Hypatia's Silence

Martin Fischer, Leon Horsten, Carlo Nicolai

October 21, 2017

Abstract

Hartry Field distinguished two concepts of type-free truth: scientific truth and disquotational truth. We argue that scientific type-free truth cannot do justificatory work in the foundations of mathematics. We also present an argument, based on Crispin Wright's theory of cognitive projects and entitlement, that disquotational truth can do justificatory work in the foundations of mathematics. The price to pay for this is that the concept of disquotational truth requires non-classical logical treatment.

1 Introduction

Can the concept of type-free truth play an essential role in justifying new mathematical knowledge? This question is clearly of philosophical importance, but it is also ambiguous. As argued in [Field 1994], there are (at least) *two* concepts of truth. There is no consensus in the literature about the exact content of these two concepts, nor is the terminology used to mark the distinction uniform. But there is some agreement on the existence of a salient distinction along the lines that Field suggests, and on the acceptability of the following minimal characterisation of the two concepts. The first is a concept of truth that plays some role in scientific explanations – e.g. explaining communication by specifying truth-conditions for some natural language expressions; we call this theoretical notion *scientific truth*. The second is a notion of truth that is governed by rules of semantic ascent and descent, and we call it *disquotational truth*.[1]

In this article, we leave the scientific concept of truth mostly aside and focus on a concept of truth characterized by the unrestricted rules of disquotation. We claim, with McGee, that disquotational truth can play a justificatory role, but we disagree with his reasons for *why* it can do this. McGee's account of the justificatory role of disquotational truth hinges on the admissibility of *stipulatively introducing* a concept of disquotational truth in certain circumstances. Against this, we will argue that the introduction of disquotational truth by a stipulative act compromises its potential for playing a justificational role. Instead, we will give an alternative account of the way in which disquotational truth can play a role in the justification of mathematical knowledge.

We will describe how disquotational truth can lead one from a justified mathematical theory to justification of a stronger mathematical theory.

We will concentrate on a *type-free* concept of disquotational truth and take the essence of disquotational truth to be that of a device for *unrestricted quotation and disquotation*. On our view, disquotational truth is a concept that is governed by the unrestricted *Truth-Introduction* ('From φ, infer $\mathsf{T}\ulcorner\varphi\urcorner$', where T is our disquotational truth predicate) and *Truth-Elimination* ('From $\mathsf{T}\ulcorner\varphi\urcorner$, infer φ') rules. McGee focuses in his discussion of disquotationalism on *typed* disquotational theories, whereas Field has in his recent work concentrates on type-free disquotational truth [Field 2008]. There are good reasons for preferring a type-free concept of truth over the Tarskian typing strategy: they have been thoroughly defended elsewhere – see for instance [Kripke 1975] and [Field 2008, Ch. 3,§1]. Concepts of type-free truth are often compared with respect to their treatment of paradoxical sentences such as Liar sentences. Our notion of type-free truth will be articulated inferentially, and it is *silent* with respect to Liar-like sentences (see [Horsten 2011]).

[1]Throughout the article we accept Field's distinction in this rough-and-ready way without arguing for it. Observe that one can accept Field's distinction while being sceptical of the credentials of one of the concepts. Horwich, for instance, agrees that correspondence notions of truth aim at being useful in science, but he disputes that these notions can live up to their promise [Horwich 1998]. Field himself has over the years also become sceptical about the usefulness of what we call the scientific concept of truth. McGee accepts Field's distinction, but argues that only disquotational truth can play a fundamental role in justifying new mathematical principles [McGee 2005a], [McGee 2005b]. Horwich, on the other hand, holds that disquotational truth can play no essential justificatory role [Horwich 1998].

Our discussion of the justificational role of disquotational truth is framed in the context of Wright's *cognitive projects* [Wright 2004a]: accepting the scientific notion of truth, and accepting the disquotational concept of truth as a justificatory device, are two distinct cognitive projects. These cognitive projects are in some sense in tension with each other. Science uses classical logic throughout, so scientific truth operates in a context of classical logic. Full type-free disquotational truth, however, can only function in a logic that is weaker than classical logic. So the question which concept of truth to use (for a given purpose) is intertwined with the discussion what the correct logic is.

Our account will make use of Feferman's theory of *implicit commitment*.[2] On this view, if one fully accepts a theory S, then one is implicitly committed to accepting a *reflection principle* for S, which says, roughly, that *everything that S proves, is true*. We propose to view the epistemological import of Feferman's theory of implicit commitment in the light of the distinction between the notions of entitlement and justification (see again [Wright 2004a]). The view will be that if one is entitled fully to accept a theory S, then one is *entitled* explicitly to accept a reflection principle for S. This results in a stronger theory. This process can be repeated, so that we eventually become justified in accepting much stronger theories obtained by iteration of reflection.

Putting these elements together, we arrive at an account of the justificatory force of disquotational truth that can be outlined as follows. Suppose that we are justified, to start with, in believing a given mathematical theory S, governed by classical logic. Then we are warranted to extend our conceptual repertoire with an unrestricted type-free disquotational concept of truth, governed by principles of some non-classical logic. This results in a theory S' in which we are entitled to believe. Then we are implicitly committed to iterated reflection principles, and we are entitled to accept them explicitly. Thus we eventually come to accept a much stronger theory $R(\ldots R(S')\ldots)$. The mathematical part of this theory will again be governed by classical logic, and it will be a theory that is significantly stronger than the mathematical theory S with which we started.

[2]See [Feferman 1962] and also [Franzén 2004].

As a particular case study of this pattern we will sketch how from a justified belief in a fragment of arithmetic, disquotational truth leads to justified belief in a substantial fragment of classical analysis.

Let us now look at the details of how all of this works.

2 Two concepts of truth

When looked at from the outside, and in a somewhat superficial manner, contemporary research in theories of type-free truth appears to be divided into two communities. The first community of researchers concentrates on truth theories formulated in *classical logic*. The second community focuses on truth theories formulated in the context of some *non-classical logic*.[3]

Against explorations of truth in the context of non-classical logic, the following concern can be raised: *can withdrawing from classical logic ever be a sound methodological move?*[4] In particular, the following argument is proposed. The concept of truth plays a role in scientific argumentation. The concept of truth plays a fundamental role in formal semantics, for instance, which is part of linguistics, and in the foundations and philosophy of logic. The concept of truth also plays a role in the foundations of mathematics. For instance, it plays a key role in one of the neatest presentations of Predicative Analysis offered by Solomon Feferman [Feferman 1991].[5] Classical logic is the one and only logic that governs scientific reasoning. Therefore classical logic governs the concept of truth.

At least partly in reaction to concerns of this kind, it has been argued that there are *two concepts of truth* [Field 1994], [McGee 2005b]. There is disagreement in the literature about the precise content of these two concepts, which shows that it is not easy to get the intended distinction into sharp focus. Here we give our own take on what the distinction amounts to.

[3]For a survey of the theories studied by the first community, see [Halbach 2011]. For a survey of the non-classical theories, see [Field 2008].

