Available online at www.sciencedirect.com
SCIENCE DIRECT®
Theoretical Computer Science 317 (2004) 167–190
Theoretical Computer Science
www.elsevier.com/locate/tcs

The modal argument for hypercomputing minds

Selmer Bringsjord*, Konstantine Arkoudas

Department of Computer Science, Department of Cognitive Science, Rensselaer AI & Reasoning Laboratory, Rensselaer Polytechnic Institute (RPI), Troy, NY 12180, USA

Received 14 July 2003; received in revised form 21 October 2003

Abstract

We now know both that hypercomputation (or super-recursive computation) is mathematically well-understood, and that it provides a theory that according to some accounts for some real-life computation (e.g., operating systems that, unlike Turing machines, never simply output an answer and halt) better than the standard theory of computation at and below the "Turing Limit." But one of the things we do not know is whether the human mind hypercomputes, or merely computes—this despite informal arguments from Gödel, Lucas, Penrose and others for the view that, in light of incompleteness theorems, the human mind has powers exceeding those of TMs and their equivalents. All these arguments fail; their fatal flaws have been repeatedly exposed in the literature. However, we give herein a novel, formal *modal* argument showing that since it's mathematically *possible* that human minds are hypercomputers, such minds *are* in fact hypercomputers. We take considerable pains to anticipate and rebut objections to this argument.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Computationalism; Hypercomputation; Incompleteness theorems

1. Introduction

Four decades ago, Lucas [50] expressed supreme confidence that Gödel's first incompleteness theorem (= Gödel I) entails the falsity of computationalism, the view that human persons are computing machines (e.g., Turing machines). Put barbarically, Lucas' basic idea is that minds are more powerful than Turing machines. Today, given our understanding of hypercomputation in theoretical computer science, and given the absolute consensus reigning in cognitive science that the human mind is, at least in large part, *some* sort of information-processing device, we know enough to infer that if Lucas is right, the mind is a hypercomputer. However, Lucas' arguments have

* Corresponding author.
E-mail addresses: selmer@rpi.edu (S. Bringsjord), koud@ai.mit.edu (K. Arkoudas).

0304-3975/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2003.12.010

not proved to be compelling (even when updated: [51]). Even the arguments given by Lucas' illustrious and indefatigable defender, the mathematician/physicist Roger Penrose, have failed. Penrose's first attempt to vindicate Lucas was a Gödelian attack on computationalism articulated in *The Emperor's New Mind* [55]. Penrose has admitted that this first attempt fell short—and so he went back to the drawing board, and soon thereafter came charging into the arena again, armed with a more elaborate and more fastidious Gödelian case against computationalism, expressed in Chapters 2 and 3 of his *Shadows of the Mind* [56]. Yet once again there was failure, as recent, sustained, painstakingly formal refutations apparently reveal [19,34,48,62,69]. [1]

[1] There is no need here to recapitulate the fatal problems plaguing Penrose's reasoning, but for sedulous readers we offer a glimpse of some of these problems. Writing in response to critics (e.g., the philosopher David Chalmers, the logician Solomon Feferman, and the computer scientist Drew McDermott) of his *Shadows of the Mind* in the electronic journal *Psyche*, Penrose provided an encapsulated version of his Gödelian case in *SOTM*. (The dialectic appeared in 1996 in volume 2 of *Psyche*, which can be accessed via enn http://psyche.cs.monash.edu). Here is the encapsulation, verbatim:

> We try to suppose that the totality of methods of (unassailable) mathematical reasoning that are in principle humanly accessible can be encapsulated in some (not necessarily computational) sound formal system F. A human mathematician, if presented with F, could argue as follows (bearing in mind that the phrase "I am F" is merely a shorthand for "F encapsulates all the humanly accessible methods of mathematical proof"):
> (A) "Though I don't know that I necessarily am F, I conclude that if I were, then the system F would have to be sound and, more to the point, F' would have to be sound, where F' is F supplemented by the further assertion "I am F." I perceive that it follows from the assumption that I am F that the Gödel statement $G(F')$ would have to be true and, furthermore, that it would not be a consequence of F'. But I have just perceived that "If I happened to be F, then $G(F')$ would have to be true," and perceptions of this nature would be precisely what F' is supposed to achieve. Since I am therefore capable of perceiving something beyond the powers of F', I deduce that I cannot be F after all. Moreover, this applies to any other (Gödelizable) system, in place of F." (Penrose [57] ¶ 3.2)

Unfortunately, (A) is not a sound argument. In order to see this, let us follow Penrose directly and set

$$\psi = \text{"}F \text{ encapsulates all the humanly accessible methods of mathematical proof"}$$

and

$$F' = F \cup \{\psi\}.$$

What the hypothetical human mathematician can now conclude, on the strength, as Penrose tells us, of Gödel's work, is that on the assumption that ψ,

$$G(F') \text{ is true.} \tag{1}$$

$$F' \nvdash G(F') \text{ and } F' \nvdash \neg G(F'). \tag{2}$$

The idea is really quite simple. It is that there is a contradiction arising from the fact that the hypothetical mathematician, i.e. F, can conclude that (1) $G(F')$ is true on the one hand, and yet (2), which "says" that F cannot conclude $G(F')$, is true on the other. But wait a minute; look closer here. Where is the contradiction, exactly? There is no contradiction. The reason is that (1) is a *meta*-mathematical assertion; it is a claim about *satisfaction*. More precisely, where I is an interpretation of the relevant type, (1) is just

$$I \models G(F') \text{ is true.} \tag{1'}$$

And for all we know, F can prove (1′) while being bound by (2)! So we see here one of the chief errors dating back to (Lucas [50]): Penrose conflates proofs within a fixed system with meta-proofs (as explained in detail in "Chapter VII: Gödel" of Bringsjord [10]).

The upshot is that as we pass into the new millennium, Lucas' confidence is looking more and more like an emotional quirk, and less and less like an attitude born of careful ratiocination. On the other hand, Lucas and Penrose are not alone. Gödel [37] himself, in his Gibbs lecture [37], and in correspondence, made remarks revealing that he believed his incompleteness results do imply that minds can exceed standard computing machines.[2] This should give pause to those positive that Lucas and Penrose are not only wrong, but irremediably so. Is it possible that there exists a formal and fundamentally new technical Gödelian argument for the view that minds are more than standard computing machines? The answer is: "Yes." In what follows we provide such an argument: a novel *modal* argument showing that since it is mathematically possible that human minds are hypercomputers, such minds *are* in fact hypercomputers.

The plan in the remainder is as follows. In Section 2, after explaining that there are two general avenues for constructing a sound Gödelian argument against computationalism, we fix some needed idealizations and take the second of these avenues, which leads to a *modalized* Gödelian refutation of computationalism that has its roots in "Chapter VII: Gödel" of *What Robots Can and Can't Be* [10], wherein the kernel of a modal argument against computationalism is presented.[3] As we explain, an immediate and obvious consequence of this argument is that minds hypercompute. In Section 3 we defend the key premise in the modal argument. In Section 4 we consider and rebut some objections to our modal argument.

2. The modal argument

2.1. The two avenues

In general, we see two avenues that are in principle viable as a way to pass from mathematical results in the realm of incompleteness to a denial of computationalism, and an affirmation of the view that minds hypercompute.

In the first avenue, taken elsewhere by Bringsjord [13,21], one first isolates and exploits mathematical reasoning that seems to be explicitly and irreducibly infinitary. The best example of such reasoning that we are aware of is found in *infinitary* mathematical logic. The key idea is to find mathematical cognition that is provably *beyond* computation. Such cognition seems to be exhibited by logicians and mathematicians

[2] Actually, Gödel believe that minds are more powerful than the likes of Turing machines for a *number* of reasons. For a nice treatment of some of these additional reasons, see Wang [79].

