LEON HORSTEN

PROVABILITY IN PRINCIPLE AND CONTROVERSIAL CONSTRUCTIVISTIC PRINCIPLES

ABSTRACT. New epistemic principles are formulated in the language of Shapiro's system of Epistemic Arithmetic. It is argued that some plausibility can be attributed to these principles. The relations between these principles and variants of controversial constructivistic principles are investigated. Special attention is given to variants of the intuitionistic version of Church's thesis and to variants of Markov's principle.

1. INTRODUCTION

There has been a long discussion among constructivistic mathematicians about the acceptability of certain "controversial" constructivistic statements. Among these are Markov's principle, constructivistic variants of Church's thesis, and choice principles of various sorts. It seems that the debate will not soon come to an end.

Extending a construction of Gödel, Shapiro has formulated a system of Epistemic Arithmetic (EA).[1] EA is a formalization of arithmetic which contains axioms governing the notion of absolute knowability or absolute provability. Heyting Arithmetic (HA) can be faithfully translated into EA by means of an extension of Gödel's translation of intuitionistic propositional logic to modal logic. It seems to Shapiro that his translation comes close to being meaning-preserving, and this leads him to claim that EA "integrates classical and intuitionistic arithmetic". Roughly, EA can be seen as expressing the theory of a classical mathematician who understands constructivistic arithmetic under something like Gödel's translation.

The present paper is an attempt to look at some of the controversial principles of constructivistic arithmetic from the framework of EA. Borrowing an expression of Kreisel, we can say that we perform an exercise of *informal rigour*.[2] Although the term "informal rigour" was first used by Kreisel, it undoubtedly has its roots in the writings of Gödel. In [14], Gödel emphasizes the open-endedness of current axiom systems: new axioms might be discovered by reflecting on the meaning of the basic concepts of a discipline ([14, pp. 260–261]). This method was used to formulate new putative axioms for constructivistic mathematical theories by Kreisel, Myhill, Troelstra and others.[3] Kreisel tried to reflect on the

Journal of Philosophical Logic **26:** 635–660, 1997.

notion of constructive proof, add notation concerning this notion to the language of constructive mathematics, and add new axioms concerning this notion to traditional constructive systems. But he is sceptical about the outcome of this program ("nothing rewarding has come of this" [21, p. 81]). We want to suggest that our chances of success are somewhat higher if we reflect on the informal notion of classical provability, in a classical context.

We prove logical properties of (and relations between) variants of these controversial principles on the assumption of certain epistemic principles that are expressible in the language of EA, but not in the language of HA. It is suggested that these epistemic principles have an epistemological status similar to that of the controversial constructivistic principles. In other words, we claim some degree of plausibility for them. We look for extrinsic support for them in their consequences in the domain of constructive arithmetic. But we also claim a reasonable degree of intrinsic plausibility for them. One could even imagine a classical mathematician arguing that the intuitive support attaching to certain controversial constructivistic principles derives from the plausibility of nonconstructive epistemic principles from which they can be derived (in EA). In this scenario, the conflicting intuitions dividing constructivists over the acceptability of certain constructivistic principles are ultimately intuitions about nonconstructive statements.

The scope of the paper is modest. On a technical level, we restrict our attention to problematic constructivistic principles that have been proposed in the course of the investigation of the *lawlike* intuitionistic arithmetical universe (and the lawlike continuum). In particular, we are interested in intuitionistic versions of Church's thesis and in versions of Markov's principle. Since controversial principles of the lawlike universe can usually be given a first-order expression, we restrict our discussion to first-order theories. Many controversial constructivistic principles have been proposed in the investigation of the intuitionistic continuum enriched with lawless objects (such as choice sequences). These principles can only be given a higher-order or set-theoretic formulation, and are outside the scope of this paper.[4] On a conceptual level, we have no illusion that all the epistemic principles that we propose "force themselves upon us as being true" ([14, p. 268]). We do hope that they are no more dubious than the variants of the controversial constructivistic statements to which we relate them.

The paper is organized in the following way. In the next section we briefly revisit the theory EA. Subsequently we formulate in the language of EA schematic epistemic principles for which we claim a reasonable degree of plausibility (Section 3). We want to proceed carefully. Therefore we first formulate principles with parameters ranging over the language of Peano Arithmetic (Section 3.1). Then we investigate to which extent these schemes can be strengthened by allowing their parameters to range over the full language of EA (Section 3.2). In the main section of the paper (Section 4), the logical relations between these epistemic principles and versions of familiar constructivistic principles are described. From an epistemic principle of Section 3.1 and an intuitionistic version of Church's thesis we derive Markov's principle. From the same epistemic principle and an epistemic generalization of the intuitionistic version of Church's thesis we derive an epistemic generalization of Markov's principle. And from an epistemic principle of Section 3.2 we derive a generalization to the language of EA of a theorem of HA. For most of the extensions of EA that are discussed in this paper we also prove faithfulness theorems, to ensure that they do not prove sentences that are not merely controversial but outright false. In Section 5 it is shown that most of these extensions of EA have a version of the disjunction property and of the numerical existence property. We conclude the paper with some philosophical remarks and directions for further research.

2. EPISTEMIC ARITHMETIC

2.1. *The Language of EA*

The language of EA ($\mathcal{L}_{EA}$) contains all the symbols of the formal language of classical first-order arithmetic ($\mathcal{L}_{PA}$), where we take $\rightarrow, \neg, \exists$ and $=$ as primitives, plus an epistemic sentential operator K. So the only nonlogical symbols are the individual constant $\underline{0}$, a one-place function symbol $\underline{s}$ (the successor function), and the two-place function symbols $\underline{+}$ and $\underline{\times}$. The identity predicate ($\underline{=}$) is taken as a logical constant. We also assume that the language contains names for all total recursive functions.

Terms, formulas and sentences of the language of EA ($\mathcal{L}_{EA}$) are defined in the usual manner. Expressions of $\mathcal{L}_{EA}$ which contain no occurrences of K are called *arithmetical* expressions. When we speak about sentences which (perhaps) do contain occurrences of K, we call them *epistemic* expressions.

The intended interpretation of the sentential operator K is described in the literature as "is absolutely knowable", or "is provable in principle". These notions are notoriously vague, and we do not pretend to resolve this vagueness. Nevertheless, for the purposes of this paper we have to be a little more precise. First, we prefer to think of the intended interpretation of K as "provability in principle" rather than as "absolute knowability", because we want only sentences that are the result of *nonempirical* arguments in the extension of K.[5] These arguments are logical derivations from a priori principles (which may be synthetic, like the induction axiom, or somehow analytic, like "provability in principle entails truth"). This is not very precise, as long as it is not indicated what the necessary and sufficient conditions are for something to count as an a priori principle. In this paper, this vagueness is not removed. Second, the sort of idealization that we are intending is provability by a *finite* mathematician (finite memory, finite degree of complexity of calculations she can carry out, finite number of computations per time interval, ...), where there are not intended to be fixed bounds on the capacities of this mathematician. The suggestion is that *a sentence KA is true on the intended interpretation iff there might have been a mathematician ("finite, but potentially unbounded") who has a proof (in the sense outlined above) of A*. Again, this is only a partial characterization of the notion of provability in principle. A more detailed characterization goes beyond the scope of this paper.

2.2. *The Theory EA*

The theory EA contains the Peano axioms for elementary arithmetic (as defined in [2, p. 182]), with its recursive axioms for addition and multiplication. We define EA as the smallest theory which contains these axioms, in which the absolute provability operator K is governed by the (Barcan-free) $S4$ axioms and rules, and which is closed under (classical) first-order logic. We will assume that this theory is given a Hilbert-style formalization, with Modus Ponens and the Necessitation Rule ("from A, infer KA") as only rules of inference. To be a little more precise, we let the predicate logical basis be as in [24, p. 165] (except that we rewrite the axioms governing $\forall$ in the obvious way in terms of $\neg$ and $\exists$), and let the modal axiom schemes be $KA \rightarrow A$, $KA \rightarrow KKA$, $KA \rightarrow (K(A \rightarrow B) \rightarrow KB)$.