[4]This worry goes back at least to [McGee 1991, Objection 3, p. 102–106].

[5]Similarly, it plays a fundamental role in Aczel's reconstruction of Frege's logicism [Aczel 1980].

The first is the concept of *scientific truth*.[6] This is the concept of truth that is used in scientific theories that are first and foremost concerned with explanations of non-semantic facts; it is governed by classical logic. Scientific truth is for instance employed in trying to understand how 'human beings communicate by language',[7] or to understand which arithmetical statements one ought to accept if one has accepted the basic axioms and rules of elementary arithmetic.[8] The scientific concept of truth is a *theoretical concept*, like the concept of force in classical mechanics, for instance. It is related to our pre-theoretical, ordinary language concept of truth. But there is no reason to think that it does or should coincide with it, just as there is no reason to expect the scientific concept of force to coincide with our pre-theoretical concept of force.

The second is the concept of *disquotational truth*.[9] This notion of truth intends to be a device of full quotation (*semantic ascent*) and disquotation (*semantic descent*). Indeed, a core part of the meaning of the truth predicate is given by inference rules that are known as *Truth-Introduction* and *Truth-Elimination*:

$$\frac{A}{\mathsf{T}\ulcorner A \urcorner} \qquad \frac{\mathsf{T}\ulcorner A \urcorner}{A},$$

where T stands for the truth predicate, and A ranges over sentences that can include T.

The liar paradox teaches us that *if* there is a coherent concept of type-free disquotational truth, then it is governed by non-classical logic. Many philosophers – especially outside the literature on paradox – have argued that *Truth-Introduction* and *Truth-Elimination* are fundamentally correct rules about truth. But our approach will not rely on this: we follow [Wright 2004a] and consider the coherence of the concept of disquotational truth as a *presupposition* in the cognitive project of a truth-theoretic justification of mathematical knowledge.

[6]McGee calls this notion *correspondence truth*; Field calls it *inflationary truth*.

[7]Cf. [McGee 2010, p. 423].

[8]See [Feferman 1991, p. 2].

[9]Field speaks of *deflationary truth*; at times McGee also uses this term. In the literature this notion is often labeled as *transparent truth* or *naive truth*.

It is often intimated that disquotational truth *is* our ordinary language notion of truth. But there really is little evidence to support this. At any rate, we shall not take it to be so in this article.[10]

3 Entitlement to cognitive project

Over the past decades, the distinction between *entitlement* and *justification* has become prominent in epistemology.[11]

The notion of entitlement has been used in certain philosophical accounts of *knowledge transfer*, which appeal to entitlement to rely on basic patterns of reasoning for which one has no justification [Boghossian 2003], [Wright 2004b], [Burge 2011]. The idea is that one can be *justified* in believing a conclusion that one has inferred by means of basic logical steps from a collection of premises for which one has justification, even if one does not have justification for the claim that the basic logical steps are valid.

Consider the following scenario:

> **Antigone** knows (and thus is justified in believing in) a proposition P. From this premise, using the logical pattern of Disjunction Introduction, she infers $P \vee Q$. Antigone does not have sufficient logical training and even does not have a sufficiently rich conceptual repertoire to justify the validity of the rule of $\vee$-*Introduction*.

Boghossian claims that Antigone has an entitlement to *blind* logical reasoning that is knowledge-transferring: in the scenario under consideration, Antigone's reasoning suffices to come to *know* the proposition $P \vee Q$.

Wright defines the notion of *entitlement of cognitive project* along the following lines [Wright 2004a, 191–192]:

[10] The ordinary language concept of truth may perhaps be seen as a *third* truth concept. This concept may be *inconsistent* [Burgess & Burgess 2011].

[11] We assume that the reader is somewhat familiar with this distinction. Two seminal articles are [Burge 1993] and [Wright 2004a].

> . . . an entitlement of cognitive project [. . .] may be proposed to be any presupposition P of a cognitive project meeting the following additional two conditions:
>
> (i) We have no sufficient reason to believe that P is untrue
>
> (ii) The attempt to justify P would involve further presuppositions in turn of no more secure a prior standing. . .

Wright argues that relying on the validity of certain logical rules of inference fulfils the condition for being an entitlement of cognitive project [Wright 2004b, Section IV]. We will not go into the details of Wright's argumentation, but assume for the purposes of our discussion that his account is *basically* correct.

There are, however, a few questions that are left open by Wright's account that turn out to be important for our discussion. First, *which* rules of inference are we entitled to rely on in logical reasoning? Wright argues that Modus Ponens, for instance, is among them. But for many putative such rules (such as Disjunctive Syllogism), he is silent about this question. Second, what is the *strength* of our entitlement to logical reasoning? In particular, are we entitled to rely on logical reasoning involving sentences of any admissible extension of the language that we are currently using? Or are we entitled to use them only for the language that we are currently using, leaving it open that we may not be entitled to rely on them for certain future language extensions? For instance, might one be entitled to rely on classical logic in mathematics but not when the language of mathematics is extended by vague predicates? Another way of putting this is the following. If we agree to regard logical inference rules as *schematic*, then what is the *substitution rule* that we are entitled to use when we instantiate the rules in concrete arguments? These questions will be addressed in due course.

4 The justificatory role of truth

We will now relate the distinction between scientific and disquotational truth to the question to which extent the concept of truth can play a *justificatory role*. We focus on the role that the concept of truth plays in justification in mathematics.

It has been claimed that truth is a *logical* notion.[12] If there is something in this slogan, then one may wonder whether, like for the first-order logical connectives, we can be *entitled* to principles and rules governing the concept of truth without having *justification* for them. If the answer to this question is yes, then truth might be able to play a role in justificatory processes that is similar to the role that logical reasoning plays in them.

4.1 Scientific truth and justification

If the concept of *scientific* truth is understood as a foundational device for empirical sciences – e.g. in giving a good and coherent account of linguistic meaning –, the answer to our question must surely be negative. The theoretical notion of truth is part of a *package*, which is a scientific theory (and we have seen that classical logic is part of this package). The package as a whole is judged, as Quine has taught us, by the extent to which it is *successful* – in giving a good model of communication via truth conditions, for instance. Derivatively, this then also holds for the truth principles or rules which belong to this package. Under this reading, there is no room for *entitlement* (or 'warrant for nothing', in Wright's terms) to theoretical truth principles or rules: they can only be to a smaller or greater extent *justified*. This, however, does not entail that scientific truth cannot play a justificatory role in non-empirical sciences such as logic itself or mathematics.