[3] The present paper is intended to give a formal argument for the following promissory note:

> I should point out that my [just-given] Gödelian argument against the proposition that persons are automata can apparently be modalized ... In this modal argument the central claim would be only that Ralf is such that it is *logically possible* that he acts as I have described him acting above via the Fixed Point Theorem, whereas no Turing machine is such that it is logically possible that it acts in this way. (Bringsjord [10, p. 264].)

who prove things in and about *infinitary* logics (which of course arose in no small part as a way to "surmount" Gödel I; see Barwise [5] for a readable discussion of this point), such as the logic $L_{\omega_1\omega}$.[4]

There is a second avenue available; it is the one we take herein. In broad strokes, the route runs like this. First, suitable idealizations of the central concepts (minds, machines, etc.) are made so as to allow for precision sufficient to enable intelligent appeal to theorems, and a formally valid *modal* argument inspired by the Gödel–Lucas–Penrose direction is constructed. Next, the key premise in this argument is defended by looking more closely at an aspect of mathematical cognition which looks to be beyond the reach of computation. This aspect is the ability of a mathematician to understand through thought-experiments what's it's like to *very rapidly* experience some process or operation—or, in other words, to understand a so-called "supertask." We will explain this aspect of mathematical thinking momentarily, in connection with Turing's own thinking about the machines that now bear his name. But first, as planned, we turn now to the needed idealizations, and then to the modal refutation of computationalism itself.

2.2. *Idealizing to make formal reasoning possible*

We have referred repeatedly to the doctrine of computationalism, but gave only a very intuitive definition of this doctrine at the start of the present paper. What exactly *is* this doctrine? It is easy to find published material that one would expect to be helpful in the attempt to answer this question, for propelled by the writings of innumerable thinkers (this touches but the tip of a mammoth iceberg of relevant writing: [4,10,30,35,40,42,46,54,58,63,68,71,72]), computationalism has reached every corner of, and indeed energizes the bulk of, contemporary AI and cognitive science. The view has also touched nearly every major college and university in the world; even the popular media have, on a global scale, preached the computational conception of mind. Despite all this, despite the fact that computationalism has achieved the status

[4] The basic idea behind $L_{\omega_1\omega}$ is straightforward. This system allows for infinite disjunctions and conjunctions, and hence allows for infinitely long derivations, where these disjunctions, conjunctions, and proofs are no longer than the size of the set of natural numbers. (We use ω to denote the "size" of the set of natural numbers: the niceties of cardinal numbers needn't detain us here.) Here is one simple formula in $L_{\omega_1\omega}$ which is such that any interpretation that satisfies it is finite:

$$\bigvee_{n<\omega} \exists x_1 \cdots \exists x_n \forall y (y = x_1 \vee \cdots \vee y = x_n).$$

It is a well-known fact that the proposition captured by this formula cannot be captured by a formula in a system at or below Turing machines. Since the behavior of some logicians and mathematicians centers around infinitary reasoning that can be accurately described only by formalisms that include such formulas (i.e., formalisms like $L_{\omega_1\omega}$), computationalism is threatened. Version 1 of the full argument (which makes use of the theorem that TM computation can be perfectly recast as deduction in first-order logic) can be found in Bringsjord [13]. A refined version of the argument can be found in Bringsjord and Zenzen [20]. For an example of irreducibly infinitary reasoning, see Bringsjord's ω-rule-based proof in Bringsjord and van Heuveln [18].

of a Kuhnian paradigm, the fact is that the doctrine is maddeningly vague. Myriad one-sentence versions of this doctrine float about; e.g.,

- Thinking is computing.
- Cognition is computation (or is comput*able*).
- People are computers (perhaps with sensors and effectors).
- People are Turing machines (perhaps with sensors and effectors).
- People are finite automata (perhaps with sensors and effectors).
- People are neural nets (perhaps with sensors and effectors).
- Cognition is the computation of Turing-computable functions.

From the standpoint of symbolic logic and mathematics, the vagueness of computationalism is unacceptable. As long as it remains vague, how can we rationally hope to determine the truth-value of computationalism on the basis, at least in part, of theorems in theoretical computer science? The only solution is that we must idealize the concepts of minds and machines to provide, as Shapiro [69] puts it, an "interface" between mathematical results and the doctrine of computationalism. Without such an idealization, we should probably simply close up shop and agree with George Boolos' pessimistic impressions of Gödel's original comments on the matter: "... it is certainly not obvious what it means to say that the human mind, or even the mind of some human being, *is* a finite machine, e.g., a Turing machine" [8].

Fortunately, the idealization is not hard to come by. In fact, it has already been firmly established on the machine side. Theoretical computer science provides us with a precise conceptual apparatus for making sense of 'computer' and 'computation.' For example, we can identify 'computer' with Turing machine (or, for that matter, with register machine, abaci, etc.), and 'computation' with Turing machine computation. (And, as we do below, we can turn to mathematics for models of 'hypercomputer' and 'hypercomputation.') This familiar identification allows us to rigorously refer to machines that never run out of memory or working space, and that infallibly obey instructions.

What about persons, or minds? What are we to take them to be? Here, of course, there is no convenient formalism to appeal to. However, it is clear that we must assume, for starters, that persons or minds can be quantified over along with Turing machines, and that they can be viewed as entities capable of taking in inputs and returning outputs that reflect decisions with respect to these inputs. It is also clear from contemporary cognitive science that the mind is *some* sort kind of information-processing device. Textbook after textbook and paper after paper in psychology, cognitive science, and cognitive neuroscience makes this plain as day (e.g., see: [2,3,73]). This is at least enough to proceed, but we will need to say more about the abstract structure of minds later on.

Now to the modal argument itself.

2.3. *The modal proof itself*

Recall that Gödel's first incompleteness theorem can be expressed in "Turingish" terms as the fact that this problem is algorithmically unsolvable: Given an arbitrary

Turing machine m and input u, does m halt on u?[5] Here is how this theorem, and a related one which in our experience many computer scientists tend to forget, are expressed on page 278 of the classic text *Elements of the Theory of Computation* by Lewis and Papadimitriou [49].[6]

Theorem 6.1.5.

(a) There is no algorithm that can determine, given a Turing machine m and an input string w, whether m accepts w.

(b) For a certain fixed Turing machine m_0, there is no algorithm that can determine, given an input string w, whether m_0 accepts w.

Intuitively, (a) says that there is no Turing machine that can "crack" every Turing machine (as to halting or non-halting); and (b) says that there is an impenetrable Turing machine: a machine that no other machine can crack. Notice that the sense of 'can' here, given that we are dealing with mathematical theorems, is very strong. Specifically, to say in this context that a Turing machine can't Φ is to say that it is mathematically or logically impossible for the Turing machine to Φ.[7] Accordingly, if we use the three-place predicate $Dmm'i$ for 'Turing machine m determines whether Turing machine m' halts on input i', and if we assume to ease exposition a sorted calculus for Turing machines, inputs, and persons (or minds), we can draw upon quantified modal logic (QML) and symbolize (b) as[8,9]

$$\forall m \exists i \neg \Diamond D m m_0 i. \tag{3}$$

[5] Actually, as many forget, Turing [77] was concerned with computable numbers; the halting problem, in the customary contemporary form we have just expressed it in, falls out as a consequence. A referee pointed out that Gödel I can be deduced from the unsolvability of the halting problems for TMs. But we leave aside characterization of the relationship between Gödel I and Turing's (more general) results.

[6] Of course, this version of the theorem presupposes the Church–Turing Thesis. If we follow Burgin [23] and accept that Inductive TMs (at least those of the first order) are algorithms, then the CTT-based version of the theorem here isn't true, because ITMs can solve the halting problem. See the next note.

[7] Recall that algorithms and Turing machines are interchangeable by the Church–Turing Thesis. If you happen to be one of the few people on the planet who are skeptical about this thesis, that is okay: simply start with a version of Theorem 6.1.5 in which 'algorithm' is supplanted with 'Turing machine.' For an argument against the Church–Turing Thesis, see "Chapter 5: A Narrative-Based Refutation of Church's Thesis" in Bringsjord and Ferrucci [17].

[8] Were we to make all the predicates explicit, with Mx for 'x is a Turing machine,' Px for 'x is a person,' and Ix for 'x is input for a Turing machine,' (3) would become either[9]

$$(3') \quad Mm_0 \wedge \forall y(My \rightarrow \exists u(Iu \wedge \neg \Diamond D(y, m_0, u)))$$

or

$$(3'') \quad \exists x(Mx \wedge \forall y(My \rightarrow \exists u(Iu \wedge \neg \Diamond D(y, x, u)))).$$

[9] We use the $\Diamond$ of QML in a generic way: we mean broadly logical or mathematical possibility; i.e., $\Diamond\phi$ if and only if it is logically/mathematically possible that ϕ. There is no need for us to express our reasoning within a particular system of QML (for reasons we explain below), but we could do so.

