Antonutti Marfori, M., & Horsten, L. F. M. (2019). Human-effective computability. *Philosophia Mathematica*, *27*(1), 61-87. [nky011]. https://doi.org/10.1093/philmat/nky011

Human Effective Computability

Marianna Antonutti Marfori (Ludwig-Maximilians Universität München)
Leon Horsten (University of Bristol)

Abstract

We analyse Kreisel's notion of *human effective computability*. Like Kreisel, we relate this notion to a concept of informal provability, but we disagree with Kreisel about the precise way in which this is best done. The resulting two different ways of analysing human effective computability give rise to two different variants of Church's thesis. These are both investigated by relating them to transfinite progressions of formal theories in the sense of Feferman.

Introduction

In [Kreisel 1972], the notion of *human effective computability* is outlined and distinguished from the familiar notion of *algorithmic computability*. Roughly, a total function f on the natural numbers is said to be *human effectively computable* if for any natural number m given in canonical notation, a number n in canonical notation can be found such that it can be *proved* that $f(m) = n$. The notion of provability involved in this explication is an *informal* or *absolute* notion of provability by an idealised human agent which should not without a good argument be identified with provability in some antecedently given formal system. It is thus crucial in this context that the numbers are given in *canonical notation*, where for definiteness we say that a number is given in canonical notation if it is given by a Peano-numeral, and that it is kept in mind that since functions are infinite abstract objects, human agents—even in the idealised sense—do not have epistemic access to them independently of the interpreted linguistic expressions that denote them (call these expressions *function presentations*). This implies that given two different presentations f and g of the same function, it cannot be assumed that the existence of an informal proof of a statement of the form $f(m) = n$ indicates the existence of a proof of $g(m) = n$.[1]

[1] This issue will be discussed in more detail in §3.

In this paper we follow Kreisel's suggestion that human effective computability should be explicated in terms of a notion of *provability*. More precisely, we suggest that human effective computability can be explicated in terms of the notion of a priori knowability, and that the latter notion can be fruitfully investigated in the formal framework of Epistemic Arithmetic.

However, there are several ways of explicating a notion of human effective computability in terms of a notion of a priori knowability. We will investigate two such explications. The first was suggested by Kreisel himself, and later articulated in the context of Intensional Set Theory by Myhill. The second explication was proposed by Shapiro (in the context of the framework of Epistemic Arithmetic). We will argue that Shapiro's explication yields a more robust notion of computability than the Kreisel-Myhill notion. In particular, we will see that if we model a priori knowability in terms of transfinite progressions of formal theories, the properties of the Kreisel-Myhill notion of human effective computability are less stable than the ones of Shapiro's notion of human effective computability. More specifically, one can formulate versions of Church's thesis for the two notions of human effective computability. We then see that for Shapiro's notion of human effective computability, the relevant version of Church's thesis is true in transfinite progression models. For the Kreisel-Myhill notion of human effective computability, in contrast, Kreisel already observed that the truth value of the relevant version of Church's thesis is sensitive to the details of the model.

The structure of this article is straightforward. In the next two sections we introduce Kreisel's notion of human effective computability, and elaborate on the role of the idealisations involved in the notion of informal provability in terms of which the notion of human effective computability is analysed. Subsequently, we contrast Kreisel's way of analysing human effective computability in terms of informal provability with that of Shapiro. We then follow Kreisel by 'testing' these two notions by articulating models for them in terms of transfinite progressions of formal theories, and by investigating whether versions of Church's thesis hold in these models.

In this paper, we will restrict ourselves to notions of computability of *total* functions on the natural numbers. Also, in our formal investigation of Kreisel's notion of human effective computability, we will restrict ourselves to first-order modal languages.

1 Machine-effective computability versus human effective computability

1.1 The distinction

A function f is said to be *effectively* or *algorithmically computable* if there is a routine step-by-step procedure which, for each natural number n (given in some canonical notation), in a finite number of steps, yields the function value $f(n)$ (in canonical notation). Church's thesis for algorithmic computability then asserts the following:

Thesis 1 (Church's thesis) *A function f is algorithmically computable if and only if, for every natural number m given in canonical notation, a canonically given number n exists such that the statement $f(m) = n$ is formally provable (by finite steps in a logic).*[2]

In more contemporary terms, the thesis states that *every effectively or algorithmically computable function on the natural numbers is λ-computable*, or, equivalently, that *every function that is algorithmically computable is Turing computable*.[3]

[Turing 1936] gave a conceptual analysis of the notion of algorithmic computability for an idealised human agent which proceeds in a stepwise, routine way without technological aid nor insight or ingenuity. Many have argued that Turing's analytical argument does not and can not amount to a mathematical proof of the Church-Turing thesis (henceforth, CT) because mathematical theorems seem to only connect mathematical notions, while the antecedent of CT contains the informal notion of effective or algorithmic computability (see e.g. [Folina 1998], [Horsten 2006]).[4] However, it is now widely accepted that Turing's analysis of the notion of algorithmic computability establishes that CT is indeed true.[5] In fact, CT is often invoked in informal proofs in computability theory to

[2] This formulation of the thesis is intensionally closer to Church's original formulation in §7 of [Church 1936] than the version which defines the notion of an *effectively computable function* in terms of a λ-definable function of positive integers. See the discussion and footnote below.

[3] The latter is called *Turing's thesis*. Some (see e.g. [Soare 1996]) distinguish Church's thesis from Turing's thesis on the grounds that they are intensionally distinct. Despite the fact that this article focuses on some intensional aspects of different notions of computability, we will not distinguish Church's thesis from Turing's thesis as nothing that is being argued hinges on this distinction. Thus we will just speak of *Church's thesis*. For discussions of the conceptual analysis of algorithmic computability on the natural numbers, see e.g. [Sieg 1994], [Soare 1996].

[4] This is, however, a controversial matter. Some have argued that we should instead introduce new primitives into the language that allow us to talk directly about algorithms, axiomatise their fundamental properties, and then prove the Church-Turing thesis from these axioms. See e.g. [Shapiro 1981], [Mendelson 1990], [Sieg 1994], [Sieg 1997], [Sieg 2013].

[5] Kreisel acknowledges Turing's analysis as a particularly successful exercise in informal rigour [Kreisel 1987, p. 505].

infer from the informal description of how a function can be effectively computed, the conclusion that that function is recursive.

A function f is said to be *mechanically computable* if there exists a machine that for each natural number n (given in some canonical notation) yields the function value $f(n)$ (in canonical notation) in a finite amount of time. An obvious analogue for Church's thesis can be formulated for the notion of machine computability by a Discrete Deterministic Mechanical Device ($DDMD$). Gandy has shown that given certain reasonable conditions on what it means to be $DDMD$ computable, Church's thesis for computability by discrete deterministic mechanical machines also holds [Gandy 1980].

In [Kreisel 1972], Kreisel draws a distinction between *machine effective computability* and *human effective computability*. With "machine effective computability" he does not mean mechanical computability in the sense of Gandy, but rather algorithmic computability in the sense of Turing [Kreisel 1972, p. 314 and p. 318]. Furthermore, Kreisel holds that Turing's analysis establishes something stronger than what he calls Church's thesis (the statement that each informal description of an algorithm corresponds to a function computing the output of that algorithm), namely that an "intensional equality" holds between machine effective computability on the one hand, and Turing machine computability on the other hand [Kreisel 1972, p. 316]. More precisely, Kreisel holds that Turing's analysis establishes what he calls *Church's Superthesis*: the statement that "each m-effective definition [i.e. algorithm] is intensionally equal to some program for an 'idealized' computer".[6]

Kreisel does not explain in much detail what is meant by the notion of human effective computability, and it is not easy to paraphrase in clear terms what he does say about this notion. He claims that "in ['human effective computability'], 'effective' means humanly performable and not only mechanical" [Kreisel 1972, p. 314]. He holds that there is an intimate connection between human effective computations and proofs [Kreisel 1972, p. 315]. Indeed, he regards "[human] effectively definable functions as the analogue of provable theorems" [Kreisel 1972, p. 316]. This suggestion can be cashed out as follows: whereas algorithmic computability (algorithmic enumerability) is naturally seen as reducible to formal provability, human effective computability can be naturally seen as reducible to informal or absolute provability, which is a notion of provability that is not relativised to any given formal system.[7]

Informal provability differs from formal provability—and thus human effective com-

[6] See also [Kreisel 1971, p. 177], where Kreisel identifies Church's Superthesis with the statement that each informal description of an algorithm corresponds to "a more or less specific programme, modulo trivial conversions, which can be seen to define the same computation process as the rule".