Theories of truth formulated in classical logic that may be regarded as theories of scientific truth in this broader sense come in *two kinds*. The *first kind* consists of theories

[12]Or rather, at best it is a *logico-mathematical* notion. For a discussion, see for instance [Horsten 2011, Chapter 10].

of truth in which truth is closed under the rules of *Necessitation* and *Co-Necessitation*:

$$\frac{\vdash A}{\vdash \mathsf{T}\ulcorner A \urcorner} \qquad \frac{\vdash \mathsf{T}\ulcorner A \urcorner}{\vdash A}.$$

These rules are *weaker* than the *Truth-Introduction* and *Truth-Elimination* rules, for they require their premises to be *proved*. Consequently, the notion of truth that they describe is not fully transparent: in reasoning under an assumption, for instance, it is not always possible to infer $\mathsf{T}\ulcorner A \urcorner$ from A and vice versa. Nonetheless, such theories express a notion of theoretical truth that most closely approximates a notion of transparent truth. The most famous of such theories is Friedman and Sheard's theory FS [Friedman & Sheard 1987]. The main problem with FS (and its close relatives) is that it does not preserve the intended structure of the truth bearers. It is *ω-inconsistent*, therefore it does not admit models based on the standard natural numbers. Under the assumption that natural numbers are satisfactory bearers of truth modulo isormorphism with suitable syntactic objects, this amounts to saying that FS-like theories do not apply to syntactic objects as we standardly conceive of them. This is a sufficient reason to put this first kind of theories of classical truth aside.

The *second kind* of theory of type-free scientific truth is not closed under Necessitation and Co-Necessitation. Theories of this type can be seen as axiomatisations of certain classes of classical models that result from 'closing off' a fixed point model of the kind described in [Kripke 1975]. The most famous of these theories is Feferman's theory KF, which is obtained by closing Peano Arithmetic under a natural collection of type-free truth principles in which the truth predicate never occurs in the scope of a negation symbol [Feferman 1991].

KF is based on a conception of truth that, in its essential traits, is fundamentally sound. Starting from a truth-free language $\mathcal{L}_0$, KF states that (i) atomic sentences $P(t)$ of $\mathcal{L}_0$ are true iff the value of t belongs to the extension of P, false if it does not; (ii) a disjunction is true iff at least one disjunct is true, false if both disjuncts are false; (iii) an existentially quantified sentence $\exists x\, \varphi$ is true iff $\varphi(t)$ is true for at least one t, false if there is no such t; (iv) a truth ascription $\mathsf{T}\ulcorner A \urcorner$ is true iff A is true, false if A is false. But being formulated in classical logic, KF cannot be completely faithful to the conception of truth that inspires it. Because of the Liar Paradox, the disquotational character of the truth predicate can only be formulated under the scope of the truth predicate: in KF, $\mathsf{T}\ulcorner \mathsf{T}\ulcorner A \urcorner\urcorner$ is equivalent to $\mathsf{T}\ulcorner A \urcorner$, but it is in general *not* the case that $\mathsf{T}\ulcorner A \urcorner$ is equivalent to A. Therefore KF cannot be considered to be a theory of *disquotational* truth. We will see, however, that the compositionality of truth that is encompassed in clauses (i)-(iv) can be fully vindicated.

Theories of theoretical truth do not sit well with what are called *reflection principles*. The most natural such reflection principle is called Global Reflection.[13] A *Global Reflection Principle* for a given theory S says that all theorems of S are *true*.[14] FS, due to its ω-inconsistency, is outright inconsistent if coupled with global reflection. *But also globally reflecting on* KF *results in some form of inconsistency.* This means that in theories of classical truth we cannot consistently hold that what they prove, is true. This entails that scientific theories of truth suffer the same fate, by our assumption that only theories of classical truth can be considered theories of scientific truth.

Actually, as far as KF-like systems are concerned, the situation is somewhat complicated. The most popular variant of KF asserts that there are no true contradictions. Globally reflecting on this system indeed yields inconsistency. But globally reflecting on the version of Feferman's theory KF which neither asserts that there are no true contradictions nor that there are no truth value gaps, results in a theory that is consistent (and classical!); but this theory asserts that there are *true contradictions*, i.e., we end up with a flavour of dialethism.[15]

There is an obvious way in which the above argument to the effect that scientific truth

[13]Other reflection principles (such as the so-called Uniform Reflection Schema) are put aside for the moment because our belief in them *derives* from our belief in Global Reflection.

[14]Formally, Global Reflection is defined as $\forall\sigma(\mathsf{Bew}_S(\sigma) \rightarrow \mathsf{T}(\sigma))$, where Bew_S expresses provability in the theory S in a canonical way, σ is a (code of) a sentence of the language $\mathcal{L}_S$ of S.

[15]For details, see [Fischer et al. 2017, Section 2.3].

cannot play a justificatory role can be challenged. The argument rests on the inference from accepting KF to being committed to the truth of KF. This step rests on Feferman's epistemological claim that unconditional acceptance of a theory S *implicitly commits* one unreservedly to accept reflection for S [Feferman 1962, p. 261].[16] So one way to resist the conclusion of the argument is to reject Feferman's theory of implicit commitment.[17] But is this really a viable way out? Iteration of proof-theoretic reflection principles is considered to be one of the paradigmatic cases of extensions of theories that are *intrinsically* justified on the basis of a trustworthy theory: decades of work on proof-theoretic reflection have not cast any doubt on the reliability of the process of reflection. If unwanted consequences are obtained by extending KF-like theories by reflection, one has to look for other possible sources of trouble before questioning the process of reflection.[18]

In the face of these problems, it has been suggested that we should not accept all of KF, but only those sentences which KF proves to be *true*: the collection of those sentences is called the *inner logic of* KF [Reinhardt 1986]. Much can be said in support of such a policy.[19] But withdrawing to the inner logic of KF means surrendering. The inner logic of KF is not closed under classical logic – but under a non-classical logic that we will later call BDM. Therefore the inner logic of KF cannot capture a concept of truth as it may be used in *scientific explanations*.

4.2 Entitlement to disquotational truth

Now we leave the concept of scientific truth aside and turn to the question whether we have *entitlement without justification* for the rules governing *disquotational* truth.

[16]For a discussion, see [Horsten & Leigh 2017, Section 6].

[17]For arguments to this effect, see for instance [Dean 2015] and [Cieslinski 2017].

[18]Even if Feferman's theory of implicit commitment is rejected, however, there may be other reasons to reject KF-like theories. One of them may be, roughly, that justifying startling conclusions in KF-like systems seems *too easy*. As an example, consider again the version of Feferman's KF that commits itself to there being no true contradictions – which Maudlin takes to be the correct theory of type-free truth [Maudlin 2004]. This theory is philosophically indeed remarkably strong. Stern has recently shown that this theory proves the elusive conclusion of the Lucas-Penrose argument, i.e., that *the human mind is not a formal system* [Stern 2017]. Yet Stern (rightly) does not in any way take this argument as a *justification* of the conclusion that the mind is not a Turing machine.

[19]It is defended in [Soames 1999].

Without being specific, let the minimal theory of disquotational truth consist of *Truth-Introduction*, *Truth-Elimination*, an elementary syntax theory (which is inter-translatable with an elementary arithmetical theory), in a suitable logic.