Now suppose for *reductio* that computationalism is true, i.e., that persons are (physically realized) Turing machines, which, given our idealizations, can be symbolized in such a way as to sweep away Boolos' pessimism:

$$\forall p \exists m\, p = m. \tag{4}$$

It follows from (3) and (4) that

$$\forall p \exists i \neg \Diamond D p m_0 i. \tag{5}$$

However, given that there are persons, (5) is inconsistent with

$$\forall p \forall i \Diamond D p m_0 i \tag{6}$$

so indirect proof yields $\neg(3)$, that is, the theorem—let us call it Theorem 1—that computationalism is false. However, since we know that persons are information processors of some sort, they must be more powerful than Turing machines, which is to say that they are hypercomputers. (This line of argument, we have come to learn, was pondered a while back by Rapaport [64], a computationalist.) Note that we do not consider in this paper where minds *qua* hypercomputers fall in any hierarchies covering information processing above the Turing Limit. For discussion of this issue in connection with the Arithmetic Hierarchy, see [21,47].

Three points should be made immediately about this argument (with others coming when we rebut objections).

First, it is important to grasp that the underlying modal argument is not inseparably linked to a particular formal derivation or a particular proof theory. We have presented the derivation in a manner with which we happen to be particularly comfortable, but an infinite number of variations on the core idea are possible. The point here is important because the computationalist will not be able to dodge our modal argument by hiding behind such claims as that the Barcan Formula in quantified modal logic (QML) is controversial,[10] or that identity statements cannot, in general, be necessitated. While there are formal versions of the argument that explicitly invoke BF and related aspects of QML, we are in fact most strongly drawn to formal versions of the argument that stay within first-order logic (FOL), on the strength of innocent readings of "... can possibly determine whether ... ." For example, the following two equations, where $D^{\Diamond}xyz$ abbreviates the merging of $\Diamond$ with D to make an FOL-representable predicate meaning that x can (logically possibly) determine whether y halts on input z, are true, as can be verified by hand or by supplying them as input to a standard theorem prover:

$$\{\exists x(Mx \wedge \forall y(My \rightarrow \exists u(Iu \wedge \neg D^{\Diamond}(y,x,u)))), \forall x(Px \rightarrow \forall y \forall u((My \wedge Iu) \rightarrow D^{\Diamond}(x,y,u))),$$

$$\forall x(Px \rightarrow \exists y(My \wedge y = x)), \exists x Px\} \vdash \phi \wedge \neg\phi,$$

[10] For the classic discussion see Hughes and Cresswell [45].

$$\{Mm_0 \wedge \forall y(My \rightarrow \exists u(Iu \wedge \neg D^{\Diamond}(y, m_0, u))), \forall x(Px \rightarrow \forall y \forall u((My \wedge Iu) \rightarrow D^{\Diamond}(x, y, u))),$$

$$\forall x(Px \rightarrow \exists y(My \wedge y = x)), \exists x Px\} \vdash \phi \wedge \neg\phi.$$

The second point is that once one sees the essence of our modal proof, it is possible to return to the idealizations that underlie it and modify them in various ways. For example, some may desire to interpret the 'are' in 'People are (standard) TMs' not as the 'are' of '=,' but rather as some concept of "instantiation," so that people *instantiate* TMs. This will simply lead to a new version of our argument.[11] This is so in light of the fact that, for all instantiated standard TMs, it is impossible that they hypercompute. (After all, theorems establishing that abstract, mathematically defined information-processing devices are unable to Φ entail that when these devices are built (i.e., instantiated), they cannot Φ. For example, we know that no instantiated standard Turing machine can solve the halting problem.)

Here is the third point. Clearly, our underlying modal argument can be based on any number of related problems. Obviously, it can be based on Theorem 6.1.5(a). But it just as obviously can be based on many other problems. For example, we know that no Turing machine can determine whether an arbitrary arithmetical formula is true on the standard interpretation of arithmetic; and, again, this implies that no Turing machine can (mathematically or logically speaking) *possibly* make this determination. Yet it certainly seems logically possible for a person to make this determination. (Remember: we momentarily provide an argument for this proposition.) The same can be said for trying to determine, in general, whether or not a first-order formula is a validity. And so on. However, we have for certain specific reasons based the above instantiation of the underlying modal argument on Theorem 6.1.5(b). These reasons pertain to the apparent fact that many logicians and mathematicians have a hard time accepting that *particular* problems are such that it is logically impossible for humanity to solve them. An impressionistic but seminal treatment of this matter is carried out by Minsky, who points out that while many are willing to accept that humanity cannot possibly find a uniform procedure for solving an infinite class of problems, "Some mathematicians accept these uniform unsolvability results but prefer to believe that there can be no analogue for single problems" [51]. Gödel was an example of such a thinker. Though this may make a difference only to some readers, we point out that we herein chose a particular instantiation of our modal argument that appeals to a *particular* machine, m_0. There are any number of other ways to "particularize" our modal argument.

3. Why the key premise is true

At first glance, the modal proof is likely to look like a piece of legerdemain. After all, why is (6), the key premise, true? Well, actually, the truth of (6) can be seen via

[11] We might even allow for computationalism to consist in the view that each person, at each time, instantiates some standard TM. Again, no damage is done to our modal argument.

two steps: by attending to the mathematics of information processing (understood to include both the recursive and super-recursive realms), and by unflinchingly attending to the underlying cognition that has *produced* this mathematics.

For the start of step one, note that we now understand well that there *are* information-processing machines that can exceed the Turing Limit (e.g., they can solve the halting problem), such machines just are not standard TMs and the like. There are in fact now many such machines. Indeed, just as there are an infinite number of mathematical devices equivalent to Turing machines (first-order theorem provers, register machines, the λ-calculus, abaci, ...; many of these are discussed in the context of an attempt to define standard computation in [11]), there are an infinite number of devices beyond the Turing Limit. While Burgin [23] is correct that the first fleshed-out account of such a machine (**trial-and-error machines**, as Kugel [47] aptly calls them) appears to have been provided at the same time by Putnam [61] and Gold [38], and while we agree with his claim [23] that his own more powerful **inductive TMs** (ITMs), introduced in 1983, have the distinct advantage of giving results in finite time, today's **infinite time TMs** (ITTMs) [39], as we will shortly see, can trace their lineage back to seminal theoreticians working well before 1965.[12] ITTMs result from extending the operation of TMs to infinite ordinal time, and produce a "supertask theory of computability and decidability" [39]. This should come as no surprise, because supertasks have long been discussed in theoretical treatments of computation. For the present paper, it is sufficient to save time space and by referring not to ITTMs, but only to intuitive supertask machines called **Zeus machines** (ZMs) by Boolos and Jeffrey [9].[13]

Zeus machines are based on the character Zeus, described by Boolos and Jeffrey [9]. Zeus is a superhuman creature who can enumerate **N**, the natural numbers, *in a finite amount of time*, in 1 s, in fact. He pulls this off by giving the first entry, 0, in 1/2 s, the second entry, 1, in 1/4 s, the third entry in 1/8 s, the fourth in 1/16 s,..., so that, indeed, when a second is done he has completely enumerated the natural numbers. Obviously, it is easy to adapt this scheme so as to produce a Zeus machine that can solve the halting problem: just imagine a machine which, when simulating an arbitrary Turing machine m operating on input u, does each step faster and faster ... (There are countably many Turing machines, and those that do not halt are trapped in an unending sequence of the same cardinality as **N**.) If, during this simulation, the Zeus machine finds that m halts on u, then a 1 is returned; otherwise 0 is given. Put in the abbreviatory first-order terms we availed ourselves of above, where $z, z', \ldots$ range over Zeus machines, we have

$$\forall z \forall i \Diamond Dzm_0 i. \tag{7}$$

[12] It is interesting to note that introducing a relevant topology in the set of strings of symbols, we obtain that ITTMs are special cases of Burgin's [22,24] limit TMs, a formal model that appeared at least ten years earlier than ITTMs.

[13] We thus leave aside discussion of other hypercomputational machines, such as: **analog chaotic neural nets** [70], artificial neural nets allowed to have irrational numbers for coefficients; and **dial machines** [16], etc.

Now for step two in the case for (6), in which we will find that the mathematical cognition underlying the likes of (7) supports (6). How will this support be generated, exactly? From the logical point of view, we will see that the mathematical reasoning in question implies this conditional:

$$(7) \rightarrow (6). \tag{8}$$

Proposition (6) will thus follow by *modus ponens* from (7) and (8).

