC + "there exist infinitely many Woodin cardinals" is consistent;

graded or credal states – e.g. high credence that PA is consistent;

conditional beliefs/credences – e.g. belief/high credence that, conditional on the inconsistency of PA, the proof of an inconsistency will essentially involve an instance of induction;

comparitive credences – e.g. a higher credence that, conditional on the inconsistency of PA, the proof of an inconsistency essentially involves induction than it does the axioms for addition.

"Coherent" needs to be understood in a relatively permissive way. Say that a position is coherent (relative to a background body of mathematical knowledge) if it could be held by a rational, reflective, mathematically sophisticated and well-informed agent (in possession of the background knowledge). In particular, I do not assume that coherence requires logical omniscience or anything analogous. In the sense intended, it can be perfectly coherent to fail to believe something that is entailed by one's other beliefs or even to have inconsistent beliefs (as long as the inconsistency is not of the kind that would be easily discovered by a mathematically sophisticated etc agent). I stress this point because it is crucial if we wish to allow the possibility that someone might coherently yet falsely believe in a theory's consistency. If T is inconsistent, then that fact will be derivable in a weak theory of syntax.[36] But we do not want to say that someone who accepts a weak syntax theory cannot coherently have false beliefs about the consistency of theories; this would not only be implausible, but would also trivialize the present investigation.[37]

Some examples may help to illustrate. Consider first the position of Edward Nelson, who believes (and has expended considerable effort attempting to prove) that weak arithmetical theories, even those as weak as PRA, are inconsistent.[38] Nelson's views may well be wrong – I believe and hope that they are – but his view nevertheless is clearly coherent relative to our current state of knowledge. In particular, it would be

[36] Any theory of syntax extending Q will suffice: Q can define the canonical provability predicate for any recursively axiomatized theory, and it is Σ_1 complete, so if T is inconsistent we will have $\mathsf{Q} \vdash \exists x \mathrm{Proof}_{\mathsf{T}}(x, \bot)$.

[37] This is the reason why I do not appeal to the usual frameworks of contemporary formal epistemology (e.g. epistemic logic and probabilism) in fleshing out the notion of epistemic space, since the most prominent formulations build in logical omniscience or equivalent.

[38] See for instance Nelson's discussion at https://mathoverflow.net/questions/142669/illustrating-edward-nelsons-worldview-with-nonstandard-models-of-arithmetic. Nelson circulated a claimed proof of an inconsistency in PA in 2015, but later withdrew it after an error was discovered by Terence Tao.

absurd to accuse him of being irrational, unreflective, mathematically unsophisticated, or poorly informed on the issue

On the other end of the spectrum, consider the kind of position held by many contemporary set theorists engaged in the search for new axioms, according to which the consistency of ZFC is virtually beyond question and the only real issue is how far into the hierarchy of large cardinal principles we are entitled to go. Hugh Woodin, for instance, has expressed high confidence in the consistency of ZFC plus 'there exists infinitely many Woodin cardinals'.[39] John Steel has done the same for ZFC plus various rank-to-rank embedding principles.[40] Perhaps these views are wrong and the relevant theories are inconsistent. But given that the mathematical community is presently in possession of no proof to that effect, it seems to me that, even if one disagrees with these positions, one must accept that they are coherent in the relevant sense.

This conception of epistemic space allows us to define a way in which a consistency proof might be diagnosed as 'trivial'. The basic idea is that a proof is substantive – non-trivial – if it rules out some coherent position in epistemic space. More fully: start by considering which positions are coherent relative to bodies of knowledge that do not include the proof; then consider the positions which are coherent once knowledge of the proof is added; finally, observe whether any formerly coherent positions are no longer coherent relative to the updated body of knowledge. The thought is then that a proof is trivial if it rules out no coherent positions in this way.

The Gentzen proof is not trivial in this sense. The reason is simply that, given the state of mathematical knowledge before the Gentzen proof was known, the following views were each coherent: (i) to believe that PA is inconsistent, but that $\mathsf{PRA} + \mathsf{TI}^{QF}_{PR}(\epsilon_0)$ is not; (ii) to believe that induction over all arithmetical conditions is 'riskier' (i.e. more likely to be inconsistent) than transfinite induction up to ϵ_0 over only primitive recursive conditions; or (iii) to have a higher credence that, conditional on PA being inconsistent, the inconsistency relies essentially on carrying out induction over complex quantified sentences involving non-primitive-recursive functions. But although these positions were each coherent before the proof was discovered, none can rationally be maintained once it is known.[41] There is a therefore a strong case to be made that the

[39] See for instance Woodin [2011], where Woodin predicts that "there will be no discovery *ever* of an inconsistency in these theories [i.e. ZFC plus 'there exists infinitely many Woodin cardinals' and others]."