We have seen that the correct ambient logic for disquotational truth must be non-classical. But there is no agreement about what the correct ambient logic for a theory of transparent truth is. Some advocate a paracomplete logic, others a paraconsistent logic. Here we assume a background logic – that we call *Basic De Morgan logic* (BDM)[20] – which is *neutral* regarding the choice between paracompleteness and paraconsistency. The main feature of our logic, which harmonises perfectly with the conception of truth given on page 9, is that it only involves rules of introduction and elimination for *monotone* connectives: informally, this means that truth values of complex sentences are preserved or 'increased' when we consider other complex sentences whose compounds have truth-values no smaller than the original compounds.

The monotonicity of BDM also explains why it does not take a stance on the existence of truth-values gaps or gluts, which would require at least one of the introduction or elimination rules for negation and implication, which are clearly non-monotone connectives. In particular, a feature that our logical system shares with paracomplete approaches is that the classical logical rule of *Conditionalisation* must be restricted. But no one wants to abandon Conditionalisation *completely*. Typically, the rule of Conditionalisation for material implication is restricted to *truth-determinate* premises as follows [Halbach & Horsten 2006], [Field 2008]:

$$\dfrac{\begin{array}{c} \\ \vdots \\ P \vee \neg P \end{array} \qquad \begin{array}{c} P \\ \vdots \\ Q \end{array}}{P \rightarrow Q}$$

In BDM, Conditionalisation is indeed restricted as indicated above, but the basic structural rules are preserved. Such theories of disquotational truth can be unproblematically

[20]For a description of *Basic De Morgan logic*: see [Fischer et al. 2017].

closed under iterated Global Reflection [Fischer et al. 2017, Section 2.3]. In particular, and in sharp contrast with KF, if one starts out with a theory S that does not *assert* that there are truth value gaps or truth value gluts, then (iterated) Globally Reflecting on S does not result in asserting that there are truth value gaps or gluts.

> Consider **Theano**, who has *not* committed herself to *full* schematic classical logic[21], but only to classical logic for the concepts that she already possesses. She leaves the possibility open that she may acquire concepts that are not governed by full classical logic. However, Theano does not remain completely neutral about the logical rules governing future concepts: she does commit herself to applying the rules of the minimal logic BDM to any future concepts.

Theano is entitled to rely on the rules of logic in the way that she does. Can she go on to acquire an entitlement to rely on the disquotational rules *Truth-Introduction* and *Truth-Elimination*?

One might argue that Theano can introduce a notion of disquotational truth by *stipulation*. The idea would be that, in Theano's situation, we are permitted to introduce a new predicate T, and to stipulate that *Truth-Introduction* and *Truth-Elimination* hold for it. These stipulations are to be seen as *meaning stipulations* or implicit definitions for a newly introduced concept.[22]

It is well-known that we are not always entitled to introduce a new concept by laying down inference rules for it. The stipulations for introducing Prior's *Tonk* [Prior 1960], for example, do not succeed in introducing a concept: we are not entitled to follow these stipulations. Belnap proposes that a condition for introducing a new concept C by stipulation is that the resulting theory is *proof-theoretically conservative* over the C-free fragment of the language [Belnap 1962]. But this is not sufficient to generate the required

[21]The distinction between a fully schematic theory and a schematic theory for a fixed signature is also used in [Feferman 1991].

[22]The suggestion for stipulating the meaning of a truth notion by giving a partial implicit definition of it is not without precedent. Soames, for instance, claims that the meaning of our truth predicate is given by axioms of a system in the KF family, and that these axioms can be taken to be a *partial implicit definition* of the meaning of our notion of truth [Soames 1999].

entitlement. At least we should insist on *semantic conservativeness*, which is a matter of not excluding possibilities.[23] Proof theoretical conservativeness is a weaker requirement than semantic conservativeness. So it seems, *pace* Belnap, that we should at least insist on semantic conservativeness of the proposed stipulation that introduces a new concept. Indeed, for natural choices of non-classical logic, introducing the *Truth-Introduction* and *Truth-Elimination* rules (against the background of a syntax theory) results in a theory that is *semantically conservative* over the truth-free part of the language: every model for the original theory in the original language (which does not contain the truth predicate) can be expanded to a model of the full truth theory [Fischer et al. 2017]. This means that in the context of such a non-classical logic, the introduction of fully disquotational truth does not exclude any possibilities; rather, possibilities are fleshed out more fully with the aid of the notion of disquotational truth.

McGee argues that a compositional notion of disquotational truth can stipulatively be introduced, and that its semantic conservativeness guarantees that this notion of truth can then do justificatory work for us [McGee 2005a, Section 5, p. 94]:

> So what justifies a disquotationalist in accepting the compositional theory of truth?
>
> Again, a one-word answer: [semantic] conservativeness.

But even semantic conservativeness is not enough to guarantee that the stipulatively introduced truth concept can function in justification, as can be seen as follows. Suppose we start with Peano Arithmetic, formulated in the language of arithmetic (without a truth predicate). Now consider a theory S, consisting of the axioms of Peano Arithmetic with the truth predicate not allowed in instances of the induction scheme, and classical logic extended to sentences involving the notion of truth. Moreover, S contains one further axiom:

$$M \vee \exists\varphi\, \neg \mathsf{IND}(\mathsf{T}\varphi),$$

[23]A theory S' in a richer language is semantically conservative over a theory S in the background language iff every model of S can be expanded to a model of S'.

where M is some very strong arithmetical principle (asserting the consistency of ZFC plus a large cardinal axiom, say, such that even the consistency of the resulting theory is not beyond doubt), $\mathsf{IND}(\mathsf{T}\varphi)$ is the instance of the induction scheme for $\mathsf{T}\varphi$ with φ a (code of a) formula of $\mathcal{L}_0$ with one free variable. Then a routine model expansion argument shows that S is semantically conservative over Peano Arithmetic. So by McGee's argument, it is admissible *stipulatively* to introduce the predicate T in this manner. Moreover, it is easy to see that S plus induction for the extended language (including the truth predicate[24]) proves M. Therefore, in Wright's terminology, we *cannot* be entitled to the presupposition given by stipulatively introducing T in this manner. Contrary to the recipe given by Wright (see page 6), in fact, there are good reasons to doubt our presupposition as viciously circular: in the act of justifying M, we presupposed a conditional $A \rightarrow M$ with an antecedent A that is clearly true, given our other presupposition on the presence of truth in induction.[25]

In sum, model-theoretic conservativeness – let alone proof-theoretic conservativeness – is not sufficient to underwrite an entitlement to rely on reasoning principles governing an introduced notion. But the concept of disquotational truth is not affected by the problems just sketched, and we may embark in an epistemologically blameless way on a new cognitive project of justifying mathematical knowledge by presupposing the validity of the rules of *Truth-Introduction* and *Truth-Elimination*.[26] If all is well, then there is a fully disquotational truth concept, governed by non-classical logical rules, that we are entitled to rely on in our reasoning —it is just that conservativeness cannot serve as a guarantee that all is well. Nonetheless, we can be entitled to rely on the validity of *Truth-Introduction* and *Truth-Elimination* in our arguments without having *justification* for it. This absence

[24]McGee holds that our commitment to mathematical induction is open-ended, and, anyway, reflection principles to which we are implicitly committed take one from induction without the truth predicate to mathematical induction for the whole language including the truth predicate.