To isolate the mathematical cognition we have in mind, return to certain aspects of Turing's original cognition when he was laying down the foundational concept of a Turing machine. Turing did *not* start with some mathematical description of a machine of some sort; he started with the concept of a *person*—or, to use his term, a 'computist'—carrying out primitive operations. [14]

Turing himself wrote about the cognitive processes that took him from intuitions about computists to specified Turing machines. In his dissertation at Princeton University [74, later published as 75] Turing distinguished between "intuition" and "ingenuity" in logic and mathematics. He wrote: [15]

> Mathematical reasoning may be regarded rather schematically as the exercise of a combination of two faculties, which we may call *intuition* and *ingenuity*. The activity of the intuition consists in making spontaneous judgments which are not the result of conscious trains of reasoning ... The exercise of ingenuity in mathematics consists in aiding the intuition through suitable arrangements of propositions, and perhaps geometrical figures or drawings [75, pp. 214–215].

As Copeland, no doubt the thinker who knows the mind and history of Turing best, explains, Turing's intuition was of "a human mathematician who is unaided by any machinery save paper and pencil, and who is working in accordance with 'mechanical' methods, which is to say, methods set out in the form of a finite number of exact instructions that call for no insight ... on the part of the person carrying them out" [28]. As Turing himself explicitly said: "A man provided with paper, pencil, and rubber, and subject to discipline, is in effect a universal [Turing] machine" [76]. [16] In light

[14] In his inaugural writings on isomorphic points (independent, by the way, of Turing's), Post [60] spoke of mindless "workers," humans whose sole job was to slavishly follow explicit, excruciatingly simple instructions. Likewise, Charles Babbage modeled the calculating cogs in his Difference Engine (and, for that matter, in the never-built-by-him Analytical Engine) on an army of hairdressers hired by Baron Gaspard Riche de Prony, director of the École des Pont et Chaussées, to transform data tables into decimal form when France went decimal in 1799: see [44].

[15] We do not mean to suggest that for Turing (who was as much of a computationalist as Gödel was not) "intuition" involved hypercomputational thought. Nor do we mean to suggest that Turing's distinction constitutes an argument for the view that some mathematical thinking is hypercomputational. Rather, "intuition," as a spontaneous non-inferential judgment, supports the *possibility* of hypercomputational thought. Turing's distinction is an accurate "surface level" description of mathematical cognition—whether or not the intuitive part of that thinking is hypercomputational.

[16] As Jack Copeland recently explained to Bringsjord in conversation at a conference on Turing, Turing practiced what he preached here: Long before IBM's Kasparov-beating *Deep Blue* came on the scene, there was Turing operating *qua* chess-playing machine! Before there was a computer to run programs on, Turing wrote a program for playing chess, and slavishly followed the program to play chess against a "normal" human opponent. The match was broadcast on the radio in England.

of the intuition/ingenuity distinction, and in light, specifically, of the mathematician's process of moving from the former to the latter by moving from what an imaginary person can do to what a corresponding machine can do, it would seem that we can safely infer from such facts as that a Turing machine, given a natural number n, can determine whether n is even, that it is logically possible for a person to make the same determination. Likewise, it would seem that we can safely infer from the fact that a Zeus machine can determine whether a Turing machine m will halt on input i, that it is logically possible for a person unaided by but (enough) paper and pencil to make the same determination—the person just has to work faster and faster. We strongly suspect that Hamkins and Lewis [39] started with strong intuitive conceptions of the infinitary capability of ITTMs before producing the formal notation used to specify them in third-person terms.[17] In fact, we would be willing to wager that *you* apprehend the truth of (7), in the absence of all the mathematical details, precisely because you started to imagine the *person* Zeus described above doing what we said he does: he lists 1, then lists 2, then 3, and so on, moving faster at each step. Some of our readers, we are even willing to bet, imagined *themselves* operating in this way.[18]

It is important to realize that many other thinkers have independently affirmed proposition (6) on the strength of the technique of gedanken-experiment, in which they imagine what it is like for a person to behave in the relevant manner; this is why we say that precursors to ITTMs were introduced long ago. It is the logician Bertrand Russell who seems to be have been the first to grasp the essence of Zeus machines. In a lecture in Boston in 1914 he said about Zeno's paradox involving the race-course: "If half the course takes half a minute, and the next quarter takes a quarter of a minute, and so on, the whole course will take a minute" [65]. And later, when lampooning finitism as championed by Ambrose, Russell wrote:[19]

> Ambrose says it is *logically* impossible [for a man] to run through the whole expansion of π. I should have said it was *medically* impossible. ... The opinion that the phrase 'after an infinite number of operations' is self-contradictory, seems scarcely correct. Might not a man's skill increase so fast that he performed each operation in half the time required for its predecessor? In that case, the whole infinite series would take only twice as long as the first operation [65, pp. 143–144].

[17] In fact, there is even textual evidence for this on p. 568 of Hamkins and Lewis [39].

[18] It is interesting to note that Turing initially (in 1936) did not think it was *physically* possible for a Turing machine to be built (see Copeland [28, p. 130]). It would be a mistake to infer from the mathematical fact that a Turing machine can determine whether an arbitrary natural number is even that it is physically possible for a person to make this determination on an arbitrary number.

[19] The chief point of this quote is to show that Russell would have affirmed (6). The point is not that there is an *argument* given here for (6), on which we are foolishly pinning our hopes. However, there can be no denying that Russell's distinction is one computationalists have tended not to made. They have in mind what is humanly possible now (medical), not what is logically possible. (For a good recent example, see Anderson and Lebiere [1].)

A number of other thinkers have conceived of that which our proposition (6) encapsulates. For example, Ralf Blake conceived this scenario quite a while ago,[20] and more recently Salmon [67] and Copeland [26,27] have imagined (and formalized) related "Zeusian" scenarios in the direction of ITTMs. Actually, the less colorful but more historically accurate name for the machines in question is "Weyl Machine," for in 1927 Hermann Weyl considered a machine able to complete

> an infinite sequence of distinct acts of decision within a finite time; say, by supplying the first result after 1/2 min, the second after another 1/4 min, the third 1/8 min later than the second, etc. In this way it would be possible ... to achieve a traversal of all natural numbers and thereby a sure yes-or-no decision regarding any existential question about natural numbers [80, p. 42].

Other names for Zeus machines and their ITTM-like counterparts are in the literature. For example, Copeland [26,27] has recently referred to Zeus machines as 'accelerated Turing machines' and 'super Turing machines.'

What about Gödel himself? And what about Turing? Gödel seems to have been of the opinion that the human mind *in fact* enters an infinite number of distinct states. Turing explicitly argued for the view that the human mind is capable of only a finite number of distinct states [79], and though Gödel rejected this argument because it presupposed a materialist conception of mind (Turing assumed that the mind equals brain; Gödel wrote and said on many occasions that the mind includes non-physical powers and parts), he also specifically wrote that "there is no reason why the number of states of the mind should not converge to infinity in the course of its development" [79]. It is safe to say that Gödel, at least during certain stages of his life, would have regarded our (6) as obviously true.

Now, finally, to Turing. As we have just indicated, he believed that the mind was "capable" of only a finite number of states. But 'capable,' modally speaking, is slippery. Did Turing believe that it is not even *logically possible* for a person to perform as Zeus? Would he have rejected (6), if this proposition were presented to him independent of the argument in which we have situated it? Apparently not. In fact, it seems likely that Turing would have *affirmed* (6). The reason is bound up with the apparent fact that contemporary computer science embraces a now-familiar mathematical scheme that would seem to entail (6), and with the fact that this scheme is due to none other than Turing himself. In his dissertation [74,75], Turing pondered the possibility of so-called *oracle machines*. These machines are architecturally identical to Turing machines, but are assumed to be augmented with an oracle which, upon being consulted about a Turing machine m and input i, returns a correct verdict as to

[20] E.g., Blake wrote:

> A process is perfectly conceivable, for example, such that at each stage of the process the addition of the next increment in the series $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, etc., should take just half as long as the addition of the previous increment. But ... then the addition of all the increments each to each shows no sign whatever of taking forever. On the contrary, it becomes evident that it will all be accomplished within a certain definitely limited duration. ... If, e.g., the first act ... takes $\frac{1}{2}$ second, the next $\frac{1}{4}$ second, etc., the [process] will ... be accomplished in precisely one second. (Blake [7, pp. 650–651].)

whether m halts on i. Oracle machines, part of the mathematical canon of computer science today, are typically the door through which students pass to enter into the study of hypercomputation. For example, here is how a recently updated textbook on computability and uncomputability introduces oracles:

> Once one gets used to the fact that there are explicit problems, such as the halting problem, that have no algorithmic solution, one is led to consider questions such as the following:
>
> Suppose we were given a "black box" or, as one says, an *oracle*, which can tell us whether a given Turing machine with given input eventually halts. Then it is natural to consider a kind of program that is allowed to ask questions of our oracle and to use the answers in its further computation [29, p. 197].

