Another Look at Reflection

Martin Fischer¹

Received: 31 July 2019 / Accepted: 21 December 2020 / Published online: 10 March 2021
© The Author(s) 2021

Abstract

Reflection principles are of central interest in the development of axiomatic theories. Whereas they are independent statements they appear to have a specific epistemological status. Our trust in those principles is as warranted as our trust in the axioms of the system itself. This paper is an attempt in clarifying this special epistemic status. We provide a motivation for the adoption of uniform reflection principles by their analogy to a form of the constructive ω -rule. Additionally, we analyse the role of informal arithmetic and the conception of natural numbers as an inductive structure, also with regard to extra conceptual resources such as a primitive truth predicate.

1 Background

Ever since Gödel's famous incompleteness results there has been an ongoing discussion about the status of reflection principles. On the one hand these principles are independent of the underlying theory. On the other hand they seem to be on a par with its theorems when it comes to questions of acceptability.

To get a better understanding of the issues involved it proves fruitful to distinguish formal arithmetical theories from informal or intuitive arithmetic. Whereas the former are understood as recursively enumerable sets of sentences closed under a notion of derivability, the latter is supposed to capture our, possibly implicit, mathematical knowledge.

Consider a formal arithmetical theory Σ , such as Peano arithmetic, PA, as a first approximation to capture explicitly our arithmetical knowledge. Under the assumption that PA is contained in our mathematical knowledge, we also consider it to be sound. Therefore it is reasonable to understand the soundness of PA as also part of our mathematical knowledge. On this understanding it appears warranted to expand our theory PA with the explicit reflection principles expressing this soundness. This relation has sometimes been claimed to be a form of implicit commitment, i.e. if somebody is

✉ Martin Fischer
M.Fischer@lrz.uni-muenchen.de

¹ Ludwig-Maximilians-Universität München, Munich, Germany

willing to accept an arithmetical theory (as sound), then she is thereby also committed to accept reflection principles for this theory.

In the following, we will focus on the epistemic issues involved in motivating the idea that we are warranted to expand our basic formal theories of arithmetic by suitable proof-theoretic reflection principles. However, we cannot avoid touching other issues as well that have been influential in the discussion.

In the remainder of the introduction, we will first provide a short historical overview of the discussion. Then we will distinguish several questions and point out the focus of this paper, which will be an argument for the acceptability of uniform reflection for arithmetical theories under specific circumstances. In the second section we prepare the ground for our proposal and take a look at our reasons to accept weak forms of arithmetical reflection in an arithmetical setting. We sketch the background assumptions involved and the mathematical conceptions in which reflection is most naturally acceptable. In the third section, we will discuss one of the central principles of reflection, uniform reflection. We will provide an original form of motivation that builds on an understanding of uniform reflection as a formalised and constructive ω -rule. The fourth section is then concerned with the underlying conception of natural numbers as a structure in which induction holds in an open-ended fashion. We will provide an informal argument, why under such a conception adding uniform reflection is warranted. The last section contains an outlook on pursuing this reflection strategy by allowing the extra conceptual resources not only in the informal part, but by introducing the relevant metamathematical notions, such as truth within the formal theories and reflections on those.

1.1 Historical Overview

Gödel's incompleteness theorems have several important consequences. Not only do they refute Hilbert's program(s) at least in its original form, but they also show that it is impossible to formulate a formal system that is axiomatizable and correct and that contains all of the correct arithmetical statements. So according to the second incompleteness theorem it appears impossible to prove the consistency of infinitary mathematics with finitary means. However, there are also positive aspects of Gödel's theorems, as they give rise to natural expansions of theories.

At the time it was quite surprising that the incompleteness already arises at the level of Π_1^0 -sentences, i.e. sentences with a single universal quantifier, such as the consistency statement. Those independent Π_1^0 -sentences are in several respects of interest. They have the form $\forall x B(x)$, where $B(x)$ is a formula with one free variable and no unbounded quantifiers. In a sense they are natural universal versions of all its substitution instances $B(\bar{n})$ for all $n \in \mathbb{N}$. The Π_1^0 -sentences are closer to being finitarily acceptable than their counterparts, the Σ_1^0 -sentences, sentences with a single existential quantifier in front. Another disanalogy is that for the Σ_1^0 -sentences completeness is possible, basically every sufficiently strong arithmetical theory is Σ_1^0 -complete. Combining these two facts one realizes that the relevant independent statements $\forall x B(x)$ cannot be false, in cases where the formal system is correct. Moreover, if all the instances $B(\bar{n})$, for all $n \in \mathbb{N}$ are derivable, but not their universal closure, then there

are good reasons to believe in the correctness of the independent principles. If the starting system is correct, then so will be the expansion of the system by its consistency statement.

Gödel in his Gibbs lectures interprets his theorem as follows:

It is *this* theorem [the second incompleteness theorem] which makes the incompleteness of mathematics particularly evident. For, *it makes it impossible that someone should set up a certain well-defined system of axioms and rules and consistently make the following assertion about it: All of these axioms and rules I perceive (with mathematical certitude) to be correct, and moreover I believe that they contain all of mathematics.* If someone makes such a statement he contradicts himself. For if he perceives the axioms under consideration to be correct, he also perceives (with the same certainty) that they are consistent. Hence he has a mathematical insight not derivable from his axioms. (Feferman et al. 1995, p. 309)

In his work on axioms of infinity he states it even clearer¹:

Let us consider, e.g., the concept of demonstrability. It is well known that, in whichever way you make it precise by means of a formalism, the contemplation of this very formalism gives rise to new axioms which are exactly as evident and justified as those with which you started, and that this process of extension can be iterated into the transfinite. (Feferman et al. 1990, p. 151)

In this paper, we try to argue in a novel way that for a certain conception of natural numbers it is warranted to expand a correct formal theory of arithmetic by suitable reflection principles. Whereas the argument will be innovative, it heavily relies on the work of Feferman, who was very much interested in questions of reflection and also sketched an interesting conception of natural numbers. Feferman characterizes reflection principles in the following way:

By a reflection principle we understand a description of a procedure for adding to any set of axioms A certain new axioms whose validity follow from the validity of the axioms A and which formally express, within the language of A , evident consequences of the assumption that all the theorems of A are valid. (Feferman 1962, p. 274)

We will consider four versions of reflection, distinguishing local from global and conditionals from rules. Starting with a suitable formal theory of arithmetic Σ (possibly EA, PRA, PA) containing enough arithmetic to develop a reasonable theory of syntax, we employ standard proof and provability predicates, proof_σ and Pr_σ , with a standard coding and with an adequate Σ_1 -representation σ of the axioms of Σ . The formal consistency statement con_σ is an abbreviation of $\neg \exists x \text{proof}_\sigma(\ulcorner 0 = 1 \urcorner, x)$ or $\neg \text{Pr}_\sigma(\ulcorner 0 = 1 \urcorner)$.

¹ In his remarks in 1972 Gödel is primarily concerned with large cardinal axioms motivated by set theoretic reflection principles. There he understands the set theoretic axioms of infinity as a way of deciding more and more Π_1^0 -sentences and not the continuum hypothesis as originally envisioned.

Local reflection principle:

$$\text{Pr}_\sigma(\ulcorner A \urcorner) \rightarrow A \quad (\text{Rfn}_\Sigma)$$

Uniform reflection principle:

$$\forall x (\text{Pr}_\sigma(\ulcorner A \dot{x} \urcorner) \rightarrow A(x)) \quad (\text{RFN}_\Sigma)$$

Local reflection rule:

$$\frac{\vdash \text{Pr}_\sigma(\ulcorner A \urcorner)}{\vdash A} \quad (\text{Rfn}_\Sigma^r)$$

Uniform reflection rule:

$$\frac{\vdash \text{Pr}_\sigma(\ulcorner A \dot{x} \urcorner)}{\vdash A(x)} \quad (\text{RFN}_\Sigma^r)$$

Besides the technical questions concerned with reflection, where a lot of progress has been achieved, we focus on some more philosophical questions that are still unresolved. These questions are interdependent and in the following we will state some of the problematic issues that we address in this paper.

1.2 Status of the Reflection Principles

The first issue concerns questions about the interpretation of reflection principles. Reflection principles are sometimes understood as making explicit our implicit commitments in accepting a theory. There are different versions, how to understand the role of reflection principles. In a radical formalist reading they are just independent statements from the point of view of a formal system Σ . Only the knowledge of the formal system is not sufficient for knowledge of the reflection principles. In this form there seems to be no pressure whatsoever to accept the reflection principles as correct.

In a second version we are able to see the correctness of the reflection principles, but in order to see this we need a justification. The principles should be derivable in stronger formal theory Σ^+ , which is itself acceptable. Mostowski (1952) can be interpreted in this way. According to Mostowski it is ‘the insufficiency of the methods of proof admitted in S^2 ’ that is the reason for the independency. If we allow for additional resources we can see the correctness of the reflection principles for example by deriving them in a second-order arithmetic or assuming transfinite induction for ϵ_0 .

The version that we want to argue for takes the reflection principles as basic principles, because there are no more basic principles that would justify them. They are as evident and justified as other axioms. Gödel’s formulation suggests that the reflection principles have the status of axioms. They are intrinsically motivated in analogy to other axioms, by our understanding of the structure of the natural numbers. What we

² See Mostowski (1952, p. 107).

do not intend to adopt is Gödel's understanding of structure and our access to it. So whereas we think that our warrant to accept the reflection principles stems from our conception of the natural numbers we want to connect it to a form of conceptual-structuralism that allows for a more naturalistic interpretation.

There is an alternative view of taking reflection principles as basic principles, namely as presuppositions or cornerstone principles of a cognitive project to which we are entitled without having a justification. This is a version based on Wright's understanding of entitlement as in Wright (2004a).³ We do not intend to pursue this line here for the case of arithmetical reflection, as we think that it is difficult to understand reflection principles as presuppositions for arithmetical knowledge. Rather, they are better understood as being intrinsically warranted by our conception of the structure of natural numbers.

The issue of reflection principles has sometimes been framed as the statement of an implicit commitment. For example Dean's formulation of the implicit commitment thesis in Dean (2014) is as follows:

(ICT) Anyone who accepts the axioms of a mathematical theory S is thereby also committed to accepting various additional statements Δ which are expressible in the language of S but which are formally independent of its axioms. (Dean 2014, p.32)

The implicit commitment thesis is a thesis just like any other and has to be argued for. The arguments for the implicit commitment thesis itself usually depend on background assumptions that are not shared by everybody. Some of the key proponents of the thesis share a predicativist view. Dean has convincingly argued in Dean (2014)⁴ that there are alternative foundational views that do not share some of the crucial assumptions for the implicit commitment thesis and Feferman seems to admit that much.⁵

The different readings of reflection appear to be connected to convictions about the foundational status of mathematical knowledge. This dependence makes it unlikely that there will be a general answer to the question whether one ought to accept the reflection principles. Especially the question of implicit commitment seems to be not independent of foundational issues as Dean correctly observes and states as 'there seems to be little reason why we should expect ICT to hold for arbitrary agents'.⁶

³ Fischer et al. (2019) for example connect reflection with Wright's story of cognitive projects.

