Mathematical Objects Are Non-Physical, So We Are Too

Selmer Bringsjord & Naveen Sundar Govindarajulu

version 0930210840NY

Abstract

Dualists since at least Descartes have insisted that mental states such as *fearing ghosts*, as well as the bearers of such states (i.e., persons, or minds), are immaterial (= non-physical). But a different class of candidates for immateriality is to be found in the formal sciences. These candidates are logico-mathematical objects ranging from the familiar and conceptually simple to the exotic and complex. In this chapter we focus on two sub-classes of the familiar type of such objects: (1) algorithms (such as Quicksort, discovered by Tony Hoare); and (2) inference schemata, such as *modus tollens*. If we suppose for the sake of argument that such objects as algorithms and inference schemata are in fact non-physical, does it then follow that since humans interact with these objects they are *themselves* non-physical? Yes. We defend this answer herein; the defense makes use of one of the main arguments for the untenability of so-called "Strong" AI (the view that the artificial agents produced by AI can literally replicate human cognition and consciousness). This defense requires some analysis of and a response to the eponymous Benacerraf-Field Problem, which in a word says that we can't fathom how our justified belief in propositions regarding logico-mathematical objects could ever be explained. We supply this response herein. We end with brief remarks about more exotic logico-mathematical objects; specifically, infinite cardinal numbers (= cardinals), and in particular the two first such, $\aleph_0$ and $\aleph_1$.

Contents

1 Introduction

Chimpanzees, the chair in which we presently sit, and the chunk of aged cheddar cheese and wine before the first author on the table before him; these things are physical, clearly. Are there any *non*-physical things? Even those who would answer this question with an adamantine negative, if reflective, will agree that perhaps the best candidates for this category are not mental states had while enjoying such cheddar with fine Carménère (states which dualists since Descartes have long insisted are non-physical, since they are bearers of so-called — to use modern terminology — "qualia'[1]'), but instead logico-mathematical objects with which plenty of human persons, from their very earliest elementary-school years, are acquainted. These immaterial objects range from the familiar to the exotic, and are the targets of study in the formal) sciences.[2] We focus herein on two familiar and elementary classes of such logico-mathematical objects: (1) algorithms (such as the famous but simple Quicksort, discovered by Tony Hoare); and (2) inference schemata that form the foundation of rigorous reasoning in the formal sciences (such as *modus tollens*, that from two declarative propositions 'if ϕ then ψ' and 'not-ψ' one can validly deduce 'not-ϕ).' Inference schemata form this foundation because the formal sciences are theorem-driven, theorems are obtained by proofs, and proofs are sequences of propositions linked by inferences that must be sanctioned by such schemata (though often the schemata employed are left implicit and not called out by name). Of course, as the reader might imagine, may propositions *themselves* are logico-mathematical objects central to the formal sciences. For instance, the proposition that there are infinitely many primes, first proved by Euclid, would be such a specimen. Below, we shall also have occasion to discuss this proposition, and the general class into which it falls.

Given the context created by the logico-mathematical objects referred to in the previous paragraph, we can now inform the reader that the overarching structure of our case for the proposition that human persons[3] are immaterial will have two steps. In Step 1 we adapt and focus prior reasoning from James Ross (1992) in order to show that such objects as algorithms and inference schemata are non-physical (= immaterial). Then, in Step 2, we show that humans interact with these objects in a certain crucial way: viz., we *understand* them; specifically, we understand that we frequently validly implement them. We then argue, in part by appeal to Bringsjord's prior refinement of John Searle's (1980) famous Chinese Room Argument (CRA) for the proposition that such understanding can't be achieved by standard computing machines, that such understanding entails that we must ourselves be non-physical.[4] Of course, inevitably some will want to resist our ultimate conclusion. Accordingly, we consider and rebut some objections, including one based on the eponymous Benacerraf-Field Problem, which in a word says that we can't fathom how our justified belief in propositions regarding logico-mathematical objects could ever be explained. When we wrap up the paper, we briefly point to some more exotic formal objects than those routinely in play in K–12 math education; for example, infinite cardinals. Such exotica, it seems to us, is almost on its face immaterial. To keep things brief and simple, we refer only to the the smallest infinite cardinals: $\aleph_0$ and $\aleph_1$.

2 Logico-mathematical Objects, in General

Some readers may find the phrase 'logico-mathematical object' to be a bit of a mouthful, and perhaps even pedantic. Actually, the idea is quite straightforward, and the objects in question are encountered and reasoned over by even very young schoolchildren, who usually continue in this regard for many years, and are along the way introduced to more and more such objects of increasing complexity. One of the first logico-mathematical objects young children come across in the classroom is $\mathbb{N}$: the set of all natural numbers

$$\{0, 1, 2, \ldots\}.$$

This object is referred to via a "number line" shown graphically, and of course before this object is introduced, the young mind will have been told about the numbers 0, 1, 2, and so forth with help from such physical things as fingers, and will often have been introduced to the arithmetic functions of addition and subtraction applied to natural numbers. In public education in the U.S. State in which the first author resides, New York, Grade-4 mathematics instruction introduces students to a new logico-mathematical object that sometimes causes a bit intellectual turbulence: $\mathbb{Q}^+$: the positive rational numbers.[5] This introduction happens, of course, once these students are taught fractions, and how to add, multiply, and divide them. While $\mathbb{N}$ and $\mathbb{Q}^+$ are of the same size, a bit later students are introduced to the *real numbers*, $\mathbb{R}$. If they are lucky enough to reach study of the differential and integral calculus in Grade 11 or 12, our student is taught how motion and change can be understood with help from functions over $\mathbb{R}$. In addition, in today's world, it's likely that our young student will be introduced as well to logico-mathematical objects that are part and parcel of computer programming and computer science — objects such as *arrays*, *algorithms*,[6] and so on. In the course of learning mathematics, or computation, inevitably the student will also be introduced to the basic Boolean operators of *and*/$\wedge$, *or*/$\vee$, *not*/$\neg$, *if* ... *then* .../$\rightarrow$, and ... *if and only if* .../$\leftrightarrow$.[7] And, finally, our student will be taught how to check and create some proofs, since this is standard fare in Algebra 2 and Geometry, both required in 47 of the 50 U.S. States (in public education). The role of proof and proof creation in secondary mathematics education in the U.S. can be clearly seen by turning to the Common-Core textbooks for Algebra 2; see for example (Bellman, Bragg & Handlin 2012). We mention all of this just to ensure that the reader understands that logico-mathematical objects, and human interaction with them, are routine, extensive, and persistent.

Although we shall need to be more specific below, the classes of logico-mathematical objects cited to this point will pretty much suffice for the remainder of the present essay.