[7] The connection between calculability (see §3) and proofs seems to have first been drawn by [Gödel 193?]; see particularly pp. 166–168. For an analysis of Gödel's views see [Sieg 2006]. However, a discussion of the points of contact and divergence with Gödel's view are beyond the scope of this paper.

putability from machine computability—in that it contains what Kreisel takes to be a *non-deterministic* element (such as the introduction of new axioms). Kreisel warns in this context against the uncritical presupposition that only a deterministic concept of provability (and thus machine computability) can be fruitfully investigated in a precise manner [Kreisel 1972, p. 319].[8]

At the moment, we do not seem to have resources available that allow us to argue conclusively that the extension of the notion of informal provability is captured by some formal notion.[9] This does not imply that a rigorous account of the notion of informal provability cannot be given. In the spirit of [Myhill 1960], we may try to treat the notion of informal provability as primitive and capture its logical properties axiomatically, or we may try to construct an informative class of models that reflects how the extension of the notion of informal provability is generated.[10]

At a first approximation, one might say that a function f is human effectively computable if and only if for every natural number m given in canonical notation, a canonically given number n exists such that the statement $f(m) = n$ is informally provable.[11] The question then presents itself whether Church's thesis for human effective computability also holds. Kreisel takes Church's thesis for human effective computability to amount to the thesis that every total function which has a human effective definition is recursive, so the question is whether for every function that is humanly computable, there is a Turing machine that computes it. Kreisel attributes to Turing the claim that Church's thesis for human effective computability holds for the same reasons that underlie the support for Church's thesis for machine effective computability. He believes that this is mistaken, because machine effective computability and human effective computability are distinct notions [Kreisel 1972, p. 319]. Moreover, he thinks it of cardinal theoretical importance that these two notions are not run together [Kreisel 1972, p. 314–315].

In sum, Kreisel held that Turing provided an intensional analysis of the notion of algorithmic computability which established that CT holds for this notion of computability, and that by doing so, Turing precisely characterised the limitations of the notion of algorithmic computability (i.e. the extension of what can be computed in a stepwise, routine fashion by an idealised computer). Similarly, we might hope that a rigorous analysis of the notion of human effective computability might establish or refute Church's thesis for human effective computability, thus giving us information about the scope and limitations

[8] This point will be discussed in more detail in the following subsection.

[9] Kreisel noted that already for certain more restricted notions of provability, such as finitist provability or predicative provability, we are in this position [Kreisel 1967, p. 157].

[10] A brief discussion of Kreisel's distinction between *genetic* and *axiomatic* theories for an informal notion—such as informal provability—in relation to informal rigour follows in §1.2.

[11] In §4 we will slightly revise this thesis.

of mathematical reasoning [Kreisel 1972, p. 316].

1.2 Informal rigour

The notion of *informal rigour* was first put forward in [Kreisel 1967] and subsequently expanded in later works such as [Kreisel 1972] and [Kreisel 1987], which particularly focuses on Church's thesis as a paradigmatic example of informal rigour.

Kreisel strongly opposes the idea that it is impossible to give a rigorous analysis of intuitive notions, such as the notions of *humanly performable instruction* or the notion of *informal* or *absolute provability*. For many such epistemological notions that are omnipresent in mathematics it is difficult to find the right level of idealisation, thus making many reluctant to the idea that a precise relation can be established between mathematical and non-mathematical (informal) notions. The difficulties do not concern the concepts themselves, but rather appear when explicit hypotheses are made concerning those notions [Kreisel 1972, p. 318]. This, however, does not exclude the possibility of a theory that rigorously characterises these notions intensionally, or possibly, extensionally; on the contrary, it should constitute an incentive to look into formal frameworks that could be used to "organise" the subject [Kreisel 1972, p. 331]. There is of course no guarantee that such theory will be found. But on Kreisel's view, we should resist the claim that such informal notions as human effective computability cannot have a well-determined idealisation: informal rigour can be used to make an informally characterised notion rigorous. In Kreisel's words,

> Informal rigour wants (i) to make this analysis as precise as possible (with the means available), in particular to eliminate doubtful properties of the intuitive notions when drawing conclusions about them; and (ii) to extend this analysis, in particular not to leave undecided questions which can be decided by full use of evident properties of these intuitive notions. [Kreisel 1967, p. 138–139]

When an informal or intuitive notion is made precise without appeal to arbitrary conventions, knowledge of informal notions can be combined with formal tools and used to establish new mathematical properties of those notions. This can produce the understanding needed to recognise whether and in what way the precise definitions of the informal notions in question contribute to solving problems that present themselves in the area to which the notions belong, and can establish the precise extent of that area [Kreisel 1987, p. 500].

Famously, [Kreisel 1967] presents what is now known in the literature as a *squeezing argument*. Suppose that we have an informally characterised concept. Then we have a strategy for making it rigorous if we can show that there are two precisely defined concepts which provide respectively necessary and sufficient conditions for falling under the informally characterised concept, and the two precisely defined concepts have the same extensions. As a paradigmatic example, informal rigour can be used to argue that the pre-theoretic concept of validity for first-order languages is captured by the model-theoretic concept of logical consequence, which can then be shown (by the completeness theorem) extensionally to coincide with the notion of classical first-order derivability. In other words, the extension of the informal concept of validity is precisely identified by being "squeezed" between the extension of the concept of soundness and the extension of the concept of completeness.

The squeezing argument is not the only strategy for achieving informal rigour. According to Kreisel, another chief example of a successful application of informal rigour is Turing's intensional analysis of the informal notion of algorithmic computability which is extensionally captured by the precisely defined class of Turing computable functions. The precise extension of the informal concept in question is not, in this case, determined by a squeezing argument, but it is established by the analysis of the idealisations involved in the informal notion. In this paper, we will not apply informal rigour to the notion of informal provability in the sense of the squeezing argument, but we will pursue a strategy closer to one that Kreisel describes with respect to CT.

A crucial factor in the success of Turing's analysis of algorithmic computability consists in identifying the right idealisations that are involved in the notion of stepwise computations on the natural numbers. The informal notion of computability and the mathematical one differ significantly from each other with respect to the idealisations of the agent or system for which the computation rules are to be effective [Kreisel 1987, p. 501]. The problem with applying informal rigour to the notion of human effective computability lies in the fact that it is much more difficult to identify the right idealisations [Kreisel 1972, p. 317]:

> Any [...] theory [of human effective computability] would seem to need an idealisation *far removed from our ordinary experience* (of human performances in mathematics). Consequently, we have not one, but two difficulties. If experience presents itself in such a way that the proper idealisation is difficult to *find* then, for the same reason, the idealisation may be difficult to *apply* even if it is found. In particular, there will now be a genuine problem of formulating *principles of evidence* or *adequacy conditions* for the validity of idealisations. Besides when idealisations are difficult to find there will, in general, be competing theories and hence the problem of *discovering* (observational) consequences which can be used to decide between different theories. *(emphasis in the original)*

In this article, we will apply informal rigour to the notion of human effective computability by following Kreisel's suggestion that human effectively definable functions constitute the analogue of provable theorems, and will attempt to explicate the notion of human effective computability in terms of a notion of provability. The picture is roughly as follows. At every moment in time, the extent of our a priori knowledge is generated by a finite set of basic principles.[12] As time progresses, more basic principles may come to be a priori known. There may be systematic aspects about this process. For instance, at a given point in time t' one may come to find a formal provability predicate for the extension of informal provability as it was a moment earlier at t, and come to realise in an a priori way that the extent of a priori knowability at time t is consistent (as expressed using this provability predicate). But there may also be unsystematic aspects about this process. This happens when a completely new axiom is adopted on a priori grounds, as was perhaps the case with the Axiom of Choice in set theory in the third decade of the twentieth century. It is especially the unsystematic aspect of this process that makes it difficult to draw general conclusions about the extension of the notion of a priori knowability in general.

Since the aim is that of providing a rigorous characterisation of Kreisel's notion of what an idealised but human mathematician can compute along the lines of the picture described above, what is needed is a notion of provability that captures the notion of what an idealised but human mathematician can prove *in principle*. Settling in advance for a notion of provability in some particular formal system would prejudge matters: there are some truths—e.g., the truth of a Gödel sentence—that we cannot prove in a specific formal system, but that we can nonetheless *prove* in the informal sense of the word 'prove'.[13] What is needed, therefore, is a rigorous way of characterising an *informal* or *absolute* notion of provability.[14]

The best possible outcome would be if we were able to formulate a *genetic theory of mathematical provability*, where the notion of a *genetic theory* is described by Kreisel as follows:

> The old aim [of the theory of genetic provability] was not merely to find an F such that every theorem (in the language of F) which can be proved at all should also possess *some* derivation in F, but we should be able to see *how*

[12] Some such principles may be schematic, for example the induction scheme for first-order arithmetic. Since the idealised agent can only entertain a finite number of principles in her mind, she cannot know (at any given moment in time) infinitely many instances of a scheme. In such cases, knowledge of the schematic principle allows her to come to know individual instances of the scheme, by recognising them as such.