[25]A similar but more sophisticated example involves the axioms of the theory of truth KF discussed above: they are model theoretically conservative in the absence of full induction, but remarkably stronger in its presence.

[26]Belnap ([Belnap 1962]) would emphasise that because the rules introducing disquotational truth do not pin down the reference of T uniquely, we have not introduced *the* concept of disquotational truth, but only *a* concept of disquotational truth, which is fair enough.

of justification for our truth rules does not prevent us from gaining knowledge of the conclusions we reach by relying on them and also claiming knowledge for them.

We will argue in the following section that such a disquotational truth concept is indeed suitable for playing a key role in genuine justificatory processes in mathematics.

5 Truth and mathematics

We now finally explain how disquotational truth can play a justificatory role in the foundations of mathematics.

> **Hypatia** works in the foundations of mathematics. Her epistemic commitments are like those of Theano, except that she is in addition happy to rely on full disquotational truth in her reasoning. She is persuaded that at least a portion of arithmetic can be fully justified. Like Hilbert and many others, she seeks to extend her justification of arithmetic to richer areas of mathematics.

We can be a bit more specific and suppose that Hypatia is justified in believing in the truth of a weak fragment of Peano Arithmetic. The question of what exactly the principles are that characterise Hypatia's commitments should not bother us too much. For our purposes, she might regard Primitive Recursive Arithmetic PRA as acceptable, or she might consider primitive recursion, and even exponentiation as non-acceptable operations (like [Parsons 1998], for instance) and therefore opt for weaker systems such as *Elementary Arithmetic* EA or one of the sub-exponential arithmetical systems such as $\mathsf{I}\Delta_0 + \Omega_1$ or Buss' S^1_2 (see [Hájek and Pudlák 1998]) respectively. The details of these systems do not matter: what matters is that Hypatia can freely choose a very weak arithmetical theory as her basic standard for mathematical justification: we only assume that the weaker she goes, her commitments become more and more uncontroversial. The question now before us is: from her epistemic vantage point, can Hypatia come to be justified in believing a portion of mathematics that non-trivially surpasses her initial commitments?

For the sake of definiteness, let's assume that Hypatia's justified arithmetical beliefs contain the axioms of EA. As for her logical background, she is entitled to believe in the validity of the principles of BDM logic in a fully schematic form so that she knows any arithmetical sentence that can be seen to follow from the axioms of EA. Now she is warranted to introduce a notion of disquotational truth. As we have seen, she cannot justify the validity of the rules of *Truth-Introduction* and *Truth-Elimination*; nonetheless, she is entitled to embark on a cognitive project that involves adopting them. Thus Hypatia comes to accept the theory EA formulated in the language expansion with a truth predicate and closed under BDM logic and the disquotational rules for truth: call this theory TS_0. When she does so, she is *implicitly* committed to believing that all the theorems of this theory are true.[27] She comes to accept the stronger theory obtained by reflecting on the basic disquotational theory TS_0. If all is well, she is *entitled* to believe in the validity of reflection rules for TS_0. If she does, then she is justified in believing all the mathematical theorems of this stronger theory. Moreover, she is now implicitly committed to believe in the truth of everything that the extension of TS_0 with reflection proves. Hypatia is then again entitled to the reflection principles for the stronger theory and justified in accepting all the (mathematical) theorems of a further iteration of reflection over TS_0.

As already mentioned, there are several reflection principles discussed in the literature. The most natural candidate in a setting with the truth predicate is global reflection in the form $\forall\sigma(\mathsf{Bew}_T(\sigma) \to \mathsf{T}\sigma)$.[28] This reflection principle talks about theorems and is sufficient for a classical setting.[29] In our non-classical BDM-setting we lack an object linguistic conditional to transform provable sequents into theorems. Therefore we have additional principles to take into account. The first one is a reflection principle for

[27] In [Tait 1981, p. 545], Kreisel's analysis of finitism is criticised on the grounds that a finitist cannot recognise the validity of PRA because she cannot rely on the notion of function. But our situation is different. We do not claim that Hypatia is a finitist and so she can come to accept the theory TS_0 in full generality.

[28] An alternative is the scheme of uniform reflection $\mathsf{Bew}_T(\ulcorner A\dot{x}\urcorner) \to A(x)$, for all formulas $A(v)$.

[29] In a classical setting we have the expressive resources to rewrite a sequent $\Gamma \Rightarrow \Delta$ as an equivalent theorem $\Rightarrow \bigwedge\Gamma \to \bigvee\Delta$.

provable sequents:

$$(\mathsf{r}_T) \qquad \frac{\Rightarrow \mathsf{Pr}_T(\ulcorner \Gamma \Rightarrow \Delta \urcorner)}{\Gamma \Rightarrow \Delta}$$

It states that if our background theory can formally recognise that the sequent $\Gamma \Rightarrow \Delta$ is provable in T, then this sequent holds. The second reflection rule involves admissible rules:[30]

$$(\mathsf{R}_T) \qquad \frac{\Rightarrow \mathsf{Pr}^2_T(\ulcorner \Gamma \Rightarrow \Delta \urcorner, \ulcorner \Theta \Rightarrow \Lambda \urcorner) \qquad \Gamma \Rightarrow \Delta}{\Theta \Rightarrow \Lambda}$$

It states that if we can formally recognise that the rule ${}^{\Gamma\Rightarrow\Delta}/{}_{\Theta\Rightarrow\Lambda}$ is admissible in T and $\Gamma \Rightarrow \Delta$ holds, also $\Theta \Rightarrow \Lambda$ holds. It is clear that the first principle can be derived from the second, and it is in fact the latter that will be mostly employed to unfold Hypatia's commitments in what follows.

Unlike the classical theories *à la* KF discussed in section §4.1, there is no incoherence resulting from extending a fully disquotational theory of truth such as TS_0 with a global reflection principle. Therefore the formulation of the rules (R_T) and (r_T) in schematic form (uniform) and without explicit mention of the notion of truth can be regarded as a *technical* choice and not as a *conceptual* one; by contrast, in a classical setting such as the one considered in [Horsten & Leigh 2017] in which one iterates reflection over theories that are not fully disquotational, uniform reflection rules such as (r_T) and global reflection rules provably come apart. As we have already discussed in §4.1, the latter forces internal or external contradictions, whereas the former do not.[31] In fact, as we shall now see, iteration of application of these rules starting from TS_0 yields mathematically sound systems that are interesting both from a truth-theoretic and from a proof-theoretic perspective.