3 Narrowing the Focus to Two Simple Exemplars

To make matters more concrete, let us focus on just a few particular logico-mathematical objects, and then anchor subsequent discussion to them. What specimens should we select as exemplars? Well, mildly put, there are quite a few logico-mathematical objects. How many? Any serious attempt to answer this question would overwhelm all the space available in the present chapter. Let's rest content with the simple observation that the universe of such objects is infinite, and with the helpful follow-on observation that this universe can be to a degree rationalized by approaching it in accordance with the sub-parts of the universe that are associated with particular disciplines within the formal sciences, and sub-parts of these disciplines. We now pass to our exemplars.

3.1 Exemplar 1, an Algorithm: Quicksort

Quicksort itself, which we denote by '$\mathcal{Q}$,' and Hoare's (1961) discovery[8] of it, are both deservedly well-known in the computational formal sciences, and this shall be our first exemplar. The algorithm itself is at bottom a simple recursive one. (There are now numerous variants, but we ignore this for efficiency.) The algorithm is to receive an array of ordered objects, for example

$$\left\langle \fbox{5} \quad \fbox{9} \quad \fbox{10} \quad \fbox{7} \quad \fbox{4} \quad \fbox{3} \quad \fbox{11} \quad \fbox{8} \quad \fbox{6} \right\rangle,$$

and to then produce as output the sorted version of this input, which in this case is:

$$\left\langle \fbox{3} \quad \fbox{4} \quad \fbox{5} \quad \fbox{6} \quad \fbox{7} \quad \fbox{8} \quad \fbox{9} \quad \fbox{10} \quad \fbox{11} \right\rangle.$$

So, what's the algorithm? In order to answer this question, we can't avoid resorting to what we can call *embodiments* or *tokens* of the general and abstract *type* $\mathcal{Q}$.[9] This terminology, and the associated concrete practice, is easy to grasp. For example, here's one high-level embodiment/token $\hat{\mathcal{Q}}_1$ of $\mathcal{Q}$ that views the algorithm as a three-stage one:[10]

I Pick the rightmost element in the array as the *pivot*.

II Partition the array so that all elements in the array less than the pivot are before it, and all elements greater than the pivot are placed after it.

III Recursively apply both I and II to the sub-array now before the pivot, as well as to the sub-array now after the pivot.

This is said to be 'high-level' for obvious reasons. $\hat{\mathcal{Q}}_1$ doesn't tell us how to carry out partitioning, and it relies on an understanding of what recursion means — or at least what it means in this context. But no worries: Stage II can be further specified by saying that we simply move to the left one entry at a time, and decide whether to move an entry to the right of our pivot, or else leave it where it is. And how to decide? Simple: If what we find is greater than our pivot, append it to whatever sub-sequence is to the right of the pivot; otherwise just leave what we find alone. Using a double-box to indicate our pivot, the result of executing Stage I and then Stage II in $\hat{\mathcal{Q}}_1$ on the initial input array will result in this configuration:

$$\left\langle \fbox{5} \quad \fbox{4} \quad \fbox{3} \quad \fbox{\fbox{6}} \quad \fbox{8} \quad \fbox{11} \quad \fbox{7} \quad \fbox{10} \quad \fbox{9} \right\rangle.$$

Now the algorithm calls for Stage III in $\hat{\mathcal{Q}}_1$, which means that the sub-array to the left of $\fbox{\fbox{6}}$ with $\fbox{3}$ as the pivot of this sub-array is processed; ditto for the sub-array to the right of $\fbox{\fbox{6}}$ with $\fbox{9}$ as the pivot of this sub-array. In the case of the right sub-array, here's the result of running Stage I, which is to be passed to Stage II to be processed (we once again indicate the pivot by a double-box):

$$\left\langle \fbox{8} \quad \fbox{11} \quad \fbox{7} \quad \fbox{10} \quad \fbox{\fbox{9}} \right\rangle.$$

Stage II applied to the input to it immediately above then results in this:

$$\left\langle \fbox{8} \quad \fbox{7} \quad \fbox{\fbox{9}} \quad \fbox{10} \quad \fbox{\fbox{11}} \right\rangle.$$

We continue in this way until we reach sub-arrays composed of but one element, which are by definition sorted, and hence processing is guaranteed to terminate.

It should be obvious to the reader that an infinite number of embodiments or tokens of Quicksort are available.[11] Many of these embodiments call upon programming languages used today. We shall assume, going forward, that $\hat{\mathcal{Q}}_2$ refers to an embodiment of Quicksort $= \mathcal{Q}$ that is expressed in the modern functional programming language known as Clojure.[12]

3.2 Exemplar 2, an Inference Schema: ***Modus Tollens***

Next, we use a variant of the famous "Wason Selection Task" (WST) (Wason 1966) to anchor our presentation of *modus tollens* = *MT*, which can be referred to by way of the following oft-used token:

$$\frac{\phi \rightarrow \psi,\ \neg\psi}{\neg\phi}\ \widehat{MT}_1$$

This token tells us that if we have two formulae of the form indicated by the two expressions above the horizontal line (the first a conditional and the second the negation of the consequent of that conditional), then the inference schema in question allows us to infer what's below the horizontal line, namely that the antecedent in the conditional can be negated.

Now here's our selection-task challenge: Imagine that, operating as a teacher of mathematics trying to transition one of our students to proof (from mere calculation), we have a deck of cards, each member of which has a digit from 1 to 9 inclusive on one side, and a majuscule Roman letter A, B, ..., K on the other. From this deck, we deal onto a table in front of one of our students the following four cards:

E	T	4	7
c1	c2	c3	c4

Now suppose that we inform the student that the following rule R is absolutely guaranteed with respect to the entire deck, and hence specifically also for the four cards c1–c4 now lying in front of the student: "Every card with a vowel on one side has an even positive integer on the other side." Next, we issue the student the following challenge:

C Does card4 have a vowel on its other side? Supply a proof to justify your answer.

What should the student do in order to succeed? It should be clear that the student should answer in the negative, and provide a proof that makes use of *modus tollens*, such as in the following sequence, which we trust will be readily understood by all our readers, after a bit of inspection:

Line #	**Proposition**	**Justification**
1.	$\forall c(\mathit{Vowel}(c) \rightarrow \mathit{Even}(c)]$	Rule R
2.	$\neg \mathit{Even}(c4)$	from observation
3.	$\mathit{Vowel}(c4) \rightarrow \mathit{Even}(c4)]$	from 1.
4.	$\neg \mathit{Vowel}(c4)$	from 2. & 3. by $\widehat{MT}_1$

In this proof, the final step to yield line 4., as indicated, makes use of *modus tollens*. Note that match of lines 2., 3., and 4. with the token of *modus tollens* that is $\widehat{MT}_1$, given above.