⁴ Dean uses the examples of finitism as understood via Tait's thesis and first-orderism, a view that takes PA to be complete with respect to first-order arithmetical statements (Isaacson's thesis). Both of these examples are not clear cut. Finitism is a difficult topic that we do not intend to make a substantial contribution. Just two minor remarks. Finitism only becomes an inherently unstable position if it is understood as incorporating some completeness claim. In contrast to Dean there are other positions that take also in the case of finitism some form of reflection as appropriate. Whereas Kreisel's view is not very convincing, more subtle versions such as Ignjatović's 1994 are not unattractive. Also in the case of first-orderism some reservations are appropriate. One could also take ICT to undermine the proposed division of first-order and second-order content, as one of the referees pointed out.

⁵ In Feferman's introduction in Feferman (1991), it reads: "To what extent can mathematical thought be analyzed in formal terms? Gödel's theorems show the inadequacy of single formal systems for this purpose, except in relatively restricted parts of mathematics." It is reasonable that Feferman included in these restricted parts, finitism as PRA and maybe first-orderism of PA . It is also interesting to note that Feferman himself seems not to have used the term 'implicit commitment', as far as is known to the author.

⁶ See for Dean (2014, p. 35).

In my opinion it would be helpful to relativize the implicit commitment thesis as follows: If a (...) accepts a theory S , then she ought to accept reflection principles for S , where the blank is to be filled with different foundational positions, such as finitist, predicativist, set-theoretic platonist and others.⁷

Rather than point-wise answering the question for each position, we try to analyse sufficient conditions for a shared conception of natural numbers that allows for a positive answer. Moreover, we will replace the ‘ought to’ by a less committing ‘is warranted to’. Also here the question arises in which cases a warrant is given. We try to provide some tentative answers by sketching a coherent version of informal mathematical knowledge in which acceptance of the reflection principles is warranted. The coherence claim is intended to include restrictions on the side of informal arithmetic, such that no special abilities on the side of our mathematician are required. So we do not assume some special access to ‘the natural numbers’. In the terminology of Button and Walsh (2018) we require a form of moderation.

Another problematic issue with ICT as a general claim is that there are different forms of acceptance of a formal theory.⁸ A very modest form of acceptance is taking the theory to be consistent. Acceptance of Σ as consistent is the assumption that there is no sentence A for which $\Sigma \vdash A$ and $\Sigma \vdash \neg A$. This reading of acceptance does not support ICT and therefore will not be our intended reading of acceptance. Rather we want our acceptance of a theory to be understood as acceptance as sound, correct or valid.⁹

Arithmetical theories are of special interest. On the one hand they are based within a setting in which we have quite strong intuitions about the structure of natural numbers that might allow for an informative answer to our main question. On the other hand the reflection principles are substantial extensions at least of the proof-theoretic strength of the systems under consideration. With these caveats we reformulate the issue as a question of implicit warrant: Which minimal assumptions (on our implicit arithmetical knowledge I) are sufficient to warrant an extension of a formal theory Σ by a reflection principle? In Sect. 4 we are more explicit about these background assumptions by considering a specific conception of natural numbers.

1.3 Mathematical Knowledge Versus Explicit Theories

Authors like Mostowski, Gödel and others distinguish between our mathematical knowledge and formal systems. The formal systems can be understood as attempts to make our mathematical knowledge explicit. We will follow this distinction to see whether it helps to resolve some problematic issues. We will call the mathematical knowledge also implicit, informal and intuitive. Implicit in the sense that we do not

⁷ Implementing a relativization to predicativism in Feferman’s quote would be a possibility to make Feferman’s remarks coherent, especially as he seems to admit that there are positions that are not committed to the acceptance of reflection principles. Dean seems to take this interpretation as he writes ‘Feferman’s use of normative terms like ‘ought’ and ‘should’ to describe the reflective closure of a theory should be interpreted from the standpoint of predicativism’ (p. 48).

⁸ Compare the discussion in Franzen (2004).

⁹ We could try to make use of Wright’s (2004b) explanation of acceptance (of a proposition) as ‘trusting it (or the theory) or to act on the assumption that it is true’.

require awareness of this knowledge. Informal, because we think of it as not necessarily recursive in the relevant sense, but rather semi-formal. Intuitive only in so far as we think that the knowledge stems from a concept of natural number and not in the sense of being quasi-concrete as for example Parsons uses the notion. So we will use I to denote our implicit mathematical knowledge.

Whereas the formal systems are bound to some very specific criteria, such as recursive axiomatizability and others, the mathematical knowledge is less rigorously restricted. We still take mathematical knowledge to be propositional knowledge, a form of ‘knowledge that’ in contrast to ‘knowledge of’, so that we can still think of our mathematical knowledge as collections of sentences. However, in contrast to formal theories we will allow for closure principles that enable us to consider also extensions of theories. We will consider different semi-formal principles that could be assumed for the implicit knowledge side.

Another difference is that for the implicit knowledge we do not need to be able to claim this knowledge, whereas in the case of the explicit formal theory we do want to be able to state it as a form of knowledge that we realize to have. This opens the possibility that we are only entitled, in Burge’s sense of entitlement, to rely on the mentioned closure principles, without requiring a justification for claiming their validity. However, in the case of reflection we intend to add the reflection principles to our formal theory as acceptable principles. In order to be warranted to do this we do not only need to know the principle as a first-order claim, but we should be able to claim this knowledge, as a second-order statement that we know that we know.

One of the usual requirements for this implicit knowledge is that it is true. A seemingly attractive explication of the notion of truth involved is to understand it as truth in the standard model. However, there are several reasons why we should not proceed this way and we will later on explain these reasons. At this point, we will leave open what understanding of truth is assumed. The importance of the notion of truth is obvious as it is deeply entrenched with our implicit knowledge and the issue of accepting a theory as correct. For the acceptance as correct or sound we have several options, but the most straightforward is that if $\Sigma \vdash A$, then A . But already here we have two readings. In the first reading we take it as being allowed in hypothetical reasoning in the other case not. In case where it is important to distinguish the two readings we will refer to the first as the conditional form and the second the admissible rule form. We will subsume all these versions for the time being as acceptance as correct without more information on the use of correctness.

2 The Basic Setting: Pure Reflection

Before looking at expansions we first choose a purely arithmetical setting as our background and starting point. There are several reasons for using this framework. First of all it is suited for the metamathematical considerations at hand. Secondly, we seem to have a clearer understanding of the relevant notion of truth in this setting. This is connected with the conviction that the structure of natural numbers is a unique structure. Already in the purely arithmetical setting we have different forms and formalizations of reflection. There will be even more options if we allow for additional resources,

such as employing a notion of truth. In this section, we focus on purely arithmetical principles without a truth predicate.

Because of the limited expressive resources in a first-order arithmetical setting the unrestricted principles are usually formulated schematically. As formalizations of our soundness statements the schematic principles of local and uniform reflection correspond to our conditional soundness assumptions that allow for applications within hypothetical reasoning. The rules on the other hand correspond to the soundness rules. Whereas the local rule is conservative and admissible the uniform rule is a real strengthening and equivalent to the uniform reflection principle itself. The uniform rule is sometimes also called Kleene's rule.¹⁰

Our first task is to consider different argument forms that try to establish the correctness of the reflection principles.¹¹ We distinguish between strategies that will provide an informal argument for the plausibility of the principles and formal derivations of the principles.

In the first case we look at possible semi-formal principles that could be assumed for the intuitive mathematical knowledge and that would suffice to argue for the correctness of reflection. These will be mostly metamathematical assumptions.

In the second case we look at extensions of the formal system Σ that are sufficient for a derivation of the relevant principles. The metamathematical character of the reflection principles and the restricted expressive resources in first-order arithmetic point towards expansions in order to accommodate natural derivations.

2.1 Motivations of Reflection

A naive version of the argument for accepting con_Σ is as follows: by accepting Σ we implicitly believe in the correctness and also in the consistency of Σ . Since con_σ expresses that Σ is consistent, we should make this implicit commitment explicit by extending our theory Σ with the formal principle con_σ . Similarly $\text{Pr}_\sigma(\ulcorner A \urcorner) \rightarrow A$ expresses the correctness (or soundness) of Σ . So we should also accept $\text{Pr}_\sigma(\ulcorner A \urcorner) \rightarrow A$ and add it to our formal theory.

This appears to be a very naive story. Various examples in the history of logics and mathematics show that such a method is in general not reliable, just think of naive set theory or naive truth. However, there is a difference between the settings. In our setting we start with a basic assumption of correctness and the claim is not that we are guaranteed to have a correct system in the end, but rather that the addition of the relevant statement does not decrease the reliability of the original system. The plausibility of the story is then clearly dependent on a suitable reading of 'correctness', but not completely naive anymore.

Let us look at a well known complication for the reading of correctness as simple consistency. In the case that we only presuppose that Σ is consistent, we have no guarantee that the resulting system that we get by adding the consistency statement will be also consistent, because there are consistent systems that are not correct. More

¹⁰ One of the first papers concerning such a rule according to Ignjatović is Rosser's paper Rosser (1937).

¹¹ In this respect the status of the consistency statement con_σ is also relevant. But the equivalence to Π_1^0 -reflection allows us to subsume this under the heading of reflection.

concretely, if we consider the system $\Sigma + \neg\text{con}_\sigma$, then the assumption that Σ is consistent implies that also $\Sigma + \neg\text{con}_\sigma$ is consistent. If we now look at the extension of $\Sigma + \neg\text{con}_\sigma$ by its consistency statement $\text{con}_{\sigma+\neg\text{con}_\sigma}$ we arrive at an inconsistent theory. So the acceptance of a theory as consistent is no guarantee for the consistency of its reflective extension. The consistency statement $\text{con}_{\sigma+\neg\text{con}_\sigma}$ itself is not the source of the problem. It is correct and the addition should also be fine, provided that the theory we extend is itself correct. The problem is that $\Sigma + \neg\text{con}_\sigma$ is not correct, assuming that Σ is correct.

The strangeness of the example disappears, if we read correct as Σ_1 -sound in the common model-theoretic sense, i.e. every provable Σ_1 -sentence is true in the standard model of arithmetic. By our assumption that Σ is consistent $\neg\text{con}_\sigma$ is a Σ_1 -sentence not true in the standard model. The theory $\Sigma + \neg\text{con}_\sigma$ is therefore only satisfied in a non-standard model. The lesson to draw is that if we employ a form of acceptance of Σ as mere consistency, then we cannot be sure that the consistency extension is acceptable as well.¹² In order to guarantee the preservation of reliability, we have to understand acceptance as correct in a stronger form. At this point we do not want to draw other conclusions, especially not from our reference to the standard model. This leaves us with the more important task of singling out sufficient reasons for accepting the consistency statement.

A second point of discussion concerns the assumption that con_σ ‘expresses’ the consistency of Σ .¹³ The sentence con_σ is short for $\neg\exists x\text{proof}_\sigma(\ulcorner 0 = 1 \urcorner, x)$ stating that there is no proof of a contradiction in Σ . Since the relation of being a proof is decidable, and in $\mathcal{Q}$ all the decidable predicates are strongly representable the proof predicate itself should be unproblematic and express the relation asked for. We will provide a more detailed argument without relying on the notion of standard model later on. Moreover, we assumed a standard representation σ of the axioms of Σ and therefore the problem of extensional versus intensional appears not to be relevant for our case.