[30]Here the two-place provability predicate $\mathsf{Pr}^2_T(\ulcorner \Gamma \Rightarrow \Delta \urcorner, \ulcorner \Theta \Rightarrow \Lambda \urcorner)$ represents the fact that it is admissible in T to infer $\Theta \Rightarrow \Lambda$ from $\Gamma \Rightarrow \Delta$. For details see [Fischer et al. 2017].

[31]Actually in the context of fully disquotational truth one can even show that suitable rules of global and uniform reflection coincide: see [Fischer et al. 2017, Prop. 1].

The compositional conception of truth introduced on page 9 can now be fully recovered by Hypatia in two iterations of the rule R_T over the basic disquotational theory TS_0, a theory that we call $\mathsf{R}^2(\mathsf{TS}_0)$. In fact, there is a sense in which $\mathsf{R}^2(\mathsf{TS}_0)$ can achieve even more than what KF has to offer.[32] In the Kripke-Feferman theory the truth predicate *cannot* commute with negation: the principle

$$\forall\sigma : \mathsf{T}(\neg\sigma) \leftrightarrow \neg\mathsf{T}\sigma,$$

where σ is again a variable ranging over (the code of) a sentence of $\mathcal{L}_\mathsf{T}$, cannot be added to KF *sine contradictione*. $\mathsf{R}^2(\mathsf{TS}_0)$ on the other side contains commutation principles for negation. The price is that this commutativity is only obtained in *rule-form* via the sequents

$$\mathsf{T}(\neg\sigma) \Rightarrow \neg\mathsf{T}\sigma, \qquad\qquad \neg\mathsf{T}\sigma \Rightarrow \mathsf{T}(\neg\sigma).$$

If both the coherence of Hypatia's entitlement to reflection and the presence of general forms of compositionality amount to compelling evidence that her cognitive project is acting on the right presuppositions, what still remains to be seen is how these new truth theoretic principles can lead to her new arithmetical knowledge. We first notice that the theory $\mathsf{R}(\mathsf{TS}_0)$ obtained by closing TS_0 under R_T also gives us the *full induction schema* for the language $\mathcal{L}_\mathsf{T}$. Moreover, we obtain by an additional reflection step the principle of transfinite induction up to and including the ordinal ω^ω stating that a property $P(x)$ expressed by a predicate of $\mathcal{L}_\mathsf{T}$ holds of all ordinals smaller than ω^ω if it can be naturally iterated over a standard well-ordering of the ordinals, i.e., if when $P(x)$ holds for all ordinals $\beta < \alpha$, $P(x)$ also holds of α.[33] This is already *more than* what full Peano arithmetic formulated in the language $\mathcal{L}_\mathsf{T}$ can give: by a result of [Halbach & Horsten 2006], it can only prove transfinite induction for $\mathcal{L}_\mathsf{T}$ up to any ordinal of the form ω^n, with n a natural

[32] See [Fischer et al. 2017, Lemma 4].
[33] For details, see [Fischer et al. 2017, Proposition 3].

number.

Still, however, there is no clear indication of how Hypatia's presupposition of disquotational truth might substantially contribute to her cognitive project of justifying mathematical claims: after all it is well-known that the step between EA and PA, even when formulated in the expanded language, can be obtained by means of reflection alone. But the combination of transfinite induction and full compositional truth just introduced and that one can reach in iterations of R_T over TS_0 enables her to significantly exceed the mathematical content of Peano Arithmetic. In particular Predicative Analysis, as characterised by [Feferman 1964], for instance, can be seen as a hierarchy of comprehension principles over Peano Arithmetic of the form

$$(\mathsf{CA}^{\alpha}) \qquad \exists X^{\alpha} \forall u (u \in X^{\alpha} \leftrightarrow B(u))$$

where α is smaller than the Feferman-Schütte ordinal Γ_0 and where $B(x)$ is a formula that can contain quantification only over sets of level $\beta < \alpha$. (CA^{α}) essentially enables us to define sets of natural numbers via previously defined ones and without quantifying over totalities yet to be defined.

Now full compositional truth and transfinite induction up to an ordinal α enable us to recover (CA^{α}) for all $\beta < \alpha$ via a hierarchy of typed truth predicates:[34] the basic idea is that the set X^{α} is interpreted as (the code of) an arithmetical formula $\varphi(x)$ with one free variable and $x \in X^{\alpha}$ as $\mathsf{T}\varphi(x)$ (see for instance [Feferman 1991]). Therefore this general pattern yields that in accepting EA, the logic BDM and full disquotational truth, Hypatia is entitled to accept all consequences of $\mathsf{R}^2(\mathsf{TS}_0)$ that, as we have just seen, include iterations of (CA^{α}) up to ω^{ω}. And the latter theory is substantially stronger than Peano Arithmetic.

Two iterations of the rule R_T, however, is not the end of Hypatia's implicit commitment. She goes on repeatedly to reflect on her commitments. This results in her accepting

[34] For a general approach on how to obtain typed truth predicates in theories of disquotational truth in BDM we refer to [Nicolai 2017].

ever larger fragments of Predicative Analysis. Eventually, she accepts all of it.[35] She has moved from a very modest almost finitist position in the foundations of mathematics to a full-blown Predicativist position.

If all is well, then the result of this process is Hypatia *knowing* the theorems of Predicative Analysis in the 'true justified belief' sense, where 'all is well' means that Hypatia was *justified* in her belief in EA in the first place, is *entitled* to rely on BDM logic, is *entitled* to rely on the inference rules that govern the concept of disquotational truth, and is *entitled* to rely on the reflection rule R_T for a suitable T. The concept of disquotational truth plays a crucial role in this process: the nominalising function of disquotational truth (semantic ascent) allows formulas to be treated as objects (sets) that can be quantified over.

The reflective process that we have described is not the way in which the bounds of mathematical knowledge are *typically* explicitly extended. Attempts in the foundations of mathematics to extend these bounds often invoke 'strong principles of infinity', or, alternatively, strong combinatorial principles. Such principles, if they can be justified, extend the limits of mathematical knowledge in much more dramatic ways than iterated reflection does.

Thus there are also other ways in which Hypatia may come to accept Predicative Analysis. For instance, she may straightaway, i.e., without going through the iterative reflection process described above, acquire a belief in Zermelo-Fraenkel set theory, perhaps by coming to understand and accept a version of the iterative conception of set. If ZF can indeed be justified from the iterative conception, then Hypatia can in this way come to know a mathematical theory that is much stronger than Predicative Analysis. This way of extending the scope of our mathematical knowledge differs structurally from extension by reflection. In order to accept a new axiom (strong principle of infinity, combinatorial principle), we need to do justificatory work, whereas no new justification is needed to adopt the global reflection principle for a theory that you are already justified to believe

[35] In [Fischer et al. 2017] it has been observed that the reflection process can be carried on, such that $\mathsf{R}^{\omega_n+1}(\mathsf{TS}_0) \vdash \mathsf{TI}_{\mathcal{L}_T}(\omega_n)$. Following Feferman's strategy of autonomous progressions we arrive at $\mathsf{TI}_{\mathcal{L}_T}(<\Gamma_0)$.

in. The global reflection principle for a theory is *exactly as evident* as the theory itself.