3.3 Additional Exemplars: Proof-by-Cases Schema, and a Theorem Via It

While below we devote a dedicated section (§6) to anticipating and disarming objections to the main theses we advance in the present essay, we now take a few moments to quickly dispose of a weak objection that will inevitably come to the minds of some readers. The objection is simply that *modus tollens* is in fact not explicitly used in the proofs given by practicing formal scientists. While we fail to see why a failure to call out *modus tollens* explicitly by name raises any question for our case for the immaterial nature of logico-mathematical objects, let us simply admit that It's true that while *modus tollens* is routinely taught in introductory formal logic,[13] rarely do professional formal scientists cite it explicitly. But this is irrelevant, for two reasons. The first reason is that the relevant professionals *do* in fact use *modus tollens* — they just don't cite it by name. This practice is the same as that followed when theorems having the form of a biconditional $\phi \leftrightarrow \psi$ are proved, because often the proofs in question are divided into two phases, one the so-called "left to right" direction in which ϕ is assumed, and then the so-called "right to left" or "converse" direction in which ψ assumed and, eventually, ϕ deduced. The second reason why *modus tollens* is perfectly fine as an exemplar is that we could just as well use inference schemata that are more robust, and which *are* explicitly called out and used in more sophisticated formal-science deduction. One nice exemplar in this regard is "proof by cases." Unlike *modus tollens*, proof by cases is a schema employed in some fairly famous, indeed in some cases even "immortal," deductive reasoning. The schema says that if we have some disjunction, and in addition perceive that each disjunct leads by some reasoning to our goal γ, then we can conclude from the disjunction itself that γ holds. Let us call this expression of the inference schema '$\widehat{PBC}_1$.'[14]

Let us end this section by referring to another type of logico-mathematical object, one that will be necessary to have in play when we arrive at §6: viz., axioms and theorems. All readers will be well-acquainted with the fact that the formal sciences make crucial use of both of these things. In early math education, when students first learn arithmetic, they are essentially learning how to calculate on the basis of axioms and theorems in nothing less than the branch of mathematics called *number theory*, but such calculation is rarely described in terms of axioms and theorems. However, all students who pay attention and progress through basic secondary mathematics education are explicitly exposed to Euclid's axioms for plane geometry, and are asked to deduce some simple theorems from these axioms. What we need for present purposes is just a simple, single exemplar from this category, and without loss of generality we choose what is commonly called "Euclid's Theorem," a rather famous result that says that there are infinitely many prime numbers. Conveniently, part of what's clever in the Euclidean proof of the theorem is a use of proof by cases, and as a matter of fact *modus tollens* too. For use below, let us denote the theorem by '**ET**,' and leave the clever-but-not-difficult proof itself in an endnote.[15]

4 The Two-Step Argument for the Immateriality of Us

As the reader will recall, our overarching argument for the proposition that humans (more precisely, to remind the reader, human *persons*; see note 3) are immaterial is a two-step one. In Step 1 we show that algorithms and inference schemata (and, as an obvious consequence, also axioms and theorems) are non-physical, by building atop some seminal prior work by James Ross. In Step 2, we then show that the humans who (sometimes) correctly use such algorithms and schemata are themselves non-physical (= immaterial). Let's move directly to Step 1 now.

4.1 Step 1: Algorithms and Inference Schemata are Non-Physical

Step 1 builds upon an insightful argument given by James Ross (1992) for a related proposition .[16] This related proposition is that "formal thinking" isn't a physical process. Here is an encapsulation of Ross's argument for this proposition in his own words:

> Some thinking (judgment) is determinate in a way no physical process can be. Consequently, such thinking cannot be (wholly) a physical process. If all thinking, all judgment, is determinate in that way, no physical process can be (the whole of) any judgment at all. Furthermore, "functions" among physical states cannot be determinate enough to be such judgments, either. Hence some judgments can be neither wholly physical processes nor wholly functions among physical processes. (Ross (1992), p. 137)

As you can see, Ross's objective is to establish that the process of thinking about, and thinking guided by, formal structures isn't physical. (As seen in the previous quote, he terms this process "judging.") In contrast, our sub-objective (= the objective of Step 1) is to establish that the *things* bound up with such formal thinking are non-physical (and of course the overarching goal, again, is to show — in Step 2 — that *we* are not physical things). However, it's easy enough to adapt the argumentation that Ross gives, and to sharpen it. We do this now.

To begin, recall the exemplars and related things that we have at our disposal at this point: the Quicksort algorithm $\mathcal{Q}$ and any number of embodiments thereof (we have $\hat{\mathcal{Q}}_1$ and $\hat{\mathcal{Q}}_2$ on hand; the latter embodiment is the Clojure code we provide in the relevant endnote); the inference schema *modus tollens* $= MT$ and any number of embodiments thereof ($\widehat{MT}_i$, given by us above, is on hand for use); and in addition we have on hand other inference schemata of greater complexity as needed (in particular *proof by cases* and the embodiments $\widehat{PBC}_1$ and $\widehat{PBC}_2$, and the theorem **ET**, with a token of this theorem shown in footnote 15. Next, here is the proposition that we establish in Step 1 (where of course Smith is an arbitrary stand-in for any human agent):

($\star$) If Smith validly implements $\mathcal{Q}$ via for instance the aforementioned Clojure program $\hat{Q}_2$, or validly instantiates inference schema MT via in for instance a proof having an inference conforming to $\widehat{MT}_1$, then in neither case is this validity due to satisfaction of some relation R holding between Smith and some physical embodiment of the abstract types in question.

We can introduce a more perspicuous variant of this proposition by representing the ternary validity relation in it by *Val*, and the referenced relation posited to constitute satisfaction of the *Val* relation by R_1 in the case of our algorithm, and R_2 in the case of our inference schema. Then we have two more precise propositions derived from ($\star$), respectively. Note that we give an English reading in both of these sub-cases, which immediately follow, and which make use of elementary logic, with the existential quantifier $\exists$ — as our English renditions show — for "there is at least one ...." The structure of the sub-cases is to assert that the validity relation holds exactly when some other relation between our agent and a material embodiment holds.

($\star'_1$) $\neg\exists R_1[\mathit{Val}(s, \mathcal{Q}, \hat{\mathcal{Q}}_1)]$ if and only if $R_1(s, \hat{\mathcal{Q}}_1)$

English: It's not the case that there is a relation R_1, holding *only* between Smith and the material embodiment $\hat{\mathcal{Q}}_1$ (of $\mathcal{Q}$), in virtue of which agent Smith validly implements Quicksort $= \mathcal{Q}$.

($\star'_2$) $\neg\exists R_2[\mathit{Val}(s, MT, \widehat{MT}_1)$ if and only if $R_2(s, \widehat{MT}_1)]$

English: It's not the case that there is a relation R_2, holding *only* between Smith and the instantiation $\widehat{MT}_2$ (of MT), in virtue of which agent Smith validly instantiates *modus tollens* $= MT$.