Let HA stand for the usual formalization of first-order Heyting Arithmetic, and let $\mathcal{L}_{HA}$ be the language of this theory.

Shapiro then inductively defines a translation V: $HA \rightarrow EA$. Indicating by means of a subscript i ("intuitionistic") that a formula belongs to $\mathcal{L}_{HA}$, we can express this definition as follows:

DEFINITION 1 (the translation V).

- For atomic formulas:

$$V(A_i) = KA,$$

- for complex formulas:

$$
\begin{aligned}
V(A \wedge B)_i &= K(V(A_i)) \wedge K(V(B_i)) \\
V(A \vee B)_i &= K(V(A_i)) \vee K(V(B_i)) \\
V(A \rightarrow B)_i &= K(K(V(A_i)) \rightarrow K(V(B_i))) \\
V(A \leftrightarrow B)_i &= K(K(V(A_i)) \leftrightarrow K(V(B_i))) \\
V(\neg A)_i &= K(\neg K(V(A_i))) \\
V(\forall x A(x))_i &= K \forall x V(A(x))_i \\
V(\exists x A(x))_i &= \exists x K V(A(x))_i
\end{aligned}
$$

The fact that this definition of the translation V closely mirrors Heyting's proof interpretation of the intuitionistic logical connectives is the strongest confirmation of the thesis that the meanings of the intuitionistic arithmetical sentences can at least in part be expressed in $\mathcal{L}_{EA}$. But a necessary condition for having *any* confidence in this thesis is the existence of a faithfulness theorem for V:

THEOREM 1 (Shapiro, Goodman, Mints). *For al* $A_1, \ldots, A_n$, $A \in \mathcal{L}_{HA}$:

$$A_1, \ldots, A_n \vdash_{HA} A \Leftrightarrow V(A_1), \ldots, V(A_n) \vdash_{EA} V(A).$$

Proof. [34, 26, 15] □

By far the most elegant proof of this theorem was given by Flagg and Friedman ([8]).

Shapiro showed that EA has the disjunction property and the numerical existence property: ([34, pp. 17–19]):

THEOREM 2 (disjunction property). *For all* $A \in \mathcal{L}_{EA}$, *if* $\vdash_{EA} KA \vee KB$, *then either* $\vdash_{EA} KA$ *or* $\vdash_{EA} KB$.

THEOREM 3 (numerical existence property). *For all* $A \in \mathcal{L}_{EA}$, *if* $\vdash_{EA} \exists x KA(x)$, *then* $\vdash_{EA} KA(\underline{n})$ *for some natural number* n.

3. Epistemic principles

Shapiro argued that the $S4$ principles are sound for the interpretation of K as provability in principle.[6] On the other hand, it is easy to see that axioms that are in some sense much stronger than $S4$ (such as the $S5$ scheme $\Diamond\Box A \rightarrow \Box A$, or the $S4.2$ scheme $\Diamond\Box A \rightarrow \Box\Diamond A$) are unsound for the intended interpretation of K. But the question whether there exist nevertheless epistemic principles which are independent of EA, but sound for the intended interpretation of $\mathcal{L}_{EA}$, has not been given serious consideration. In this section some candidates for such principles are proposed. We will do our best to motivate them as well as possible. But it must be stressed that since they are to be taken as putative *basic principles*, the force of these motivations will be limited.

The natural inclination is to look straight-away for principles with schematic letters ranging over $\mathcal{L}_{EA}$. But since the exact mathematical content of such principles is often hard to judge (due to the expressive power of $\mathcal{L}_{EA}$), we will be more cautious. First epistemic principles about sentences of $\mathcal{L}_{PA}$ are formulated. Subsequently we investigate to what extent these principles can be generalized to epistemic principles about sentences of $\mathcal{L}_{EA}$.

3.1. *Epistemic Principles about Arithmetical Sentences*

Let $\vec{x}$ stand for a finite sequence of variables, and let $Q\vec{x}$ stand for a finite string of quantifiers (the variable of the ith quantifier being the ith variable of the string $\vec{x}$). Then the strongest epistemic axiom concerning arithmetical sentences that we propose is the following:

AXIOM 1 (S). *For all $A_1(\vec{x}),\ldots,A_n(\vec{x}),A(\vec{x}) \in \mathcal{L}_{PA}$:*

$$KQ\vec{x}(\neg KA_1(\vec{x}) \wedge \cdots \wedge \neg KA_n(\vec{x})) \\ \rightarrow Q\vec{x}K(\neg A_1(\vec{x}) \wedge \cdots \wedge \neg A_n(\vec{x})).$$

Its motivation is the main concern of this section.

On an abstract level, it is motivated by the following (strong) thesis:

THESIS 1 (superthesis). *The negative epistemic properties that can be established of arithmetical sentences supervene on the nonepistemic properties that can be established of them.*

This thesis can be seen as a closure principle, somewhat analogous to the reflexivity principle $KA \rightarrow A$. The problem with the superthesis is that it is very abstract and somewhat vague, and that *its* motivation is not clear. But the underlying idea is this. Suppose we have an arithmetical sentence A. Then there are epistemic properties that can be established of A and that are not dependent on nonepistemic properties that can be established of A. For instance, we can prove $K(KA \rightarrow A)$ purely on the basis of epistemic logic, even if A is neither provable nor refutable in principle. The claim of Thesis 1 is that this is not so for negative epistemic properties of A, such as $\neg KA$. To prove $\neg KA$, one would have to in one way or another, refute A. Now it is not clear from Thesis 1 how this class of negative epistemic properties is to be characterized, whether in these negative epistemic properties iterations of K are allowed to occur, and exactly on which establishable nonepistemic properties these establishable epistemic properties of A supervene. Axiom S is a way of (partially) answering these questions, and thereby of giving content to Thesis 1.

Let us illustrate this on the basis of a special case (it is not hard to see how this illustration generalizes to all instances of axiom S). Suppose that the antecedent of an instance of axiom S, with $Q\vec{x} = \forall x \exists y$, and $i = 2$ holds. Then an epistemic property of $A_1(x,y)$, $A_2(x,y)$ can be established, namely:

$$\forall x \exists y(\neg KA_1(x,y) \wedge \neg KA_2(x,y)).$$

Since $\neg KA_1(x,y)$ and $\neg KA_2(x,y)$ are negative epistemic properties of $A_1(x,y)$, $A_2(x,y)$, respectively, $\neg KA_1(x,y) \wedge KA_2(x,y)$ is also a negative epistemic property, and presumably so is $\forall x \exists y(\neg KA_1(x,y) \wedge \neg KA_2(x,y))$. Thesis 1 says that in such a situation, nonepistemic properties of the arithmetical sentences $A_1(x,y)$, $A_2(x,y)$ must be provable in principle. But this is exactly what axiom S entails: for every natural number a, there is a natural number b such that $\neg A_1(\underline{a},\underline{b}) \wedge \neg A_2(\underline{a},\underline{b})$ can be established. In this sense, the provability in principle of $\forall x \exists y(\neg KA_1(x,y) \wedge \neg KA_2(x,y))$ supervenes on the fact that for every natural number a, there is a natural number b such that $\neg A_1(\underline{a},\underline{b}) \wedge \neg A_2(\underline{a},\underline{b})$ is provable in principle.[7]

We produce additional motivation for S, and for the thesis that motivates it, in an indirect way. We look at some consequences of S, and try to give a more precise motivation for them (this more precise motivation is intended also to provide support for the superthesis). Of course, having "nice" consequences is no guarantee for the validity of an axiom. So we also look at consistency and faithfulness properties of S. First of all, it seems that the principle ought to be consistent with EA. It is easily seen that S satisfies this requirement. For if we had a proof of $\perp$ (falsum) in $EA+S$, then deleting all occurrences of K would result in a proof of $\perp$ in PA. Another question is whether the new principle allows one to prove, in the context of EA, translations in EA of constructivistic sentences that one cannot prove in EA alone. If the answer to this question would be affirmative, then there are two possibilities. Either these "new" constructivistic theorems are sound, or they are (obviously or not so obviously) unsound. In the former case it is likely that they have already come up in the literature about constructivistic mathematics. In the latter case the new epistemic principle should be rejected. If the answer to the question of conservativeness over HA is affirmative, then this is an indication that the principle may be not too strong. But it is not a guarantee: the principle may have undesirable consequences about epistemic sentences that are not translations of constructivistic sentences. Later in this section we show that the principle S, when added to EA, does not allow us to derive new constructivistic statements (even though in the context of constructivistic versions of Church's thesis, it does).