[40] See Steel [2014], p. 156. Steel is more confident still in the consistency of axioms for which an inner model theory can be constructed: roughly as high as (but not including) the existence of a Woodin limit of Woodin cardinals.

[41] It is interesting to assess the so-called 'semantic argument' for the consistency of a theory T in these terms. The argument runs roughly as follows: the axioms of T are true, the inference rules of logic preserve truth, hence all of the theorems of T are true; but if T were inconsistent, it would have false theorems; so it is consistent. The semantic argument cannot be formalized in T itself but can be in a theory T^+ which adds certain compositional truth axioms to T. Very plausibly, the semantic argument is diagnosed as trivial by the criterion above: for if there is some coherent position that is ruled out by the argument, it must be one

Gentzen proof is not trivial in the sense under discussion.

5. Real Vindication?

If what I've said so far is right, the circularity argument against the Gentzen proof's ability to transmit justification is flawed and the proof is non-trivial in the sense that it rules out certain antecedently coherent positions about how an inconsistency in arithmetic might arise. But there is an important question remaning. Our notion of a coherent position in epistemic space was a fairly permissive one: a position that someone could hold, given a body of mathematical knowledge, without thereby being guilty of irrationality. But simply because a position is coherent in these terms does not mean that it is *positively justified* or *well-motivated* in a more demanding sense. It's plausibly true that before the Gentzen proof, someone could have coherently accepted $\mathsf{PRA} + \mathsf{TI}^{QF}_{PR}(\epsilon_0)$ while failing to accept $\mathrm{Con}(\mathsf{PA})$. But is there any positive justification for inhabiting such a position?

A full answer to this question would require a comprehensive theory of positive mathematical justification – well beyond the scope of this paper. Instead, I'll offer a brief sketch of a view – perhaps even the orthodox view – which, I believe, will allow progress to be made. I suggest that we consider the most prominant examples of what might be called *foundational stances*: informal conceptions of various domains of mathematical objects (e.g. the natural numbers, syntactic objects, the universe of sets) or modes of reasoning (e.g. constructive or finitistic proof), corresponding to a principled position or foundational program in the philosophy of mathematics.[42] The proposal is that these stances are, in the first instance, the primary sources of positive mathematical justification.

Foundational stances as just described do not immediately give rise to positively well-motivated positions in epistemic space, for foundational stances are informal and positions in epistemic space as we are understanding them concern attitudes towards *formal* mathematical theories. To bridge this gap, we need what I'll call *foundational equivalence theses*: claims to the effect that a foundational stance is extensionally equivalent to a formal mathematical theory.[43] A theory which does not enjoy the privilege

that endorses T^+. But if so, it presumably also endorses $\mathrm{Con}\mathsf{T}^+$. However, since T is a subtheory of T^+, $\mathrm{Con}\mathsf{T}^+$ implies $\mathrm{Con}\mathsf{T}$ (and indeed this will be known by a reflective agent). So, any position in epistemic space which endorses the premises of the semantic proof must, plausibly, rationally already endorse its conclusion. If so, the proof fails to rule out any coherent position.

[42] See for instance the essays in Benacerraf and Putnam [1964] Part I. Further examples of foundational stances, helping to clarify the notion, will be given shortly below.

[43] Foundational equivalence theses are examples of the larger class of what might be called *informal equivalence theses*, such as the Church-Turing thesis – characteristic examples of the method of 'informal rigour' discussed in Kreisel [1967].

of being on the right hand side of a foundational equivalence thesis is thus, in a nice phrase of Walter Dean's, epistemically unstable: there is no positive rationale for accepting that theory and only that theory; either it has no motivation at all, or it is motivated only derivatively, via its relation to a more powerful system whose axioms are themselves well-motivated.[44]

Our question then becomes: is there some foundational stance that motivates $\mathsf{PRA} + \mathsf{TI}^{QF}_{PR}(\epsilon_0)$ but does not immediately motivate ConPA? The question can be sharpened further still. Granted the additional epistemic premise that someone who endorses T is rationally committed to its consistency claim ConT, what we need to find is some stance that motivates $\mathsf{PRA} + \mathsf{TI}^{QF}_{PR}(\epsilon_0)$ but does not motivate PA itself.[45]

My suspicion, which I will defend in the remainder of this section, is that this cannot be done.