More important than that Turing would probably have affirmed the likes of (6) is the fact that computer science itself commits to this proposition. As the quote from Davis et al. indicates, first oracles are imagined, not as mathematized, mechanical devices, but rather as intuitive objects. *After this*, oracles can be cashed out (as we have said) via Zeus machines, trial-and-error machines, analog chaotic neural nets, and so on. And of course since hypercomputation can itself be studied from the standpoint of relative computability, once (e.g.) ITTMs are given as a way to specify oracles of the type that Davis et al. [29] introduce, one can consider halting problems for ITTMs, etc.; see [39]. So here we see the same sequence that supports our argument for (6): human imagination and intuition first, followed by formal ingenuity.

4. Objections

4.1. Objection 1: "You're Begging the Question"

The first objection is simply: "It strikes me that your modal proof is unnecessary. I say this because the final premise therein—(6)—is just a denial of computationalism. You are begging the question."

First, some points of logic in reply. Computationalism is expressed in the modal proof as proposition (4), which says that persons are Turing machines. The denial of (4) does not follow from (6); this can be certified by an easy proof showing the consistency of (4) and (6).[21] Perhaps Objection 1 presupposes a version of computationalism not captured by our (4). To quickly make explicit this possibility, we can for the sake of analysis legislate that computationalism includes the claim that no person can solve a problem that no Turing machine cannot. However, this kind of claim is invariably a non-modal one. At best, the claim is a naive one, with 'can' and 'cannot' within it left ambiguous. Computationalists just never have claimed that their view includes the proposition that it is *logically impossible* that a person solve a Turing-unsolvable

[21] A disproof of the purely extensional first-order version of the proof can be done with in standard fashion, i.e., build a model satisfying (6) and (4). The modal version requires use of possible worlds in standard ways.

problem. Our modal proof therefore doesn't beg the question. Rather, the proof reveals and exploits a key distinction that has hitherto gone unnoticed.

4.2. Objection 2: "You're Misrepresenting Computationalism"

The second objection runs as follows: "You identify computationalism with the claim that people are Turing machines, but computationalists needn't maintain any such thing. It suffices for the computationalist to claim that people can be *simulated* by computational systems (e.g., by Turing machines). After all, Turing's test for machine mentality, now known simply as the 'Turing Test,' requires for a 'passing grade' that the computer in question be conversationally indistinguishable from a person; the test does *not* insist that a computer *be* a person, a bearer of such things as subjective awareness."

This objection conflates "Strong" and "Weak" versions of computationalism and AI, and thus flies in the face of a fundamental distinction in AI and the philosophy of AI. The view that computing machines can simulate the behavior of persons is known as "Weak" AI; the view that people quite literally *are* computing machines, and hence that AI can eventually *replicate*, not just simulate, persons, is known as "Strong" AI. We aren't attacking Weak AI. Doing so, for formal reasons one of us has explained elsewhere, would be foolish: Weak AI is almost certainly true.[22] We seek to refute the kind of position nicely described by John Haugeland:

> What are minds? What is thinking? What sets people apart, in all the known universe? Such questions have tantalized philosophers for millennia, but ... scant progress could be claimed ... until recently. For the current generation has seen a sudden and brilliant flowering in the philosophy/science of the mind; by now not only psychology but also a host of related disciplines are in the throes of a great intellectual revolution. And the epitome of the entire drama is *Artificial Intelligence*, the exciting new effort to make computers think. The fundamental goal of this research is not merely to mimic intelligence or produce some clever fake. Not at all. AI wants only the genuine article: *machines with minds*, in the full and literal sense. This is not science fiction, but real science, based on a theoretical conception as deep as it is daring: namely, we are, at root, *computers ourselves* [41, p. 2].

As the reader will recall, we gave above an extensive list of propositions offered by thinkers as encapsulations of Strong AI. An even longer list of candidate propositions for expressing the core of the Strong computational conception of mind is presented in (Bringsjord and Zenzen [20]), with the "People are Turing machines" distillation chosen in the end there as well, as accurate and fair.

Nonetheless, perhaps Objection 2 can be sustained as follows: "Very well, I agree to talk not just of simulation, but replication; point well-taken. But I do think it makes a vast difference whether we say that persons are Turing machines or 'real life' embodied computers. Specifically, computationalism as the claim that people are physical computers is insulated from your modal argument; in fact, your argument is a non-starter once talk of Turing machines is supplanted with talk of such computers. To see

[22] An explicit argument for the "obviousness" of Weak AI is given by Bringsjord [15].

this, note first that computationalism, encapsulated by your (4), now becomes, where c ranges over embodied computers,

$$\forall p \exists c\, p = c. \tag{4'}$$

Now, in order for there to be a parallel modal argument here, (3) must be replaced with

$$\forall c \exists i \neg \Diamond Dcm_0 i. \tag{3'}$$

But this proposition is false! Take a particular "real life" computer, say the Powerbook G4 on which you are working at present; let us dub it W. Obviously, it is logically possible for W to crack m_0: you have showed us how above. What W needs to do is simply work in Zeusian fashion on m_0 and some input i. The scenario is easy to imagine—just as easy as it is to imagine that some *person* works in this fashion. Since (3′) is false, the modal argument fails."

The reply to this objection is straightforward. Computationalism is the view that people are *standard* computers—that is to say, computers that instantiate a formal scheme equivalent to that defined by Turing machines. Computationalism is *not* the view that people are hypercomputers, that is, information-processing devices capable of feats beyond the Turing Limit. After all, the holy grail for those thinkers seeking to overthrow computationalism—Gödel, Lucas, Penrose, ...—has long been some cognition that exceeds the Turing Limit. Penrose, in this regard, is right on key: his is a search for mathematical cognition that involves information processing beyond the Turing Limit. In sum, then, the modal argument, as the previous objection makes plain, does not the slightest harm to the view that the mind is (at least in part) a hypercomputer.[23]

4.3. Objection 3: "But Then Computers Aren't Computers!"

"Your argument can't possibly be right. For if we know anything about computation, we know that Macintosh Powerbooks are physically instantiated Turing machines.[24] But observe what happens when we replace reference to persons in your modal argument with reference to Powerbooks, a class we can assume b to range over. Proposition (4) becomes the fact that Powerbooks are Turing machines, that is,

$$\forall b \exists m\, b = m. \tag{4''}$$

It now follows from (3) and (4″) that

$$\forall b \exists i \neg \Diamond Dbm_0 i. \tag{5'}$$

But this proposition, given that there are Powerbooks, is inconsistent with

$$\forall b \forall i \Diamond Dbm_0 i. \tag{6'}$$

[23] Indeed, one of us offers a book-length defense of this view [20].

[24] Actually, since Turing machines do not have random access memory, this is not really true. It would be more accurate to say, e.g., that Powerbooks are physically instantiated register machines. A nice presentation of register machines can be found in [32].

It thus follows by indirect proof that Powerbooks are not computers, which is patently absurd!"