Let us first turn to consequences of S. A principle that is weaker than S says that the only way that a sentence of $\mathcal{L}_{PA}$ can be shown to be unprovable in principle is by refuting it. Its content can be approximated in $\mathcal{L}_{EA}$ by the following axiom:

AXIOM 2 (T). *For all* $A \in \mathcal{L}_{PA}$: $K\neg KA \rightarrow K\neg A$.

It is easily seen that T can be proved in $EA + S$.

T is motivated by the following considerations. If one were to reject T, it would have to be either on empirical grounds, or on the basis of a nonempirical argument.[8] On the empirical side, the only relevant considerations would have to be finiteness considerations (time may be finite, our memory may be finite, there may be an upper limit on the complexity of the problems we can solve, ...), or knowledge of the future development of the universe (we might know that within a few decades all people will forever lose all interest in mathematical problems). But these considerations are made irrelevant by the aspect of *absoluteness* of the intended interpretation of K (see Section 2.1). So the reasons for rejecting T would have to be of a nonempirical nature. Then the claim made by T is that only an essentially *mathematical* argument is available to us to show that an arithmetical sentence A is absolutely unprovable. Namely, one would have to refute A, and then using some epistemic principles of EA to arrive at $K\neg KA$. One might worry that there are essentially *epistemic* arguments (not involving a mathematical refutation of A) showing that A is absolutely unprovable, based on sound and somehow very basic principles concerning K itself. In reply we might challenge the objector to produce these principles and the accompanying derivation of the counterexample, and we might feel that the content of the principle T is so clear that such a counterexample cannot be derived from sound principles. Nevertheless, one can never remove the worry completely. The best we can do is probably to prove faithfulness properties.

A thesis that is weaker still can be expressed as follows: granting that there may very well be arithmetical sentences that are undecidable in principle, one cannot prove of a given sentence that it is undecidable in principle.

In $\mathcal{L}_{EA}$, this thesis can be expressed as:

AXIOM 3 (M_{PA}). *For all* $A \in \mathcal{L}_{PA}$: $\neg K(\neg KA \wedge \neg K\neg A)$.

M_{PA} is a version of McKinsey's axiom $\Box\Diamond A \to \Diamond\Box A$ (the only difference being that M_{PA}'s schematic letter is allowed to range only over nonmodal sentences), and is easily seen to be derivable from T.

There is an interesting variant of M_{PA}:

AXIOM 4 (M^+_{PA}). *For all* $A(x) \in \mathcal{L}_{PA}$: $\neg K\exists x(\neg KA(x) \wedge \neg K\neg A(x))$.

M^+_{PA} is stronger than M_{PA}, and cannot be derived from T (although it is easily seen to be derivable from S). To a large extent, our motivation for T may be repeated at this point to argue for the plausibility of M^+_{PA}. Except that we need the additional claim that the only way to show $\exists x(\neg KA(x) \wedge \neg K\neg A(x))$ for some $A \in \mathcal{L}_{PA}$ is to establish an instance of it (i.e. that no reductio proof can be given). I do not know how to motivate this claim any further. It expresses an aspect of S (and of the superthesis that motivates it) of which I am less confident than of what is contained in T.

We will now show that the principle S does not allow us to derive new constructivistic statements. We do this by transforming theorems of $EA + S$ into theorems of EA by a translation ρ that leaves translations of constructivistic sentences in $\mathcal{L}_{EA}$ unchanged. ρ is defined as $\tau(\xi)$, where τ and ξ are the following translations:

DEFINITION 2 (the translation τ).

- For atomic formulas:

$$\tau(A) = A,$$

- for complex formulas:
  1. if $* \in \{\wedge, \vee, \to, \leftrightarrow\}$:

  $$\tau(A * B) = \tau(A) * \tau(B),$$

  2. if $\circ \in \{\neg\} \cup \{\exists\zeta \mid \zeta \text{ is a variable}\} \cup \{\forall\zeta \mid \zeta \text{ is a variable}\}$:

  $$\tau(\circ A) = \circ\tau(A),$$

  3. if $A = KB$:

(a) if $B = \neg C_1 \wedge \cdots \wedge \neg C_n$ or $B = \neg KC_1 \wedge \cdots \wedge \neg KC_n$ for some $n \in \omega$, with $C_1, \ldots, C_n \in \mathcal{L}_{PA}$:

$$\begin{aligned}\tau(KB) &= (K\neg C_1 \wedge K\neg KC_n) \wedge \cdots \\ &\quad \wedge (K\neg C_n \wedge K\neg KC_n),\end{aligned}$$

(b) otherwise:

$$\tau(KB) = K\tau(B).$$

Let $Q_K(\vec{x})$ be the result of prefixing each quantifier in the string $Q(\vec{x})$ with an occurrence of K. Then the translation ξ is defined as follows:

DEFINITION 3 (the translation ξ).

- For atomic formulas:

$$\xi(A) = A,$$

- for complex formulas:
  1. if $* \in \{\wedge, \vee, \rightarrow, \leftrightarrow\}$:

$$\xi(A * B) = \xi(A) * \xi(B),$$

  2. if $\circ \in \{\neg, K\} \cup \{\exists\zeta \mid \zeta \text{ is a variable}\} \cup \{\forall\zeta \mid \zeta \text{ is a variable}\}$:
     (a) if $A = Q(\vec{x})\varphi$, where $\varphi = \neg KC_1 \wedge \cdots \wedge \neg KC_n$ for some $n \in \omega$, $C_1, \ldots, C_n \in \mathcal{L}_{PA}$:

$$\xi(A) = Q_K(\vec{x})K\varphi,$$

     (b) otherwise:

$$\xi(\circ A) = \circ\xi(A).$$

Now let T^+ be the following scheme:

$$\begin{aligned}&K(\neg KA_1 \wedge \cdots \wedge \neg KA_n) \\ &\quad \rightarrow K(\neg A_1 \wedge \cdots \wedge \neg A_n) \quad [A_1, \ldots, A_n \in \mathcal{L}_{PA}].\end{aligned}$$

Using these definitions, we can prove a couple of statements that will lead us to our theorem. We will not give their proofs in full, since they are straightforward but tedious.

LEMMA 1. *For all* $A \in \mathcal{L}_{EA}$: $EA + S \vdash A \Rightarrow EA + T^+ \vdash \xi(A)$.

Proof. Induction on the length of proofs in $EA + S$. □

To prove the next lemma, we need a simple proposition:

PROPOSITION 1. *For all* $A \in \mathcal{L}_{EA}$: $EA \vdash K\tau(A) \Rightarrow EA \vdash \tau(KA)$.

Proof. Induction on the complexity of A. □

Using this proposition, we establish:

LEMMA 2. *For all* $A \in \mathcal{L}_{EA}$: $EA + T^{+} \vdash A \Rightarrow EA \vdash \tau(A)$.

Proof. Induction on the length of proofs in $EA + T^{+}$. □

Combining the two lemmas, we are done:

THEOREM 4. *For all* $A \in \mathcal{L}_{HA}$: $EA + S \vdash V(A) \Rightarrow EA \vdash V(A)$.