Before I defend that claim, however, let me say something briefly about how the present attempt to evaluate Gentzen's proof differs from other approaches in the literature. Consider first Kleene's judgement that the status of Gentzen's proof depends on "how ready one is to accept induction up to ϵ_0 as a finitary method." While that issue is clearly relevant to our question, it cannot be the whole story. For unless we go along with Hilbertian views about the absolute epistemic security of finitary methods, there is no particular reason why they should have a privileged place in our inquiry: perhaps $\mathsf{TI}^{QF}_{PR}(\epsilon_0)$ can be informatively justified by some foundational stance other than finitism. Second, while commentators are often sensitive to the possibility that a foundational stance (usually, like Kleene, focusing on the example of finitism) undershoots relative to the proof in that it fails to motivate the required resources, I do not think the possibility and the significance of overshooting has been appreciated to the same degree. But suppose that the only well-motivated positions able to carry out the Gentzen proof are those that can already be shown to motivate PA (or ConPA) in a much more direct way. In that case it is hard to see how the proof can constitute a 'real vindication' of its conclusion, to use Gentzen's nice term: for anyone who justifiably accepted its premises, prior to the proof being known, was already in a position – via a much simpler set of reflections – to justifiably accept its conclusion.

Let us turn to the question of how PA might be motivated. The thesis that bears most directly on the question is:

[44] Dean [2015], p. 53.

[45] This additional premise is widely accepted, sometimes as a consequence of the stronger thesis that anyone who accepts a theory is rationally committed not only to the consistency of T but also to its soundness, expressed via a reflection principle. Halbach [2017], p. 308, for instance, claims that "accepting a theory without believing in its consistency strikes many logicians at least as odd if not incoherent. If one endorses a theory, so one might argue, one should also take it to be sound". For an opposing view, see Dean [2015].

Isaacson's thesis. According to Isaacson, there is a class of "purely arithmetical" truths: those that are justified on the basis of our intuitive conception of the natural numbers as the structure containing an initial element (0) and closed under a successor relation, but which do *not* rely on "higher order" or "set-theoretic" considerations.[46] Isaacson's thesis is that the purely arithmetical truths are captured precisely by Peano Arithmetic, in the sense that all and only the theorems of PA are purely arithmetical in this sense.

However, there is a subtlety. Isaacson is explicit that the purely arithmetical truths form a proper subset of those that are justified by a more fundamental, and more expansive, conception of the natural numbers. As he puts it, "I am *not* claiming that PA could itself constitute an adequate conceptual basis for our understanding of the concept of natural number. Far from it, I consider that we can only arrive at such a system on the basis of some higher-order understanding.[47]" That more expansive understanding is given by what we might call:

Dedekind's thesis. The conception of the natural numbers as the smallest structure containing an initial element (0) and closed under a successor relation is captured precisely by the formal system of second-order arithmetic Z^2.

Even if Isaacson's thesis does not give rise to an autonomous motivation for PA, it is clear that, if Dedekind's thesis is correct, PA can be motivated derivatively via Z^2. It is worth saying a little more about motivating a theory by 'reducing' it to a more expansive well-motivated theory. In the most straightforward case, when the theories are formulated in the same language, it suffices for every axiom of the 'smaller' theory to be derivable in the 'larger' theory. But sometimes theories are not stated in the same language, and here a number of formal reducibility relations have been proposed. The most prominent are *interpretability* – roughly, a systematic mapping from the theorems of the 'larger' theory to the axioms of the 'smaller' theory in such a way that logical structure is preserved – and *proof-theoretic reduction* – roughly, when there is an effective procedure for transforming a proof of ϕ in the 'smaller' theory to a proof of ϕ in the 'larger' theory, and the 'larger' theory is able to recognize this fact.[48] Plausibly, if a 'smaller' theory can be formally reduced to a 'larger' theory in any of these ways, an *epistemic* reduction is thereby effected in the sense it provides a means for someone who justifiably accepts the larger theory to justifiably accept the smaller.

[46] Isaacson [1987].

[47] Isaacson [1987], p. 209.

[48] Burgess [2005], p. 38, notes that "Philosophically, interpretation is a way of legitimizing a theory that is not legitimate when taken literally." For more on proof-theoretic reducibility, see e.g. Feferman [1993] and Feferman [2000].

So there are at least a couple of potential routes for motivating PA. It is far less clear, however, what can be said about $\mathsf{PRA} + \mathsf{TI}^{QF}_{PR}(\epsilon_0)$. My conclusions here are tentative but pessimistic: I do not think there is any stance that motivates $\mathsf{PRA} + \mathsf{TI}^{QF}_{PR}(\epsilon_0)$ but not PA. I have no knock-down argument for this claim, and I do not know how one could be constructed. Rather, I'll proceed in the only way I can see: by briefly examining the three most promising candidates and arguing that they either under- or overshoot the target, in the sense that each either fails to motivate $\mathsf{PRA} + \mathsf{TI}^{QF}_{PR}(\epsilon_0)$ or proceeds to motivate PA. The three stances we'll consider are (i) finitism, (ii) constructivism, and (iii) predicativism. Let us take each in turn.