Next, we need to articulate a Rossian argument that can do the job of establishing both $(\star_1')$ and $(\star_2')$. In this argument, we essentially rely upon what Ross relies upon (see footnote 16), but we give our own rationale to make the present essay self-contained. Here's the argument:

The Step-1 Argument

Suppose without loss of generality that our agent Smith, using paper and pen, writes down the program $\hat{\mathcal{Q}}_2$ as a physical token/embodiment of $\mathcal{Q}$, and also writes down a proof that instantiates $\widehat{MT}_1$, the token of *MT*. For fixity, but without loss of generality, the embodiments in both cases can be pieces of paper upon which Smith writes. How can we be sure that the validity of what Smith does here cannot be due to the satisfaction of some relation holding exclusively between Smith and these physical embodiments; that is, how do we apprehend that the negation of $(\star)$ doesn't hold? Well, let's temporarily suppose otherwise; that is, we assume $\neg(\star)$. Now, we ask a simple question: How many embodiments of $\mathcal{Q}$ and *MT* are there? Our parable has featured only one, and our prose above have in the case of both Quicksort and *modus tollens* presented a total of two — but obviously there are many, many more. How many, then? Clearly, there are an *infinite* number of such embodiments. Minimally, there are as many embodiments as there are natural numbers, so we can index embodiments to these numbers. But in each case, the embodiment E_j, where $j \in \mathbb{N}$ is different, indeed often radically so. So the relation *Val* must be one holding between our Smith, the particular embodiment $\hat{\mathcal{Q}}_i$, and $\mathcal{Q}$, in virtue of some *ternary* relation, one that includes ranging over $\mathcal{Q}$ itself/MT itself! For if this is not the case, what serves to unite the infinite embodiments E_j? So, now, are we to take $\mathcal{Q}$ and *MT* to be physical, or non-physical? It cannot be the former. For then we are right back to where we started from, since we then have merely a relation between our agent and a particular embodiment, say E_k, by definition (since every physical thing is embodied). We thus arrive at having to affirm $(\star)$, the opposite of what embroiled us in this trouble.

4.2 Step 2: Why *We* Are Immaterial

Now we come to Step 2. This is the harder, more intricate step. In particular, it's one that involves an understanding of the limitations of machine intelligence, or what is called 'AI,' for 'artificial intelligence.' In particular, we refer here to limitations on so-called "Strong" AI, which aims not only to produce intelligent machines able to *behave* like human persons, but machines of this type to quite literally *be* persons (with cognition and consciousness at and possibly above the human level). The distinction between Strong vs. Weak AI is discussed at length in (Bringsjord 1992), but we take the distinction at this point to be sufficient for present purposes, which we now continue to advance.[17]

To begin Step 2, we assert that whether or not the vast majority of scientists whose professional business it is to study human persons and their brains believe that these persons are physical things, the vast majority of such thinkers *say* that they believe this, when in conventional scientific venues.[18] For the sake of argument in the present paper we shall take these scientists at their word. But what word, exactly? We doubt we can find a better case in point than John Anderson, and the mentor he venerates, Allen Newell, one of the illustrious founders of AI, and a substantive contributor at the famous Dartmouth conference that marks the birth of modern AI (as explained e.g. in Bringsjord & Govindarajulu 2018). Specifically, we can consider Anderson's *How Can the Human Mind Occur in a Physical Universe* of 2009. The title of this book should send a clear message to our readers. The core, driving idea in the book is the two-part one that, first, whatever the mind (i.e., human persons; see again footnote 3) is and however it might work, it's certainly physical; and so, second, we need an account of how the mind works that at least fits what physics tells us, and preferably an account that itself is, if you will, "physics-ish."

Step 1 in our overarching argument relied significantly upon analysis and argument given by Ross (1992). Step 2 relies upon insights from John Searle (1980) , first given in his landmark paper "Minds, Brains, and Programs." There, Searle introduces his famous "Chinese Room Argument" (CRA), which we have already introduced above. The Chinese Room is a room (some refer to it as a "box") he enters, accompanied only by a "rulebook" that allows him (in concept: he has to work fast!) to output symbol strings in response to such strings coming into the room — but all the while Searle has no idea what the symbols in question mean, because they are in Chinese, a language he doesn't at all understand. Subsequently, Bringsjord refined these insights and provided the relevant argumentation in (Bringsjord 1992), which he then further refined in (Searle 1980). There is insufficient space to rehearse the Searlean argument in question in any detail. It suffices to report here simply that the argument's conclusion is that human understanding of what symbolic inscriptions mean cannot possibly be achieved by a standard computing machine. Such a machine, for instance a modern high-speed digital computer, can process all sorts of symbolic inscriptions, but it cannot have human-level understanding of these inscriptions. Why? The key part of the Searlean argument given in its most mature form in (Searle 1980) is the answer to this question, which in essence is: "Because when we consider whether understanding is conveyed by our own mere symbol manipulation à la computing machine (of any symbols, with processing governed by any program), we see that we have no understanding thereby whatsoever. But we *would* have such understanding if a computing machine could achieve understanding by its mere computing." For a more recent defense of this line, in connection with whether modern robotics can endow a robot with human-level understanding, see (Bringsjord 2015*b*).

Now, given this brief background, what is the Step-2 argument for the conditional that if logico-mathematical objects are immaterial, we are as well? Efficiently put in the interests of obvious space constraints, we articulate it as follows.

The Step-2 Argument

As humans, we can *understand* Quicksort = $\mathcal{Q}$ and *modus tollens* = MT, clearly. In fact, you, the reader, are in this fortunate group of understanders. This understanding, as we appreciate in the light of Searlean argumentation (see above), happens only when we do more than move around particular symbols and diagrams (that can be used to token the relevant types). These things are all mere embodiments, by definition.[19]

Now, if the immateriality of the logico-mathematical objects in question don't imply that we are ourselves immaterial, what are we? We would be exactly what John Anderson says we are: biological, and hence physical, computers.[20] But then even as such we can understand Quicksort and *modus tollens*, this despite that as such computing machines we are restricted to interaction consisting of our manipulation of *particular* embodied symbols and diagrams. Indeed, such manipulation is the essence and full reach of what a computing machine is. In other words, we all as humans on this line become nothing more than Searles-in-the-room! But this is inconsistent with what is possible for things-in-the-room: these things, as mere symbol manipulators, can't have understanding therefrom . Hence, the assumption of our materiality has led us to contradiction, and we therefore in fact can't be material.

5 Regarding Related Work

Menuge (2016) has written an excellent paper that relates in interesting ways to the two-step argument given above.[21] His paper is an attempt to "show that materialism is incapable of explaining a large and important area of human knowledge" (p. 7), viz. knowledge of abstracta, including specifically some of the formal objects upon which we focus here. Menuge appears not to be aware of the reasoning of Ross to which we crucially and centrally appeal in our Step-1 argument;[22] this is noteworthy because some of what Menuge writes about "understanding" a rule of inference is directly in line with Ross (e.g. see Menuge 2016, p. 23). While we avowedly rely upon Ross, Ross in turn relies upon what he says are certain long-established "jewels of analytic philosophy" (Ross 1992, p. 137).[23] To really flesh out our Step-1 argument, we would analyze and tap directly into its roots in the work of those Ross cites as those he's building upon, most prominently Nelson Goodman. The power of Ross's paper inheres in no small part in the fact that though his position on the nature of thinking is that of an immaterialist (and a theist), his support is found in many of those who are nothing of the sort (and who are often not theists).