Proof. From the two lemmas, and the fact that ρ leaves all sentences of the form $V(A)$ unchanged. □

We have tried in this section to argue for the plausibility of S. In part S is supported by a general thesis (the superthesis), in part it is supported by consistency and faithfulness properties, and by some of its consequences (see Section 4). If the reader remains in spite of this unconvinced of the soundness of S, she may still feel that one or more of its consequences (M_{PA}, M^{+}_{PA}, T), of which the content is somewhat clearer, are correct.

3.2. *Epistemic Principles about Epistemic Sentences*

We will now investigate to what extent the schematic axioms of the previous section can be generalized to range over all epistemic sentences.

Let us call these generalized principles S_{EA}, T_{EA}, M, M^{+}. At first blush, they appear to be quite natural generalizations of their restricted counterparts. But a simple argument shows that if we add T_{EA} as an extra axiom to EA (call the resulting system EAT) then the epistemic operator K collapses:[9]

PROPOSITION 2 (Fitch, Schumm). $\vdash_{EAT} A \rightarrow KA$.

Proof. Suppose $K(A \wedge \neg KA)$. Then KA and $K\neg KA$, so $KA \wedge K\neg A$, which entails a contradiction. Therefore we have $\neg K(A \wedge \neg KA)$ as a theorem of EAT, whence also $K\neg K(A \wedge \neg KA)$ (by the necessitation rule for K). We can reformulate this as $K\neg K\neg(A \rightarrow KA)$. So by T_{EA} and the reflexivity axiom, we obtain $A \rightarrow KA$. □

Let us pause for a moment to look at this proposition and its proof. The argument produces a sentence of $\mathcal{L}_{EA}$ of which it is possible to show that it is absolutely irrefutable without actually proving it, which explains why our motivation for T cannot be repeated here. The proof is not of an arithmetical nature, it consists in a manipulation of the epistemic principles of EA. And the derived consequence of taking T_{EA} as an extra axiom is clearly unacceptable. So T_{EA}, and hence also S_{EA}, are not viable as axiom candidates.

Note that the argument of Proposition 2 cannot be strengthened to a proof of $T_{EA} \vdash_{EA} A \rightarrow KA$. In the argument, the necessitation rule for K is applied to a formula which is itself obtained by an application of T_{EA}. This is possible only because T_{EA} is a theorem of EAT. Nevertheless, it is easy to see that for all $A \in \mathcal{L}_{HA}$: $T_{EA} \vdash_{EA} V(\neg\neg A)_i \rightarrow V(A_i)$. In other words, from T_{EA} we can derive the translation of the intuitionistic reading of the law of excluded third. This surely makes T_{EA} (and a fortiori S_{EA}) suspect even as a hypothesis.[10]

What then are we to say about M and M^+? We maintain that some plausibility can be attributed to these principles, although the support that we adduce for them is somewhat weaker than that for T.

Let us begin by looking at M and M^+ from a conceptual point of view. The argument about T_{EA} shows that the fact that they appear to be natural generalizations of M_{PA} and M^+_{PA} can be very misleading. But this is not all that we can say. If we let $\neg_i$ be the translation of the intuitionistic negation in L_{EA} (i.e., $\neg_i A =_{\text{def}} K\neg KA$), then the scheme M is equivalent to the scheme $\neg_i A \rightarrow \neg\neg_i\neg A$. Shapiro regards this as a "plausible statement of consistency between the classical and the intuitionistic negation", but hastens to add that nevertheless the precise mathematical content of M is not easy to determine ([35]). Most significantly, perhaps, we will show in the next section that $EA + M^+$ proves a very natural generalization of a theorem of HA, which apparently cannot be derived in EA alone. This adds to the plausibility of M^+ (and hence also of M). Altogether, however, this conceptual support is somewhat indirect, and we will have to rely more than before on faithfulness properties.

Using the same trick as before, we easily show that M and M^+ are consistent with EA. It is a little more difficult to show that they cannot be used to prove constructivistic statements that cannot be proved in EA alone.

LEMMA 3. *For all* $A \in \mathcal{L}_{EA}$: $EA + M \vdash V(A) \Rightarrow EA \vdash V(A)$.

Proof. This can be shown using the method of [8]. Inspection of their proof makes it clear that it suffices to show that the reverse translation (see [8, p. 56]) of the instances of M can be derived in HA. This property can be verified to hold. □

DEFINITION 4 (the translation ς). Let ς be the translation which transforms any sentence of the form

$$\neg K\exists x\neg(\neg KA(x) \rightarrow K\neg A(x))$$

into

$$\neg\exists x K \neg(\neg K A(x) \rightarrow K \neg A(x)),$$

and leaves all other sentences unaffected.

Using this shallow translation we prove the following theorem:

THEOREM 5. *For all* $A \in \mathcal{L}_{HA}$: $EA + M^{+} \vdash V(A) \Rightarrow EA \vdash V(A)$.

Proof. By induction on the length of proofs in $EA + M^{+}$ it is shown that ς translates theorems of $EA + M^{+}$ into theorems of $EA + M$. Then the result follows from the observation that ς leaves V-images unchanged and the previous lemma. □

Pankrat'ev claims that even when we extend EA with the Grz scheme

$$K(K(A \rightarrow KA) \rightarrow A) \rightarrow A,$$

no new translations of constructivistic statements can be proved [29].[11] This suggests that we investigate the plausibility of the Grz axiom. But since I find it hard to obtain intuitions concerning this scheme, I prefer to defer judgement.

There may exist mutually incompatible principles such that when one of them is added to EA, no new constructivistic statements can be proved. Given the vagueness surrounding the notion of provability in principle, it may in such a case be a complicated matter to judge which one of them should be rejected. We may come to the conclusion that some such principles just do not have a determinate truth value, because the concept of provability in principle *in se* is not well enough determined to decide the matter (as it is argued to be the case with the notion of set). Perhaps M^{+} (or even S) is already in that category.

In sum, we see that the motivation of generalized versions of the principles discussed in the previous section is a much trickier affair. The generalized versions of S and T are not viable axiom candidates, and the support that we can give for M and M^{+} is weaker than the support for their restricted versions.

4. Controversial Constructivistic Principles

4.1. *Markov's Principle and the Intuitionistic Version of Church's Thesis*

Markov's principle is usually stated as follows (in $\mathcal{L}_{HA}$):

$$(\forall x(A(x) \vee \neg A(x)) \wedge \neg\neg\exists x A(x)) \rightarrow \exists x A(x).$$

Let us abbreviate this principle as MP. MP is often discussed in the context of Constructive Recursive Mathematics (the Russian version of constructivism). This constructivistic school identifies proofs with Markov algorithms (or Turing machines, we may say). The justification which they give for MP is this. If $A(x)$ is constructively decidable, then there must be a Turing machine e which decides it. If we also know that there can be no proof of $\neg\exists x A(x)$, then in particular e cannot be such a proof. Hence if we start e on the numbers $0, 1, 2, \ldots$, eventually we find a number n for which e tells us that it has the property $A(x)$.

On Heyting's proof interpretation of the logical connectives, the content of MP can be expressed roughly as follows: **p** *is a proof of (an instance $A(x)$ of) Markov's principle just in case* **p** *is a mathematical procedure which successfully converts a proof* **q** *that $A(x)$ is intuitionistically decidable and that it is impossible for $A(x)$ to hold universally into a number* $\mathbf{p}(\mathbf{q})_1$ *and a proof* $\mathbf{p}(\mathbf{q})_2$ *such that the latter is a proof that $A(x)$ holds of the former.* Even though on certain constructive interpretations of the logical connectives its acceptability is well-known, the acceptability of Markov's principle as a *general* principle of constructive arithmetic is notoriously hard to decide [40, p. 204].