The reflective process that we have described in this section is not restricted to finitist arithmetic but also applies (in essentially the same way) to stronger theories. In particular, it applies to our most encompassing justified mathematical theory.[36] In this way, disquotational truth plays not just a justificatory role in the foundations of mathematics, but even a *foundational* role: however many principles of infinity we have come to accept, we are always implicitly committed to more than what can logically be derived from them.

6 Claiming knowledge

The conclusion of the foregoing is that Hypatia can, from a starting point where she is justified in believing the consequences of a weak theory of arithmetic, by reflection and relying on disquotational truth, arrive at justified belief in Predicative Analysis. By Wright's lights ([Wright 2004b, section VIII]), if Hypatia can in addition come to *know* that she is justified in believing in elementary arithmetic, then she can come to *know* that she knows what follows from the axioms of Predicative Analysis, i.e., she can "claim knowledge" of theorems of Predicative Analysis.

Before we tackle the question of justifying the logical laws we address a possible worry. According to Glanzberg, type-free truth predicates face a problem with explaining away strengthened liar reasoning: by the very lights of Hypatia's truth theory, the reflection principle $\mathsf{R}_{\mathsf{TS}_0}$ should already be part of her truth theory TS_0, and thereby her position is unstable. Glanzberg's argument goes as follows. The theory TS_0 is silent about the truth value of the liar sentence, and therefore does not classify it as true. The type-free truth theorist tries to dissolve the threat of strengthened liar reasoning by emphasising that only what is asserted by her truth theory is to be taken as true. So, in particular, the statement that the liar sentence is not true, should not itself be taken to be true. So

[36] *Pace* [Dean 2015, p. 56]. For details of how this phenomenon generalises in straightforward ways to stronger theories such as second-order number theory or set theory, see [Fujimoto 2012].

for Hypatia, the reflection principle is part of the explanation of what goes wrong in the strengthened liar reasoning. But then, Glanzberg says [Glanzberg 2004, p. 294]:[37]

> this principle [$\mathsf{R}_{\mathsf{TS}_0}$] must be properly assertible. The norms of assertion require us to only assert what we take to be true. But by the very view being offered, the only ground for truth there can be is the provability of truth in [TS_0]. Hence, the explanation *requires* the provability of truth of [$\mathsf{R}_{\mathsf{TS}_0}$] in [$\mathsf{TS}_0$] for the explanation to be acceptable by its own lights.

Of course, for familiar Gödelian reasons, TS_0 cannot contain $\mathsf{R}_{\mathsf{TS}_0}$.

Against this, we maintain that Hypatia is not forced to accept that 'the only ground for truth there can be is the provability of truth in a theory', i.e., she does not and should not believe that *only* what is proved by TS_0, is acceptable. Indeed, she implicitly has the resources for acquiring more truths: she is implicitly committed to $\mathsf{R}(\mathsf{TS}_0)$, and this goes beyond the explicit content of TS_0.

Later in his paper Glanzberg acknowledges the possibility of reflection as being only implicit in the formulation, but he takes this to reveal the hierarchical nature of truth although he also maintains that it is still the same concept. So we agree with Glanzberg that the closure under reflection principle is crucial, but we disagree that this is only available for hierarchical approaches. In the end, reflection as implicit commitment is perfectly compatible with Hypatia's silence about the truth value of the liar sentence.

The question whether and how Hypatia can also come to know that her logical inference rules are valid, and that the reflection process and the disquotation truth rules are reliable, is more delicate. According to [Wright 2004b], in order for Antigone to know, for instance, that an instance of the Disjunction Introduction rule is valid, she would have to prove the corresponding material conditional. Clearly such proofs will typically be circular, but, according to Wright, not viciously so [Wright 2004b, section VIII, p. 173].

Antigone, however, has only signed up unrestrictedly to BDM logic. This means that

[37] Glanzberg is focusing on a type-free truth theory that is somewhat different from TS_0, but this does not affect the argument.

Conditionalisation is not unrestrictedly available to Antigone (or Hypatia): so she is not able, according to Wright, to claim the validity of Disjunction Introduction. Similar remarks hold, *mutatis mutandis*, for the reliability of reflection and of the disquotational truth rules.

What Hypatia can do, is to prove the reliability of Conditionalisation for the arithmetical instances of her inference rules. Moreover, she can do this, using the truth predicate, in a *uniform* manner. For instance, Hypatia can show:

$$\forall \sigma, \tau \in \mathcal{L}_0 : \mathsf{T}(\sigma) \rightarrow \mathsf{T}(\sigma \vee \tau).$$

This also works for all the other rules used in the classical mathematical theory of Predicative Analysis.

We do not have to stop here: we can step by step expand the range of sentences for which the classical rules are provably valid. Basically, we can prove more and more sentences to be *grounded*. The fragment of the language for which we can do this corresponds to the amount of transfinite induction for the language containing the truth predicate is provable. We will have Conditionalisation exactly for these initial segments of the minimal fixed point.

7 Two cognitive projects

We have discussed the entanglement of two types of cognitive project: one about logic, another about truth.

The first project, in its boldest form, involves the *acceptance of full classical logic* in open-ended schematic form. Some have argued that we are entitled to rely on, and *must* rely on, particular unrestricted inference rules governing particular logical concepts because we could not have the concepts without relying on the rules [Boghossian 2003]. In particular, our entitlement to rely on Conditionalisation has been defended in those terms. However, there are strong reasons for rejecting this line of reasoning. For any logical concept and any logical rule governing it, it is possible to understand the concept without accepting the logical inference rule [Williamson 2003]. It seems then that acceptance of logical inference rules is never *inevitable*; one can never be *forced* to do so. But this does not mean that one ought not to fully engage in this cognitive project: we may well be entitled to accept, in an epistemologically blameless way, full classical logic in open-ended schematic form.[38]

The second project consists in *fully embracing a notion of type-free disquotational truth*. We have argued that from a place where one has not yet signed up to Conditonalisation in open-ended schematic form, one can come to accept such a notion.

Nonetheless, the two cognitive projects clash with each other. One cannot fully rely on classical material implication and on type-free disquotational truth at the same time —even though there is no problem whatsoever in *understanding* both concepts at the same time. Having signed up to one of these two projects, one can of course always reconsider, retrace one's steps, and embark on the other project instead. But it is impossible to *exercise* or *practice* both concepts at the same time. In this light it might be interesting to investigate, in more detail than has been done so far, the *relations* between cognitive projects in general (and the entitlements that go with them).

It does not follow from anything that we have said that one of the two projects is somehow flawed, or epistemologically blameworthy. It is just that engaging in a cognitive project imposes limitations: *choosing is losing*. Both the practicer of material implication and the user of disquotational truth may well have their own 'warrants for nothing'. But if you open-endedly rely on classical material implication, then you cannot also use full disquotational truth. If you rely on a notion of full disquotational truth, then you cannot fully rely on the inference patterns of material implication.