One possible reason for the doubt might be that in the formal consistency claim the universal quantifier is unbounded. So whereas in the informal claim we would only consider standard proofs with standard lengths, we have to ensure that we only quantify over natural numbers. In the model-theoretic terminology we have to take into account non-standard proofs, that we get by overspill. A general worry towards universal statements not being able to express their informal counterparts appears exaggerated. However, we admit that an argument is required that warrants the claim that the universal quantifier ranges over ‘the’ natural numbers and we provide it in Sect. 4.2.

For consistency itself it is difficult to doubt that con_σ expresses the consistency of Σ and also the requirements on the metamathematical side to enable an understanding of this fact are not very far reaching and rather basic assumptions on our understanding of language.¹⁴

¹² Compare Franzen (2004) who is very explicit on this point.

¹³ Ketland (1999) and others seem to be doubtful.

¹⁴ As far as the author is aware of, not even finitists tried to avoid Gödel’s theorems by exploiting this worry of expressiveness.

2.2 Consistency

One of the most straightforward ways to establish the correctness of the consistency claim is to use the Σ_1 -completeness of Σ . In our choice of Σ we have taken care to have some basic arithmetic so at least Q . As Q is Σ_1 -complete any extension Σ will be. Now in the case of consistency we can then argue as follows:

Assume that the consistency statement con_σ is not true. Then its negation $\neg\text{con}_\sigma$ is true. Since the consistency statement is a Π_1 -sentence the negation will be a Σ_1 -sentence and therefore by Σ_1 -completeness of Σ , $\Sigma \vdash \neg\text{con}_\sigma$. But by Gödel $\Sigma \not\vdash \neg\text{con}_\sigma$ under the assumption that Σ is Σ_1 -sound. Then we can conclude that $\neg\text{con}_\sigma$ is not true and therefore con_σ is true.

So in order to know that con_σ is true it is sufficient to know that Σ is Σ_1 -sound. Here we think of Σ_1 -soundness in the conditional form. In general we can conclude that Σ_1 -soundness is preserved by consistency extensions. So if we accept a theory as Σ_1 -sound, then we should also accept its consistency extension as Σ_1 -sound.

It is possible to formalize this argument by using an explicit form of Σ_1 -soundness employing the definable partial truth predicate T_{Σ_1} for Σ_1 -sentences.

$$\forall x (\Sigma_1(x) \wedge \text{Pr}_\sigma(x) \rightarrow T_{\Sigma_1}(x)) \quad (\Sigma_1\text{-GRFN}_\Sigma)$$

Then it is possible to show that $\Sigma + \Sigma_1\text{-GRFN}_\Sigma \vdash \text{con}_\sigma$.¹⁵ In this explicit form of the argument it is questionable whether we arrive at an epistemologically relevant reduction, as the explicit assumption of Σ_1 -soundness is far from being more obvious or evident than the consistency statement itself. The problem for explicit justification is that as usual the assumptions motivating the formal principles are more committing than the theorems, unless there is an equivalence. Only in the rare cases where we have a good intuition to prefer some axioms over theorems there is a real epistemological gain. However, the explicit principle of Σ_1 -soundness of a theory is not an obviously evident axiom for a first-order arithmetical mathematical theory.

Let us compare this to another way of formally deriving con_σ . It is well-known that con_σ is provably equivalent to $\Pi_1\text{-GRFN}_\Sigma$. Moreover, $\Sigma_1\text{-GRFN}_\Sigma$ is provably equivalent to $\Pi_2\text{-GRFN}_\Sigma$, a proper extension of $\Pi_1\text{-GRFN}_\Sigma$.¹⁶ So in the formal representation it would be sufficient for a justification of the consistency statement to assume that $\Pi_1\text{-GRFN}_\Sigma$ is contained in our implicit knowledge.¹⁷ But this is different from our informal argument, as can be seen by considering the informal Π_1 -soundness assumption. If we assume that our implicit knowledge is Π_1 -sound, we are not guaranteed that the consistency statement con_σ is true, but rather only the informal consistency assumption for Σ , which is not sufficient to argue that $\Sigma \not\vdash \neg\text{con}_\sigma$.

What this argument shows is that it is perfectly possible to argue for the truth of the consistency statement from informal metamathematical assumptions. So the informal argument at least circumvents the circularity problem by pushing the assumption of Σ_1 -soundness on the informal level of our mathematical knowledge. Whether this

¹⁵ Alternatively we could have also used $\text{Rfn}_{\Sigma_1}(\Sigma)$ which is stronger than con_σ .

¹⁶ Compare Smoryński (1977, p. 849).

¹⁷ Apparently Tennant is arguing in this direction, compare Dean (2014) footnote 31.

resolves the issue or just shifts the problem is up for discussion. That is we still have to answer the question how these informal soundness assumptions are motivated and on what our warrant to the corresponding closure principle is founded.¹⁸

2.3 Local Reflection

In the case of reflection we can use the strategy of Σ_1 -completeness only for restricted versions, such as Π_1 -reflection. The argument that shows that the consistency statement implies Π_1 -reflection exactly employs this strategy. In the general case we cannot use the Σ_1 -completeness directly. But what about using a stronger notion of acceptance as correct, specifically a notion of full soundness? If we think that for all $A \in I$, A is true (or $\mathbb{N} \models A$) and that $\Sigma \subseteq I$, we could try to argue for the warrant to explicitly accept also local reflection.

We know that if Σ is sound (in the conditional form), then $\Sigma + \text{Rfn}_\Sigma$ will also be sound. So if we think that $\Sigma \subseteq I$ and I is sound, then we are warranted to consider $\Sigma + \text{Rfn}_\Sigma \subseteq I$. Again we have to consider the warrant for taking I to be sound in the conditional form.

Franzen (2004 p. 207) draws the conclusion that ‘if we don’t know T to be sound, there is no obvious reason why we should accept every autonomous extension by reflection’. On the other hand, if we know that Σ is sound, then there is no obvious reason why we should not accept every autonomous extension by reflection.

One of the shortcomings here is the fact that we cannot simply state this knowledge in a single arithmetical sentence. Because of Tarski’s undefinability of truth the formalized version has to either rely on first order-approximations of the soundness claim, or we have to expand the expressive resources appropriately.

If we think of the reflection principles as adequate first-order approximations for expressing the informal soundness claim, the circularity of the argument becomes obvious. So the question remains whether the reflection principles are explicit formulations of the soundness claim and whether we have to justify these principles, or not.

In case of the local rule of reflection we are able to provide an argument, by assuming only the informal Σ_1 -soundness in its conditional form. In general, the rule is admissible for Σ_1 -sound theories Σ , although this admissibility is not provable in Σ itself. So adding the local rule of reflection to Σ will not provide new theorems.¹⁹ In contrast to the local reflection principles, we have an informative way to argue for the acceptability of the rule.

The distinction between the conditional form and the rule form is also connected with the question of correctly expressing, because in the case of reflection principles, we might have the additional concern that the material conditional gives a stronger

¹⁸ Piazza and Pulcini (2015) claim that this strategy is not satisfactory as it relies on the derivation of Σ_1 -completeness of Σ and that this derivation already uses a form of generalization on arbitrary objects. I do not see this as a general problem because the Σ_1 -completeness proof can be formalized in systems with restricted induction such as $I\Sigma_1$. So for accepted theories extending $I\Sigma_1$ the argument should be unproblematic. A closer inspection is needed for weaker theories. Horsten (2021) provides a phenomenological story for the process of reflection in the case of accepting the consistency statement.

¹⁹ Parikh (1971) showed that there will be substantial speed-up.

formalization than intended. There is also a dependence on the question, what the informal soundness claim amounts to. The informal soundness claim is usually also supposed to support hypothetical reasoning. For example in the incompleteness theorems we usually argue by the assumption that for the Gödel sentence G_σ , its negation $\neg G_\sigma$ is derivable. Under that assumption we employ the soundness claim. Although the conditional form is commonly in use, this does not preclude an understanding of acceptance of a theory in rule form. In terms of our implicit knowledge the difference would amount to our implicit knowledge containing all the conditionals $I(\text{Pr}_\sigma(\ulcorner A \urcorner) \rightarrow A)$ or being closed under the rule, i.e. $I(\text{Pr}_\sigma(\ulcorner A \urcorner)) \rightarrow I(A)$.

The conditional not only implies an unproblematic local reflection rule, which is a form of conecessitation $\vdash \text{Pr}_\sigma(\ulcorner A \urcorner) \Rightarrow \vdash A$. With the contraposition of the reflection principle we also get a rule of the following form $\vdash \neg A \Rightarrow \vdash \neg \text{Pr}_\sigma(\ulcorner A \urcorner)$, labelled $(\neg\text{Rfn}_\Sigma^r)$. Whereas $\Sigma \vdash \neg A \Rightarrow \Sigma \vdash \text{Pr}_\sigma(\ulcorner \neg A \urcorner)$ is derivable by Σ_1 -completeness of Σ , the negated version $(\neg\text{Rfn}_\Sigma^r)$ establishes new Π_1 -sentences, such as con_σ and is therefore independent. However, in the light of our consistency or correctness assumption regarding Σ , a principle such as $\neg(\text{Pr}_\Sigma \ulcorner A \urcorner \wedge \text{Pr}_\sigma \ulcorner \neg A \urcorner)$ appears plausible and this is sufficient to derive the rule $(\neg\text{Rfn}_\Sigma^r)$.²⁰

Whereas an argument for the local reflection rule seems promising, there are several reasons for not being satisfied with local reflection itself. The argument from soundness is not as convincing as one might wish. Also the full soundness statement is not fully expressible in the purely first-order arithmetical setting. Moreover, local reflection is rather weak. The expectations in case of the assumption of full soundness is much higher and should provide us with reasons to accept stronger reflection principles. Let us therefore take a closer look at uniform reflection, one of the most important reflection principles in the discussion.

3 Motivating Uniform Reflection

A direct argument for the uniform reflection principle appears to be difficult. It will inherit not only the problems we encountered in the case of local reflection, but also some additional obstacles. Luckily there is an alternative path that will bring us to the same conclusion. The idea is basically to motivate the uniform reflection rule as a generalization of the local reflection rule. For a clarification of this step, we will take a detour via the infamous ω -rule. Although this rule would clearly justify the step from the local to the uniform rule, it is impossible to add it to the formal theory. Moreover, it is problematic to assume the ω -rule as a warranted closure principle for our informal knowledge. Therefore we will focus on more constructive versions of the rule. Finally, we can employ a variant of Feferman's equivalence result, establishing that the uniform rule is equivalent to the uniform reflection principle.

²⁰ Even if we grant that soundness should support hypothetical reasoning, the conditional form might not be the best version. One of the problems with the conditional is that it is completely based on the deduction theorem and the classical setting, where the conditional works fine. For sequent systems a rule form might be preferable, such as $\Gamma, \text{Pr}_\sigma(\ulcorner A \urcorner) \Rightarrow A, \Delta$. The difference will appear substantial in the case of sentences which are neither determinately true nor false.