The following scheme is a simplification of an approximation to the translation of MP in $\mathcal{L}_{EA}$:

AXIOM 5 (MP_{HA}). *For all $(A(x))_i$ that are translations of sentences $A(x) \in \mathcal{L}_{HA}$:*

$$(K\forall x(K(A(x))_i \vee K\neg K(A(x))_i) \wedge K\neg K\neg K\exists x(A(x))_i \\ \rightarrow \exists x(A(x))_i.$$

Since Markov's principle is independent of HA, we know by Theorem 1 that MP_{HA} is independent of EA.

In $\mathcal{L}_{EA}$ it is also possible to formulate versions of the axiom scheme MP_{HA} in which the schematic formula does not range over V-translations of formulas of $\mathcal{L}_{HA}$, but over all arithmetical formulas, or over all formulas of $\mathcal{L}_{EA}$:

AXIOM 6 (MP_{PA}). *For all $A(x) \in \mathcal{L}_{PA}$:*

$$(K\forall x(KA(x) \vee K\neg KA(x)) \wedge K\neg K\neg K\exists xA(x)) \\ \rightarrow \exists xA(x).$$

AXIOM 7 (MP_{EA}). *For all* $A(x) \in \mathcal{L}_{EA}$:

$$(K\forall x(KA(x) \vee K\neg KA(x)) \wedge K\neg K\neg K\exists xA(x)) \\ \rightarrow \exists xA(x).$$

Another controversial first-order scheme is the intuitionistic version of Church's thesis. If $\varphi_e(x)$ is the formalization in $\mathcal{L}_{HA}$ (and therefore also in $\mathcal{L}_{PA}$) of the function which computes the output that is generated when the Turing machine with gödel number e is started on a natural number x, then this principle can be expressed in $\mathcal{L}_{HA}$ as ICT:

$$\forall x\exists yA(x,y) \rightarrow \exists e\forall xA(x,\varphi_e(x)).$$

In $\mathcal{L}_{EA}$, ICT is translated as:

AXIOM 8 (ICT_{HA}). *For all* $(A(x,y))_i$ *that are translations of sentences* $A(x,y) \in \mathcal{L}_{HA}$:

$$K\forall x\exists yK(A(x,y))_i \rightarrow \exists eK\forall x(A(x,\varphi_e(x)))_i.$$

ICT is a strengthening of Church's thesis. Like Church's thesis, it connects effective methods with Turing machines. But unlike Church's thesis, it does this in the form of a *constructive* implication. Again, even though the soundness of ICT on certain interpretations is well-known, its acceptability as a *general* principle of constructivistic arithmetic is controversial.

We can formulate in $\mathcal{L}_{EA}$ a generalization of ICT in which the parameter ranges over all formulas of $\mathcal{L}_{EA}$:

AXIOM 9 (ICT_{EA}). *For all* $A(x,y) \in \mathcal{L}_{EA}$:

$$K\forall x\exists yKA(x,y) \rightarrow \exists eK\forall xA(x,\varphi_e(x)).$$

By developing a realizability interpretation for EA (*Flagg realizability*), Flagg has proved the consistency of $EA + ICT_{EA}$ ([7]). A considerably simpler proof of the consistency of $EA + ICT_{EA}$ was later found by Goodman ([16]). It would be interesting to know whether $EA + ICT_{EA}$ is conservative over $HA + ICT$, i.e. whether for all $A_i \in \mathcal{L}_{HA}$, if $EA + ICT_{EA} \vdash V(A_i)$, then $HA + ICT \vdash A_i$.

4.2. *Relations with Epistemic Principles*

We now establish some connections between these variants of constructivistic principles and the epistemic principles that we discussed earlier.

PROPOSITION 3 (Shapiro). *For all* $A(x) \in \mathcal{L}_{PA}$:

$$EA \vdash (K\forall x(KA(x) \vee K\neg A(x)) \wedge K\exists x A(x)) \\ \rightarrow K\exists x KA(x).$$

Proof. [33, p. 30 (Theorem 19)]. □

LEMMA 4. $EA + T \vdash MP_{PA}$.

Proof. An easy derivation shows that MP_{PA} follows in $EA+T$ from

$$(K\forall x(KA(x) \vee K\neg A(x)) \wedge K\exists x A(x)) \\ \rightarrow K\exists x KA(x).$$

Hence the desired result follows from Proposition 3. □

So we have a natural variant of MP, of which there appears to be no proof in EA, and which follows from T.

Using this lemma, we establish a statement linking T to the intuitionistic version of Church's thesis and Markov's principle. The proof idea behind this statement is the following. Lemma 4 says that in EA we can derive MP_{PA} from T. So if we want to derive MP_{HA}, we are done if we can reduce its derivability to the derivability of MP_{PA}. This is what ICT_{HA} allows us to do. It guarantees the existence of a Turing machine $\varphi_e(x)$ with code e such that we can prove for every translation $A(x)$ of an intuitionistic predicate that $A(n) \longleftrightarrow \varphi_e(n) = 1$ and $\neg A(x) \longleftrightarrow \varphi_e(n) = 0$ for all natural numbers n. This means that the *arithmetical* formulas $\varphi_e(x) = 1$, $\varphi_e(x) = 0$, can replace the epistemic formulas $A(x)$, $\neg A(x)$, respectively, in the antecedent of MP_{HA}. This reduces this antecedent to an antecedent of an instance of MP_{PA}, and from there Lemma 4 takes over. Thus we have the following theorem:

THEOREM 6. $EA + T + ICT_{HA} \vdash MP_{HA}$.

Proof. Assume that we have $K\forall x(KA(x) \vee K\neg A(x))$ and $K\neg K\neg K\exists x A(x)$ for some formula $A(x)$ which is a translation of a formula of $\mathcal{L}_{HA}$.

$$A^*(x,y) \equiv K[(KA(x) \wedge K(y=1)) \\ \vee(K\neg A(x) \wedge K(y=0))].$$

Clearly $A^*(x,y)$ is the translation of an intuitionistic formula, and it can be simplified to

$$K[(KA(x) \wedge (y=1)) \vee (K\neg A(x) \wedge (y=0))].$$

It can easily be shown that the following two statements hold:

$$(1) \qquad K\forall x\big[(KA(x) \longleftrightarrow A^*(x,1)) \wedge (K\neg A(x) \longleftrightarrow A^*(x,0))\big],$$

$$(2) \qquad K\forall x(A^*(x,1) \longleftrightarrow \neg A^*(x,0)).$$

Now suppose $K\forall x(KA(x) \vee K\neg A(x))$. It follows by (1) that $K\forall x\exists y KA^*(x,y)$. If we then apply ICT_{HA}, it follows that $K\forall x\exists y KA^*(x,\varphi_e(x))$ for some Turing machine with Gödel number e. From this we can infer by the definition of $A^*(x,y)$ that $K\forall x(KA^*(x,1) \vee KA^*(x,0))$.
From this, in turn, we obtain

$$K\forall x(K(\varphi_e(x) = 1) \vee K(\varphi_e(x) = 0)).$$

For suppose that $KA^*(x,1)$. Then if $\varphi_e(x) \neq 1$, then either $\varphi_e(x) = 0$ or $\varphi_e(x) > 1$, so either $A^*(x,0)$ or $A^*(x,n)$ for some $n > 1$. In the former case, we contradict (2); in the latter case we are in conflict with the definition of $A^*(x,y)$. So we have $\varphi_e(x) = 1$, and since the premise is of the form $K\psi$, we obtain $K(\varphi_e(x) = 1)$. Similarly, we can prove from the supposition that $KA^*(x,0)$ that $K(\varphi_e(x) = 0)$, so we have indeed

$$K\forall x(K(\varphi_e(x) = 1) \vee K(\varphi_e(x) = 0)).$$

On the other hand, we can prove that

$$\exists x A(x) \Longleftrightarrow \exists x A^*(x,1) \Longleftrightarrow \exists x(\varphi_e(x) = 1).$$

Therefore,

$$K\big[\exists x A(x) \longleftrightarrow \exists x(\varphi_e(x) = 1)\big].$$

So from $K\neg K\neg K\exists x A(x)$, we may infer $K\neg K\neg K\exists x(\varphi_e(x) = 1)$.