5.1 Finitism

Finitism, in the sense of Hilbert and Bernays, can be roughly described as a conception or mode of mathematical reasoning about the natural numbers that does not presuppose that they constitute a completed infinite totality.[49]

How much mathematics can be carried out on this basis? A widely accepted foundational equivalence thesis concerning finitism is:

> *Tait's thesis.* Finitism is captured by the formal system of Primitive Recursive Arithmetic (PRA) in the sense that any proof within PRA is finitistically acceptable, and conversely any finitistically acceptable proof can be converted to a proof in PRA with the same conclusion.[50] It bears emphasis that (as with many foundational equivalence theses) this is an external characterization of the position: Tait does not claim that the finitist ought to endorse PRA itself, since, after all, it is a theory that involves reference to (infinitely many) total functions on the natural numbers – objects which plausibly presuppose the notion of a completed infinite totality of numbers and thus cannot be recognized by the finitist as such. Rather, as with many foundational equivalence theses, Tait's thesis characterizes finitism externally.[51]

[49] For more on philosophical aspects of finitism, see Tait [1981] and Incurvati [2015]. As with other major foundational programs, finitism admits several apparently distinct characterizations. One is historical – roughly, as whatever Hilbert and Bernays (the pre-eminent historical figures associated with the position) meant by it. There are very interesting questions concerning whether, on this historical conception, Gentzen's proof is finitistically acceptable. See e.g. Zach [2003] for more. Another characterization is explicitly epistemic – roughly, the part of mathematics that enjoys a certain kind of intuitive evidence or epistemic security. However the conceptual characterization given above seems more fundamental: presumably, if finitism is epistemically distinguished, that ought to be explained in terms of the nature of the domain or reasoning it encompasses.

[50] Tait [1981]. See also Dean [2015], p. 50-52.

[51] Tait also defends an analogous thesis for functions: that any finistically acceptable function is primitive recursive, and conversely any primitive recursive function is finistically acceptable.

Clearly, if Tait's thesis is correct, it rules out finitism as a suitable motivation for Gentzen's proof: for assuming PRA is consistent, $\mathsf{TI}^{QF}_{PR}(\epsilon_0)$ cannot be proven in PRA and therefore cannot be justified on finitist grounds.

Although Tait's thesis is the orthodox view concerning the limits of finitism, it is worth briefly considering a more expansive conception, due to Kreisel, who famously argued that "finitist results include essentially those of classical number theory".[52] More precisely, Kreisel argues that finitism corresponds to a certain quantifier-free system capable of proving analogues of the Π_2 statements as PA.[53]His argument begins with the idea that PRA is finitistically acceptable and appeals to the finitistic acceptability of an autonomous progression of theories extending it: in effect, the thought is that, given a formalization of a notion of finitist proof, it can be expanded to a more inclusive notion of finitist proof by adding reflection principles expressing its soundness; furthermore, this process can be carried out along limit ordinals *provided that there is a finitist proof of their well-ordering*.

The crux of Kreisel's disagreement with Tait thus appears to be whether finitists can accept the general notion of a total function on the natural numbers: this is implicit in the move from the claim – which Tait accepts – that the finitist can accept any *particular* primitive recursive function – to the claim – which Tait denies – that the finitist can accept a *general* principle to the effect that functions can be constructed via primitive recursion. Nevertheless, even if Kreisel's more expansive conception of finitism is correct, it does not provide a vindication of Gentzen's result: in particular, it does not yield any principles of transfinite induction beyond those available in PA. It appears, consequently, that no plausible understanding of finitism is capable of motivating the resources to carry out Gentzen's proof.

5.2 Predicativism

Predicativism is a conception of the natural numbers and sets thereof motivated by the vicious circle principle in the sense of Poincaré and Russell, according to which a set of natural numbers is acceptable insofar as they can be defined without quantifying over totalities to which they belong.[54]

The most prominent foundational equivalence thesis concerning predicativism is the:

Feferman-Schütte thesis. Predicativism is captured precisely by the formal

[52]Kreisel [1958], p. 290.

[53]See also Kreisel [1970] and Dean [2015], p. 45-7 for discussion.

[54]Again, predicativism has a rich and storied history; for more, see Feferman [2005], Hellman [2004], and Feferman [2004].

system $\mathsf{RA}_{<\Gamma_0}$ of ramified analysis up to the Feferman-Schütte ordinal Γ_0.[55]

Unfortunately, for our purposes, $\mathsf{RA}_{<\Gamma_0}$ overshoots, for it is a theory extending PA. Feferman is explicit that predicativism in the sense his thesis is intended to characterize is *predicativism given the natural numbers* – the conception of natural numbers given by PA. This is clear from the procedure used to arrive at $\mathsf{RA}_{<\Gamma_0}$. Informally, we start with PA, and allow sets that are definable by a formula of PA. Then we proceed in stages, adding sets that are definable at the previous stage (or at limits, some previous stage); and iterate this procedure into the transfinite for 'predicatively acceptable' ordinals, i.e. those expressed by ordinal notations provably well-ordered at some previous stage.