Many additional points could be written about the relationship between Menuge's paper and ours; we shall rest content with the following additional three:

1. In general, we do not at all wish to go in the "knowledge connection" direction. Notice that we are not talking about knowledge in the case of the agents in our parables. Take a look at the key propositions we set off typographically in the Step-1 argument given above. There is no reference in those to knowledge or belief in said agents. In particular, there is no mention of knowledge at all in the agents that validly follow an algorithm or an inference schemata. We see talk of knowledge as a bit of a potential quagmire, to be avoided. Note that most physicalists who are computer scientists will simply maintain, *contra* Menuge, that while it's true for instance that they know that every natural number n is greater than 0, they *don't* know that $1 > 0$, that $2 > 0$, that $3 > 0$, *ad infinitum*.[24] And most of these folks will not agree that when Jones validly codes or follows Quicksort = $\mathcal{Q}$ Jones knows that something holds of an infinite number of cases, or knows an infinite number of propositions. An argument like that is, as we see things, much better to make directly of professional mathematicians and logicians — but even then you run into epistemic finitists. We want to avoid this rabbit hole (which is not to say that we agree with such skeptics). At a minimum, because we are inclined to require formal argumentation/proof in sorting out such matters, we would need to demonstrate, formally, that an agent's knowing some proposition ϕ entails this agent's knowing at least as many propositions $\psi_1, \psi_2, \ldots$ as there are natural numbers. We confessedly find such a need daunting — despite the fact that we have worked on infinitary knowledge and belief from an at-once formal and computational perspective; see for instance (Arkoudas & Bringsjord 2004).
2. While Menuge articulates an abductive argument for the existence of a transcendant being, our two-step argument eventuates in conclusion about the nature of human persons, and stops there. One could certainly promisingly explore linking from what we conclude, to Menuge's abductive reasoning about God — but that exploration requires a separate future day.

6 Refuting Two Objections

We now anticipate and refute two objections.

6.1 "We Simply Legislate Logico-mathematical Objects!"

Here's how the present objection can be expressed: "Your case for the immateriality of human persons is vitiated by a simple fact: these persons *legislate* the logico-mathematical objections upon which your case is built. That this is so is shown by Lakoff & Nuñez (2000)."

There are at least two problems with this objection, both of which are fatal.

The first problem is that the objection is flatly self-refuting: if it's sound, it's unsound. This defect relates to the notion that logic is invincible, which can be encapsulated by way of the following short parable:

> Jones holds that belief fixation by rational agents should be based upon the construction and assessment of arguments. In particular, Jones holds that rationally believing some proposition ϕ can happen only if there is some argument for ϕ of which the rational agent in question is aware, and if, in addition, that agent understands the formal validity of the argument in question. For instance, ϕ might be some expression of this in-English proposition: "The cardinal number $\aleph_0$ is a non-physical object." Smith challenges Jones, by expressing his dripping disdain for logic, which he (Smith) regards to be worthless, or at least just plain wrong. But if Smith's reasoning succeeds, then it does so on the strength of making use of logic. It then follows that if Smith's reasoning succeeds it fails, since that reasoning is aimed at establishing that logic is worthless and so on.

The problem here can of course be expressed without a story, and instantiated to the present objection, as follows. The critic here appeals to the book *Where Mathematics Comes From: How The Embodied Mind Brings Mathematics Into Being* (= *WMCF*), by Lakoff & Nuñez (2000), in which these two authors (L&N) argue that human persons are *physical* creatures that create mathematical objects — including therefore the particular objects upon which we place weight in the present essay: i.e., algorithms and inference schemata. For ease of exposition, let's say that L&N offer only one argument for this claim, and let's label that argument α. The present objection has force only if α is formally valid; that is, only if the inferences in α conform to inference schemata that regiment normative correctness in reasoning (including, most certainly, *modus tollens*). Now suppose that L&N are correct. Then it follows that the inference schemata that undergird α have themselves been legislated by humans. But then α is not really valid at all, for what's to stop someone from advancing an argument for the falsity of the very inference schemata that L&N have employed in their α? Nothing (assuming that L&N are correct). Hence, if L&N are correct, they end up refuting themselves.

The reader should rest assured that a close analysis of *WMCF* more than fully supports our claim that the case given in this book is self-refuting. The reason, in short, is that L&N appear to be completely unaware of the fact that their own argumentation hinges on the non-arbitrariness of inference schemata. They seek to show, for instance, that what they call the "laws of arithmetic" arise from particular "cognitive mechanisms" that run in "embodied minds" — but they are blithely unaware of the fact that the inference schemata conformity to which is a *sine qua non* for the soundness of their argumentation, if merely arising directly from the physical mechanisms in question, are entirely arbitrary, and hence unsound.

Oddly enough, they do ask this question, and we quote: "And why, in formal logic, does every proposition follow from a contradiction?" (L&N 2000, p. XIII) Leaving aside the fact that there is no such thing as 'formal logic' as a single system in which the inference schema — explosion — to which L&N here allude (formal logic is instead a discipline, one that covers and invents infinitely many particular formal logics, some of which as a matter of fact don't include explosion), the question we press is: What about the "law" labeled *modus tollens*? L&N rely upon it, and upon many other such "laws." But if the laws just emerge from the particular physical things that L&N are, and the laws are themselves physical things arising adventitiously from the physical mechanism of cognition on the part of L&N, then what's to stop another agent from showing up and saying that they reject *modus tollens* in favor of some preferred schema of their own that marks a rejection of *modus tollens*?[25]

6.2 "The Benacerraf-Field Problem Refutes You!"

The objection here can be stated in compressed form as follows: "Your position, and reasoning you give to support it, presuppose a solution to an unsolvable problem: the so-called Benacerraf-Field Problem. Hence you position shouldn't be affirmed."

Let's label the problem in question, which we suspect a number of our readers may be unaware of, 'B-FP.' The reason for the hyphenation is that this problem is a refined version of the problem as originally described by Benacerraf (1973) writing alone; let's dub this 'BP.' Here's an encapsulation of Part 1 of the problem as Benacerraf summed it up:

> [O]n a realist (i.e., standard) account of mathematical truth our explanation of how we know the basic postulates must be suitably connected with how we interpret the referential apparatus of the theory. ... [But] what is missing is precisely ... an account of the link between our cognitive faculties and the objects known. ... We accept as knowledge only those beliefs which we can appropriately relate to our cognitive faculties. (Benacerraf 1973, p. 674)

Given the foregoing context we have laid out, 'postulate' can be understood to be an axiom or theorem in the formal sciences, and in particular our exemplar **ET**, Euclid's Theorem, can be used without loss of generality, and without begging any questions against Benacerraf. Now, Part 2 of BP is the claim that there can in fact not be a "suitable connection" between the "basic postulates" of mathematics and the "cognitive faculties" of human persons. But why is Benacerraf pessimistic in this regard? What supports the claim? The answer is that he insists that (i) any suitable connection must be a *causal* one, and that (ii) there can't be a causal connection between such an agent and a postulate. The idea, specifically tied to our context, would be that between you and **ET**, which you can be assumed to now fully understand (by virtue e.g. of having assimilated footnote 15), there must be some sort of causal connection. Here's a quote that confirms this interpretation, from Benacerraf himself:

> I favor a causal account of knowledge on which for X [= you] to know that S [= **ET**] is true requires some causal relation to obtain between X and the referents of the names, predicates, and quantifiers of S. ... [But] ... combining *this* view of knowledge with the "standard" view of mathematical truth makes it difficult to see how mathematical knowledge is possible. ... [T]he connection between the truth conditions for the statements of number theory [such as **ET**] and any relevant events connected with the people who are supposed to have mathematical knowledge cannot be made out. (Benacerraf 1973, pp. 671–673)

It's at this point easy to see that BP poses not the slightest problem for the view we have advanced about logico-mathematical objects, the nature of human mentation regarding them, and the nature of the agents who enjoy such mentation. How? Well, the BP is baldly based on the particular theory of the interaction between these agents and things like Euclid's Theorem and Quicksort and *modus tollens* that Benacerraf himself happens to like. For confirmation, simply look again at the quote immediately above. He informs us that he happens "to favor a causal account" of knowledge; and it's only on this account that the relation between agents and logico-mathematical objects becomes problematic. The idea seems to be that there must be some sort of *physical causal* connection between a human and, say, **ET**. Benacerraf is simply begging the question against anyone like Ross, or the two of us, who believe and seek to establish that there is no such causal connection to be had. And why not? Because we hold that the objects are immaterial, that the agents are too; and obviously then the immediate implication is that there's no causal relation between agents and the objects to be had![26]

Our case for the immateriality of human persons is not yet clear of the general idea expressed by Benacerraf. For so far we have considered only BP, the original version of the problem as specifically defined by Benacerraf; we have not yet considered and disposed of a variant of BP introduced by Field (1989): the variant B-FP. As Clarke-Doane (2017) reports (p. 20), philosophers of mathematics today invariably take what's at issue to be B-FP, not the original BP. Well then, what *is* B-FP? We do not have the luxury of giving a full presentation of the problem, and then proceeding to a detailed refutation. But no matter, for B-FP, in broad strokes, is quite easy to convey: it's the problem that "it appears in principle impossible to explain" (Field 1989, p. 233) in *any* way how it is that our beliefs align so perfectly with logico-mathematical objects. Field doesn't demand a causal explanation; he just demands an explanation, insists that there simply isn't one to ever be had, and then says that because of the absence of such an explanation our belief in mathematical entities is — to use his word — "undermined" (p. 233).

In response we offer what is as far as we are aware a new counter-objection to BP/B-FP, one rooted in work we have carried out in the intersection of self-belief and AI (e.g. Bringsjord & Govindarajulu 2020). In this work we have among other things presented axioms for characterizing, precisely, what we call *cognitive consciousness*, including in particular cognitive *self*-consciousness. From this work we only need here a sliver of one of the axioms in question: namely, that at least in the case of human self-consciousness we have beliefs about our own occurrent mental states, and those beliefs are true. For instance, all of us every now and then believe that we are angry (beyond just simply being angry), and sometimes we believe that we are angry and really shouldn't be (perhaps because of some ethic we subscribe to, but have nonetheless violated after succumbing to temptation), and so on. To focus things, consider not anger, but pain, acute pain. Suppose that Tommy believes at some time t that he is in acute pain at t. Is Tommy's belief correct? Well, how could he possibly be mistaken about such a thing? When you believe that you are in excruciating pain at some particular time, you *are* in excruciating pain.[27] This fact is what — to use a term sometimes used by philosophers of mind — makes such beliefs "incorrigible."

But notice then how the Benacerraf-Field Problem is obliterated. This happens because we absolutely, positively cannot give an explanation for why our beliefs that we are in excruciating pain are veridical. It's not like we have some argument in support of such beliefs, or empirical evidence from some experiment we have run, or the testimony of some other agent who assures us that we're indeed in pain; no, nothing of the sort, at all. When we believe that we are in pain, we are right; and the absence of some explanation for this alignment does not in any way impugn the brute fact that are beliefs are correct. It follows from this that Field's objection melts away, for we have a case where the absence of an explanation puts not the slightest dent in the reliability and correctness of our beliefs. This shows that the general premise employed in the B-FP against the immateriality positions we have presented and defended herein, the premise that lack of explanation for correct belief about abstract, seemingly non-physical things (e.g., pain) provides reason to doubt the accuracy of those beliefs, is destroyed.[28]

7 Conclusion, The Cardinals, and Beyond

So, in sum, we are immaterial — if we're right. We have little doubt that some physicalists will remain obdurate, despite the two-step argument we have given, and defended. We also have little doubt that additional objections will be brought against our case for the proposition that human persons are immaterial. Given this reality, it seems prudent for us to point out that the two-step argument given above employs only some exceedingly simple formal objects. Put in terms of mathematics education followed in all the technologized societies on Earth, everything we've done above uses no more than pre-college mathematics taught in the classrooms of these societies. We point this out because our case will grow in power as the robustness of the logico-mathematical objects to which we appeal grows. (Part of the reason in turn why this is so is that any degree of plausibility associated with the view that logico-mathematical objects can be in some way tied directly and exclusively to physical objects quickly erodes to zero, or at least — and the irony of picking the purely formal concept employed by both Leibniz and Newton for their invention of the calculus is intended — to an infinitesimal.) What, specifically, have we in mind? The most fertile area to mine in order to articulate even more powerful versions of the argument given above is likely to be the world of the very, very, very large. This means that turning to set theory should prove productive — and this move can be taken by building seamlessly upon the elementary elements introduced above. Specifically, subsequent refinements and extensions of our case can start with a more serious look at the progression of sets

$$
\begin{aligned}
&\mathfrak{0} := \{0\} \\
&\mathfrak{1} := \{0, 1\} \\
&\mathfrak{2} := \{0, 1, 2\} \\
&\mathfrak{3} := \{0, 1, 2, 3\} \\
&\vdots \\
&\mathfrak{n} := \{0, 1, 2, 3, \ldots, n\} \\
&\vdots
\end{aligned}
$$

and then two of the sets called out above that enter into childhood math education: $\mathbb{N}$ and $\mathbb{Q}^+$. Specifically, the more serious look is undertaken in order to deeply understand the *size* of some $\mathfrak{n}$ in this progression, versus the size of $\mathbb{N}$. A first step in achieving this deeper understanding is to prove and thereby understand that $\mathbb{N}$ is of an infinite size, whereas each $\mathfrak{n}$ is merely finite. A second step is to understand that the size of $\mathbb{N}$ corresponds to the first infinite size-indicating number: the cardinal $\aleph_0$. And a third step is to understand that even though the positive rationals $\mathbb{Q}^+$ has all the natural numbers as a proper subset, $\mathbb{Q}^+$ is nonetheless also of size $\aleph_0$. These first few steps in the line of inquiry we sketch here will require inference schemata rather more nuanced than *modus tollens*, and likewise algorithms a bit trickier than Quicksort![29] These are assuredly immaterial things, and, by application of the reasoning pattern given above, we as beings who understand these immaterial objects must ourselves be immaterial.