[13] For more extensive discussions of this topic, see [Myhill 1960], [Leitgeb 2009] and [Antonutti 2010].

[14] For a more extended discussion of difficulties surrounding the modelling of the notion of informal provability see [Horsten 2005, §2].

> *to get from an intuitive proof to its formalization (in F).* [...] A paradigm of a genetic theory in the sense described (for machine-effective definitions in place of proofs) is provided by Turing's analysis: each machine-effective definition is *intensionally* equal to some program for an 'idealized' computer. [Kreisel 1972, p. 316]

The distinction is also clarified by saying that genetic theories of human effective definitions "tell us what objects we are talking about", while axiomatic theories "state properties of the objects considered without providing an explicit list of them" [Kreisel 1972, p. 315].

Kreisel then goes on to point out that we lack a genetic theory even for the concept of formal provability [Kreisel 1972, p. 316]. Part of the problem is that mathematicians are always inventing new languages to express new concepts: so we cannot stick to a fixed vocabulary (as Turing machines and other formal algorithms do).[15] For the concept of informal or absolute provability, the situation is even more problematic: not only we lack a genetic theory, but we do not have a mathematical characterization of its extension even if we restrict ourselves to the language of arithmetic. Of course, it is not difficult to find a lower bound for it (Peano Arithmetic, for instance). The problem is that we have at present no convincing arguments that allow us to impose a sharp upper bound on the extension of informal provability. However, this does not mean that informal rigour cannot be applied to the notion of informal provability. When the prospects for a genetic theory of a specific notion are unpromising, we could still have an *axiomatic* theory providing axioms for the intended interpretation of the notion in question [Kreisel 1972, p. 317]. Naturally, the problem of what idealisations are embedded in the interpretation of that notion would still arise, together with the question of whether Church's thesis holds for that notion.[16] Hence, the possibility of finding a convincing theory of informal provability depends on convincingly identifying the idealisations embedded in the notion of what an idealised mathematician could prove in principle. This will be the object of the next section.

2 Idealisations

Informal provability is not itself a mathematical notion, because the notion of what an idealised but human mathematician can prove in principle concerns the concept of ide-

[15] Thank you to anonymous referee for pointing this out.

[16] As Kreisel puts it, "while it is obvious that all genetic *formal* theories of effective definitions [...] satisfy Church's thesis, axiomatic theories need not" [Kreisel 1972, p. 315].

alised human knowledge, which is a philosophical notion (whereas formal provability—or derivability in a formal system—is of course a precisely definable mathematical notion).

We will need an *iterable* notion of informal provability, i.e. one where statements involving the notion of informal provability can themselves be informally provable (or refutable). For this reason, it should not be thought of as a notion of informal *mathematical* provability, but can best be understood as a notion of *a priori knowability* over a given base language, which we will take to be the language of first-order arithmetic. It is not ill-formed to claim of a mathematical statement ϕ that it is a priori knowable that ϕ is a priori knowable. However, since informal mathematical provability or a priori knowability in mathematics is not itself a mathematical notion, but an epistemic one, it cannot be *mathematically* provable that ϕ is a priori knowable.[17] Accordingly, in the rest of the paper we will use the terms "informal provability" and "a priori knowability" interchangeably in the sense outlined above.

The formal framework of *Epistemic Arithmetic* ([Shapiro 1985a]) provides a suitable formal framework for the rigorous treatment of the notion of a priori mathematical knowability (see §3 below). Since there presently exists no suitable formal framework for *directly* axiomatising the informal notion of human effective computation, and since the notion of human effective computation seems to presuppose the notion of informal provability, Epistemic Arithmetic may be seen as a promising framework for mediately analysing an iterable notion of human effective calculability.

Even though there are difficult problems with the formalisation of informal provability are related to difficulties of identifying the right idealisations involved in the notion, there nonetheless are some idealisations for a priori knowability that seem *prima facie* reasonable:

1. **The subject of our notion of a priori knowability is the idealised human mathematical community as a whole**. It is intuitive to think about mathematical knowledge as knowledge by a mathematical community, as opposed to knowledge by a single individual. However, this assumption is unnecessary for the sake of the arguments contained in the paper. It suffices to stress that the idealised subject whose knowledge is in question does not have superhuman cognitive abilities nor computational powers, so provable or computable by God or by an oracle are not admissible interpretations of what an idealised subject of knowledge can do in this context.[18]

[17] This is compatible with the fact that (a coded version of) the statement expressing provability of a sentence *in a formal system S* can itself be formally provable in *S*.

[18] See [Horsten 2005].

2. **A priori knowability has a discretely ordered temporal structure**. Statements come to be known a priori in time. Time should be taken to be discretely ordered in the sense that for every moment, there should be a least successor moment at which new statements are proved. We may want to build a modal dimension into the time dimension, since knowa*bility* contains a modal component. This would lead us to a *branching* or tree-like rather than a linear temporal structure. Moreover, there may be reasons for imposing further constraints on the temporal relation. For instance, we may want to require that time is open-ended in the sense that it contains no last moment (while leaving it open whether or not time extends in the future direction into the transfinite). We may (or may not) want the earlier-than relation on the moments of time at which new statements are proved to form a well-ordered relation. This would allow us to assign ordinal stages to moments in time.

   In the literature on *infinite time Turing machines* [Hamkins & Lewis 2000], the time dimension along which computation is performed is a *transfinite* ordinal. The effect of such transfinite time computations can be simulated in certain models of General Relativity Theory, known as *Malament-Hogarth spacetimes*. This is so, even though there are observers in such spacetimes, who in a finite time interval of her own contains the infinite world line of some calculating machine.[19] Despite this, transfinite time is admittedly not easy to motivate. Moreover, if—and this is a very big 'if'—time extends into the transfinite future, then as far as we can tell, it might well contain non-well-ordered parts. Since a detailed discussion of this issue is beyond the scope of the paper, we will assume for the sake of the argument that the idealisation into transfinite time is legitimate in this context.

3. **The subject is finite but does not have any fixed limitations of time and memory space**. Computation over a finite domain is sufficient to guarantee that the computing procedure can be carried out by an idealised but human mathematician. However, some functions grow very fast, and while the question of what is the threshold at which a computation over an arbitrarily large but finite domain exceeds the actual capacities of the embodied mind is surely not easy to answer, the answer to this question does not seem to be to be essential to the meaning of what we regard as effective computations. In saying that a function ϕ is *effectively computable*, we do not refer to the actual ability of this or that specific mathematician to compute ϕ; rather, we refer to the computation abilities as such—we mean that ϕ is *in principle* effectively computable by a mathematician. So when we talk about what is computable

[19] For a discussion of these matters, see [Welch 2008].

in principle, we are in fact abstracting away from the time that the computing procedure will run and the way in which it is established that, if at all, the computing procedure is going to terminate. Nonetheless, at no moment in time will the knowing subject have carried out transfinitely many computation steps. If we were to admit this, then the idealised counterparts would differ from their human counterparts not only in degree, but qualitatively. Analogously, for the purpose of modelling the informal notion of a priori knowability we abstract away from the limitations concerning specific finite boundaries of the subject of knowledge. It suffices for informal provability of a sentence ϕ that the mathematical community will *eventually* have at their disposal all of the axioms and rules of inference by means of which ϕ can be finitely derived, no matter what cognitive resources need to be employed and how long the derivation process takes.[20] But whilst we impose no specific finite boundary on the number of mathematical axioms that the subject can come to know and we are imposing no finite upper bound on the discrete time structure, we do require that at each point in time the knowing subject remains finite. Thus at each moment in time the subject only knows a finite number of axiom schemes. Or, given the correspondence between axiom systems and Turing machines, we impose no fixed bound on the complexity of the Turing machine that the knowing subject can be, but we do insist that the knowing subject essentially remains equivalent to a Turing machine at every moment in time (and therefore at every given point in time, the extension of what is a priori known is recursively axiomatisable). To relax this requirement would amount to a *qualitative* difference between actual human provers and their idealised counterparts [Shapiro 1985b, p. 20].

4. **A priori knowability is cumulative, and at every given point in time it is reasonable to take what is a priori known to be closed under logical consequence**. The mathematical community is taken to be idealised in the sense that it is assumed not to make mistakes, and not to 'forget' known facts as time goes on. Moreover, if at some moment in time t, ϕ is a priori known and $\phi \rightarrow \psi$ is a priori known, then at a subsequent moment t' the mathematical community apply modus ponens and come to know a priori that ψ.[21] We will assume that the mathematical community will indeed infer all the logical consequences of what is known at any given point in time as time goes on.[22] Moreover, we might as well assume that at every given

[20] For a detailed discussion of these constraints, see [Parsons 1997].