However, it is not at all clear that predicativism must take a conception of the natural numbers *as codified by* PA as basic; in principle there is no reason why predicativism might not be developed from other starting points. To my knowledge, there are two main proposals in the literature for extracting the predicativist commitments from a given theory. Neither, I believe, provides room for optimism.

The first involves *autonomous progressions*.[56] However, it is difficult to see how autonomous progressions starting from a well-motivated arithmetical theory *weaker* than PA could yield the relevant transfinite induction principle. A finitist starting point certainly does not appear promising; the results of Kreisel discussed above show that extending finitist arithmetic with reflection principles up to autonomous ordinals will not suffice.[57] Perhaps an autonomous progression based on some other well-motivated arithmetical theory would suffice to motivate Gentzen's proof, but I do not know of any, and none to my knowledge has been proposed in the literature.

The second approach is also found in the work of Feferman.[58] Starting with a schematically axiomatized theory S, one extends it to obtain its 'unfolding' $\mathcal{U}(S)$. The details of the unfolding construction are technically complex, but they can be viewed as extending S by adding new predicates and operations obtained from those of S by a kind of generalized recursion. According to Feferman, unfolding is philosophically significant as a means of characterizing the implicit predicativist commitments of acceptance of a starting theory; indeed, one source of support for this claim is the fact that $\mathcal{U}(\mathsf{PA}) \equiv \mathsf{RA}_{<\Gamma_0}$, so that the unfolding of Peano arithmetic is (proof-theoretically

[55] See Feferman [1964] and Schutte [1965].

[56] In addition to Kreisel [1958], see Feferman [1962].

[57] Two kinds of autonomous progressions have been discussed above: one involving reflection extensions, the other involving ramified theories of sets. Kreisel's results concern the former with PRA (a finitist theory, assuming Tait's thesis) as a starting point; I do not know of any results concerning the autonomous ramified progression arising from PRA. Possibly this is because PRA is not a natural theory to consider in this context, for it is quantifier-free and thus it is not entirely clear how it could be 'extended' to a second-order theory in the way required to get ramification off the ground. Perhaps the most natural approach would be to consider the theory $QF - IA$, the conservative extension of PRA which adds standard first-order quantification and allows induction for quantifier-free sentences.

[58] See for instance Feferman et al. [1996], Feferman and Strahm [2000], and Feferman and Strahm [2010].

equivalent to) precisely the theory that arose from PA via autonomous progressions. Again, for our purposes, the interesting question is the strength of the unfolding of theories of arithmetic weaker than PA. Here the main result of significance is due to Feferman and Strahm, and applies to a theory they call *FA* for Finitist Arithmetic, so-called because it bears a close relation to PRA.[59] Their result is that $\mathcal{U}(FA)$ is proof theoretically equivalent to PRA itself. Granted the assumptions (a) that *FA* is a reasonable expression of finitist arithmetical commitments and (b) that unfolding represents the predicativistic commitments implicit in the acceptance of a theory, this result can be interpreted as saying that predicativism *given a finitist conception of the natural numbers* motivates a theory proof-theoretically equivalent to PRA. But if so, it undershoots in our sense, and thus the most natural ways of developing predicativism do not yield a theory that motivates $\mathsf{TI}^{QF}_{PR}(\epsilon_0)$ without motivating PA.

5.3 Constructivism

The last foundational stance we'll consider is constructivism: *very* roughly a view that requires, for any object claimed to exist, an explicit construction procedure to be provided.[60] It is sometimes claimed, for instance in Feferman [2000], that transfinite induction of the kind needed for Gentzen's proof may be justified on constructive grounds. Again, though, it should be emphasized that this is not exactly our question: ours is not merely whether $\mathsf{TI}^{QF}_{PR}(\epsilon_0)$ can be constructively justified but rather whether it can be done so in a way that does not itself justify PA. Perhaps even more than the other stances we've considered, constructivism is a broad church, encompassing a variety of philosophical motivations, and as before, we cannot hope to exhaustively consider all its possible forms here. Instead let us focus upon the main theories that might be motivated on a constructivist basis.