From this and

$$K\forall x(K(\varphi_e(x) = 1) \vee K(\varphi_e(x) = 0)),$$

we may by Lemma 4 infer to $K\exists x KA(x)$. □

COROLLARY 1. *For all* $A_i \in \mathcal{L}_{HA}$: $HA + MP + ICT \vdash A_i \Leftrightarrow EA + T + ICT_{HA} \vdash V(A_i)$.

Proof. The $\Rightarrow$-direction follows from Theorem 6 and the soundness direction of Theorem 1. The $\Leftarrow$-direction follows from Theorem 4. □

Like Gödel's Dialectica Interpretation ([13]), Theorem 6 gives us a way to prove MP without having to assume it beforehand. $HA+MP+ICT$

is taken to be a good axiomatization of the first-order theory of arithmetic of the Russian constructivistics ([3, pp. 16–17]).[12] Therefore $EA+T+ICT_{HA}$ can be taken as an *interpretation* of Russian constructivism. Corollary 1 says that when we look at constructivistic sentences, in the context of ICT, the principle T proof-theoretically plays the role of MP. In other words (and put rather loosely), even though T does not prove new constructivistic statements in the context of HA (Theorem 4), it does prove new constructivistic statements in the context of $HA+ICT$.

Using the same reasoning as in the proof of Theorem 6, MP_{EA} can be derived from ICT_{EA} and T:

THEOREM 7. $EA+T+ICT_{EA} \vdash MP_{EA}$.

Proof. Similar to the proof of Theorem 6. □

CONJECTURE 1. For all $A_i \in \mathcal{L}_{HA}$: $HA + MP + ICT \vdash A_i \Leftrightarrow EA+T+ICT_{EA} \vdash V(A_i)$.

The left-to-right direction follows from Theorem 7. I do not know how to prove the other direction. There is, however, a fairly straightforward proof of the weaker statement that $EA+T+ICT_{EA}$ is consistent.[13] To show this, we first introduce some notation:

DEFINITION 5 (constructivization of a formula). For any given formula $A \in \mathcal{L}_{EA}$, let A^{-K} be the result of removing all occurrences of K from A. Then the constructivization of A (denoted as $\mathcal{C}(A)$) is defined as $V(A^{-K})$.

By analogy, the constructivization of a *proof* $A_1, \ldots, A_n,\ A$ is defined as

$$\mathcal{C}(A_1), \ldots, \mathcal{C}(A_n), \mathcal{C}(A).$$

PROPOSITION 4. *$EA+T+ICT_{EA}$ is consistent.*

Proof. Suppose we have a proof in $EA+T+ICT_{EA}$ of $\underline{0}=\underline{1}$, and let HEA be the constructive fragment of EA (i.e. the theory which is just like EA except that it lacks the principle of excluded third). Then there will be a proof $\mathcal{P}$ in $HEA+T+ICT_{EA}$ of $\underline{0}=\underline{1}$ (this can be seen by a simple induction on the length of proofs). Now consider the constructivization of $\mathcal{P}$. $\mathcal{C}(\mathcal{P})$ can be considered as a proof of $\underline{0}=\underline{1}$ in $HA+ICT$. But since $HA+ICT$ is consistent, no such proof can exist. □

Note that the same reasoning yields a very simple consistency proof for $EA+ICT_{EA}$.

Our last theorem of this section connects M^+ with the generalized form of MP. By a theorem of Luckhardt, the scheme $\neg\neg MP$ (of $\mathcal{L}_{HA}$) is provable in HA [23, p. 73]. We can prove a somewhat analogous proposition in EA for the stronger scheme MP_{EA}:

THEOREM 8. $EA + M^+ \vdash \neg K \neg MP_{EA}$.

Proof. Suppose that we have a formula $A(x) \in \mathcal{L}_{EA}$ such that:

(3) $\quad K\forall x(KA(x) \vee K\neg A(x))$

(4) $\quad K\neg K\neg K\exists x A(x)$

(5) $\quad K\forall x\neg KA(x)$

Combining (3) and (4), we see that $\forall x K\neg KA(x)$.
We also know that $\neg K\exists x A(x)$. For suppose $K\exists x A(x)$. Then $\exists x A(x)$, so suppose $A(\underline{b})$. Then we have $\neg K\neg A(\underline{b})$. But from (5) we also have $\neg KA(\underline{b})$. So we have

$$\exists x(\neg K\neg A(x) \wedge \neg KA(x)).$$

Since all premises are of the form $K\psi$, we can infer to

$$K\exists x(\neg K\neg A(x) \wedge \neg KA(x)),$$

contradicting M^+.
Furthermore, we have $\neg K\neg\exists x A(x)$. For suppose $K\neg\exists x A(x)$. Then

$$K\neg K\neg K\neg\exists x A(x).$$

But from (4) we know that $K\neg K\neg K\exists x A(x)$. So we have

$$K\neg K\neg K(\exists x A(x) \wedge \neg\exists x A(x)),$$

which is easily refuted.
So we have

$$\neg K\exists x A(x) \wedge \neg K\neg\exists x A(x),$$

i.e. $\exists x A(x)$ is undecidable in principle. Since all the premises are of the form $K\psi$, we obtain

$$K(K\exists x A(x) \wedge \neg K\neg\exists x A(x)).$$

But this conflicts with M, which is a consequence of M^+. $\square$

So even though M^+ does not prove new constructivistic statements, it does prove a natural generalization of a theorem of HA of which there appears to be no proof in EA alone. We regard this as providing support for M^+.

5. The Disjunction Property and the Numerical Existence Property

In [9, pp. 27–28], Flagg lists three conditions that an epistemic framework F must meet in order to serve as a reasonable synthesis of classical and constructive mathematics: the existence of appropriate faithfulness theorems, the definability of certain classical and constructive operators (in particular: we want a disjunction and an existential quantifier which have the disjunction property and the numerical existence property, respectively), and the consistency with F of certain problematic principles of intuitionism. We have seen in Sections 3 and 4 that most of the systems discussed in this paper have the appropriate faithfulness properties. And these systems contain versions of Church's thesis and Markov's principle, so they are a fortiori consistent with them. In this section we show that they also have the disjunction property and the numerical existence property. Therefore we suggest that they give at least a partial epistemic synthesis of the lawlike intuitionistic arithmetical universe.

We show that the systems $EA + T$, $EA + S$, $EA + M^+$, $EA + T + ICT_{HA}$ have the disjunction property and the numerical existence property.

PROPOSITION 5. *$EA+T$ has the disjunction property and the numerical existence property.*

Proof. It suffices to show that $\mid K\neg KA \rightarrow K\neg A$ for all $A \in \mathcal{L}_{PA}$, where $\mid$ is Shapiro's adaptation of the Kleene slash (see [34, p. 18]). Assume $\mid K\neg KA$, i.e. $\mid \neg KA$ and $\vdash_{EA+T} \neg KA$. Now erasing all K's in a proof of $EA + T$ results in a proof of PA. So $\vdash_{PA} \neg A$, whereby trivially $\vdash_{EA} \neg A$. But by the soundness of PA, we also have $\neg A$ true in the natural number structure. So by [34, p. 18, Lemma 3], we have $\mid \neg A$. Therefore we have $\vdash_{EA} \neg A$ and $\mid \neg A$, i.e. $\mid K\neg A$. This means that if $\mid K\neg KA$, then $\mid K\neg A$, whereby $\mid K\neg KA \rightarrow K\neg A$. □

PROPOSITION 6. *$EA + S$ has the disjunction property and the numerical existence property.*

Proof. Essentially the same as the proof of Proposition 5. □

PROPOSITION 7. *$EA+M^+$ has the disjunction property and the numerical existence property.*

Proof. It suffices to show that $\mid \neg K\exists x(\neg KA(x) \wedge \neg KA(x))$. The following equivalences hold:

$$
\begin{aligned}
&\mid \neg K\exists x(\neg KA(x) \wedge \neg KA(x)) \Leftrightarrow \\
&\text{not } \mid K\exists x(\neg KA(x) \wedge \neg KA(x)) \Leftrightarrow \\
&\quad \nvdash_{EA+M^+} \exists x(\neg KA(x) \wedge \neg KA(x)) \\
&\quad \text{or not } \mid \exists x(\neg KA(x) \wedge \neg KA(x))
\end{aligned}
$$

But the first disjunct of this last equivalence has to hold, otherwise $EA + M^+$ would be inconsistent. □

In order to prove that $EA + T + ICT_{HA}$ has the disjunction property and the numerical existence property, we have to modify Shapiro's modification of the Kleene slash, since $\mid ICT_{HA}$ fails to hold. We do this by (in a way) combining Shapiro's modification of the Kleene slash with the ordinary numerical realizability interpretation for HA.