[21] For a discussion of questions of closure in a modal-epistemic setting, see [Heylen 2015].

[22] The principle of deductive closure for the extension of the notion of a priori knowability is also assumed in [Shapiro 1985b, p. 12].

moment t, the mathematical community has inferred *all* the logical consequences of what is known at t, i.e. that t is *closed under logical implication*. This is an additional idealisation, but structurally this does not really make a difference, and it simplifies the models that we are considering.

3 Formalising human effective computability

Following suggestions in [Myhill 1960], the notion of a priori knowability can be investigated in an axiomatic way. In particular, the investigation is carried out within the framework of Epistemic Arithmetic developed in [Shapiro 1985b].

The formal framework of Epistemic Arithmetic can be described as follows. The formal language $\mathcal{L}_{EA}$ consists of the first-order language of arithmetic plus an intensional propositional operator $\Box$; the arithmetical vocabulary receives its intended interpretation, and the operator $\Box$ is intended to be interpreted as a priori knowability. The axiomatic theory EA that is proposed by Shapiro as describing the laws of a priori knowability, consists of the axioms of Peano Arithmetic plus the laws of $S4$ modal logic. Note that the modal logic $S4$ contains the Necessitation rule and the axiom $\Box\phi \rightarrow \Box\Box\phi$,[23] so that $\Box$ is indeed an iterable notion: for instance, $\Box\Box(0 = 0)$, should be taken to be true.

The aim of the paper is to provide a rigorous characterisation of Kreisel's notion of effective computability by an idealised human agent, and to do this in terms of informal provability by an idealised but human mathematician. As it is well known, notions like Turing computability and formal provability can be expressed formally in $\mathcal{L}_{PA}$. In this context such notions are captured extensionally and do not involve any reference to a computing or proving subject. According to Shapiro, $\mathcal{L}_{EA}$ is a suitable language for the formalisation of epistemic notions like effective computability and informal provability in a way that is faithful to our intuitions concerning what an idealised but human mathematician could compute or informally prove in principle. Hence EA appears to be an adequate formal framework for formulating principles that establish extensional equivalences between informal, intensional notions on the one hand, and formal notions on the other hand. When this is done, and provided that the right idealisations are embedded in the principles governing the intensional notions, specific claims concerning computability and provability by an idealised subject become provable, and real progress can be made.

[23] Suppose that ϕ is a priori known at some moment in time t. Then at a subsequent moment t' the statement that ϕ is knowable can become a priori known. Idealising away from the length of time required for the idealised mathematical community to know a priori that ϕ is a priori knowable (see idealisation (3) in §1.2 above), it seems reasonable to assume that the mathematical community will eventually have a priori epistemic access to their own knowledge of ϕ (cfr. [Shapiro 1985b, p. 15]).

For example, the following property of a functional predicate $\phi(x, y)$ can be expressed in $\mathcal{L}_{EA}$:

$$\Box \forall x \exists y \Box \phi(x, y),$$

meaning that it is a priori knowable that for all x there exists a y such that y can be informally known to stand in the relation ϕ to x. Let us, using the terminology of Shapiro, call this condition the *calculability* of ϕ, where *calculability* refers to computability by an idealised but human mathematician. According to the thesis proposed in [Shapiro 1985b, p. 43], a function presentation F is *calculable* if and only if there is an algorithm A such that it is a priori knowable that A represents F. Recall that since functions are infinite abstract objects, human subjects—even in the idealised sense—do not have epistemic access to functions independently of the interpreted linguistic expressions that denote them (the *function presentations*). In fact, in the epistemic context it cannot be assumed that if F and G are different presentations of the same function, then F is calculable if and only if G is calculable. The condition $\Box \forall x \exists y \Box \phi(x, y)$ appears to respect this constraint: it seems at least *prima facie* likely that there are co-extensive functional relations $\phi(x, y)$ and $\psi(x, y)$ expressible in $\mathcal{L}_{EA}$, such that $\Box \forall x \exists y \Box \phi(x, y)$ is true whereas $\Box \forall x \exists y \Box \psi(x, y)$ is false.

All this leads us to provisionally characterise *human effective computability* as follows:

Thesis 2 (Human Effective Computability) *A function f is human effectively computable if and only if,* ***recognisably****, for every natural number m given in canonical notation, a canonically given number n exists such that the statement $f(m) = n$ is informally provable.*

In other words, it is a priori knowable (i.e., *recognisable* by an idealised agent) that for each x we can find a $y \in \mathbb{N}$ which provably (in the informal sense) stands in the relation ϕ to x. Thesis 2 will be considered in detail in the following section, but first we will compare it with another property of a functional predicate that was considered in [Myhill 1985], namely:

$$\forall x \exists y \Box \phi(x, y).$$

This expresses that for each x we can find a $y \in \mathbb{N}$ which provably (in the informal sense) stands in the relation ϕ to x. This notion differs from Shapiro's notion of calculability only in the absence of the initial occurrence of $\Box$. We then have a rivalling, stronger thesis concerning the notion of human effective computability:

Thesis 3 (Kreisel-Myhill) *A function f is human effectively computable if and only if for every natural number m given in canonical notation, a canonically given number n exists such that the statement $f(m) = n$ is informally provable.*

Both Shapiro's and Myhill's notions were originally proposed as an expression of the notion of calculability or effective computability in order to formalise the antecedent of a version of Church's thesis in an intensional context.[24] We will see shortly that there are reasons to think that Myhill's notion is probably the more faithful expression in $\mathcal{L}_{EA}$ of Kreisel's notion of human effective computability. However, we will also see that there are good reasons for holding that neither the principle proposed by Shapiro nor the one proposed by Myhill are good approximations of the content of CT. Nonetheless, we will argue that Shapiro's notion of calculability is better motivated than Myhill's notion; it captures a notion of human computability that gives rise to an interesting variant of Church's thesis. In other words, Kreisel *should have* explicated the informal notion of human effective computability in terms of Shapiro's notion rather than in terms of Myhill's notion. For this reason we will in the sequel be concerned with a variant of the Church-Turing thesis based on Shapiro's notion of calculability rather than on Myhill's notion of effective computability.

4 Epistemic Church's Thesis

Using Shapiro's notion of calculability, we can express Church's thesis for human effective computability in $\mathcal{L}_{EA}$ as follows [Shapiro 1985b, p. 31]:

Thesis 4 (*ECT*)

$$\Box\forall x\exists y\Box\phi(x,y) \rightarrow \exists e\,[e \textit{ is a Turing machine } \wedge \forall x : \phi(x,e(x))],^{25}$$

for ϕ ranging over formulae of the language of Epistemic Arithmetic.[26] This principle is called *Epistemic Church's Thesis* in the literature because it was originally proposed as an approximation of the content of Church's thesis in $\mathcal{L}_{EA}$. It should be noted that in order for the antecedent to ensure that $\phi(x,y)$ expresses a *function*, a choice principle is implicit in ECT. However, the choice principle could be eliminated by prefixing the functionality

[24] The context was *Epistemic Arithmetic* in the case of Shapiro, and *Intensional Set Theory* in the case of Myhill.

[25] The notion of being a Turing machine can be formalised in the underlying language of arithmetic in the standard way in terms of Kleene's T-predicate and the U function symbol. The versions of ECT that we consider in what follows are schematic, and each instance involves a particular function presentation (given by a formula $\phi(x,y)$). Hence, the intensional aspect of functions that is relevant here (i.e. that they are always presented through a particular interpreted linguistic expression) is taken into account at every juncture.

[26] This is not necessary, though; it suffices for the purposes of this paper that functional predicates range over formulae in the language of PA, or even over a fragment of this language.

of $\phi(x,y)$ as a condition on ECT, so that it assumes the form

$$\phi(x,y) \text{ is functional } \rightarrow ECT.$$

Shapiro takes ECT to be "a weaker version of CT [in the standard formalisation] which is closer to Church's thesis [than the intuitionistic version of CT]" [Shapiro 1985b, p. 31].[27] This is because—like in CT—the existential quantifier in the consequent of ECT is *classical*, so it does not require that any particular Turing machine can be *shown* to compute the effectively computable function described in the antecedent.

Nonetheless, there are reasons to be sceptical about the extent to which ECT approximates the content of CT in EA. Note that the antecedent of ECT does not involve the informal notion of algorithm, so it is implausible that the antecedent of ECT expresses that $\phi(x,y)$ is effectively or algorithmically computable. Indeed, there is no way to *directly* express or quantify over algorithms in the language of EA [Shapiro 1985b, p. 41–43], and as of yet, no satisfactory axiomatic treatment of the informal notion of algorithm is available. Another reason why ECT does not capture the content of CT is that the *converse* of CT is obviously true, whereas the converse of ECT is not obviously true [Black 2000, §2]. We suggest, instead, that the antecedent of ECT comes close to capturing Kreisel's notion of human effective computability.