A natural first thought is to wonder if constructivist approaches to arithmetic might fit the bill. The main equivalence thesis in the vicinity is what we might call:

> *Heyting's thesis.* Constructive arithmetic is precisely captured by the system of Heyting Arithmetic HA, i.e. the theory with the same axioms as PA but which uses an intuitionistic background logic.[61]

If Heyting's thesis is correct, however, then constructivist arithmetic will not suffice. The reason is due to the Gödel-Gentzen double-negation translation: if ϕ is a theorem of PA, then ϕ^N is a theorem of HA, where ^{N} is the translation prefixing every atomic

[59] Feferman and Strahm [2010].
[60] See e.g. Troelstra and Van Dalen [2014].
[61] Heyting [1930].

claim, disjunction, and existential claim with $\neg\neg$. Constructivist arithmetic thus undershoots, since its proof theoretic ordinal can be seen to be the same as that of PA. It also arguably overshoots, since, as Rumfitt claims, a plausible reading of the translation theorem is that it "shows how someone who accepts Heyting Arithmetic might be rationally persuaded to accept the classical Peano Arithmetic".[62] If so, then the translation provides a reduction of PA to HA in the epistemic sence, of providing a means for someone who justifiably accepts the latter to justifiably accept the former.

What about constructivist theories going beyond arithmetic? Constructive analysis comes in several varieties (for some, see Troelstra and Van Dalen [2014]), reflecting differing philosophical treatments of the construction of the real numbers. There are relatively few foundational equivalence theses connecting these views with particular formal systems – when a formal treatment of constructive analysis is needed, it usually takes place within a more powerful constructivist system (to be discussed shortly below). It is plausible, however, that any formal system justified on the basis of constructivist approaches to analysis will undershoot: Simpson [1999], p. 43, for instance claims that analysis in the mode of Errett Bishop (one of the most liberal constructivist treatments of the subject) is captured by the weak fragment of second-order arithmetic known as RCA_0, whose proof-theoretic ordinal is simply ω^ω. If that is right, then it is hard to see how any approach in the vicinity can justify enough transfinite induction to carry out the Gentzen proof.

Finally, we turn to more expansive constructivist systems. The most prominent of these are (a) constructivist set theory in the form of CST (for "Constructive Set Theory") and IZF (for "Intuitionistic Zermelo-Fraenkel") and (b) constructivist type theory in the form of the system proposed by Martin-Lof.[63] The former might be viewed as capturing certain constructivist conceptions of the universe of sets; the latter as a constructivist approach to type theory and higher-order logic. Needless to say, as one would expect from frameworks that have been proposed as foundations for constructive mathematics, a great deal of mathematics can be carried out within them. For our purposes, each appears to incorporate significant arithmetical commitments: both constructivist set theories incorporate axioms of infinity, asserting in effect that the set of natural numbers exists; Martin-Lof type theory contains a type of natural numbers, satisfying analogues of the usual axioms. In particular, these approaches each overshoot in our sense, for each is capable of interpreting HA (as well as a great deal beyond).

This concludes our discussion of foundational stances and the formal systems to which they give rise. Having examined all of the most natural foundational stances and

[62] Rumfitt [2015], p. 288.

[63] For CST and IZF see Aczel and Rathjen [2010]; for Martin-Lof type theory, see Martin-Löf and Sambin [1984].

formal equivalence theses, we have seen that none motivates the premises of Gentzen's proof in a way that does not, more directly, motivate Peano arithmetic. In light of this, it is difficult to defend the view that the proof provides a real vindication of the consistency of arithmetic.

6. Concluding Remarks

Our fundamental question has been whether Gentzen's consistency proof is a genuine proof in the epistemic sense: whether it can be used by a mathematically reflective agent to gain new or enhanced justification in the consistency of arithmetic. The first two parts of our discussion led to broadly favourable conclusions. We examined the main argument that the proof is circular, and saw that it fails due to its ability to fully appreciate the restricted scope of the required induction principles. A notion of triviality for consistency proofs was introduced – roughly, a proof is trivial when it is unable to rule out any antecendely coherent positions about the consistency/inconsistency of mathematical theories – and we saw that the Genzten proof is non-trivial in this sense. The last part of the discussion, however, led us to a less optimistic conclusion: although the proof is not circular or trivial, it fails to constitute a real vindication of the consistency of Peano Arithmetic, for there is no foundational stance which motivates the principles it uses but which does not more directly motivate Peano Arithmetic, and hence its consistency, itself. If all of this is right, then perhaps of all of the commentators on Gentzen's proof, perhaps Tarski was the closest to the truth: it should raise our confidence in the consistency of arithmetic, but only by about an epsilon.