Of course, we can move on to sets that are larger than $\mathbb{N}$ and $\mathbb{Q}$, to the size of $\mathbb{R}$, which corresponds to the next infinite cardinal, $\aleph_1$; and from here we can continue. As we do this, it will, we believe, begin to seem simply preposterous that the logico-mathematical objects in mental play are not immaterial, and we predict it will be harder and harder to see how deep human understanding of these objects can be obtained by processing that is no more than standard mechanizable manipulation of embodied symbols (i.e., no more than Turing-machine-level computation). Notice, finally, that there is empirical prediction that clearly emerges from our suggested line for extension and refinement of the line of reasoning given above. Our prediction is that AI carried out on the basis of its textbook definition now in force for nearly three quarters of a century[30] will perpetually fail to match great human achievements in the formal sciences.

Notes

[1]First introduced in the early 20th century by C.I. Lewis:

> There are recognizable qualitative characters of the given, which may be repeated in different experiences, and are thus a sort of universals; I call these "qualia." But although such qualia are universals, in the sense of being recognized from one to another experience, they must be distinguished from the properties of objects. ... The quale is directly intuited, given, and is not the subject of any possible error because it is purely subjective. (Lewis 1929, p. 121)

[2]Pure mathematics, mathematical/theoretical physics, formal logic, decision theory, game theory, theoretical computer science, etc.

[3]Notice that we specifically speak of human *persons*. Human beings have physical bodies, of course; and physical bodies are (obviously) not non-physical. Hence there is a danger in speaking simply of 'humans' being immaterial. This danger is often dodged by speaking of "the mind," or "the human mind." Because this way of speaking, as pointed out by Chisholm (1978), seems to multiply entities beyond what seems reasonable. From this point on in the present essay, we shall for purposes of easing exposition feel free to say 'humans' instead of 'human persons.'

[4]Some readers might wisely ask whether circumspection dictates that we say instead "we must ourselves, *at least in part*, be non-physical." We are of the opinion that further analysis would inevitably reveal that this more conservative language is superfluous. The reason is that since we as humans are individuated persons enjoying a unity of consciousness/thinking, we admit of parts, once our argument herein is appreciated, only insofar as we make use of things like hands or feet or eyes or brain parts. Cf. (Chisholm 1978).

[5]Public K–12 mathematics education in New York State follows so-called "Common Core" standards. This standards, which revolve around a progression of increasingly tricky logico-mathematical objects, can be scrutinized here.

[6]In mathematics, students in Grade 1, in New York State public education, are taught algorithms for addition, subtraction, and multiplication, and for simple "algebraic reasoning" in which unknown in equations are determined.

[7]Sometimes the symbols used to denote the Boolean operators/connectives will be different. E.g., one sometimes sees $\supset$ instead of $\rightarrow$ for material implication, and sometimes $\equiv$ for $\leftrightarrow$, etc.

[8]The discovery itself was in 1959.

[9]In prior work that must be left aside here, we have made considerable use of the type-vs-token distinction for rendering talk of algorithms and computer programs precise; see e.g. (Bringsjord & Govindarajulu 2017, Arkoudas & Bringsjord 2007, Bringsjord 2015*a*). In the present paper, we will use both 'embodiment' and 'token' freely, with just the general understanding that these are physical things that vary between themselves, but which all stem from the general type they refer

to.

[10]Many would classify what we give here as an "algorithm-sketch."

[11]Consider that each embodiment need only have a tiny, tiny notational difference relative to its predecessor, thus forming a sequence matching the natural numbers. Of course, we could be wrong here; but nothing substantive hinges on this particular issue.

[12]The token $\hat{\mathcal{Q}}_2$ immediately follows. This is a function defined in the Clojure programming language of today. All readers, regardless of background, will understand upon a bit of inspection that this is a definition (hence the string `defn`) of a function called `quick-sort`, and will note that this function is recursive, since it calls itself. This is in keeping with the abstract algorithm $\mathcal{Q}$ discovered by Hoare.

```
(defn quick-sort [coll]
      (if (not-empty coll)
        (let [pivot (rand-nth coll)]
          (concat (quick-sort (filter #(< % pivot) coll))
                  [pivot]
                  (quick-sort (filter #(> % pivot) coll))))))
```

[13]Where it's sometimes given a different name, e.g. "conditional elimination"; see e.g. (Barwise & Etchemendy 1999).

[14]A more precise specification is this one:

$$\frac{\begin{array}{c}\phi_1 \vee \phi_2 \vee \ldots \vee \phi_k \\ \{\phi_1\} \vdash \gamma, \ldots, \{\phi_k\} \vdash \gamma\end{array}}{\gamma}\ \widehat{PBC}_2$$

[15]Euclid's indirect proof of his theorem (= **ET**) can be couched in terms of proof-by-cases and *modus tollens*, as follows:

> **Proof**: Suppose that $\Pi = p_1 = 2, p_2 = 3, p_3 = 5, \ldots, p_k$ is a finite, exhaustive consecutive sequence of prime numbers. Next, define $\mathbf{M}_\Pi$ as $p_1 \times p_2 \times \cdots \times p_k$, and set $\mathbf{M}'_\Pi = \mathbf{M}_\Pi + 1$. Then either $\mathbf{M}'_\Pi$ is prime, or not; we thus have two (exhaustive) cases to consider. Both cases lead by *modus tollens* to the negation of our supposition:
>
> C1 Suppose $\mathbf{M}'_\Pi$ is prime. In this case we immediately have a prime number beyond any in Π, and our supposition is negated.
>
> C2 Suppose on the other hand that $\mathbf{M}'_\Pi$ is *not* prime. Then some prime p divides $\mathbf{M}'_\Pi$. (Why?) Now, p itself is either in Π, or not; we hence have two sub-cases. Supposing that p is in Π entails that p divides $\mathbf{M}_\Pi$. But we are operating under the supposition that p divides $\mathbf{M}'_\Pi$ as well. This implies that p divides 1, which is absurd (a contradiction). Hence the prime p is outside Π, and once again the starting supposition is negated.

Hence for *any* such list Π, there is a prime outside the list. That is, there are infinitely many primes. **QED**

[16]In seeking to show that some of our thinking, when we validly follow inference schemata and algorithmic functions, is immaterial, Ross appeals in his own case to prior work upon which he intends to build. E.g., he writes:

> Now we need reasons why no physical process or function among physical processes can determine "the outcome" for every relevant case of a "pure" function. Those considerations mark some of the most successful in analytic philosophy, from W. V. Quine, to Nelson Goodman, to Saul Kripke. (Ross (1992), p. 140)

Ross's reader is supposed to be familiar with what he is specifically appealing to here, and it's beyond the scope of the present paper for me explain the role that work by the trio Ross cites plays in Ross's reasoning. In the case of Goodman, the motivated reader can consult (Goodman 1955).