Furthermore, unlike algorithmic computability, which is an extensional concept, Kreisel has characterised the notion of human effective computability as an *intensional* notion. Calculability also appears to be an intensional notion because in it function presentations occur in the scope of an epistemic operator. In addition, the notion of effective calculability as expressed in the antecedent of ECT does not have built in the "routineness" restriction that is built into the notion of algorithm. Whereas this is another reason for thinking that ECT is close to Kreisel's notion of human effective computability, it does bring with it the difficulty that, as with Kreisel's notion, it is difficult to see what the right idealisations are that we should adopt for the notion of human effective calculability.

But before discussing what idealisations it is plausible to adopt with respect to the notion of calculability or human effective computability, let us turn to the question whether Myhill's version of ECT might be a better formalisation of effective calculability than Shapiro's version considered above. Myhill proposed a thesis that is stronger than ECT, which we may call ECT^{+} [Myhill 1985, p. 48]:

[27] For a discussion of the intuitionistic version of Church's thesis, see [Troelstra & van Dalen 1988].

Thesis 5 (*ECT^+*)

$$\forall x \exists y \Box \phi(x,y) \rightarrow \exists e[e \text{ is a Turing machine } \wedge \forall x : \phi(x, e(x))]$$

Informally, ECT^+ says that if for every x there is a y which can be a priori known to stand in the relation ϕ to x, then ϕ determines a recursive function. ECT^+ is just like ECT, except that it is based on Myhill's notion of effective computability (thesis 3) instead of Shapiro's notion of effective computability (thesis 2).

5 Evaluation

We will argue that ECT^+ is significantly less plausible than ECT. We will do this by evaluating these versions of Church's thesis in particular models. Following Kreisel, special attention is given to models in which the extension of informal provability is given by a path in a transfinite progression of formal theories (in the sense of [Feferman 1962]). A detailed discussion of Feferman's result and its philosophical significance for an account of the scope and limits of mathematical reasoning are beyond the scope of this paper.[28]

5.1 Unsystematic infinite progressions

Consider a non-recursive (total) function $\psi(x,y)$ (e.g., the self-halting problem). Suppose that for every m, there is an n such that $\psi(\overline{m}, \overline{n})$ can be informally proved. In such a case, such infinite collection of proofs cannot be "captured" by one single algorithm. Given the idealisations involved in a priori knowability that were discussed earlier, such a scenario is not wholly implausible. To be a little more concrete, suppose that the mathematical community lives on for an ω-sequence of years, and that after every moment in time, a new axiom (independent from all the axioms that were previously known) is discovered. (Note that this scenario does not violate any of the reasonable restrictions on idealisations involved in the notion of informal provability that were discussed in §2.) Then there seems to be no reason to think that the infinite sequence of axioms that are from some time onwards a priori known forms a recursive set. If these axioms do form a non-recursive set, then they may suffice to decide every instance of the halting problem. If that is the case, then the antecedent of ECT^+ holds, whereas its consequent fails. Since we seem to have no way of ruling out scenarios such as this, it seems that ECT^+ is at present doubtful at best.[29]

[28] For a detailed exposition of Feferman's completeness theorem see [Franzén 2004] and [Antonutti 2013].

[29] For a more detailed explanation of this issue, see [Horsten 1998].

Note, on the other hand, that while the situation just sketched would suffice to falsify ECT^+, it would not suffice to falsify ECT, for the latter is compatible with there never being an informal proof that every instance of the self-halting problem will be solved.

In sum, for "unsystematic" progressions such as these, there are simple ways of generating models in which ECT^+ is false, whereas finding such models in which ECT is false is not straightforward. More on this will be said in the following subsection.

5.2 Systematic transfinite progressions

One of Kreisel's main questions in [Kreisel 1972] was how versions of Church's thesis for human effective computability relate to *systematic* transfinite progressions of formal theories.

The background of Kreisel's discussion of a further possible reason for doubting ECT^+ is given by Feferman's famous completeness theorem in [Feferman 1962]. In this article, Feferman shows that there are paths P in $\mathcal{O}$ that prove all first-order arithmetical truths, in the following sense.[30] At stage 0, all theorems of the initial theory T_0 are proved. At a successor stage $T_{\alpha+1}$, the first-order consequences of the theory T_α plus the uniform reflection principle for T_α are proved.[31] In a uniform effective manner, at limit stages, unions are taken. Then there is a particular hyperarithmetical maximal path[32] P in $\mathcal{O}$ of length $\omega^{\omega^{\omega}+1}$ such that:

$$K \equiv \{\phi \in \mathcal{L}_{PA} \mid \exists e \in P \text{ with } e \vdash \phi\} = \{\phi \in \mathcal{L}_{PA} \mid \phi \text{ is true}\},$$

where $e \vdash \phi$ means that ϕ is proved at stage e [Feferman 1962, theorem 5.15]. In other words, Feferman's completeness result shows that there exists a path P such that for every arithmetical sentence ϕ, if ϕ is true, then ϕ is proved in some theory indexed along P.

This set K can be the basis for a model wherein a function f exists that is Myhill-effective but not Turing computable. Take any non-recursive total first-order definable function f—by $\phi(x,y)$, say—on $\mathbb{N}$. Then we have that for any $m,n \in \mathbb{N}$, $f(m) = n \Leftrightarrow \phi(\overline{m},\overline{n}) \in K$. This means that f is Myhill-effective, but not Turing computable. Hence, if K is the extension of a priori knowability of arithmetical sentences, then ECT^+ is false.

Whilst Feferman showed that there are paths in $\mathcal{O}$ that "prove" all first-order arith-

[30] Here and in what follows we presuppose familiarity with Kleene's system $\mathcal{O}$ of notations for ordinals and the ordering relation $<_{\mathcal{O}}$ on it. For a definition of these notions, see [Sacks 1990, chapter 1, section 4].

[31] A uniform reflection principle for a formal theory T is a schematic principle of the form $\forall x(Prov_T(\ulcorner\phi(\dot{x})\urcorner) \rightarrow \phi(x))$, where $Prov_T$ is the standard provability predicate for T.

[32] 'Maximal' means that there is no $w \in \mathcal{O}$ such that for all $u \in P : u <_{\mathcal{O}} w$. For a definition of the concept 'hyperarithmetical' see [Sacks 1990, chapter 1, section 1].

metical truths, Feferman and Spector showed that there are many paths *through* $\mathcal{O}$—meaning that the length of those paths is ω_1^{CK}, the first non-constructive ordinal—that yield an extension of provability that is far from being first-order complete [Feferman & Spector 1962, theorem 2.5 and theorem 4.4]. Kreisel showed that if $\phi(x)$ is an open formula (with one free variable) such that along some Π^1_1 path[33] through $\mathcal{O}$ each of its instances is decided at some stage, then the extension of ϕ is recursive [Kreisel 1972, p. 313]. This means that even the stronger variant ECT^+ of Epistemic Church's Thesis is true in models with extensions for informal provability that are generated along Π^1_1 paths through $\mathcal{O}$.

Such considerations take us back to Kreisel's worries about the idealisations involved in the notion of a priori knowability. Since we have no firm grasp on what the right idealisations are, it is extremely difficult to adjudicate whether ECT^+ is true or false. On the one hand, the fact that many paths *through* $\mathcal{O}$ make ECT^+ true might give one reason to think that ECT^+ might be true if transfinite progressions provide good models for informal provability. On the other hand, Kreisel does not see convincing reasons to dismiss the kinds of Fefermanian models in which ECT^+ comes out false. He summarises the situation as follows:

> Unless it can be shown that the progression is not included in any (legitimate) model of mathematical reasoning, we cannot establish Church's thesis (for human effective definitions). And unless it can be shown that each recursive progression on a Π^1_1 path is inadequate (as a model for mathematical reasoning), we cannot refute Church's thesis [Kreisel 1972, p. 325].

For these reasons, the discussion about the truth of ECT^+ is left by Kreisel in an unsettled state.

However, there is hope that the truth or falsehood of ECT is not as sensitive to the level of idealisation involved in the notion of a priori knowability as the truth or falsehood of ECT^+ is. Indeed, we will now show, using a realisability argument, that ECT is true in *every* model that is based on a path in a Fefermanian transfinite progression of formal theories.[34]

Unlike Myhill-computability, human effective computability as explicated by Shapiro (thesis 2) involves iterated provability. This means that our base theory should not be a theory formulated purely in the language of arithmetic (such as PA) but a theory formulated in the language of Epistemic Arithmetic. It seems then reasonable to take EA as our base theory.