I will conclude by mentioning some possible directions for further investigation. First, our discussion of whether the proof constituted a real vindication appealed to a specific – some might say, an old fashioned and 'foundationalist' – conception of mathematical justification, according to which justification arises from some underpinning foundational stance. But of course, that is not the only possible view one might have of mathematical justification; it would be interesting to know whether the Gentzen proof fares any better under any of the alternatives. Second, there exist other consistency proofs for arithmetic: most notably, Gödel's 'Dialectica Interpretation' which proves the consistency of arithmetic in a system of primitive recursive functions of higher type. I suspect that this proof is not circular or trivial for reasons similar to those presented above for Gentzen's proof, but do not know whether there is some foundational stance that motivates its premises without more directly motivating PA. Finally, since Gentzen, proof theorists have expended considerable effort on producing ordinal analyses of theories beyond PA. In addition to the mathematical information that these ordinal analyses provide, they can also be viewed as consistency proofs, since an or-

dinal analysis of a theory T yields a theorem to the effect that for some ordinal (and associated notation scheme) α, $\mathsf{PRA} + \mathsf{TI}^{QF}_{PR}(\alpha) \vdash \mathrm{Con}\mathsf{T}$. Do such theorems constitute real vindications of consistency? An investigation analogous to that of the previous section could presumably be carried out. As the theories under consideration grow stronger, it becomes harder to 'overshoot' in our sense, and so it is not at all obvious to me how such an investigation would conclude.

References

P. Aczel and M Rathjen. *CST*. 2010. Available online at https://www1.maths.leeds.ac.uk/ rathjen/book.pdf.

Mark Balaguer. *Platonism and Anti-platonism in Mathematics*. Oxford University Press, 1998.

Paul Benacerraf and Hilary Putnam. *Philosophy of Mathematics: Selected Readings*. Cambridge University Press, 1964.

John P. Burgess. *Fixing Frege*. Princeton University Press, 2005.

Walter Dean. Arithmetical reflection and the provability of soundness. *Philosophia Mathematica*, 23(1):31–64, 2015.

Fred I. Dretske. Epistemic operators. *Journal of Philosophy*, 67(24):1007–1023, 1970.

Solomon Feferman. Transfinite recursive progressions of axiomatic theories. *The Journal of symbolic logic*, 27(3):259–316, 1962.

Solomon Feferman. Systems of predicative analysis. *The Journal of Symbolic Logic*, 29(1): 1–30, 1964.

Solomon Feferman. What rests on what? the proof-theoretic analysis of mathematics. In J. Czermak, editor, *Philosophy of Mathematics*, pages 1–147. Hölder-Pichler-Tempsky, 1993.

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

Solomon Feferman. Comments on "predicativity as a philosophical position". *Revue internationale de philosophie*, (3):313–323, 2004.

Solomon Feferman. Predicativity. In Stewart Shapiro, editor, *Oxford Handbook of Philosophy of Mathematics and Logic*, pages 590–624. Oxford: Oxford University Press, 2005.

Solomon Feferman and Thomas Strahm. The unfolding of non-finitist arithmetic. *Annals of Pure and Applied Logic*, 104(1-3):75–96, 2000.

Solomon Feferman and Thomas Strahm. Unfolding finitist arithmetic. *The Review of Symbolic Logic*, 3(4):665–689, 2010.

Solomon Feferman et al. Gödel's program for new axioms: Why, where, how and what. *Gödel*, 96:3–22, 1996.

Hartry Field. *Realism, Mathematics & Modality*. Blackwell, 1989.

Gerhard Gentzen. *The Collected Papers of Gerhard Gentzen*. Amsterdam: North-Holland Pub. Co., 1969.

M. Giaquinto. *The Search for Certainty: A Philosophical Account of Foundations of Mathematics*. Oxford University Press, 2002.

Jean-Yves Girard. *The Blind Spot: Lectures on Logic*. European Mathematical Society, 2011.

Volker Halbach. *Axiomatic Theories of Truth*. Cambridge University Press, second edition, 2017.

Bob Hale and Crispin Wright. *The Reason's Proper Study: Essays Towards a Neo-fregean Philosophy of Mathematics*. Oxford University Press, 2001.

Geoffrey Hellman. *Mathematics Without Numbers: Towards a Modal-structural Interpretation*. Oxford University Press, 1989.

Geoffrey Hellman. Predicativism as a philosophical position. *Revue internationale de philosophie*, (3):295–312, 2004.

Arend Heyting. Die formalen regeln der intuitionistischen logik. *Sitzungsbericht PreuBische Akademie der Wissenschaften Berlin*, pages 42–56, 1930.

Luca Incurvati. On the concept of finitism. *Synthese*, 192(8):2413–2436, 2015.

Daniel Isaacson. Arithmetical truth and hidden higher-order concepts. In *Studies in Logic and the Foundations of Mathematics*, volume 122, pages 147–169. Elsevier, 1987.

Reinhard Kahle. Gentzen's consistency proof in context. In *Gentzen's Centenary*, pages 3–24. Springer, 2015.

Reinhard Kahle and Michael Rathjen. *Gentzen's Centenary*. Springer, 2015.

Stephen Cole Kleene. *Introduction to Metamathematics*. North Holland, 1952.