3.1 The Reflection Rule

Before we dig into the acceptability of the uniform rule we take a look at the equivalence as proved by Feferman (1962) Theorem 2.19. The central step in the argument establishes that the uniform reflection principle in the form $\forall x(\text{Pr}_\sigma(\ulcorner A\dot{x}\urcorner) \rightarrow A(x))$ is derivable from the reflection rule. The proof is a clever argument employing different facts. Besides the usual properties of formal provability predicates it relies on a formalized version of Craig's trick that allows Feferman to employ a proof predicate $\text{proof}_{\sigma'}$ that binumerates the proof relation of a recursively enumerable theory Σ represented by a formula σ . Reformulating the uniform reflection principle with this proof predicate as $\forall x\forall y(\text{proof}_{\sigma'}(\ulcorner A\dot{x}\urcorner, y) \rightarrow A(x))$ allows him to see it as the result of applying the reflection rule to the premise $\forall x\forall y(\text{Pr}_\sigma(\ulcorner \text{proof}_{\sigma'}(\ulcorner A\dot{x}\urcorner, \dot{y}) \rightarrow A\dot{x}\urcorner))$, which is itself derivable in Peano arithmetic. The proof of this latter fact, Lemma 2.18 in Feferman (1962), makes use of the formalized Σ_1 -completeness. Although the proof applies the law of excluded middle, the application would also be justified in an intuitionistic setting as its instance is concerned with the decidable proof predicate.

The equivalence proof makes only use of basic facts and relying on the equivalence should really be unproblematic. Therefore, it would be a suitable justification for the reflection principle, if we could argue for our warrant to claim the uniform reflection rule to be admissible. In a way, we can interpret the equivalence as establishing that the use of the conditional within the uniform reflection principle is unproblematic. If we are entitled to claim the admissibility of the rule for our mathematical knowledge I , then we are also warranted to accept the conditional claim.

To argue for the correctness of the uniform reflection rule and our entitlement to claim its admissibility, will be a greater challenge. If we intend to rely on the correctness of the local reflection rule we have to analyse the similarities and differences. Especially we have to explain how the step from $\text{Pr}_\sigma(\ulcorner A\dot{x}\urcorner)$ to $A(x)$ is different from the step from $\text{Pr}_\sigma(\ulcorner A\urcorner)$ to A and why it is nevertheless unproblematic. Besides an argument for the correctness of the rule we will analyse the background assumptions that are needed for the correctness argument.

To avoid a possible misinterpretation of the uniform reflection rule we remind the reader that it is not just a generalization of the local rule in allowing formulas instead of sentences. It is crucial that $\ulcorner A\dot{x}\urcorner$ is a complex term with a free variable and not a constant denoting the open formula, such as $\ulcorner A(x)\urcorner$. Officially $\ulcorner A\dot{x}\urcorner$ is an abbreviation for $\text{sub}(\ulcorner A\urcorner, \ulcorner x\urcorner, \text{num}(x))$, where sub is a function symbol representing the primitive recursive substitution function for terms in formulas and num a function symbol representing the primitive recursive function that sends a number n to $\#n$, the Gödelnumber of its numeral.

If we look at the premise of the uniform reflection rule, we see that from $\vdash \text{Pr}_\sigma(\ulcorner A\dot{x}\urcorner)$ we can derive all the 'standard' instances $\vdash \text{Pr}_\sigma(\ulcorner A(\bar{n})\urcorner)$ for numerals $\bar{n}$. Applying the local reflection rule would give us $\vdash A(\bar{n})$ for all $n \in \mathbb{N}$. Assuming an understanding that the numerals exhaust all the natural numbers we could argue that $\forall x A(x)$ has to be correct.

Clearly there are several moves in this argumentation, that are question begging, but let us try to make them more reasonable as we go along. There are two basic

problems: first of all we have infinitely many inferences, as well as an inference with infinitely many premisses. A second problem lies in the assumption that the numerals exhaust the natural numbers. This seems to presuppose a conception of the natural numbers that is in need of some further explanation. From an external perspective and the assumption of an intended interpretation in which $\forall x$ is intended to run over the natural numbers $\mathbb{N}$ the argumentation appears to be trivially sound. If we fix our interpretation as given by the standard model, then we don't have any problems with the validity of this move. However, this presupposes a prior understanding of 'standard model' or 'natural number' itself, an assumption we are reluctant to make from our more moderate perspective, as it would undermine our undertaking. We have to address this concern in detail in the next section, but for the moment we will focus on the infinity problem.

3.2 The ω -Rule

One way to argue for the plausibility of the inference step is to consider a form of ω -rule and taking uniform reflection to be a 'constructive' counterpart. Clearly the ω -rule is not adoptable in the case of our explicit formal theory. However, is it reasonable that our informal arithmetical knowledge is closed under the ω -rule? Also this version is rather implausible, as it would make unrealistic assumptions regarding the abilities of the average mathematician. By accepting or assuming the ω -rule we seem to subscribe to a pretty realist picture, which is in contrast to a more naturalistic picture that assumes some form of moderation. Not much is gained in contrast to employing a standard interpretation of the 'natural numbers'. Since this system is complete for arithmetical sentences all models of PA^ω are elementary equivalent to the standard model $\mathcal{N}$. This fact and the implausible demands on the mathematicians abilities show the unsuitability of the ω -rule as a reasonable motivation for uniform reflection.

So far we haven't considered the possibility that these strong assumptions are not necessary. Especially, we are wondering whether there is a less committing understanding of the concept of natural number that still provides sufficient reason to accept uniform reflection. Let us try to be more modest by considering alternative ways to argue for uniform reflection. The strategy that we will consider concerns more constructive forms of the ω -rule. This constructiveness requires that we have a uniform proof strategy that works for all natural numbers. This also means that we are able to establish a formalized version of it. We will argue that the premiss of our uniform reflection rule almost directly expresses this.²¹

3.3 The Constructive ω -Rule

One of the problems of the ω -rule is its infinitary character and therefore its unsuitability to function as an effective rule in a formal system. Instead of using the full ω -rule

²¹ An alternative strategy for avoiding the ω -rule could be by introducing a rule of generalization, working with arbitrary natural numbers. We will not pursue this strategy here further.

one could try to use a constructive version of the ω -rule. Piazza and Pulcini (2015) recently suggested to employ a constructive version of the ω -rule for these purposes.

So let us look at the strategy of using a constructive ω -rule. The use of the term ‘constructive ω -rule’ seems not to be uniform in the literature. For concreteness we will therefore focus our discussion on Shoenfield’s version in Shoenfield (1959). The main ingredient of the constructivization of the ω -rule is the extra requirement for the premiss, in Shoenfield’s words: that there is ‘an effective method of obtaining a proof of $A(n)$ when n is given’.

Basically we require that there is a recursive function, that for every n will provide a code of a proof of $A(\bar{n})$. So we want a recursive function f such that for all $n \in \mathbb{N}$, $(\#A(\bar{n}), f(n)) \in \text{PROOF}_\Sigma$ or in its representation as $\text{proof}_\sigma(\ulcorner A(\bar{n}) \urcorner, f(\bar{n}))$. In a first order setting we cannot existentially quantify over functions and therefore it is difficult to literally state it as a rule of inference. Even if we cannot quantify over functions in a first-order setting we can quantify over indices for partial primitive recursive functions. We can formulate it as $\exists e \text{proof}(\ulcorner A(\bar{n}) \urcorner, \{e\}(\bar{n}))$. This is closer to

$$\frac{\exists e \text{proof}(\ulcorner A\dot{x} \urcorner, \{e\}(x))}{\forall x A(x)}$$

In Shoenfield (1959) the official definition is an inductive definition of a set of sentences Z^* , closed under the constructive ω -rule. Surprisingly as it turns out Z^* contains true arithmetic, i.e. the set of first-order arithmetical sentences true in the standard model and therefore the constructive ω -rule has the same strength as the full ω -rule with respect to first-order arithmetic. Therefore, also in this case it is not plausible to assume that I is closed under the constructive ω -rule.

Despite its name the constructive ω -rule therefore inherits several of the features of the ω -rule. One of them is its informal character, we cannot employ it as an effective rule in a formal system. However, we can define the set of codes of proofs in Z^* via an inductive definition, along the notations in Kleene’s $\mathcal{O}$. Comparing it with the situation of Feferman’s completeness results for transfinite progressions it appears that this inductive definition is not only of the complexity Δ_1^1 as the set of true arithmetical sentences, but at least Π_1^1 . Moreover, it seems to inherit all the epistemological problems that apply to Feferman’s completeness results in Feferman (1962).²² It is well known that these completeness results are of limited value for the search of capturing intuitive arithmetic.²³ The problem is that employing progressions along paths through Kleene’s $\mathcal{O}$ to extend our formal theories would presuppose knowledge that undermines the approach.

These are reasons enough to also give up on ‘the’ constructive ω -rule as providing a warrant for accepting uniform reflection, at least in Shoenfield’s version. Besides the complexity issues of the definition we still have a reliance on a substantial notion of

²² Compare Franzen (2004, 378f).

²³ Turing, Feferman and others seem to agree on this. For a discussion of these issues compare also Franzen (2004).

‘natural number’.²⁴ The problem is not that relying on a notion of natural number is in principle problematic, rather it is invoking concepts or structures that are even more demanding and therefore unnecessarily strong. In this specific case it is Kleene’s $\mathcal{O}$ itself, which is a more complex structure than the inductive structure of natural numbers.

3.4 From the Constructive ω -Rule to the Uniform Reflection Rule

As we have seen Shoenfield’s constructive ω -rule is in several respects not suitable for our purposes. It is not directly formalizable as an effective rule in the first-order setting and it is not acceptable as a closure principle for our arithmetical knowledge. In order to circumvent these problems we will formalize the idea behind the rule, i.e. we will not only require that there is a recursive function that provides for any n a proof of $A(\bar{n})$, but additionally that this is recognizable or provable. There are three problems to be fixed:

First, we need to decide which proof system is represented by our proof predicate, especially we have to decide whether it includes the additional rule or not. The former option has an air of circularity and requires an inductive definition that is too complex for finding a system that is closed under a reflection rule for itself and appears to require already some form of reflective closure. The latter option on the other hand is unproblematic and it is standard in these settings to settle for this more modest option and we follow this practise. However, also in the latter version it points towards an iterative process that can be repeated and leading to a hierarchy of stronger and stronger systems. To avoid possible confusion, it is preferable to treat it as a closure principle stating that if the premiss is derivable in the base system Σ , then the system closed under the rule Σ^r contains the conclusion of the rule. If Σ is recursively enumerable, then Σ^r is as well, so we can treat it in the usual fashion as an axiomatizable theory in contrast to the closure under Shoenfield’s constructive ω -rule. Also in the formalization we will distinguish proof _{σ} and proof _{σ^r} .

Second, we need to ensure that e is a total function and moreover that it can be recognized to be total, i.e. we require it to be provably total in Σ . In order to simplify the presentation we make some unproblematic simplifications. Rather than quantifying over indices for partial recursive functions we will assume that our theories contain function symbols f for all primitive recursive functions and prove the recursive clauses for them. Basically, we assume that the theories of interest are extensions of the non-equational version of PRA, also known as QF-IA.

Third we need to make the position for n arbitrary. This is achieved with the primitive recursive *num* function.

²⁴ Piazza and Pulcini at this point give it a twist. In order to circumvent a dependence on a prior understanding of natural number, they just take the rule as basic and state that natural number is everything that conforms to such a rule. It is difficult to understand this move as it requires a primitive understanding of the rule. We have seen that it is not easy to grasp the rule itself and a characterization of the rule employs the notion of recursive function as well as that of inductive definition. Both notions are closely connected to the notion of natural number. We think it is a serious undertaking to explain the constructive ω -rule without relying on the notion of ‘natural number’.