DEFINITION 6 (modification of Shapiro's variant of the Kleene slash).

1. If A is atomic, then $\|A$ if and only if $EA + T + ICT_{HA} \vdash A$.
2. $\|A \wedge B$ if and only if $\|A$ and $\|B$.
3. $\|A \vee B$ if and only if either $A \vee B = V(C_i)$ for some $C_i \in \mathcal{L}_{HA}$ and $EA + T + ICT_{HA} \vdash V(C_i)$, or: $\|A$ or $\|B$.
4. $\|A \rightarrow B$ if and only if either: if $\|A$ then $\|B$, or $K(A \rightarrow B) = V(C_i)$ for some $C_i \in \mathcal{L}_{HA}$ and $EA + T + ICT_{HA} \vdash V(C_i)$.
5. $\|\neg A$ if and only if not $\|A$.
6. $\|\forall xA$ if and only if $\|A(\underline{n})$ for all $n \in \omega$.
7. $\|\exists xA$ if and only if either $\|A(\underline{n})$ for some $n \in \omega$, or $\exists xA = V(C_i)$ for some $C_i \in \mathcal{L}_{HA}$ and $EA + T + ICT_{HA} \vdash V(C_i)$.
8. $\|KA$ if and only if either $\|A$ and $EA + T + ICT_{HA} \vdash A$, or $KA = V(C_i)$ for some $C_i \in \mathcal{L}_{HA}$ and $EA + T + ICT_{HA} \vdash V(C_i)$.

LEMMA 5. *For all $A \in \mathcal{L}_{PA}$, $\|A$ if and only if A is true.*
Proof. Similar to the proof of [34, p. 18, Lemma 3]. □

LEMMA 6. *For all $A \in \mathcal{L}_{EA}$, if $EA + T + ICT_{HA} \vdash A$, then $\|A$.*
Proof. By induction on the length of the proof of A. □

LEMMA 7. *$HA + MP + ICT$ has the disjunction property and the numerical existence property.*
Proof. [38, p. 179]. □

THEOREM 9. *$EA + T + ICT_{HA}$ has the disjunction property and the numerical existence property.*

Proof. We only verify that the disjunction property holds (the verification of the numerical existence property is similar).

Suppose $EA + T + ICT_{HA} \vdash KA \vee KB$. By Lemma 6, it follows that $\|KA \vee KB$.

1. If $KA \vee KB = V(C_i \vee D_i)$ for some $C_i,\ D_i \in \mathcal{L}_{HA}$, then $HA + MP + ICT \vdash C_i \vee D_i$ by Corollary 1 (the faithfulness theorem for $EA + T + ICT_{HA}$). Therefore by Lemma 7, we have $HA + MP + ICT \vdash C_i$ or $HA + MP + ICT \vdash D_i$, whereby $EA + T + ICT_{HA} \vdash V(C_i)$ or $EA + T + ICT_{HA} \vdash V(D_i)$.
2. If $KA \vee KB \neq V(C_i \vee D_i)$ for any $C_i,\ D_i \in \mathcal{L}_{HA}$, then $\|KA$ or $\|KB$. Suppose $\|KA$ (the other case is similar). Then $EA + T + ICT_{HA} \vdash A$, whereby $EA + T + ICT_{HA} \vdash KA$. □

CONJECTURE 2. $EA + T + ICT_{EA}$ has the disjunction property and the numerical existence property.

This statement seems harder to prove.

6. CONCLUDING PHILOSOPHICAL REMARKS

Shapiro's theory of Epistemic Arithmetic provides a good framework for an investigation into the relations between classical and constructivistic arithmetic. This is so even if the translation V does not preserve the exact meaning of the constructivistic statements.[14] For it does seem to be the case that when classical mathematicians "learn" the meanings of the constructivistic connectives, or do constructivistic mathematics, they implicitly use an "absolute" notion of provability (as [11] suggests) and a translation function that resembles V. The claim of this paper is that logical principles concerning provability in principle play a role in the intuitions that mathematicians have concerning the acceptability of certain variants of problematic constructivistic principles.

Of course more work remains to be done. An obvious question is whether the approach of this paper can be fruitfully extended to second-order arithmetic and set theory, and to variants of constructivistic choice and continuity principles that can be expressed in these languages. Even in the first-order language of EA there may be many more epistemic principles that merit consideration, and there may be more connections with variants of constructivistic principles to be established. But all that is left for another occasion.

The notion of provability in principle is regarded as philosophically suspect, or even incoherent by many classical mathematical logicians (even though these feelings are seldom made explicit). Certainly the notion is usually regarded as less respectable than the venerable notion of truth. I am not sure that these doubts are justified, and there seem to me to be two ways to help dissolving them. First, there is a need of conceptual clarification of the notion of provability in principle. The present paper hardly makes a contribution to this task. But secondly, it needs to be shown that the notion is fruitful, that something interesting can be done with it. That is what this paper intends to show, and it thereby attempts to be a small step in the direction of making the notion of provability in principle respectable.

ACKNOWLEDGEMENTS

The research for this paper was supported by a grant from the Belgian National Fund for Scientific Research, which is gratefully acknowledged. I want to thank Stewart Shapiro, George Schumm, Tony Anderson, Natasha Kurtonina, Herman Roelants and an anonymous referee for valuable comments on earlier versions of this paper.

NOTES

[1] Around the same time, Reinhardt independently developed similar epistemic formalizations of arithmetic (see [31], [32]).

[2] The classical paper on informal rigour is [20], but see also [18, 19, 21, 22].

[3] See [37, 39; 40, Chapter 4; 28, 3, 4, 20].

[4] For a good introduction to these two research traditions in constructivistic mathematics, and their characteristic principles, see [40, Chapter 4].

[5] The notion of absolute knowability does not accord very well with this idea: one might argue on the basis of knowledge of the number of atoms in the universe that there is a fixed bound on the complexity of proofs that can be constructed, and hence that some mathematical questions are "absolutely undecidable". We do not want to count the conclusion of this argument for that reason as being in the intended extension of K, because the argument is not a *proof*, i.e. it is not a nonempirical argument.

[6] Not everyone agrees with Shapiro on this score. The soundness of the transitivity axiom $\Box A \rightarrow \Box\Box A$ for the intended interpretation of K, for instance, is questioned by Martin-Löf ([25]).

[7] Note by the way that axiom S is a quite strong way of giving content to Thesis 1. Suppose $Q\vec{x} = \exists x$, and $i = 1$. Then axiom S states

$$K\exists x \neg K A_1(x) \rightarrow \exists x K \neg A_1(x).$$

In other words, axiom S states that a form of *de dicto* provability supervenes on a corresponding *de re* provability, namely on the provability of a particular natural number that it has the property $\neg A_1(x)$.