Both cognitive projects have their benefits and drawbacks. On the one hand, the absence of full Conditionalisation undeniably makes mathematical argumentation cum-

[38]This question is too large for us to tackle in this article.

bersome and restricts its power, whereas mathematically reasoning in classical logic is perfectly natural.[39] On the other hand, the concept of scientific truth can do no justificatory work in the foundations of mathematics, whereas we have argued that Hypatia's silences allow her to acquire a fully disquotational truth concept which can do justificatory work for her in the foundations of mathematics.

References

[Aczel 1980] P. Aczel, Frege Structures and the Notions of Truth and Proposition, in J. Barwise and H. J. Keisler and K. Kunen (eds.), *The Kleene Symposium*, North-Holland, 1980.

[Beall & Armour-Garb 2005] J.C. Beall & B. Armour-Garb (eds.) *Deflationary truth.* Open Court, 2005.

[Belnap 1962] Belnap, N. *Tonk, plonk and plink.* Analysis **22**(1962), p. 130–134.

[Boghossian 2003] Boghossian, P. *Blind reasoning.* Proceedings of the Aristotelian Society Supplemental Volume LXXVII, p. 225–248.

[Burge 1993] Burge T. *Content preservation.* Philosophical Review **102**(1993), p. 457–488.

[Burge 2011] Burge, T. *Self and self-understanding.* The Journal of Philosophy **108**(2011), p. 287–383.

[Cieslinski 2017] Cezary Cieslinski, *The Epistemic Lightness of Truth: Deflationism and its Logic*, Cambridge University Press, 2017.

[Dean 2015] Dean, W. *Arithmetical reflection and the provability of soundness.* Philosophia Mathematica **23**(2015), p. 31–64.

[Burgess & Burgess 2011] Burgess, A. & Burgess J. *Truth.* Princeton University Press, 2011.

[39]See [Halbach 2011, chapter 20], [Halbach & Nicolai 2017], [Nicolai 2017].

[Feferman 1962] Feferman, S. *Transfinite recursive progressions of axiomatic theories.* Journal of Symbolic Logic **27**(1962), p. 259–316.

[Feferman 1964] Feferman, S. *Systems of predicative analysis.* Journal of Symbolic Logic **29**(1964), p. 1–30.

[Feferman 1991] Feferman, S. *Reflecting on incompleteness.* Journal of Symbolic Logic **56**(1991), p. 1–49.

[Feferman 2008] Feferman, S. *Axioms for determinateness and truth.* Review of Symbolic Logic **1**(2008), p. 204–217.

[Field 1994] Field, H. *Deflationist views of meaning and content.* Mind **103**(1994), p. 249–285.

[Field 2008] Field, H. *Saving truth from paradox.* Oxford University Press, 2008.

[Fischer et al. 2017] Fischer, M., Nicolai, C., & Horsten, L. *Iterated reflection on full disquotational truth.* Online First version: https://doi.org/10.1093/logcom/exx023

[Franzén 2004] Franzén, T. *Inexhaustibility*, ASL Lectures Notes in Logic, CRC Press.

[Friedman & Sheard 1987] Friedman, H. & Sheard, M. *An axiomatic approach to self-referential truth.* Annals of Pure and Applied Logic **33**(1987), p. 1–21.

[Fujimoto 2012] Fujimoto, K. *Classes and truths in set theory.* Annals of Pure and Applied Logic, **163**(2012), p. 1484–1523.

[Glanzberg 2004] Glanzberg, M. *Truth, reflection, and hierarchies.* Synthese **142**(2004), p. 289–315.

[Hájek and Pudlák 1998] Hájek, Petr and Pudlák, Pavel. *Metamathematics of First–Order Arithmetic*. Springer Verlag, 1998.

[Halbach 2011] Halbach, V. *Axiomatic theories of truth.* Cambridge University Press, 2011.

[Halbach & Horsten 2006] *Axiomatizing Kripke's theory of truth.* Journal of Symbolic Logic **71**(2006), p. 677–712.

[Halbach & Nicolai 2017] Halbach, V. & Nicolai, C. *On the cost of non-classical logic.* Journal of Philosophical Logic, to appear.

[Horsten 2011] Horsten, L. *The Tarskian Turn. Deflationism and axiomatic truth.* MIT Press, 2011.

[Horsten & Leigh 2017] Horsten, L. & Leigh, G. *Truth is simple.* Mind **216**(2017), p. 195–232.

[Horwich 1998] Horwich, P. *Truth.* Second edition. Clarendon Press 1998..

[Kripke 1975] Kripke, S. *Outline of a theory of truth.* Journal of Philosophy **72**(1975). Reprinted in [Martin 1984, p. 53–81].

[Martin 1984] Martin, R. (ed). *Recent essays on truth and the liar paradox.* Oxford University Press, 1984.

[Maudlin 2004] Maudlin, T. *Truth and paradox. Solving the riddles.* Oxford University Press, 2004.

[McGee 1991] McGee, V. *Truth, Vagueness, and Paradox.* Hackett, 1991.

[McGee 2005a] McGee, V. *Two conceptions of truth? — Comment.* Philosophical Studies **124**(2005), p. 71–104.

[McGee 2005b] McGee, V. *Afterword: Trying (with limited success) to demarcate the disquotation-correspondence intuition.* In [Beall & Armour-Garb 2005, p. 142–152].

[McGee 2010] McGee V., *Field's logic of truth.* Philosophical Studies **147**: 421–432, 2010.

[Nicolai 2017] Nicolai, C. *Provably true sentences across axiomatizations of Kripke's theory of truth.* Studia Logica, online first, https://doi.org/10.1007/s11225-017-9727-y.

[Parsons 1998] Parsons, C. (1998). *Finitism and intuitive knowledge.* In M. Schirn (Ed.), *The philosophy of mathematics today* (pp. 249–270). Oxford: Oxford University Press.

[Prior 1960] Prior, A. *The runaway inference ticket.* Analysis **21**(1960), p. 38–39.

[Reinhardt 1986] Reinhardt, W. *Some remarks on extending and interpreting theories with a partial predicate for truth.* Journal of Philosophical Logic **15**(1986), p. 219–251.

[Soames 1999] Soames, S. *Understanding truth.* Oxford University Press, 1999.

[Stern 2017] Stern, J. *Proving that the Mind is not a Machine*? To appear in *Analysis.*

[Tait 1981] Tait, W. *Finitism.* Journal of Philosophy **78**(1981), p. 524–546.

[Williamson 2003] *Understanding and inference.* Proceedings of the Aristotelian Society Supplemental Volume LXXVIII, p. 249–293.

[Wright 2004a] Wright, C. *Warrant for nothing* (*and foundations for free*)? Proceedings of the Aristotelian Society Supplemental Volume LXXVIII, p. 167–212.

[Wright 2004b] Wright, C. *Intuition, entitlement and the epistemology of logical laws.* Dialectica **58**(2004), p. 155–175.