We therefore focus our attention on rules of the form:

$$(UR) \frac{\Sigma \vdash \forall x \text{proof}_\sigma(\ulcorner A\dot{x}^\neg, f(x) \urcorner)}{\Sigma' \vdash \forall x A(x)} \text{ for } f \text{ a primitive recursive function}$$

For simplification we only mentioned the rules for 1-ary primitive recursive functions, but we assume the rules for arbitrary arity. Officially, the theory Σ' is the deductive closure of $\Sigma \cup \{\forall x A x \mid \Sigma \vdash \forall x \text{proof}_\sigma(\ulcorner A\dot{x}^\neg, f(x) \urcorner) \text{ for some primitive recursive function } f\}$.²⁵

Since $\Sigma \vdash \forall x \text{proof}_\sigma(\ulcorner A\dot{x}^\neg, f(x) \urcorner)$ implies $\Sigma \vdash \text{Pr}_\sigma(\ulcorner A\dot{x}^\neg \urcorner)$ our rules (UR) are contained in the rule of uniform reflection. In the other direction we show that we can derive all the instance of uniform reflection in Σ' . We proceed by following the structure of Feferman's equivalence proof. We can replace the provability predicate by a suitable proof predicate for a specific primitive recursive function. For the crucial step employing the formalized Σ_1 -completeness, we rely on the fact that for a primitive recursive formula $A(x, y)$, there is a primitive recursive function f , such that $\Sigma \vdash A(x, y) \rightarrow \text{proof}_\sigma(\ulcorner A\dot{x}, \dot{y}^\neg, f(x, y) \urcorner)$.²⁶

Let $B(x, y)$ be an abbreviation for the formula $\text{proof}_{\sigma+}(\ulcorner A\dot{x}^\neg, y \urcorner \rightarrow A(x))$, where $\text{proof}_{\sigma+}$ is the proof-predicate that we get by using Craig's trick. Then the argument runs as follows:

$$\begin{array}{ll} \text{proof}_{\sigma+}(\ulcorner A\dot{x}^\neg, y \urcorner) \rightarrow \text{proof}_{\sigma+}(\ulcorner A\dot{x}^\neg, f(x, y) \urcorner) & \text{for } f \text{ p.r. with } f(x, y) = y \\ \text{proof}_{\sigma+}(\ulcorner A\dot{x}^\neg, y \urcorner) \rightarrow \text{proof}_{\sigma+}(\ulcorner B\dot{x}, \dot{y}^\neg, f'(x, y) \urcorner) & \text{by weakening} \\ \neg \text{proof}_{\sigma+}(\ulcorner A\dot{x}^\neg, y \urcorner) \rightarrow \text{proof}_{\sigma+}(\ulcorner \neg \text{proof}_{\sigma+}(\ulcorner A\dot{x}^\neg, \dot{y} \urcorner)^\neg, g(x, y) \urcorner) & \text{by } \Sigma_1\text{-completeness} \\ \neg \text{proof}_{\sigma+}(\ulcorner A\dot{x}^\neg, y \urcorner) \rightarrow \text{proof}_{\sigma+}(\ulcorner B(\dot{x}, \dot{y})^\neg, g'(x, y) \urcorner) & \text{by weakening} \end{array}$$

Since $\text{proof}_{\sigma+}(\ulcorner A\dot{x}^\neg, y \urcorner)$ is decidable, we can construct a primitive recursive function h by cases from f', g' to get:

$$\Sigma \vdash \forall x \forall y \text{proof}_{\sigma+}(\ulcorner B(\dot{x}, \dot{y})^\neg, h(x, y) \urcorner).$$

Applying our primitive recursive rule for h we receive

$$\Sigma' \vdash \forall x \forall y (\text{proof}_{\sigma+}(\ulcorner A\dot{x}^\neg, y \urcorner \rightarrow A(x)),$$

which is provably equivalent to an instance of uniform reflection, so we get

$$\Sigma' \vdash \forall x (\text{Pr}_\sigma(\ulcorner A\dot{x}^\neg \urcorner) \rightarrow A(x)).$$

The conclusion of this argument is that Shoenfield's constructive ω -rule is not needed for the uniform rule of reflection, but a much more modest form, such as our

²⁵ Ignjatović (1994) uses similar rules with additionally restricting it to primitive recursive formulas A , which is a substantial restriction. He also remarks that the other restriction, that the proofs have to be given by a primitive recursive function is an inessential difference to Feferman's uniform reflection rule. Compare his remark in fn. 24 on p. 340.

²⁶ See for example Lindström (1997 Fact 9, p. 17).

(UR) suffices. At best it can be seen as a heuristic intermediary step, but it plays no substantial role, similar to the full ω -rule. Once we have understood the motivation for (UR) we can drop these intermediary steps that are much more demanding on their presuppositions. Whereas the ω -rule presupposes a standard understanding of numerals, the constructive ω -rule presupposes the impredicative notions of inductive definitions. Our (UR) on the other hand seems only to presuppose an understanding of provably recursive function and the totality of the *num* function. But let us take a closer look at the presuppositions.

We have on the one hand the presuppositions for the equivalence, which are as we have seen minimal. They include formalized Σ_1 -completeness. More pressing are questions concerning our constructive reflection rules (UR). As we want to employ it for claiming the acceptability of the uniform reflection principles, we should argue that we can claim the admissibility of (UR) for our informal knowledge. The strategy will be to argue that there are coherent conceptions of natural number that not only provide a warrant for the acceptability of an arithmetical theory Σ , but moreover for the admissibility of (UR).

A specific concern has been put forward by Franzen (2004, p. 217), who claims that the uniform rule of reflection is not an expression of trust. The reason he provides is supported by an example of a derivation of the consistency of a theory from the theory together with the uniform reflection rule. In the example the consistency is derived via a specific assumption and the rule appears to be used in an unintended way. Franzen takes this as the basis for the claim that every use of the rule is unintended. However, there are examples in the literature, where such a rule is employed in the intended way and moreover plays an important role. A specific example is given by Schwichtenberg (1977) in his proof of Theorem 5.2 on p. 892f., where he uses an instance of the rule to establish the schema of transfinite induction up to ϵ_0 , TI_{ϵ_0} . Together with the fact that transfinite induction for ϵ_0 is equivalent to uniform reflection over PA , this also shows that our rules (UR) are equivalent to uniform reflection.

4 Informal Arithmetic and the Natural Numbers

In the previous section we have argued that there are suitable rules (UR) that warrant the adoption of uniform reflection principles in the arithmetical setting. In other words, if we assume that we are warranted to claim that our informal arithmetic is closed under the rules (UR), then it is reasonable to explicitly add the uniform reflection principles to our best formal theories of arithmetic that we have. Now it is time to look at our warrant to claim the admissibility of those reflection rules. Since we rely on a strong form of acceptance, which is not forced on us in general and as the status of the reflection principles is not independent of the more general views concerning mathematics and metamathematics we have to take a closer look at these foundational conceptions.

We have argued that an explicit justification of reflection principles in form of derivations from epistemically more basic principles is unmotivated. The formal attempts of justification provide some technical insights into the strength of the principles itself and possibly also connections to other principles, such as induction

principles. However, it appears unlikely that there is an informative way to reduce the reflection principles to principles that are epistemically more basic. The acceptance of the principles that allow for a derivation of the reflection principles is at least as committing as accepting the reflection principles themselves. This is especially telling in the pure arithmetical setting, where the close analogy to transfinite induction points towards an understanding of the status of the reflection principles in close analogy to other axioms.²⁷

The reflection principles are not obviously true. It takes a moment of ‘reflection’ and some ingenuity on Gödel’s side to see their truth.²⁸ Not all axioms are obviously true. For example the set theoretic axioms are often motivated via the iterative conception of set. In the case of arithmetic the concept of ‘natural number’ is involved in the formulation of axioms or axiom schemata. It is one of the tasks of a philosopher to clarify what assumptions on the side of informal arithmetic suffice to motivate reflection principles.

4.1 The Inductive Conception of ‘Natural Number’

There are two extreme positions that will not concern us here further. On the one hand there are foundational views taking the connection between the formal system and the implicit mathematical knowledge to be very close or even identifying the two. Sometimes their understanding of arithmetic is tied to specific formal systems, such as PRA, PA or others. In such a case there is no reason to ask of the holder of such a view to be committed to reflection principles for that system. Therefore the strong version of ICT as a general claim is implausible.²⁹ These conceptions are not sufficient to provide a warrant for reflection, namely those that are based on the view that everything that we know about the natural numbers has to be expressed in a fixed first-order arithmetical language.³⁰ On the other hand there are positions that trivialize the role of reflection, i.e. positions that presuppose a special access to ‘the’ unique structure of the natural numbers. In order to have a point of reference we follow partially the distinctions used in Button and Walsh (2018). In their terminology we consider ‘moderate’ conceptions. Therefore we exclude very realistic or platonistic versions of the concept of natural number.³¹

A coherent conception of natural numbers providing a warrant to accept reflection is conceptual-structuralism, as argued for by Parsons and Feferman.³² The structure of natural numbers is understood as an ω -sequence or an inductive set. Our knowledge of

²⁷ Although labelling such an analogy as ‘justificatory equivalence’, as Dean (2014) does, appears to be in need of an argument.

²⁸ Compare Horsten (2021) for an interesting analysis of this notion of ‘reflection’.

²⁹ Dean (2014) and Horsten (2021) seem to share this view.

³⁰ It is not obvious whether those ‘formalist’ conceptions are very attractive and their coherence has also been questioned, but it is not the place to provide an argument.

³¹ Also here there are different shades of realism. Bernays (1935) was one of the first to distinguish different forms of Platonism. Whereas Bernays sees strong forms of Platonism threatened by paradoxes, he considers weaker forms of Platonism. One of the ingredients is the acceptance of the natural numbers as a completed totality, in contrast to constructive conceptions.

³² We distinguish it from object-structuralist positions.

the concept of the natural numbers goes beyond a fixed first-order theory in a language with a fixed signature. A conception of the natural numbers as an ω -sequence allows us for example to claim induction not only for formulas of the specific language of arithmetic built on 0, Sc, +, $\times$, but for any definite property of the natural numbers expressible in a suitable language expansion.³³ This is a coherent position, even if we do not have a clear delineation of the concept of definite property. It is sufficient, to accept the induction principle for a predicate P , if we come to accept the predicate as expressing a definite property of the natural numbers that we could not express with the standard PA-signature. These expansions need for every case a judgement of the mathematician and it is therefore itself not given as a definite collection.³⁴ In order to understand this we do not need to presuppose the set of all subsets as a ‘given totality’. This would undermine our undertaking as it requires an understanding of a more complex structure than the structure of the natural numbers itself.

The intuition that there is only one structure of natural numbers is very strong. Arithmetic is a ‘univocal’ theory in contrast to algebraic ones and there is good reason for this kind of uniqueness and it also plays part in the motivation for reflection. In the following, we will see the connection more clearly, but let us first mention one of the central issues for conceptual-structuralism: the doxological challenge³⁵, i.e. the challenge to explain how it is possible to access a ‘unique’ structure of natural numbers. The challenge is a real challenge because in the light of moderation our access to the structure is restricted, so that we can assume only access by description and not by acquaintance.³⁶

Another obvious task is to provide an explication of structure itself. An attractive option is to conceive of structures as model-theoretic structures, a position which we label ‘modelism’, again following Button and Walsh. In this case, uniqueness boils down to a class of isomorphic models and the doxological challenge in the case of arithmetic is to find a way to single out the class of structures that are isomorphic to the standard model.