This objection surreptitiously conflates two different senses in which Powerbooks (and indeed any class of physical computers) can be said to be Turing machines (or some other idealized type of machine). The first sense is an *architectural* sense only, while the second is an architectural sense *and* a temporal sense. For a physical machine to be a Turing machine (or some other idealized machine) merely architecturally means that the structures of the two correspond. Turing machines, for example, have read/write heads that move over tapes divided into squares, so it is possible to take a model railroad set and build a physical computer that "is" a Turing machine, by allowing the model railroad track to function as a tape, and so on. But this leaves the temporal issues completely open. The fact of the matter is that a (standard) Turing machine not only has a certain structure; it also has an inherent temporal limitation: viz., it can take only a finite number of steps in any finite time (as can be verified by looking at any relevant formalization; e.g., see again [49]). The entire area of computational complexity in theoretical computer science is based upon this limitation. The problem is that (6′) is true only if it construes Powerbooks as devices allowed to carry out an infinite number of operations in a finite time (in which case they are not standard Turing machines, i.e., architecturally and temporally standard Turing machines), while (4″) is true only when Powerbooks are interpreted as *both* architecturally and temporally an instantiation of the Turing machine scheme. One traditional way to begin to mathematize those devices beyond the Turing Limit, by starting with Turing machines, is to drop the inherent temporal limitations on Turing machines: this is exactly the seminal move Putnam [61] and Gold [38] made.[25]

At this point someone might object that our criticism of Objection 3 applies to our own modal argument: "Proposition (4) should be interpreted to say that persons are, temporally *and* architecturally speaking, Turing machines. However, in proposition (6), you secretly drop the temporal restriction. So, the very distinction you introduced in response to Objection 3 serves to derail your modal argument."

There is a simple reason why this objection fails. Standard Turing machines, in all (to use the terminology of the standard formal semantics of modal logic) possible worlds, are temporally bound. What it *means* to be a (standard) Turing machine, among other things, is that processing happens one discrete step at a time, and that a computation is a *finite* sequence of such steps. Personhood is rather different. Every standard account of personhood is quite consistent with a person, in some possible world, carrying out a Zeusian task (e.g., see the account provided in Bringsjord [12]). Personhood (of the human variety) is traditionally taken to consist in capacities to communicate in a language, reason, form n-order beliefs, set long-term goals, and so on. None of these attributes block the logical possibility of Zeusian activity. To assume such blockage is to beg the question against us.

[25] Note that Burgin [23] provides *empirical* evidence for the view that modern physical computers are not (standard) TMs.

4.4. Objection 4: "Zeus Merely Seems Coherent"

The fourth objection is inevitable: "You have said that the logical possibility of a Zeus machine solving the halting problem, which I agree to be a mathematical fact, commits one to the view that it is logically or mathematically possible for a *person* to solve this problem by working faster and faster. But at most you have shown that it is apparently *coherent* that a person perform in this way."

For rebuttal we rely on a familiar point, one made recently by David Chalmers [25], namely, when some state of affairs ψ seems, by all accounts, to be perfectly coherent, the burden of proof is on those who would resist the claim that ψ is logically or mathematically possible.[26] Specifically, those who would resist us need to expose some contradiction or incoherence in the activity of Zeus, or the activity of persons as described by Russell, Weyl, Blake, Copeland, and so on. No such thing has ever been exposed.[27]

4.5. Objection 5: "As a Matter of Fact, Zeus is Not Coherent!"

The fifth objection: "As a matter of fact, some thinkers have argued that Zeus' behavior is incoherent. These are people who don't seem to share the attitude of Russell, Weyl, Blake, Copeland, and others you have conveniently listed. For example, Moore's attitude about expanding π, expressed in connection with a form of Zeno's paradox, runs rather counter to Russell's:

> Suppose that Achilles runs for half a minute, then pauses for half a minute, then runs for a quarter of a minute, then pauses for a quarter of a minute, and so on *ad infinitum*. At the end of two minutes he will have stopped and started in this way infinitely many times. Yet there is something repugnant about admitting this possibility, even as a conceptual—let alone a physical—possibility. For example, suppose that each time he pauses he performs a task of some kind, there being no limit to how quickly he can do this. Then at the end of two minutes he will have performed infinitely many of these tasks. He might, say, have written down the complete decimal expansion of π (3.141592...), for which he needs only a finite sheet of paper and the ability to write down digits that get smaller without limit (see [Fig. 1]). We are loath to admit this as a conceptual possibility, though we seem bound to do so [53, p. 4].

Moreover, such an expansion of π has been called in the literature a 'supertask,' and Moore [53] has provided a rather famous argument for the view that supertasks are incoherent."

Moore's statement that "We are loath to admit this as a conceptual possibility" is mystifying. Who does the 'we' refer to? Not to the authors: Moore wrote *The Infinite* by himself. The idea must be that *thinkers in general* are reluctant to admit this conceptual possibility—but this just isn't the case. In our experience, most thinkers

[26] Chalmers gives the case of a mile-high unicycle, which certainly seems logically possible. The burden of proof would surely fall on the person claiming that such a thing is logically impossible.

[27] For more on these issues in connection with the computational conception of mind, see Bringsjord [14].

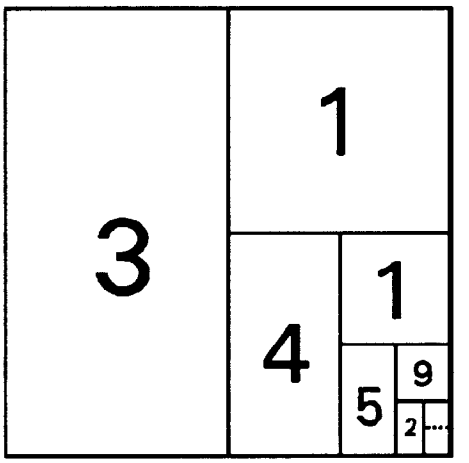

Fig. 1. Expanding π as a supertask.

rendering a verdict on whether the expansion is a conceptual possibility, believe, with Russell and all the others, that it is.

Indeed, it now seems clear that supertask computation is not only logically/mathematically possible, but *physically* possible as well, as shown by certain general relativistic models [31,43,59].

We should probably mention that even the simple notion of a limit, central to elementary calculus, according to many (e.g., Salmon [67]), presupposes the mathematical coherence of supertasks.[28] Even children are frequently taught that supertasks are perfectly coherent, because they are prepared early on, in mathematics, for calculus down the road. For example, see Fig. 2, which is taken from page 268 of Eicholz et al. [33]. A typical case recently involved Bringsjord's son, Alexander, who in the 7th grade was asked to determine the "percent pattern" of the outer square consumed by the ever-decreasing shaded squares. The pattern, obviously, starts at 1/4, and then continues as 1/16, 1/64, 1/256, ... When asked what percent "in the limit" the shaded square consumes of the original square, Alexander was expected to say "Zero"—but the notion of a limit was a bit tricky for him (perhaps understandably). When asked what percentage the shaded square would "get down to" if someone could work faster and faster, and smaller and smaller, at drawing the up–down and left–right lines that make each quartet of smaller squares, Alexander said zero. This is anecdotal, yes, but what it indicates is something in keeping with the process we discussed earlier: some humans may start with an intuitive picture of a supertask, and move from there to the formalisms in question (i.e., those seen in elementary calculus). It would be interesting to systematically poll students about these matters.

[28] Please note that the coherence of supertasks *simpliciter* most certainly does *not* imply that there is always a "final step" in a supertask. There can be a result in the limit without there being a last step. For a crisp, recent demonstration of this point, see the pair: [6,36].

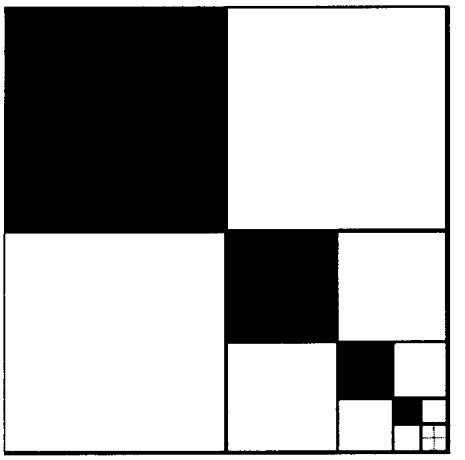

Fig. 2. Picture of supertask from seventh grade math.