Stephen Cole Kleene. *Mathematical Logic*. Dover Publications, 1967.

G Kreisel. Formal rules and questions of justifying mathematical practice. *Konstruktionen versus Positionen*, pages 99–130, 1979.

Georg Kreisel. Ordinal logics and the characterization of informal concepts of proof. In *Proceedings of the international congress of mathematicians*, volume 14, page 21, 1958.

Georg Kreisel. Informal Rigour and Completeness Proofs. In Imre Lakatos, editor, *Problems in the Philosophy of Mathematics*, pages 138–157. North-Holland, 1967.

Georg Kreisel. Principles of proof and ordinals implicit in given concepts. In *Studies in Logic and the Foundations of Mathematics*, volume 60, pages 489–516. Elsevier, 1970.

Per Martin-Löf and Giovanni Sambin. *Intuitionistic Type Theory*, volume 9. Bibliopolis Naples, 1984.

George Edward Moore. Proof of an external world. *Proceedings of the British Academy*,

25(5):273–300, 1939.

Ram Neta. Easy Knowledge, Transmission Failure, and Empiricism. *Oxford Studies in Epistemology*, 4:166, 2013.

James Pryor. What's wrong with moore's argument? *Philosophical Issues*, 14(1):349–378, 2004.

James Pryor. When Warrant Transmits. In Crispin Wright and Annalisa Coliva, editors, *Mind, Meaning, and Knowledge: Themes From the Philosophy of Crispin Wright*. Oxford University Press, 2012.

Michael Rathjen. Recent advances in ordinal analysis: $\pi^1_2 - ca$ and related systems. *Bulletin of Symbolic Logic*, 1(4):468–485, 1995.

Michael Rathjen. The realm of ordinal analysis. *London Mathematical Society Lecture Note Series*, pages 219–280, 1999.

H. Rogers. *Theory of Recursive Functions and Effective Computability*. MIT Press, 1987.

Ian Rumfitt. *The Boundary Stones of Thought: An Essay in the Philosophy of Logic*. Oxford University Press, 2015.

K. Schütte. *Proof Theory*. Springer Verlag, 1977.

Kurt Schutte. Predicative well-orderings. In *Studies in Logic and the Foundations of Mathematics*, volume 40, pages 280–303. Elsevier, 1965.

Stewart Shapiro. *Foundations Without Foundationalism: A Case for Second-order Logic*. Oxford University Press, 1991.

Stewart Shapiro. *Philosophy of Mathematics: Structure and Ontology*. Oxford University Press, 1997.

Stephen G. Simpson. *Subsystems of Second-order Arithmetic*. Springer-Verlag, 1999.

Peter Smith. *An Introduction to Gödel's Theorems*. Cambridge University Press, 2012.

J.R. Steel. Gödel's program. In Juliette Kennedy, editor, *Interpreting Gödel: Critical Essays*, pages 153–179. Cambridge University Press, 2014.

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

Gaisi Takeuti. *Proof Theory*. Elsevier, 1987.

Anne Sjerp Troelstra and Helmut Schwichtenberg. *Basic Proof Theory*. Number 43. Cambridge University Press, 2000.

Anne Sjerp Troelstra and Dirk Van Dalen. *Constructivism in Mathematics*, volume 2. Elsevier, 2014.

Jonathan Vogel. Reliabilism leveled. *Journal of Philosophy*, 97(11):602–623, 2000.

Alan Weir. *Truth through Proof: A Formalist Foundation for Mathematics*. Oxford University Press, 2010.

W Hugh Woodin. The transfinite universe. In *Horizons of truth. Kurt Gödel and the foundations of mathematics*, pages 449–474. 2011.

Crispin Wright. Facts and certainty. In *Proceedings of the British Academy, Volume 71:*

1985, pages 429–472. Published for the British Academy by Oxford University Press, 1986.

Crispin Wright. About "the philosophical significance of gödel's theorem": Some issues. In Brian McGuiness and Gianluigi Oliveri, editors, *The Philosophy of Michael Dummett*, pages 167–202. Kluwer Academic Publishers, 1994.

Crispin Wright. (anti-)sceptics simple and subtle: G. e. moore and john mcdowell. *Philosophy and Phenomenological Research*, 65(2):330–348, 2002.

Richard Zach. The practice of finitism: Epsilon calculus and consistency proofs in hilbert's program. *Synthese*, 137(1-2):211–259, 2003.

Richard Zach. Hilbert's program then and now. In Dale Jacquette, editor, *Philosophy of Logic*, pages 411–447. Amsterdam: North Holland, 2007.

José L. Zalabardo. Wright on moore. In Annalisa Coliva, editor, *Mind, Meaning, and Knowledge: Themes From the Philosophy of Crispin Wright*, pages 304–322. Oxford University Press, 2012.