It is well-known that there are several properties that make the standard model special. First of all, it is the initial segment of all PA-models and therefore unique. Secondly, according to Tennenbaum’s result it is the only model in which addition and multiplication are computable.³⁷ There are more ways to characterize the standard model. This seems to be a good indication that there are ways to differentiate between the intended structure and non-standard models.

These characterizations require not only additional resources that go beyond a fixed first-order arithmetic describing the concept of the natural numbers, but also introduce

³³ Views in this direction have been expressed by Parsons (2008), Feferman (2014), Lavine (1994) and others.

³⁴ Compare Feferman (2014).

³⁵ The term doxological is introduced by McGee and taken over by Button and Walsh (2018).

³⁶ Excluded are therefore forms of platonism as expressed for example by Gödel, as well as forms of intuitive arithmetic as expressed by Parsons (2008). Even if the story of Parsons concerning intuitive arithmetic might have some appeal, this kind of access is not sufficient to provide a warrant for reflection principles, because only a limited amount of primitive recursive functions is intuitible.

³⁷ Halbach and Horsten (2005) sketch a form of computational structuralism employing Tennenbaum’s theorem.

an additional dimension to the structure of natural numbers, such as the domain of second-order quantifiers. In principle, we think that it is not a problem to allow for additional expressive resources, representing properties of natural numbers and this will also be the way that we proceed. However, to single out the standard interpretation among all the PA -models it is not sufficient to add predicates, but additional structural components or semantic restrictions are required. This is reason enough to dismiss modelism, at least for questions concerning reflection. Improving our understanding of the natural numbers by excluding some non-standard interpretations by the addition of new predicates is possible. For the motivation of reflection principles, however, it is not necessary to uniquely determine the standard model.

A similar conclusion is reached by considering an argument given in Button and Walsh (2018). They argue convincingly that there is a serious problem for moderate modelism. The argument has well-known predecessors and in simplified form it runs as follows. In order to fix an isomorphism type for arithmetic we would have to employ ‘full’ second-order logic, in the semantic sense, which is not justified by moderation.³⁸ Button and Walsh conclude that one would have to give up on moderation or modelism. In our case giving up moderation seems to grant also reflection. For us therefore, the interesting case is giving up on modelism.

In the case of arithmetic there are additional considerations that provide reasons for giving up on modelism. By looking at the presuppositions of model-theory itself we realize that already on the language side of model-theory we are making use of an inductive structure, such as the formulas, conceived as types. These inductive structures are structurally equivalent to the structure of natural numbers. Model-theory therefore itself presupposes an understanding of an inductive structure, so that it is not suitable to provide a reduction in the relevant sense.³⁹

This leaves us with two open problems: Explicating the structure of natural numbers and the uniqueness issue. For the first we follow Parsons and others in characterizing the structure of the natural numbers as one that satisfies induction (and recursion) in an open-ended way. Therefore the natural numbers are such that for any definite property that is expressible via some predicate P in some suitable expansion of the arithmetical language, we have induction for P .

According to Button and Walsh, there is also a solution to the second problem, at least if we allow for second-order arithmetical theories, not requiring a standard semantics but also allowing for a Henkin semantics. By internalizing the categoricity result we can argue within second-order arithmetic even without restricting attention to ‘full’ second order logic in the semantic sense. The internal categoricity result is less demanding⁴⁰ and therefore also the conclusion of the internal categoricity is weaker than the external categoricity theorem. According to Button and Walsh, it

³⁸ Isaacson (2011) on the other hand seems to give up on moderation and argues for the adoption of full second-order logic as the only option to get an external categoricity result and with it uniqueness.

³⁹ A similar point can be made with respect to set theoretic reductionism.

⁴⁰ The requirements depend on the formulation of the internal categoricity result. What is definitely sufficient, is the axiomatic theory of second-order arithmetic, without assuming a ‘full’ second-order semantics, see for example Väinänen and Wang (2015). Simpson and Yokoama show that for a specific version, that WKL_0 is sufficient, whereas Feferman argues that his version is directly formalizable in a conservative extension of PRA .

is still sufficient to argue for a univocal structure of natural numbers and provide a solution to the doxological challenge. They label the relevant corollary to the internal categoricity, intolerance. Intolerance states that for an arbitrary arithmetical sentence either all the internal structures satisfying the axioms of PA^2 make the sentence true or all structures make it false. Even if we cannot externally establish categoricity the non-standard models do not really pose a threat for a non-modelist.⁴¹

There is a rather peculiar point about the use of the second-order version of arithmetic. Even if we do not presuppose ‘full’ second-order in the semantic sense, we still quantify over ‘arbitrary’ subsets of the natural numbers. Conceptually this appears to be a non-trivial step in addition to the concept of natural numbers as an ω -sequence. From a foundational point of view it is coherent, if not plausible, to assume that the concept of natural numbers is understandable without presupposing an understanding of the concept of arbitrary subset of natural numbers. We focus on positions that take arithmetic as the very starting point, i.e. positions that take the structure of the natural numbers $(\omega, \dots)$ as more basic than the structure of second-order arithmetic $(\omega, M, \dots)$. The latter structures are implicit in employing impredicative second-order arithmetic and with it second-order quantifiers. These reservations about impredicative comprehension are not connected to questions of semantics or the missing of compactness, but rather about a second-order quantifier as presupposing the concept of arbitrary subset of natural numbers. These second-order quantifiers rely on a collection of subsets of natural numbers as a completed totality. Even if someone does not share these predicativist worries, employing second-order arithmetic is not useful in our setting, because the axioms go way beyond adding uniform reflection principles to PA. Already $\Pi_1^1\text{-CA}$ is far stronger than the reflective closure of PA.

The alternative route that we proceed, is to give up on the formality and allow for an informal understanding of arithmetic and induction as an open-ended schema.⁴² The price we have to pay is that we do not have one formal theory of arithmetic, but a series of acceptable theories. In this picture the acceptance of reflection principles emerges as very natural.

From the perspective of open-ended theories we have to answer at least two challenges.⁴³ Even if we do not pose the problem of uniqueness as a question of standard versus non-standard models of arithmetic, we have to address the related issue in form of the doxological challenge, solving the conflict between our intuition or desire that there is a uniqueness or definiteness about the structure of natural numbers and our

⁴¹ Button and Walsh (2018) call the intermediate position internalism, which is close in spirit to what we sketch here. However, they connect internalism with a claim that disallows the internalist semantic reflections in form of semantic ascent. We do not follow this identification. Whereas we do not allow for a model-theoretic reductionist position we still allow for ‘semantic reflections’ in form of semantic ascent, at least as provided by a deflationary notion of truth.

⁴² Whether this open-endedness should be made explicit along the lines of Feferman’s conception of open-endedness as captured by a rule of substitution as for example in Feferman (1991) is a different question and will not be answered here.

⁴³ A challenge that we will not address here is the claim that schematic induction is not understandable without the second-order induction axiom, a claim made for example by Kreisel (1967). For foundational goals such a view appears to be problematic, because it presupposes an understanding of second-order quantification, which is relying on an understanding of ‘arbitrary’ subset of natural numbers that is again question begging.

abilities to single out a unique structure. The situation is clearly different to the algebraic case where we do not have the uniqueness, but also to the case of set theory.⁴⁴ Since we cannot use the strategy of Button and Walsh via second-order arithmetic and intolerance we are forced to give a different answer more in line with Parsons. Parsons asks us to imagine two mathematicians, Kurt and Michael, who share some arithmetical background assumptions for their specific arithmetical vocabulary. Assuming an understanding of the open-endedness that allows Kurt to use induction and recursion on the concepts of Michael and vice versa, enables Parsons to use an internal categoricity result. Whether this internal categoricity result is acceptable is up to debate.⁴⁵

Even if we can solve the doxological issue in this way, it is not clear whether the argument provides sufficient reason for the acceptance of an analogue of intolerance. In order to formulate intolerance for the full language and for internal structures of schematic theories of arithmetic we would have to expand our expressive resources. The appropriate notion appears to be truth itself, but as long as we are not able to establish the definiteness of all our arithmetical notions we are not warranted to introduce a truth predicate for the full language.

4.2 Why We Can Agree on Correctness Statements

Luckily we do not have to settle this question in full, but a partial answer will suffice. Moreover, we will not need the full schematic version of intolerance for all the sentences. It is unproblematic to argue for the absoluteness of the Δ_0 -formulas. The escape route is provided by our restriction of the premise in our rules (UR), especially by our requirement of the primitive recursive function. All mathematicians sharing the picture of the natural numbers as given by open-ended induction and recursion should agree on the acceptability of (UR) and with it uniform reflection.

The starting point is the simple observation that we have an internal absoluteness for Δ_0 -formulas A : If we have two vocabularies that agree on at least the axioms of Q , then we have to agree on the recursive concepts.

Proposition 1 *If $A(\vec{x})$ is an arithmetical formula in which all the quantifiers are restricted, then for any vocabularies (N_1, z_1, S_1) , (N_2, z_2, S_2) , if they satisfy Q , then they agree on the status of A . If we allow for induction and recursion for both vocabularies, then we can prove:*

$$Q(N_1 z_1 S_1) \wedge Q(N_2 z_2 S_2) \rightarrow (A^{(N_1 z_1 S_1)} \wedge A^{(N_2 z_2 S_2)}) \vee (\neg A^{(N_1 z_1 S_1)} \wedge \neg A^{(N_2 z_2 S_2)})$$

Here $Q(N_1 z_1 S_1)$ stands for the conjunction of the axioms of Q formulated for the vocabulary $(N_1 z_1 S_1)$ and $A^{(N_1 z_1 S_1)}$ for the replacement of $N, 0, Sc$ in A by N_1, z_1, S_1 . For simplification we will just write A^1 .

⁴⁴ Feferman for example takes the structure of natural numbers to be clear and definite, whereas for set theory there is no determinate truth-value. The question whether set-theory has a unique definite structure will not be addressed here.

⁴⁵ For a criticism of the appropriateness of using induction and recursion in such a way see Field (2001). Field argues that it is not appropriate for Kurt to use induction to establish the surjectiveness of the function from his numbers to the numbers of Michael. Feferman establishes the bijectiveness by additionally defining a function from Michael's numbers to Kurt's numbers by recursion on Michael's numbers.

The argument basically follows Parsons' and Feferman's suggestions.⁴⁶ Since we allow recursion for both vocabularies, we can define two primitive recursive functions $g : \mathbb{N}_1 \rightarrow \mathbb{N}_2$ with $g(z_1) = z_2$ and $g(S_1(x)) = S_2(g(x))$ as well as $h : \mathbb{N}_2 \rightarrow \mathbb{N}_1$ with $h(z_2) = z_1$ and $h(S_2(x)) = S_1(h(x))$. Then we can show that $g(h(x)) = x$ and $h(g(y)) = y$, establishing basically the existence of an isomorphism, between the two structures of natural number. Based on this it is easy to show that for an arithmetical formula $A(\vec{x})$ with restricted quantifiers $A^1(\vec{x}) \rightarrow A^2(g(\vec{x}))$ and $A^2(\vec{x}) \rightarrow A^1(h(\vec{x}))$. Therefore, it is reasonable that mathematicians are able to agree on Δ_0 -formulas without relying on the concept of the standard model, but on a concept of natural numbers based on induction and recursion.