Though Moore would seem to be singular in his belief that supertasks are not mathematically or logically possible, as the end of Objection 4 intimates, his ammunition includes more than a suspicious straw poll, against which we could certainly pit our own: He does have one *argument* for the incoherence of supertasks; this argument is part of a larger case for finitism (a view which denies even the existence of N, a set we have of course explicity invoked). Needless to say, we do not have the space to present and resolve one of the most profound issues in philosophy of mathematics, and if it takes an affirmation of finitism to ward off our modal argument, then our argument has all the force we could ever hope it to have (after all, if finitism is true, most of theoretical computer science is incoherent). But we can focus specifically on Moore's argument for the incoherence of what the two of us, following Russell, Weyl, Blake, Copeland, and others, find transparently coherent. Here, then, is Moore's argument:

> If it did make sense to say that I had just constructed all of the natural numbers in a minute, by the Zenonian procedure, then it would also make sense to say this: while I was constructing them, my constantly increasing speed of performance meant that time seemed to be going more and more slowly to me; it seemed that I was constructing them at a steady rate. Yet there is nothing that could count for me as a retrospective grasp of such an experience, in its apparent endlessness. (I could not have an apparently endless experience, apparently followed by further experience.) I must, subsequently, have forgotten all but an initial segment of it. How can this be? Surely what we have here is symptomatic of the fact that nothing could ever count, for anyone, as a grasp of an infinite reality. The grammar of 'infinity' is not geared to this. The special problems that arise when we envisage time seeming to go more slowly merely serve to make graphic an incoherence that is there to be acknowledged anyway—an incoherence that crept in at the very

> beginning of the story. It does not make sense to say that I have just performed infinitely many tasks of any kind, nor to say that anything is infinite in any respect [53, p. 213].

Put in a more explicit form that can be evaluated, and ignoring the grandiose claim that finitism is vindicated, the argument runs like this:

> If supertasks are coherent, then it could seem to someone who (conceptually speaking) performs such a task that each action along the way took the same amount of time (say n seconds). (9)

> If it could seem to someone who performs a supertask that each action along the way took n seconds, then someone could retrospectively experience, or relive, each action for n seconds—and then proceed to have other, normal life experiences. (10)

> No one could retrospectively experience, or relive, each action in a supertask for n seconds—and then proceed to have other, normal life experiences. (11)

> Supertasks are incoherent [(9), (10), (11)]. (12)

Is this argument any good? Well, the argument appears to be formally valid, an instance of hypothetical syllogism and *modus tollens*; at any rate, we are prepared to concede that the argument is valid. Premise (11) certainly seems to be true. After all, $n + n + n + \cdots$ will eventually exceed the amount of time a human has to operate with, whereas a supertask is based on some such sequence as $(1/2^n)$, $n = 1, 2, \ldots$. Premise (9) seems extremely plausible. For suppose that Jones is a sprinter who runs 100 m races in 9 s. Could not it nonetheless seem to Jones that he runs the race in 15 s? Could not it in fact seem to Jones that he runs the race in 30 s? And why not a minute? After all, we are talking here about Jones' subjective perspective. He could hallucinate during his sprint, or go into some kind of wild dream that seems to span 10 years. In general, then, when a human performs an action having property F, it may not seem to the human that action has F; the human may perceive the action to have a radically different property G.

So where does this leave the argument? It leaves it hinging on premise (10). But this premise is at best controversial; at worst, the premise begs the question against us. The problem is that it does not follow from the fact that a series of actions can *seem* to have certain properties to someone, that that person can in any way genuinely experience actions having these properties. It may seem to me that a moment ago I jumped a tall building in a single bound (perhaps I had a dream), but it does not follow from this that I can experience jumping a tall building in a single bound. I may *think* I am superman, but I am not. Likewise, it may seem to someone that they just squared the

circle, but they cannot experience squaring the circle.[29] (If one experiences squaring the circle, squaring the circle can be pulled off—but it *cannot* be pulled off.)

Finally, why do we say that premise (10) may make the argument in question circular? We imagine that once one has (conceptually) performed a supertask, the only way to retrospectively experience this task is for the retrospective experience to itself be a supertask. To the extent that premise (10) rules out by fiat this method of retrospective experience, it preemptively rules on precisely what's at issue: Moore's argument becomes a *petitio*.

Acknowledgements

For comments along the road to refining our argument, we're indebted to: Leslie Burkholder, David Chalmers, Jack Copeland, Martin Davis, Jim Fahey, Ken Ford, Pat Hayes, Andrew Irvine, Bob McNaughton, Marvin Minsky, Bill Rapaport, Kelsey Rinella, Stuart Shapiro, Chris Welty, and Michael Zenzen. Bringsjord would like to express special thanks to two minds he very much suspects do a lot of hypercomputing: Roger Penrose, for extemporaneous debate and conversation free of written text of the sort that, by our lights, has hampered his Gödelian case against the computational conception of mind; and Mark Burgin, for both seminal theory in the area of hypercomputation and leadership in organizing a symposium on this topic for the American Mathematical Society.

References

[1] J.R. Anderson, C. Lebiere, The Atomic Components of Thought, Lawrence Erlbaum, Mahwah, NJ, 1998.
[2] M. Ashcraft, Human Memory and Cognition, Harper Collins, New York, NY, 1994.
[3] R. Baron, M. Kalsher, Psychology, Allyn & Bacon, Boston, MA, 2000.
[4] A. Barr, Artificial intelligence: cognition as computation, in: F. Machlup (Ed.), The Study of Information: Interdisciplinary Messages, Wiley-Interscience, New York, NY, 1983, pp. 237–265.
[5] J. Barwise, Infinitary logics, in: E. Agazzi (Ed.), Modern Logic: A Survey, Reidel, Dordrecht, The Netherlands, 1980, pp. 93–112.
[6] R. Black, Solution of a small infinite puzzle, Analysis 65 (4) (2002) 345–346.
[7] R. Blake, The paradox of temporal process, J. Philos. 23 (1926) 675–684.
[8] G.S. Boolos, Introductory note to: some basic theorems on the foundations of mathematics and their implications, in: Collected Works III, Oxford University Press, Oxford, UK, 1995, pp. 290–304.
[9] G.S. Boolos, R.C. Jeffrey, Computability and Logic, Cambridge University Press, Cambridge, UK, 1989.

[29] Notice that it will do no good for Moore to replace (10) with

> If it could seem to someone who performs a supertask that each action along the way took *n* seconds, then it could seem to someone *both* that each task took *n* seconds *and* (once the supertask was over) that something else happened subsequently. (10′)

This proposition seems quite plausible. But since the consequent of this proposition itself seems plausible—again, it can even seem to someone that he/she has performed a mathematically impossible action like squaring the circle—, (10′) is of no help to Moore.