[33] For a definition of the concept of being Π^1_1, see again [Sacks 1990, chapter 1, section 1].

[34] The connection between realisability and uniform reflection in the context of EA was first explored in [Halbach & Horsten 2000].

We start by defining a *transfinite recursive progression based on iteration of uniform reflection* defined (roughly) according to the following clauses:

(C1) $T_0 = EA$;

(C2) $T_{\alpha+1} = T_\alpha + RFN(T_\alpha)$;

(C3) $T_\lambda = \bigcup_{\beta<\lambda} T_\beta$ for λ a limit ordinal.

(Here $RFN(U)$ is the uniform reflection principle for U.)

Instead of working with constructive ordinals, as Feferman does, we will work with primitive recursive orderings. The difference is largely technical, but working with primitive recursive orderings is more convenient [Beklemichev 1995, p. 29]. Also, we will not work with Feferman's transfinite progressions but with Beklemichev's *smooth transfinite progressions*, allowing us to formulate our arguments in a more perspicuous manner.[35] More importantly, smooth progressions have a monotonicity property which is of crucial significance in the context of informal provability, but which is not known to hold for the more traditional way of defining transfinite progressions.

A *primitive recursive well-ordering* $(P, <)$ is a relative interpretation of the theory of linear orderings in the theory of Primitive Recursive Arithmetic (PRA) with domain D and where the relation $<$ well-orders the set D in the standard model of arithmetic. We assume that each theory T comes together with a primitive recursive formula $Ax_T(x)$ numerating the set of Gödel numbers of mathematical axioms of T, from which a primitive recursive formula expressing the proof-in-T relation can be constructed, from which in turn provability in T is defined in the standard manner and denoted as $Prov_T(x)$.

A primitive recursive formula $Ax_{EA}(z;x)$ is called a *smooth numeration*[36] *based on iteration of uniform reflection along* $(D, <)$ *applied to* EA if and only if PRA proves

$$\forall z, x : Ax_{EA}(z;x) \leftrightarrow [Ax_{EA}(x) \vee \exists u \in D(u < z \wedge \exists v \in \mathcal{L}_{EA}(x = \ulcorner \forall y : Prov_{T_{\dot{u}}}(\dot{v}(\dot{y})) \dot{\to} v(\dot{y}) \urcorner))].$$

This means that we are now explicitly adding at each successor stage not just the uniform reflection principle for the previous theory, but uniform reflection for *each* earlier theory [Franzén 2004, p. 380], which makes the definition uniform between successor and limit stages. This has the form of a fixed point equation. Since the existential quantifiers on the right-hand side can be bounded by x, the solution of this equation must be equivalent to a primitive recursive formula. Then by metatheoretical transfinite induction one can

[35] We thank an anonymous referee for suggesting that we formulate this argument in terms of smooth progressions.

[36] See [Beklemichev 1995, section 2.2]. In what follows we rely heavily on [Beklemichev 1995].

show that $(T_u)_{u \in D}$ is a strictly increasing sequence of theories satisfying (C1)–(C3), and PRA-provably satisfies the formalised versions of (C1)–(C3) [Beklemichev 1995, p. 30]. Let $T_D \equiv \bigcup_{u \in D} T_u$.

Let greek variables $\alpha, \beta, ...$ from now on be assumed to range over ordinals, that is, over the domain D. Smooth numerations enjoy a desirable robustness property [Beklemichev 1995, lemma 2.2]:

Proposition 1 *Take any two smooth enumerations $Ax_T(z;x)$ and $Ax'_T(z;x)$ along one and the same primitive recursive well-ordering and satisfying the same initial conditions. Then $PRA \vdash \forall \alpha \forall x : x \in T_\alpha \leftrightarrow x \in T'_\alpha$.*

Moreover, smooth enumerations are *monotone*:

Lemma 1 $\forall \alpha, \beta : \alpha < \beta \to \forall x : Prov_T(\alpha, x) \to Prov_T(\beta, x)$.
Proof. *[Beklemichev 1995, p. 30].* ■

Recall that we have explicitly committed ourselves to monotonicity as a desirable idealisation (idealisation 1 in §2). Remarkably, this idealisation is *not* known to hold for Feferman's way of defining transfinite progressions.[37]

Let $\mathfrak{M}_D$ be the model generated by D in the sense that

$$\mathfrak{M}_D \models \Box\phi \equiv \exists e \in D : T_e \vdash \phi.$$

So $\mathfrak{M}_D$ is simply the model in which T_D is taken to be the interpretation of $\Box$. Our question then is whether ECT holds in $\mathfrak{M}_D$.

It is known that EA has what is called the *numerical existence property*, which means that if $EA \vdash \forall x \exists y \Box\phi(x,y)$, then for every natural number m there is a natural number n such that $EA \vdash \phi(\overline{m}, \overline{n})$. This follows from a realisability argument [Shapiro 1985b, p. 19]. We will show that the numerical existence property in fact holds for all T_d such that $d \in D$ [Shapiro 1985b, p. 19]:[38]

Definition 1 (Kleene's slash) *For any theory T and for any sentence $\phi \in \mathcal{L}_{EA}$, we define $T \mid \phi$ ("T realises ϕ") as follows:*

1. *$T \mid \phi$ iff ϕ is atomic and true;*

[37] For a discussion of this aspect of the difference between 'traditional' progressions and smooth progressions, see [Franzén 2004, p. 380–381]. Thanks to an anonymous referee for drawing our attention to the significance of this monotonicity property and its connection with smoothness.

[38] In a similar way, it can be shown that the *disjunction property* holds for all such T_d, i.e., that if $T_d \vdash \phi \vee \psi$, then $T_d \vdash \phi$ or $T_d \vdash \psi$.

2. *$T \mid \phi \wedge \psi$ iff $T \mid \phi$ and $T \mid \psi$;*

3. *$T \mid \neg\phi$ iff not: $T \mid \phi$;*

4. *$T \mid \forall x \phi(x)$ iff for all $n \in \mathbb{N}$: $T \mid \phi(\overline{n})$;*

5. *$T \mid \Box\phi$ if and only if $(T \vdash \phi)$ and $(T \mid \phi)$*

We start by proving the following slight strengthening of theorem *TB* in [Shapiro 1985b, p. 18]:

Theorem 1 *For any theory T: $T \supseteq EA \Rightarrow T \mid EA$.*
Proof. *The proof is a routine induction on the length of proofs in EA. The key case is the inductive case for the necessitation rule.*

Suppose that ϕ is derived on line n, and $\Box\phi$ appears on the next line. The inductive hypothesis entitles us to assume that $T \mid \phi$. Since ϕ is derived in T, we have $T \mid \Box\phi$. ■

Lemma 2 *For any theory T, if $T \mid T$, then T has the numerical existence property.*
Proof. *Straightforward.* ■

We now seek to establish that $\forall e \in D : T_e \mid T_e$. In order to do this, we need two simple lemmas:

Lemma 3 *For all $e \in D$: $(\forall f < e : T_e \mid T_f) \Rightarrow T_e \mid T_e$.*
Proof. *By induction along D.*

1. $e = 0$, and $T_0 = EA$. This holds by theorem 1.

2. Let us assume, for a reductio, that there is an $e \neq 0$ for which the property does not hold. Since D is a well-ordering, there is then a $<$-least such e, for which $\forall f < e : T_e \mid T_f$ but not $T_e \mid T_e$. $T_e = EA + \{RFN(T_d) : d < e\}$.

Take any $d < e$. $T_e \mid RFN(T_d) \Leftrightarrow$ for all $\varphi(y) \in \mathcal{L}_{EA} : T_e \mid RFN(T_d, \varphi(y))$. However,

$$T_e \mid \forall y : Prov_{T_d}\ulcorner \varphi(\dot{y}) \urcorner \rightarrow \varphi(y) \Leftrightarrow$$

for all $n \in \mathbb{N} : T_e \mid Prov_{T_d}\ulcorner \varphi(\overline{n}) \urcorner \rightarrow \varphi(\overline{n}) \Leftrightarrow$ for all $n \in \mathbb{N} : T_e \mid Prov_{T_d}\ulcorner \varphi(\overline{n}) \urcorner \Rightarrow T_e \mid \varphi(\overline{n})$. We know that $T_e \mid Prov_{T_d}\ulcorner \varphi(\overline{n}) \urcorner \Leftrightarrow \varphi(\overline{n}) \in T_d$, since $Prov_{T_d}\ulcorner \varphi(\overline{n}) \urcorner$ is arithmetical. So it suffices to prove that for all $n \in \mathbb{N} : \varphi(\overline{n}) \in T_d \Rightarrow T_e \mid \varphi(\overline{n})$, i.e. $T_e \mid T_d$, which we have since $\forall f < e : T_e \mid T_f$. By theorem 1, $T \mid EA$, so we conclude that $T_e \mid T_e$ after all. ■

Lemma 4 *For all $e \in P$: $\forall f < e : T_e \mid T_f$.*
Proof. *Fixing e, we establish the lemma by an induction along $<$.*

1. $f = 0$. This is the case of EA, which we know to be covered by theorem 1.