So far we made a case for the agreement on Δ_0 -formulas and especially the premise of our constructive reflection rules (UR). Let us try to connect this with our informal arithmetic as given by the open-ended schematic understanding of induction and recursion. Let us assume that for a formula $A(x)$ and a primitive recursive function f , we have established $\vdash \text{proof}_\sigma(\ulcorner A\dot{x}^\top, f(x) \urcorner)$.

This does not only provide us with a reason to believe in the correctness of $\text{Pr}_\sigma(\ulcorner A(\bar{n})^\top \urcorner)$ for all our 'natural numbers', but that anybody having a similar inductive conception of natural numbers will agree that for all his numbers $\text{Pr}_\sigma(\ulcorner A(\bar{n})^\top \urcorner)$ will be the case.

We would like to have an inductive argument for the correctness of $\forall x(\mathbb{N}(x) \rightarrow A(x))$. Let us try to argue informally in a metatheory and assume that $f(z)$ is given by the primitive recursive function $g(z)$ and $f(\text{Sc}(x))$ by $h(f(x))$:

1. $\vdash \text{proof}_\sigma(\ulcorner A\dot{x}^\top, f(x) \urcorner)$
2. $\vdash \text{proof}_\sigma(\ulcorner A\dot{z}^\top, g(z) \urcorner)$
3. $\vdash A(z)$
4. $\vdash \text{proof}_\sigma(\ulcorner A\overline{\text{Sc } z}^\top, h(g(z)) \urcorner)$
5. $\vdash A(\text{Sc } z)$
- ...
6. $\vdash \text{proof}_\sigma(\ulcorner A\overline{\text{Sc } \text{Sc } \dots z}^\top, h(f(\text{Sc } \dots z)) \urcorner)$
7. $\vdash A(\text{Sc } \text{Sc } \dots z)$

By 'reflecting' on the argument it appears that we can turn it into an informal inductive argument. The primitive recursive function allows us to argue inductively that $A(n) \rightarrow A(\text{Sn})$. In contrast to a metatheoretic argument that establishes $A(n)$ for all $n \in \mathbb{N}$, we do not have to assume a specific understanding of $\mathbb{N}$, but the inductive structure allows us to argue for all structures of natural numbers that agree on the inductive build up. If we try to phrase the crucial steps in structure theoretic terms we can argue as follows:

1. $\vdash \text{proof}_\sigma(\ulcorner A\dot{x}^\top, f(x) \urcorner)$,
2. for my structure $(\mathbb{N}, \text{Sc}, z)$ for all n in $\mathbb{N}$, $(\mathbb{N}, \text{Sc}, z) \models \text{proof}_\sigma(\ulcorner A\dot{x}^\top, f(x) \urcorner)[n]$,
3. for my structure $(\mathbb{N}, \text{Sc}, z)$ $(\mathbb{N}, \text{Sc}, z) \models \forall x A(x)$,
4. since there is an (primitive recursive) isomorphism i from my natural numbers to any other inductive structure $(\mathbb{N}', S', z')$ we have: $(\mathbb{N}', S', z')$ models $A'(z')$ as well as $A'(n')$ implies $A'(S'n')$,
5. so for all n' in $\mathbb{N}'$ $(\mathbb{N}', S', z') \models A'(x)[n']$,

⁴⁶ See for Parsons (2008) and Feferman (2014).

6. and since (N', S', Z') is an arbitrary inductive structure: $\forall x(A(x))$ is true.

This argument establishes that our entitlement to the open-ended induction and recursion principles provides reasons for the admissibility of our (UR)-rules and with it the uniform reflection principles. A conclusive argument for the entitlement to these principles based on the inductive conception of natural numbers however, is beyond the scope of this paper.

5 Additional Conceptual Resources

To carry out the previous argument as a formal argument, we are missing the expressive resources. This is mainly due to our restriction to first-order arithmetical theories. Now as a further layer we could make these missing expressive resources explicit by formulating some axioms for a primitive notion of correctness or truth. The reflection principles are naturally based within metamathematical investigations. We have argued that modelism is not attractive for our purposes and for this reason we do not want to frame the discussion in model-theoretic terms. Alternatively giving up on any kind of ‘semantical’ considerations seems premature.⁴⁷ A promising alternative is to introduce some primitive notions, such as truth in order to state the correctness claim explicitly. Understood as expressions of soundness, it is natural to formulate the reflection principles with the help of extra conceptual metamathematical resources. So again, rather than using second-order resources it is convenient to accept these metamathematical concepts as primitive and employ them in the formulation of reflection principles. The most obvious candidate is truth itself. However, it might be interesting to consider also notions of validity, and intensional notions.

So far our observation of internal absoluteness of recursive concepts warrants the expansion by a typed truth predicate for these primitive recursive notions. We would like to expand this notion to a wider class of sentences. At first glance, it might appear to be a small step from intolerance to definite truth for all arithmetical statements. However, as often, things are much more complicated and it is not obvious that we have a good argument for the introduction of a full blown typed truth predicate in the fashion of the compositional theory CT.⁴⁸ If we are not relying on an interpretation of truth as truth in the standard model, we have to give a different argument for the definiteness of the introduced notion. It might be possible that Parsons’ strategy via an internal categoricity can be turned into a reliable argument, also in the case of schematic first-order theories. We have already mentioned several critical points concerning Parsons’ strategy. In order to circumvent these problems we stick to a less committing notion of truth with more modest principles than CT itself.

We assume the following principles for our primitive notion of truth:

1. (T-in) $B(x) \rightarrow T(\ulcorner B\dot{x} \urcorner)$ for $B \in \Delta_1^0$;
2. (U-Inf) $\forall x(N(x) \rightarrow T(\ulcorner A\dot{x} \urcorner)) \rightarrow T(\ulcorner \forall x(N(x) \rightarrow A) \urcorner)$;
3. (Conec) $\frac{\vdash T(\ulcorner A \urcorner)}{\vdash A}$ for A an arithmetical sentence.

⁴⁷ Button and Walsh connect their internalism with a rejection of semantic considerations.

⁴⁸ Compare Button and Walsh ch. 12 for more on this.

Now we could reformulate our reflection rules as:

$$(UR_T) \frac{N(x) \rightarrow T(\ulcorner \text{proof}_\sigma(\ulcorner A\dot{x}^\neg, f(\dot{x}) \urcorner)^\neg \urcorner)}{N(x) \rightarrow (T(\ulcorner A\dot{x}^\neg \urcorner) \rightarrow T(\ulcorner AS\dot{c}x^\neg \urcorner))}$$

In this case the assumptions about our new predicate are minimal. (T-in) is justified by internal absoluteness. Since we use a typed truth predicate, (Conec) appears to be harmless, but we have to be careful as this rule is for sentences of arbitrary complexity. Considering our possibilities of introducing $T(\ulcorner A^\neg \urcorner)$ either by (T-in) or (U-Inf) via (UR_T) there seems to be no real issue, as we have no other compositional principles. More demanding is the (U-Inf) principle, but for a truth predicate for arithmetic it appears motivated.⁴⁹ Overall this provides enough reasons to accept this weak notion of truth as definite and therefore allow it within the induction schema.

The introduction of the truth predicate illustrates the motivation for the adoption of (UR). We are not trying to hide the fact that we still rely on a strong notion of correctness of Σ in form of the rules (UR_T) . What is now more explicit is that for every natural number we have a witness, namely its proof, for the truth of its substitution instance in A . Moreover, these witnesses are not provided in a piecemeal fashion for each n , but they are given by a primitive recursive function along the inductive structure of our natural numbers. In order to argue now for the truth of $A(x)$ for arbitrary substitution instances, we need a form of the correctness assumption. The local reflection rule alone is not sufficient, but our understanding of the natural numbers as an inductive structure and the way that the witnesses are provided, we have good reasons to accept $T(\ulcorner A\dot{x}^\neg \urcorner) \rightarrow T(\ulcorner AS\dot{c}x^\neg \urcorner)$. Then we are able to argue by induction to get $\forall x T(\ulcorner A\dot{x}^\neg \urcorner)$. By (U-Inf) and (Conec) we receive $\forall x A$. This concludes our argument for the acceptability of our reflection rules (UR) for suitable arithmetical theories Σ relying on a conception of the natural numbers that transcends Σ and is captured as our informal arithmetical knowledge I .

5.1 Truth as an Explicit Concept

So far the notion of truth played a minor role in motivating arithmetical reflection. The notion of truth appears on many other occasions in the debate on reflection. On the one hand the notion of truth appears in the informal unrestricted soundness statements as well as their formal counterparts the global reflection principles. On the other hand truth functions crucially in a well-known argument for a justification of reflection. We

⁴⁹ This appears to be much less demanding than Feferman's understanding of reflection principles as based on a notion of validity in Feferman (1962)

(3) A concept of validity is available for the sentences of K , such that

- (a) for any $\phi \in St_{K_0}$, ϕ is valid if and only if it is true,
- (b) for any $\phi \in Fm_K$, if $Fv(\phi) = \{x\}$, then $(\forall x)\phi(x)$ is valid if and only if $\phi(\bar{n})$ is valid for each n , and
- (c) if $\phi \in A_1$, then ϕ is valid.

(Feferman 1962, p. 268)

take the opportunity to distinguish our use of truth here from other applications and sketch possible extensions of our view.

In our case the introduction of the truth predicate was motivated by the desideratum of having an expressive device that allows us to make an informal argument for the plausibility of the rules (UR) more explicit. The role here is a very modest one and does not presuppose that all arithmetical sentences have a definite truth value. In this case the truth predicate has a purely instrumental role, we only reflect on our arithmetical knowledge and use a truth predicate to express this. In this case it suffices to trust our mathematical theory and accept the additional resources as expressive devices. The situation is similar if we use global reflection principles, such as $\forall x(\text{sent}_A(x) \wedge \text{Pr}_\sigma(x) \rightarrow T(x))$, to express our trust in the underlying formal theory. Also in this case, the relevant function of the truth predicate is primarily its expressive one. We do not think that we can only understand the schematic version of uniform reflection, because we understand the universal statement of global reflection, rather our mathematical knowledge is such that we can make sense of schemata directly as our disposition to accept the instances once we realise that the relevant preconditions are satisfied.⁵⁰

The situation is different in the case of a fully compositional truth predicate for arithmetic, especially if we aim for a justification of reflection via a formal derivation. It is well-known that CT derives the global reflection principle as well as all the uniform reflection principles. However, in order to motivate the introduction of such a truth predicate, we have to argue for the definiteness of arithmetical truth which is a non-trivial undertaking. It is implausible to take this route in order to justify reflection as it is more demanding to justify CT itself, than reflection for PA. CT itself is much stronger than just adding uniform reflection to PA. The theory CT is as strong as ACA, more concretely CT is mutually interpretable with ACA preserving the arithmetical vocabulary. The theory CT does not only prove the uniform reflection principles for PA, but also iterations of uniform reflection. So expanding the resources via a typed truth predicate along the lines of CT is much more than is really needed. To clarify the point, the previous is not an argument that CT itself is unacceptable. It might well be that there is a convincing argument, maybe even internal, for the definiteness of arithmetical truth and CT. However, this has to be argued for, presumably by relying on a conception of natural numbers. Along the way we will also be able to argue for the acceptability of reflection, as it is only a small step on the way to definite truth.