[10] S. Bringsjord, What Robots Can and Can't Be, Kluwer, Dordrecht, The Netherlands, 1992.
[11] S. Bringsjord, Computation, among other things, is beneath us, Minds Mach. 4 (4) (1994) 469–488.
[12] S. Bringsjord, Abortion: A Dialogue, Hackett, Indianapolis, IN, 1997.
[13] S. Bringsjord, An argument for the uncomputability of infinitary mathematical expertise, in: P. Feltovich, K. Ford, P. Hayes (Eds.), Expertise in Context, AAAI Press, Menlo Park, CA, 1997, pp. 475–497.
[14] S. Bringsjord, The zombie attack on the computational conception of mind, Philos. Phenomenol. Res. 59 (1) (1999) 41–69.
[15] S. Bringsjord, Review of John Searle's the mystery of consciousness, Minds Mach. 10 (3) (2000) 457–459.
[16] S. Bringsjord, In computation, parallel is nothing, physical everything, Minds Mach. 11 (2001) 95–99.
[17] S. Bringsjord, D. Ferrucci, Artificial Intelligence and Literary Creativity: Inside the Mind of Brutus, a Storytelling Machine, Lawrence Erlbaum, Mahwah, NJ, 2000.
[18] S. Bringsjord, B. van Heuveln, The mental eye defense of an infinitized version of Yablo's paradox, Analysis 66 (1) (2003) 61–70.
[19] S. Bringsjord, H. Xiao, A refutation of Penrose's Gödelian case against artificial intelligence, J. Exp. Theoret. Artificial Intelligence 12 (2000) 307–329.
[20] S. Bringsjord, M. Zenzen, Cognition is not computation: the argument from irreversibility, Synthese 113 (1997) 285–320.
[21] S. Bringsjord, M. Zenzen, Superminds: People Harness Hypercomputation, and More, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.
[22] M. Burgin, Universal limit Turing machines, Notes Russian Acad. Sci. 325 (4) (1986) 328–654.
[23] M. Burgin, How we know what technology can do, Commun. ACM 44 (11) (2001) 83–88.
[24] M. Burgin, Topological algorithms, in: Proc. ISCA 16th Internat. Conf. Computers and their Applications, Los Angeles, CA, 2001, pp. 61–67.
[25] D. Chalmers, The Conscious Mind: In Search of a Fundamental Theory, Oxford University Press, Oxford, UK, 1996.
[26] B.J. Copeland, Even Turing machines can compute uncomputable functions, in: J. Casti (Ed.), Unconventional Models of Computation, Springer, London, UK, 1998, pp. 150–167.
[27] B.J. Copeland, Super Turing machines, Complexity 4 (1998) 30–32.
[28] B.J. Copeland, Turing's O-machines, Searle, Penrose and the brain, Analysis 58 (2) (1998) 128–138.
[29] M. Davis, R. Sigal, E. Weyuker, Computability, Complexity, and Languages: Fundamentals of Theoretical Computer Science, Academic Press, New York, NY, 1994.
[30] E. Dietrich, Computationalism, Soc. Epistemol. 4 (2) (1990) 135–154.
[31] J. Earman, J. Norton, Forever is a day: supertasks in Pitowski and Malament-Hogarth spacetimes, Philos. Sci. 60 (1) (1993) 22–42.
[32] H.D. Ebbinghaus, J. Flum, W. Thomas, Mathematical Logic, Springer, New York, NY, 1984.
[33] R.E. Eicholz, P.G. O'Daffer, R.I. Charles, S.I. Young, C.S. Barnett, S.R. Clemens, G.F. Gilmer, A. Reeves, F.L. Renfro, M.M. Thompson, C.A. Thornton, Grade 7 Mathematics, Addison-Wesley, Reading, MA, 1995.
[34] S. Feferman, Penrose's Gödelian argument, Psyche 2.1. This is an electronic publication. It is available at http://psyche.cs.monash.edu.au/psyche/volume2-1/psyche-95-2-7-shadows-5-feferman.html, 1995.
[35] J. Fetzer, Mental algorithms: are minds computational systems?, Pragmatics Cognition 2 (1) (1994) 1–29.
[36] K. Friedman, A small infinite puzzle, Analysis 65 (4) (2002) 344–345.
[37] K. Gödel, Some basic theorems on the foundations of mathematics and their implications, in: Collected Works III, Oxford University Press, Oxford, UK, 1951/1995, pp. 304–323.
[38] M. Gold, Limiting recursion, J. Symbolic Logic 30 (1) (1965) 28–47.
[39] J.D. Hamkins, A. Lewis, Infinite time Turing machines, J. Symbolic Logic 65 (2) (2000) 567–604.
[40] S. Harnad, Other bodies, other minds: a machine incarnation of an old philosophical problem, Minds Mach. 1 (1) (1991) 43–54.
[41] J. Haugeland, Artificial Intelligence: The Very Idea, MIT Press, Cambridge, MA, 1985.
[42] D. Hofstadter, Waking up from the Boolean dream, in: Metamagical Themas: Questing for the Essence of Mind and Matter, Bantam, New York, NY, 1985, pp. 661–665.

[43] M. Hogarth, Does general relativity allow an observer to view an eternity in a finite time? Found. Phys. Lett. 5 (1992) 173–181.
[44] J. Holt, The Ada perplex: how Byron's daughter came to be celebrated as a cybervisionary, The New Yorker March 5 (2001) 88–93.
[45] G. Hughes, M. Cresswell, An Introduction to Modal Logic, Methuen, London, UK, 1968.
[46] P. Johnson-Laird, The Computer and the Mind, Harvard University Press, Cambridge, MA, 1988.
[47] P. Kugel, Thinking may be more than computing, Cognition 18 (1986) 128–149.
[48] G. LaForte, P. Hayes, K. Ford, Why Gödel's theorem cannot refute computationalism, Artificial Intelligence 104 (1998) 265–286.
[49] H. Lewis, C. Papadimitriou, Elements of the Theory of Computation, Prentice-Hall, Englewood Cliffs, NJ, 1981.
[50] J.R. Lucas, Minds, machines, and Gödel, in: A.R. Anderson (Ed.), Minds and Machines, Prentice-Hall, Englewood Cliffs, NJ, 1967, pp. 43–59 (Lucas' paper is available online at http://users.ox.ac.uk/~jrlucas/mmg.html).
[51] J.R. Lucas, Minds, machines, and Gödel: a retrospect, in: P. Millican, A. Clark (Eds.), Machines and Thought: The Legacy of Alan Turing, Oxford University Press, Oxford, UK, 1996, pp. 103–124.
[52] M. Minsky, Computation: Finite and Infinite Machines, Prentice-Hall, Englewood Cliffs, NJ, 1967.
[53] A.W. Moore, The Infinite, Routledge, New York, NY, 1990.
[54] A. Newell, Physical symbol systems, Cognitive Sci. 4 (1980) 135–183.
[55] R. Penrose, The Emperor's New Mind, Oxford University Press, Oxford, UK, 1989.
[56] R. Penrose, Shadows of the Mind, Oxford University Press, Oxford, UK, 1994.
[57] R. Penrose, Beyond the doubting of a shadow: a reply to commentaries on Shadows of the Mind, Psyche 2 (3) (1996), This is an electronic publication. It is available at http://psyche.cs.monash.edu.au/v2/psyche-2-23-penrose.html.
[58] R.S. Peters (Ed.), Body, Man, and Citizen: Selections from Hobbes' Writing, Collier, New York, NY, 1965.
[59] I. Pitowsky, The physical Church thesis and physical computational complexity, Iyyun 39 (1990) 81–99.
[60] E. Post, Recursively enumerable sets of positive integers and their decision problems, Bull. Amer. Math. Soc. 50 (1944) 284–316.
[61] H. Putnam, Trial and error predicates and a solution to a problem of Mostowski, J. Symbolic Logic 30 (1) (1965) 49–57.
[62] H. Putnam, Review of Roger Penrose's shadows of the mind, Bull. Amer. Math. Soc. 32 (3) (1995) 370–373.
[63] W. Rapaport, How minds can be computational systems, J. Exp. Theoret. Artif. Intell. 10 (1998) 403–419.
[64] W. Rapaport, Syntactic semantics: foundations of computational natural-language understanding, in: J. Fetzer (Ed.), Aspects of Artificial Intelligence, Kluwer, Dordrecht, The Netherlands, 1988, pp. 81–131.
[65] B. Russell, Our Knowledge of the External World as a Field for Scientific Method in Philosophy, Open Court, Chicago, IL, 1915.
[66] B. Russell, The limits of empiricism, Proc. Aristotelian Soc. 36 (1936) 131–150.
[67] W.C. Salmon, Space, Time and Motion: A Philosophical Introduction, Dickenson, Encino, CA, 1975.
[68] J. Searle, Minds, brains and programs, Behavioral Brain Sci. 3 (1980) 417–424, This paper is available online at http://members.aol.com/NeoNoetics/MindsBrainsPrograms.html.
[69] S. Shapiro, Mechanism, truth, and Penrose's new argument, J. Philos. Logic 32 (1) (2003) 19–42.
[70] H. Siegelmann, E. Sontag, Analog computation via neural nets, Theoret. Comput. Sci. 131 (1994) 331–360.
[71] H. Simon, Cognitive science: the newest science of the artificial, Cognitive Sci. 4 (1980) 33–56.
[72] H. Simon, Study of human intelligence by creating artificial intelligence, Amer. Sci. 69 (3) (1981) 300–309.
[73] N. Stillings, S. Weisler, C. Chase, M. Feinstein, J. Garfield, E. Rissland, Cognitive Science, MIT Press, Cambridge, MA, 1995.

[74] A. Turing, Systems of Logic Based on Ordinals, Ph.D. Dissertation, Princeton University, Princeton, NJ, 1938.
[75] A. Turing, Systems of logic based on ordinals, Proc. London Math. Soc. (Ser. 2) 45 (1939) 161–228.
[76] A. Turing, Intelligent machinery, in: B. Meltzer, D. Michie (Eds.), Machine Intelligence, Edinburgh University Press, Edinburgh, UK, 1969, pp. 1–24.
[77] A.M. Turing, On computable numbers with applications to the entscheidung-problem, Proc. London Math. Soc. 42 (1936) 230–265.
[78] H. Wang, From Mathematics to Philosophy, Keagan Paul, London, UK, 1974.
[79] H. Wang, On computabilism and physicalism: some sub-problems, in: J. Cornwell (Ed.), Nature's Imagination: The Frontiers of Scientific Vision, Oxford University Press, Oxford, UK, 1995, pp. 161–189.
[80] H. Weyl, Philosophy of Mathematics and Natural Science, Princeton University Press, Princeton, NJ, 1949.

R00082606_TCS_5000