[8] The distinction between empirical and non-empirical (i.e. a priori) reasoning has been criticized by Quine (see [30]) and his followers. But the last decade has seen a growing consensus that there is a clear and useful distinction here. First and perhaps foremost, Kripke deserves to be credited for clearly distinguishing pairs of notions that were often used in the literature as though they are interchangeable (see [17, pp. 34–41]), which made it seem as if criticism of any of these pairs of notions (e.g. a questioning of the usefulness of the distinction between the analytic and the synthetic) tells equally against any of the other distinctions (e.g. the empirical/a priori distinction). The empirical/a priori distinction is argued to be respectable and is further analyzed in [1] and in [5].

[9] This argument was given to me by George Schumm. It is a version of an argument in [6].

[10] There appears to be an analogy between T_{EA} and the naive comprehension axiom, or the axiom of full determinacy in set theory. Even though these principles have an initial plausibility, they can on closer inspection be seen to be false. But restricted versions of them might well be true.

Whether they are viable axiom candidates is yet a further question. Since determininacy principles have recently been shown to follow from large cardinal axioms, they are of course no longer viable axiom candidates (if they ever were).

[11] Since $EA + Grz$ is stronger than $EA + M$, this implies that M does not allow us to derive new constructivistic statements. We have not been able to verify Pankrat'ev's claim that the "reverse translation" of the Grz axiom in provable in EA (the computation seems complicated).

[12] According to Troelstra, most Russian constructivists accept a somewhat stronger form of the intuitionistic version of Church's thesis ([40, p. 202]).

[13] We cannot use the technique of [16] to prove the consistency of $EA + T + ICT_{EA}$, since T is not Flagg realizable.

[14] For an argument to this effect, see [36]. The arguments given on both sides of this issue seem to me to be inconclusive; I am not even sure how to formulate the problem in a way that makes an unequivocal answer possible.

REFERENCES

1. Anderson, C.A.: Toward a logic of a priori knowledge, Forthcoming in *Philosophical Topics*.
2. Boolos, G.S. and Jeffrey, R.C.: *Computability and logic*, U.P. Cambridge, 1989.
3. Beeson, M.: *Problematic principles in constructivistic mathematics*, in: D. Van Dalen, et al. (eds), *Logic Colloquium '80*, North-Holland, 1982, 11–55.
4. Beeson, M.: *Foundations of constructive mathematics. Metamathematical studies*, Springer, 1985.
5. Burge, T.: Content preservation, *Philosophical Review* 102 (1993), 457–488.
6. Fitch, F.B.: A logical analysis of some value concepts, *Journal of Symbolic Logic* 28 (1963), 135–142.
7. Flagg, R.: *Church's thesis is consistent with epistemic arithmetic*, in: S. Shapiro (ed), *Intensional Mathematics*, North-Holland, 1985, 121–172.
8. Flagg, R. and Friedman, H.: Epistemic and intuitionistic formal systems, *Annals of Pure and Applied Logic* 32 (1986), 53–60.

9. Flagg, R.: Integrating classical and intuitionistic type theory, *Annals of Pure and Applied Logic* 32 (1986), 27–51.
10. Friedman, H. and Sheard, M.: The equivalence of the disjunction and existence properties in modal arithmetic, *Journal of Symbolic Logic* 54 (1989), 1456–1459.
11. Gödel, K.: An interpretation of the intuitionistic propositional calculus. Reprinted in: S. Feferman, et al. (eds), *Kurt Gödel. Collected works. Volume I*, U.P. Oxford, 1986, 296–302.
12. Gödel, K.: Some basic theorems on the foundations of mathematics and their implications. Reprinted in: S. Feferman, et al. (eds), *Kurt Gödel. Collected works. Volume III*, U.P. Oxford (forthcoming).
13. Gödel, K.: On a hitherto unutilized extension of the finitary viewpoint. Reprinted in: S. Feferman, et al. (eds), *Kurt Gödel. Collected works. Volume II*, U.P. Oxford, 1990, 217–251.
14. Gödel, K.: What is Cantor's continuum problem? Reprinted in: S. Feferman, et al. (eds), *Kurt Gödel. Collected works. Volume II*, U.P. Oxford, 1990, 254–270.
15. Goodman, N.: Epistemic arithmetic is a conservative extension of intuitionistic arithmetic, *Journal of Symbolic Logic* 49 (1984), 192–203.
16. Goodman, N.: Flagg realizability in epistemic arithmetic, *Journal of Symbolic Logic* 51 (1986) 387–392.
17. Kripke, S.: *Naming and necessity*, U.P. Harvard, 1980.
18. Kreisel, G.: Mathematical logic: what has it done for the philosophy of mathematics? in: R. Schoenman (ed), *Bertrand Russell. Philosopher of the century*, Little, Brown and Co, 1967, 201–272.
19. Kreisel, G.: The formalist-positivist doctrine of mathematical precision in the light of experience, *L' age de la science* 3 (1970), 17–46.
20. Kreisel, G.: Informal rigour and completeness proofs, in: I. Lakatos (ed), *Problems in the philosophy of mathematics. Proceedings of the international colloquium on the philosophy of science (London, 1965), volume 1*, North-Holland, 1972, 138–186.
21. Kreisel, G.: Frege's foundations and intuitionistic logic, *Monist* 67 (1984), 72–91.
22. Kreisel, G.: Church's thesis and the ideal of informal rigour, *Notre Dame Journal of Formal Logic* 28 (1987), 499–519.
23. Luckhardt, H.: Über das Markov-Prinzip, *Archiv für Mathematische Logik* 18 (1976), 73–80.
24. Mates, B.: *Elementary Logic*, Oxford University Press, 1972.
25. Martin-Löf, P.: Truth of a proposition, evidence of a judgement, validity of a proof, *Synthese* 73 (1987), 407–420.
26. Mints, G.E.: On Novikov's hypothesis, in: G.E. Mints (ed), *Selected papers in proof theory*, Bibliopolis and North-Holland, 1992, 147–151.
27. Myhill, J.: Some remarks on the notion of proof, *Journal of Philosophy* 57 (1960), 461–471.
28. Myhill, J.: Notes towards an axiomatization of intuitionistic analysis, *Logique et Analyse*, 35–36 (1966), 280–297.
29. Pankrat'ev, N.: Predicate calculus and arithmetic with Grzegorczyk, *Moskow University Mathematics Bulletin* 44.5 (1989), 83–85.
30. Quine, W.V.O.: Two dogmas of empiricism. Reprinted in: W.V.O. Quine (ed), *From a logical point of view*, MIT Press, 1953, 20–46.
31. Reinhardt, W.: Absolute versions of incompleteness theorems, *Noûs* 19 (1985), 317–346.

32. Reinhardt, W.: Epistemic theories and the interpretation of Gödel's incompleteness theorems, *Journal of Philosophical Logic* 15 (1986), 427–474.
33. Shapiro, S. (ed): *Intensional Mathematics*, North-Holland, 1985.
34. Shapiro, S.: Epistemic and intuitionistic arithmetic, in: S. Shapiro (ed), *Intensional Mathematics*, North-Holland, 1985, 11–45.
35. Shapiro, S.: *Personal communication* (letter dated June 9, 1991).
36. Smorynski, C.: Review of S. Shapiro (ed): Intensional Mathematics, *Journal of Symbolic Logic* 56 (1991), 1496–1499.
37. Troelstra, A.S. (ed): *Metamathematical investigation of intuitionistic arithmetic and analysis*, Springer, 1973.
38. Troelstra, A.S.: Notes on intuitionistic second-order arithmetic, in: A. Mathias and H. Rogers (eds), *Cambridge summer school in mathematical logic*, Springer, 1973, 171–205.
39. Troelstra, A.S.: Aspects of constructive mathematics, in: J. Barwise (ed), *Handbook of mathematical logic*, North-Holland, 1977, 973–1052.
40. Troelstra, A.S. and Van Dalen, D.: *Constructivism in mathematics. An introduction. Volume I*, North-Holland, 1988.

Center for Logic and Philosophy of Science
Institute of Philosophy, University of Leuven
B-3000 Leuven, Belgium
E-mail: Leon.Horsten@hiw.kuleuven.ac.be