2. Let us assume, for a reductio, that there is an $f \neq 0$ for which the property does not hold. Since D is a well-ordering, there is then a $<$-least such e, for which it is not the case that $T_e \mid T_f$. Since EA is covered by theorem 1, there must be a $d < f$ such that it is not the case that $T_e \mid RFN(T_d)$. But this cannot happen. As in the proof of the previous lemma, we see that

$$T_e \mid \forall y : Prov_{T_d} \ulcorner \varphi(\dot{y}) \urcorner \rightarrow \varphi(y) \Leftrightarrow$$

for all $n \in \mathbb{N} : \varphi(\overline{n}) \in T_d \Rightarrow T_e \mid \varphi(\overline{n})$, i.e., $T_e \mid T_d$. But this cannot be true, for $d < f$, and f was assumed to be the least for which it is not the case that $T_e \mid T_f$. ■

Lemma 5 *For all $e \in P : T_e \mid T_e$.*
Proof. *Follows directly from lemma 3 and lemma 4.* ■

Given lemma 2, for all $e \in P$, the theory T_e has the numerical existence property.

Theorem 2 $\mathfrak{M}_D \models ECT$.
Proof. *Take any $\phi(x, y) \in \mathcal{L}_{EA}$, and assume that $\mathfrak{M}_D \models \Box\forall x \exists y \Box \phi(x, y)$ and that $\mathfrak{M}_D$ thinks that $\phi(x, y)$ is functional. Then*

$$\exists e \in D : T_e \vdash \forall x \exists y \Box \phi(x, y).$$

However, we know that T_e has the numerical existence property, so for all m there is an n such that $T_e \vdash \phi(\overline{m}, \overline{n})$. So T_e can be regarded as a Turing machine that computes ϕ. ■

In sum, whereas the truth value of ECT^+ is sensitive to the details of the transfinite progressions model in which it is investigated, the truth value of ECT is not. We have seen that Kreisel takes the truth value of Church's thesis for human effective computability to be sensitive to the details of the transfinite progressions model in which it is evaluated [Kreisel 1972, p. 325]. This can be taken to be evidence for the thesis that he leans towards Myhill's explication of human effective computability rather than to Shapiro's explication of it. However, since Myhill's notion of effective computability is less robust than Shapiro's notion, we prefer the latter notion.

Moreover, we can show that $\mathfrak{M}_D$ does in fact satisfy a minimal condition for being a reasonable model:

Theorem 3 $\mathfrak{M}_D \models EA$.
Proof. *Induction on the length of proofs in EA.*

The only non-trivial case is the inductive case where ϕ is an instance of the scheme $\Box\psi \to \psi$. All these instances can be shown to hold by an induction on the complexity of ψ, as can be seen as follows. The atomic case is unproblematic. For the inductive cases, we consider:
(1) $\psi = \exists x\theta(x)$. Suppose $\mathfrak{M}_D \models \exists x\theta(x)$. Then, $\exists x\theta(x) \in T_d$ for some $d \in D$. However, by the numerical existence property, this means that $\theta(\overline{n}) \in T_d$ for some $n \in \mathbb{N}$. Then, by the induction hypothesis, $\theta(\overline{n})$, and therefore also $\exists x\theta(x)$, is true in $\mathfrak{M}_D$.
(2) $\psi = \theta \vee \mu$. Like (1), except that we now use the disjunction property.
(3) Suppose that ψ is a negated formula. Since we can "push the negation signs inwards" (and cancel double negations), we may assume that each negation sign either prefaces an atomic formula, or prefaces a formula of the form $\Box\theta$. The former is unproblematic, so we concentrate on the latter. Suppose $\mathfrak{M}_D \models \Box\neg\Box\theta$. Then $\neg\Box\theta \in T_d$ for some $d \in D$. Now assume also that $\mathfrak{M}_D \models \Box\theta$. Then θ, and therefore also (by Necessitation) $\Box\theta \in T_e$ for some $e \in D$. So both $\neg\Box\theta$ and $\Box\theta$ are in $max(T_d, T_e)$. But we know that $max(T_d, T_e)$ is consistent, so this cannot be the case. ■

Note that $\mathfrak{M}_D$ does not make the *necessitations* of all instances of ECT true. Constructing transfinite progression models that not only make the theorems of EA and ECT true, but that in addition take all instances of EA and ECT to be a priori knowable, would require a more complex construction that lies beyond the scope of this paper. However, we conjecture that the techniques developed by Carlson in [Carlson 2000] do indeed yield such models.

The importance of the numerical existence property in the evaluation of variants of Church's thesis was anticipated by Kreisel. When discussing the *intuitionistic* version of Church's thesis, he writes:

> [...] present knowledge does not exclude that some $\forall x\exists yR(x,y)$ is intuitionistically valid (where R is logically complex), but no recursive function f satisfies $\forall xR(x, fx)$... Evidently, under such conditions, $\forall x\exists yR(x,y)$ could not be proved in any (sound) formal system F which satisfies the principle of demonstrable numerical instantiation of existential theorems. [Kreisel 1972, p. 327]

The arguments above show that Kreisel's point generalises to the *classical* variants of Church's thesis that we are concerned with in this article.

5.3 Diagonalising out?

In addition to the formal notion of a recursive function, there is also the notion of a *provably* recursive function of a system T, where provable recursiveness is relative to an axiomatic theory and is therefore a formal notion.

The logical structure of the definition of human effective computability closely resembles the structure of a provable recursiveness claim. Indeed, the T-provable recursiveness of ϕ can be expressed as

$$Prov_T(\ulcorner\forall x\exists!y\phi(x,y)\urcorner),$$

which under fairly general circumstances (and assuming that T is sound) is equivalent to $Prov_T(\ulcorner\forall x\exists!yProv_T\ulcorner\phi(\dot{x},\dot{y})\urcorner\urcorner)$.

There is a standard way of diagonalising out of the class of provably recursive functions for every recursively enumerable theory T. We can effectively list all T-provably recursive functions, and then, by diagonalisation, produce a function that is (total) recursive but not T-provably recursive. So the question arises, can we in a similar way diagonalise out of the class HEC of human effective computable functions on the natural numbers?

Suppose that $A(x)$ defines $HEC(\mathcal{L}_{EA})$, which is the class of human effective computable functions that are expressible in the language of Epistemic Arithmetic. Then we can define:

$$\phi(x,y) \equiv A(x) \wedge T(x(\dot{x},\dot{y}-1)),$$

where T is a truth predicate for $\mathcal{L}_{EA}$. Then clearly ϕ differs from every function in $HEC(\mathcal{L}_{EA})$ by design.

Moreover, given that every $\ulcorner\psi\urcorner \in A$ defines an element of HEC and that $A(\ulcorner\psi\urcorner) \leftrightarrow \Box A(\ulcorner\psi\urcorner)$, we have that $\forall x\exists y\Box\phi(x,y)$. Thus, we have diagonalised out of the class of the Myhill computable functions (thesis 3): by going to a language extension, we can produce new Myhill computable functions.

There is, however, no guarantee that for the property ϕ that we have constructed above, we have that $\Box\forall x\exists y\Box\phi(x,y)$, so our attempt to diagonalise out of HEC fails. We conclude, again, that human effective computability appears to be a more robust notion than the class of Myhill-computable functions.

6 Conclusion

When Kreisel investigated the concept of human effective computability, it seemed to him a notion that is hard to get a firm grip on. Not only its extension is hard to determine, but also its logical properties seemed to Kreisel somewehat non-robust.

We have reached a different conclusion; when properly analysed, the notion of human effective computability is more robust than Kreisel took it to be. Even though the notion of informal provability in terms of which human effective computability is explicated is admittedly somewhat unclear, some of the key properties of human effective computability cannot be easily influenced by the details of the model in which it is considered. In particular, a form of Church's thesis holds for human effective computability in a wide class of models, namely, in those models that are based on transfinite progressions of formal theories.

Funding

This work was partly supported by a Jacobsen Fellowship awarded to the first author by the Royal Institute of Philosophy, whose support is gratefully acknowledged.

Acknowledgements

Earlier versions of this article were presented at Birkbeck College, University of Manchester, University of Notre Dame, University of Oxford, University of Lund, IHPST (Paris), and Ecole Normale Superieure (Paris): we thank the audiences for valuable reactions. We are grateful to Roy Cook, Walter Dean, Benedict Eastaugh, Kentaro Fujimoto, Jan Heylen, Harold Hodes, Peter Koellner, Pawel Pawlowski, Richard Pettigrew, Sam Roberts, Stanislav Speranski, Sean Walsh, Philip Welch, and Tim Williamson for valuable comments. We are also grateful to two anonymous referees for their helpful suggestions.