So far we have only been concerned with purely arithmetical theories. Once we have accepted a truth predicate, expanded our expressive resources and introduced a truth theory connected with it, the question arises, whether we are warranted to accept reflection principles also for the theory of truth. This clearly depends on the notion of truth as we have argued that we need a strong form of acceptance as correct to motivate reflection for the underlying theory. This requires a certain trust in our notion of truth. In the typed case, there are no extra complications in addition to the required argument for the univocal structure and the definite truth values. If we accept Parsons' argument and Feferman's claim of determinate truth for arithmetics, then we could also argue for reflections on typed truth theories, such as UTB or CT. For example, we can use

⁵⁰ The issue is not easily resolved here and deserves more attention. Nevertheless, there is something deeply puzzling about the view that our understanding in schemata has to be based on an understanding of a universal statement. A comparison with our understanding of Modus Ponens might be relevant here.

uniform reflection for UTB for the full language, and derive all the principles of CT.⁵¹ In this case the extra resources of truth significantly expand our basic mathematical knowledge.

At this stage we encounter a slight incoherence. We introduced a typed truth predicate in order to make the correctness assumption explicit. Then we realised that the theory of truth is itself correct. Unfortunately, the typed truth predicate cannot be used to make the correctness claim for our theory of truth explicit. In the case of global reflection the problem is obvious, because sentences containing the truth predicate are not allowed to occur within the scope of the truth predicate in the consequent of global reflection. Employing the ramified truth hierarchy is subject to the usual points of criticism attached to the hierarchical strategy.

The introduction of a type-free truth predicate would be more coherent. In this case, the global reflection principle could be employed to make the correctness claim explicit. However, we have an additional concern and this is the Liar. The correctness assumption becomes non-trivial, especially if we have abandoned modelism. Referring back to a model-theoretic construction for a soundness proof becomes unavailable.⁵² Thus, to close a type-free theory of truth under reflection we have to provide good reasons for accepting it as correct. At this point it is also relevant that we did not take reflection as an implicit commitment, but rather as a warrant for acceptance. In every case we have to decide whether the type-free theory of truth is acceptable in the sense of being correct.

One possible continuation of this reasoning could be as follows. If we trust our mathematical theory Σ and if we trust in our metamathematical theory $T[\Sigma]$, then we should be warranted to adopt the following global reflection principle for it:

$$\forall x(\text{sent}_{T[\Sigma]}(x) \wedge \text{Pr}_{T[\Sigma]}(x) \rightarrow T(x))$$

If we follow this reasoning, then we can also make use of the contrapositive claim. If we cannot close our theory of truth under global reflection, then we should not accept the correctness claim for the theory. Several classical truth theories are not compatible with their own global reflection principle, such as the usual formulations of KF or FS.⁵³ In the simple case of FS we have a theory known to be ω -inconsistent. Even without referring to the standard model, the inadequacy of FS is evident. Just by adding uniform reflection, it already becomes inconsistent. FS is therefore not only incompatible with a standard interpretation of arithmetic, but also with an open-ended conception of arithmetic as an inductive structure.

Not all the cases are as clear as the case of FS. In order to resolve these issues in general, we would need an informative answer to decide whether we should reject the truth theory or the strong reflection principle, which we do not attempt in this paper. To pursue this line of reflection, it appears more promising to reflect on positive type-

⁵¹ This is based on results by Halbach (2001) and a strategy along this line has been proposed by Horsten and Leigh (2017).

⁵² Alternatively, we could try to provide proof-theoretic consistency proofs via cut-elimination.

⁵³ Feferman's use of type-free truth in Feferman (1991) appears to be more pragmatic. It appears to be a way to finitely characterize predicativism. Feferman exploits the type-free truth predicate to simulate enough uniform reflection. In his case the theory of truth is not reflected on.

free T-biconditionals.⁵⁴ An alternative strategy would be to retreat to non-classical principles, such as those used in Fischer et al. (2019).

5.2 More Resources

One of the best suited conceptual resources for a strategy of reflection expansions is a primitive notion of truth. The truth predicate allows for explicit formulations of soundness statements. The truth predicate is not the only notion that could be useful for reflection. Also intensional notions such as mathematical necessity could play an important role in the expansions of sound theories. We will just briefly state two examples for a fruitful use of intensional notions for reflection.

The first example concerns a non-classical setting. Consider the application of the logic of HYPE to truth, as suggested in Leitgeb (2019). In this case, we have an intensional conditional and a truth predicate that is local as it does not interact appropriately with the conditional. In the intended interpretation we have a complete lattice of Kripkean fixed-point, where truth is interpreted locally at a state, whereas the conditional is interpreted via the structure of the states. It appears natural to consider expansions by intensional notions that are non-local and interact with the conditional in a suitable way.

The second example concerns potentialism. In the potentialist literature reflection principles are in use that employ modal notions. For example Parsons (1983) and Linnebo (2013) invoke such principles to recover modal versions of ZF. But also Reinhardt (1980) introduces modal reflection principles to obtain large cardinals. These principles are employed in a set theoretical setting and closer to set theoretic reflection principles and therefore not directly transferable to the arithmetical setting and proof-theoretic reflection principles. Nevertheless, there are potentialist versions of arithmetic. It is possible to understand second-order arithmetic as not fully determined and give predicativism for example a potentialist reading. Without a convincing argument for the definiteness of arithmetical truth, there might even be room for first-order potentialist versions. An investigation of modal reflection principles in the setting of arithmetic especially second-order arithmetic is a fruitful field for further research.

Acknowledgements Thanks to Mario Piazza and his support in Pisa within the project “Implicit commitments of intensional notions”. Thanks also to Carlo Nicolai and Matteo Zicchetti for helpful discussions, as well as two anonymous referees. Thanks also to the DFG for the financial support: “Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) - Projektnummer 407312485”.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

⁵⁴ See for Leigh (2016).

References

- Bernays, P. (1935). Sur le platonisme dans les mathématiques. *L'enseignement mathématique*, 34, 52–69.
- Button, T., & Walsh, S. (2018). *Philosophy and model theory*. Oxford: Oxford University Press.
- Dean, W. (2014). Arithmetical reflection and the provability of soundness. *Philosophia Mathematica*, 23(1), 31–64.
- Feferman, S. (1962). Transfinite recursive progressions of axiomatic theories. *The Journal of Symbolic Logic*, 27(3), 259–316.
- Feferman, S., Dawson, J., Kleene, S., Moore, G. Solovay, R., Heijenoort, J. (Eds.). (1990). *Kurt Gödel: Collected Works, Publications 1938–1974* (Vol. II). Oxford: Oxford University Press.
- Feferman, S. (1991). Reflecting on incompleteness. *The Journal of Symbolic Logic*, 56, 1–47.
- Feferman, S., Dawson, J., Goldfarb, W., Parsons, C. Solovay, R. (Eds.). (1995). *Kurt Gödel: Collected works, unpublished essays and lectures*. Oxford: Oxford University Press.
- Feferman, S. (2014). Logic, mathematics and conceptual structuralism. In P. Rush (Ed.), *The metaphysics of logic* (pp. 72–92). Cambridge: Cambridge University Press.
- Field, H. (2001). Correspondence truth, disquotational truth, and deflationism. In M. P. Lynch (Ed.), *The nature of truth: Classical and contemporary perspectives* (pp. 483–503). Cambridge: Cambridge University Press.
- Fischer, M., Horsten, L., & Nicolai, C. (2019). Hypatia's silence truth justification and entitlement. *Noûs*. <https://doi.org/10.1111/nous.12292>.
- Franzen, T. (2004). *Inexhaustibility. Lecture Notes in Logic*. Boca Raton: CRC Press.
- Halbach, V. (2001). Disquotational truth and analyticity. *The Journal of Symbolic Logic*, 66, 1959–1973.
- Halbach, V., & Horsten, L. (2005). Computational structuralism. *Philosophia Mathematica*, 13, 174–186.
- Horsten, L. (2021). On Reflection. *The Philosophical Quarterly*, <https://doi.org/10.1093/pq/pqaa083>.
- Horsten, L., & Leigh, G. E. (2017). Truth is simple. *Mind*, 126(501), 195–232.
- Ignjatović, A. (1994). Hilbert's program and the omega-rule. *The Journal of Symbolic Logic*, 59(1), 322–343.
- Isaacson, D. (2011). The reality of mathematics and the case of set theory. In Z. Novák & A. Simonyi (Eds.), *Truth, reference, and realism* (pp. 1–75). Budapest: Central European University Press.
- Ketland, J. (1999). Deflationism and Tarski's paradise. *Mind*, 108, 69–94.
- Kreisel, G. (1967). Informal rigour and completeness proofs. In I. Lakatos (Ed.), *Problems in the philosophy of mathematics* (pp. 138–186). Amsterdam: North Holland.
- Lavine, S. (1994). *Understanding the infinite*. Harvard: Harvard University Press.
- Leigh, G. E. (2016). Reflecting on truth. *IfCoLog Journal of Logics and their Applications*, 3(4), 557–594.
- Leitgeb, H. (2019). Hype: A system of hyperintensional logic (with an application to semantic paradoxes). *Journal of Philosophical Logic*, 48, 305–405.
- Lindström, P. (1997). *Aspects of incompleteness, Lecture Notes in Logic* (Vol. 10). Berlin: Springer.
- Linnebo, O. (2013). The potential hierarchy of sets. *Review of Symbolic Logic*, 6(2), 205–228.
- Mostowski, A. (1952). Sentences undecidable in formalized arithmetic: An exposition of the theory of Kurt Gödel. North Holland.
- Parikh, R. (1971). Existence and feasibility in arithmetic. *Journal of Symbolic Logic*, 36, 494–508.
- Parsons, C. (1983). Sets and modality. In *Mathematics in philosophy*. Cornell University Press.
- Parsons, C. (2008). *Mathematical Thought and its Objects*. Cambridge: Cambridge University Press.
- Piazza, M., & Pulcini, G. (2015). A deflationary account of the truth of the Gödel sentence G. In G. Lolli, et al. (Eds.), *From logic to practice, Boston Studies in the Philosophy and History of Science* (pp. 71–90). Berlin: Springer.
- Reinhardt, W. N. (1980). Necessity predicates and operators. *The Journal of Philosophical Logic*, 9, 437–450.
- Rosser, B. (1937). Gödel theorems for non-constructive logics. *The Journal of Symbolic Logic*, 2(3), 129–137.
- Schwichtenberg, H. (1977). Proof theory: Some applications of cut-elimination. In *Handbook of mathematical logic* (pp. 867–896). Amsterdam: North Holland.
- Shoenfield, J. (1959). On a restricted ω -rule. *Bulletin de L'Academie Polonaise des Sciences*, 7, 405–407.
- Smoryński, C. (1977). The incompleteness theorems. In Barwise, J. (Ed.), *Handbook of Mathematical Logic* (pp. 821–865). Amsterdam: North Holland.
- Väänänen, J., & Wang, T. (2015). Internal categoricity in arithmetic and set theory. *Notre Dame Journal of Formal Logic*, 56(1), 121–134.
- Wright, C. (2004). Intuition, entitlement and the epistemology of logical laws. *Dialectica*, 58(1), 155–175.

Wright, C. (2004). Warrant for nothing (and foundations for free)? In *Proceedings of the Aristotelian Society, supplementary volumes* (Vol. 78, pp. 167–245).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.