References

[Anderson 1983] Anderson, C.A. *The paradox of the knower.* Journal of Philosophy **80**(1983), p. 338–355.

[Antonutti 2010] Antonutti, M. *Informal provability and mathematical rigour.* Studia Logica **96**(2010), p. 261–272.

[Antonutti 2013] Antonutti, M. *Theories of absolute provability.* PhD dissertation, University of Bristol, 2013.

[Barwise et al 1980] J. Barwise; H. Keisler; K. Kunen (eds.) *The Kleene Symposium.* North-Holland, 1980.

[Beklemichev 1995] Beklemichev, L. *Iterated local reflection versus iterated consistency.* Annals of Pure and Applied Logic **75**(1995), p. 25–48.

[Black 2000] Black, R. *Proving Church's Thesis.* Philosophia Mathematica **8**(2000), p. 244–258.

[Carlson 2000] Carlson, T. *Knowledge, machines, and the consistency of Reinhardt's strong mechanistic thesis.* Annals of Pure and Applied Logic **105**(2000): p. 51–82.

[Carlson 2012] Carlson, T. *Can a machine know that it is a machine?* Unpublished manuscript.

[Church 1936] Church, A. *An Unsolvable Problem of Elementary Number Theory.* American Journal of Mathematics **58**(1936): p. 345–363.

[Enderton 2001] Enderton, H.B. *A Mathematical Introduction to Logic*, Second Edition. Academic Press, 2001.

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

[Feferman & Spector 1962] Feferman, S. and Spector, C. *Incompleteness along paths in progressions of theories.* Journal of Symbolic Logic **27**(1962), p. 383–390.

[Folina 1998] Folina, J. *Church's Thesis: Prelude to a Proof.* Philosophia Mathematica **6**(1998), p. 302–323.

[Franzén 2004] Franzén, T. *Transfinite Progressions: A Second Look at Completeness.* The Bulletin of Symbolic Logic, **10**(2004), p. 367–389.

[Gandy 1980] Gandy, R. *Church's Thesis and principles for mechanisms.* In [Barwise et al 1980, p. 123–148].

[Gödel 1933] Gödel, K. *Eine Interpretation des intuitionistischen Aussagenkalküls.* In: S. Feferman et al. (eds) *Kurt Gödel. Collected Works. Volume I: Publications 1929–1936.* Oxford University Press, 1986, p. 300–302.

[Gödel 193?] Gödel, K. *Undecidable Diophantine Propositions.* [193?] In: S. Feferman et al. (eds) *Kurt Gödel. Collected Works. Volume III: Unpublished Essays and Lectures.* Oxford University Press, 1995, p. 164–175.

[Gödel 1951] Gödel, K. *Some Basic Theorems on the Foundations of Mathematics and Their Implications.* [1951] In: S. Feferman et al. (eds) *Kurt Gödel. Collected Works. Volume III: Unpublished Essays and Lectures.* Oxford University Press, 1995, p. 304–323.

[Goodman 1986] Goodman, N., *Flagg realizability in Epistemic Arithmetic.* Journal of Symbolic Logic **51**(1986), p. 387–392.

[Halbach & Horsten 2000] Halbach, V. & Horsten, L. *Two proof-theoretic remarks about* $EA + ECT$. Mathematical Logic Quarterly **46**(2000), p. 461–465.

[Hamkins & Lewis 2000] Hamkins, J. & Lewis, A. *Infinite time Turing machines.* Journal of Symbolic Logic **65**(2000), p. 567–604.

[Heylen 2015] Heylen, J. *Closure of A Priori Knowability Under A Priori Knowable Material Implication.* Erkenntnis **80**(2015), p. 359–380.

[Horsten 1997] Horsten, L., *Provability in principle and controversial constructivistic principles.* Journal of Philosophical Logic **26**(1997), p. 635–660.

[Horsten 1998] Horsten, L. *In defense of Epistemic Arithmetic.* Synthese **116**(1998), p. 1–25.

[Horsten 2000] Horsten, L. *Models for the logic of possible proofs.* Pacific Philosophical Quarterly **81**(2000), p. 49–66.

[Horsten 2005] Horsten, L. *Remarks on the Content and Extension of the Notion of Provability.* Logique & Analyse **189–192**(2005), p. 15 –32 .

[Horsten 2006] Horsten, L. *Formalizing Church's Thesis.* In: A. Olszewski et al (eds.), *Church's Thesis after 70 years.* Ontos Verlag, 2006.

[Kaplan and Montague 1960] Kaplan, D. and Montague, R. *A paradox regained.* Notre Dame Journal of Formal Logic **1**(1960), p. 79–90.

[Kreisel 1967] Kreisel, G. *Informal rigour and completeness proofs.* In I. Lakatos (ed.), *Problems in the Philosophy of Mathematics.* North-Holland, p. 138-186.

[Kreisel 1971] Kreisel, G. *Some reasons for generalizing recursion theory.* In R. Gandy et al. (eds.) *Logic Colloquium '69.* North-Holland, 1971, p. 139–198.

[Kreisel 1972] Kreisel, G. *Which number theoretic problems can be solved in recursive progressions on* Π^1_1*-paths through* $\mathcal{O}$*?* Journal of Symbolic Logic **37**(1972), p. 311–334.

[Kreisel 1987] Kreisel, G. *Church's Thesis and the ideal of informal rigour.* Notre Dame Journal of Formal Logic **28**(1987), p. 499–519.

[Leitgeb 2009] Leitgeb, H. *On Formal and Informal Provability.* In O. Linnebo and O. Bueno (eds.), *New Waves in Philosophy of Mathematics.* Palgrave Macmillan, p. 263–299.

[Mendelson 1990] Mendelson, E. *Second Thoughts about Church's Thesis and Mathematical Proofs.* The Journal of Philosophy **87**(1990), p. 225?233.

[Myhill 1960] Myhill, J. *Some remarks on the notion of proof.* Journal of Philosophy **57**(1960), p. 461–471.

[Myhill 1985] Myhill, J. *Intensional set theory.* In: [Shapiro 1985a, p. 47–61].

[Parsons 1997] Parsons, C. *What Can We Do "in Principle"?* In M. L. Dalla Chiara *et al.* (eds.), Logic and Scientific Methods, Kluwer Academic Publisher, 1997.

[Reinhardt 1986] Reinhardt, W. *Epistemic theories and the interpretation of Gödel's incompleteness theorems.* Journal of Philosophical Logic **15**(1986), p. 427–474.

[Sacks 1990] Sacks, G. *Higher recursion theory.* Springer, 1990.

[Shapiro 1981] Shapiro, S. *Understanding Church's Thesis.* Journal of Philosophical Logic **10**(1981), p. 353–365.

[Shapiro 1985a] Shapiro, S. *Intensional Mathematics.* North-Holland, 1985.

[Shapiro 1985b] Shapiro, S. *Epistemic and Intuitionistic Arithmetic.* In: [Shapiro 1985a, p. 11–46].

[Sieg 1994] Sieg, W. *Mechanical procedures and mathematical experience.* In A. George (ed.), *Mathematics and Mind*, Oxford University Press, 1994, p. 71–117.

[Sieg 1997] Sieg, W. *Step by recursive step: Church's analysis of effective calculability.* Bulletin of Symbolic Logic **3**(1997), p. 154–180.

[Sieg 2006] Sieg, W. *Gödel on computability.* Philosophia Mathematica **14(III)**(2006), p. 189–207.

[Sieg 2013] Sieg, W. *Axioms for Computability: Do they allow a proof of Church's Thesis?* In H. Zenil (ed.), *In A Computable Universe – Understanding and Exploring Nature as Computation*, World Scientific Publishing: Singapore, 2013, p. 99–123.

[Soare 1996] Soare, R. I. *Computability and recursion.* Bulletin of Symbolic Logic **2**(1996), p. 284–321.

[Troelstra & van Dalen 1988] Troelstra, A. & van Dalen, D. *Constructivism in Mathematics. An introduction. Volume 1* North-Holland, 1988.

[Turing 1936] Turing, A. M. *On Computable Numbers, with an Application to the Entscheidungsproblem.* Proceedings of the London Mathematical Society **42**(1936), p. 230–265.

[Welch 2008] Welch, P. *Turing Unbound: on the extent of computation in Malament-Hogarth spacetimes.* British Journal for the Philosophy of Science **59**(2008), p. 659–674.

[Williamson 2000] Williamson, T. *Knowledge and its Limits.* Oxford University Press, 2000.