[17]Of late, some have taken to using 'AGI' (for 'artificial general intelligence') to refer to Strong AI. I herein stick with the original, older terminology, because in my experience some today mean by their use of 'AGI' to refer to a category of artificial agents that have general-purpose cognitive powers cutting across many (perhaps all) human-relevant domains, but *not* necessarily to artificial agents that are subjectively aware/conscious. The phrase 'Strong AI' unmistakably refers to creatures that have full-blown — as it's called — phenomenal consciousness.

[18]In our personal experience, outside such venues and in the flow of real discussion in real life, at the lunch or dinner table and not in the "official" environments of academic papers and presentations (which are undeniably more than colored by careerist ambition), the belief in question is often admitted, sometimes unwittingly, to be, minimally, weak, and maximally, simply absent.

[19]In our experience as educators, a hallmark of a human's understanding an algorithm such as $\mathcal{Q}$ is that he/she can grasp that various particular embodiments $\hat{\mathcal{Q}}_1$, $\hat{\mathcal{Q}}_2$, and so on of $\mathcal{Q}$ all token the same type $\mathcal{Q}$.

[20]For another book-length defense of the view that this is what we are, see (Pinker 1997).

[21]We are indebted to reviewers of an earlier version of our paper for bringing Mengue's paper to our attention.

[22]Our Step-2 argument of course marks a debt to Searle's CRA and subsequent improvement achieved by Bringsjord, and the CRA line of reasoning is not involved in Mengue's paper/reasoning.

[23]Among these jewels, for technical reasons too far afield for treatment herein, the "jewel" that is most relevant to our two-step case is the "Grue Paradox" (= GP) seminally introduced by Goodman (1955); this is work that Ross specifically cites. In the view of the first author, the only escape from GP is to require of the gemologist featured in Goodman's famous parable that he specify use of any and all inference schemata used to support conclusions about scientific laws regarding emeralds — but once these schemata are specified, it's the following of them as algorithms that becomes central, and allows GP to be avoided. This is philosophically in line with purely mathematical treatments of

inductive logic, even when such treatment involves no argumentation, but — following Carnap, the founder of formal inductive reasoning — uses the machinery of probability calculation to adjudicate competing scientific hypotheses; see, for coverage, (Paris & Vencovská 2015).

[24]For exploration of exactly these issues from the perspective of computational formal logic and AI, see (Arkoudas & Bringsjord 2004).

[25]The fatal problem of arbitrariness infecting the case given by L&N for the embodied and legislative nature of mathematics is related in interesting ways to the argument given by C.S. Lewis (1947) against naturalism. A great place for interested readers to start is (Reppert 2003) — though we inform, and indeed caution, our readers that while Reppert (2003) affirms for the assessment of arguments a "Bayesian model with a subjectivist theory of prior probabilities" (p. 346), we wholly reject Bayesian frameworks, since they are insufficiently expressive [they have base formal languages that are purely extensional (e.g., zero-order and first-order logic)], and are committed to standard Kolmogorovian probability calculi interpreted subjectively, a position we take to be decisively overthrown by e.g. Pollock (2006). We consider any premeditated framing of competing arguments in natural theology within Bayesianism and/or its underlying formalisms to almost instantly be fatal to natural theology. Along this line, see (Bringsjord & Sundar Govindarajulu 2020).

This is as good a place as we can find to inform the reader that while we see (as just indicated) a connection between the view that mathematics is legislated and the argument originated by Lewis [and the argument as improved by Reppert (2003)], the connection between our two-step argument for the immateriality of human persons and what Reppert calls "The Argument from Reason" in his wide-ranging paper so titled (Reppert 2009), a family of arguments members of which include the ones articulated by Lewis (and the improved-upon-Lewis versions from Reppert himself), is one we find painfully obscure. We say this knowing full well that in his (Reppert 2009), Reppert classifies the argument of James Ross (upon which we of course heavily rely herein) in this very family (see pp. 365–366 Reppert 2009). Of note for motivated readers and scholars is the fact that Reppert's explicit distillation of Ross's argument (Reppert calls the distillation a "formalization") says that this argument's ultimate conclusion is that some human mental states are non-physical. Verbatim, Reppert says that Ross's ultimate conclusion is that "the mental states involved in mathematical operations are not and cannot be identical to physical states" (p. 366). As we have made clear, the key proposition yielded by Step 1 in our two-step argument is that certain *objects* are non-physical, and as the title of the present paper makes plain, the idea is that if these objects are non-physical, the objects that are us are too.

[26]Further investigation of the relevant literature on a causal account of knowledge is by way not helpful to Benacerraf. As Clarke-Doane (2017) points out, arguably the leading proponent of such causal accounts in the 20th century, and someone in fact credited with originating such accounts, Alvin Goldman, is disinclined to apply them in the case of knowledge of logico-mathematical propositions; see e.g. (Goldman 1967).

[27]Someone might somehow doubt this, though we can't fathom how. But if so, we can simply retreat to something that will get the job of overthrowing B-FP done just fine: We can focus on such self-beliefs as that one *seems* to be in acute pain. No one can possibly be mistaken in a belief that one seems to be in acute pain — and this despite the fact that no one can rigorously explain such perfect.

[28]While we don't need it for our case to be defended, as a matter of fact, it seems to us, for what it's worth, that such beliefs as that $\mathcal{Q}$ applied to some array $\langle \boxed{2}\boxed{1}\boxed{6} \rangle$ is guaranteed to produce $\langle \boxed{1}\boxed{2}\boxed{6} \rangle$ is correct — and yet it's very hard (impossible?) to explain why, without coming to rest on logico-mathematical things that are indubitable, directly and without proof.

[29]We shall need to employ, e.g., the inference schemata *mathematical induction*, and from its use another inference schemata often referred to as *the pigeonhole principle*. Nice coverage is provided in (Goldrei 1996).

[30]According to which AI is the field devoted to building artificial agents that compute functions from what they perceive to the actions they take, where these functions are Turing-computable ones that match in their nature what Searle-in-the-room has as a resource. For such textbooks, see e.g. (Russell & Norvig 2020, Luger 2008).

References

Anderson, J. (2009), *How Can the Human Mind Occur in the Physical Universe?*, Oxford University Press, Oxford, UK.

Arkoudas, K. & Bringsjord, S. (2004), Metareasoning for Multi-agent Epistemic Logics, *in* 'Proceedings of the Fifth International Conference on Computational Logic In Multi-Agent Systems (CLIMA 2004)', Lisbon, Portugal, pp. 50–65.
URL: *http://kryten.mm.rpi.edu/arkoudas.bringsjord.clima.crc.pdf*

Arkoudas, K. & Bringsjord, S. (2007), 'Computers, Justification, and Mathematical Knowledge', *Minds and Machines* **17**(2), 185–202.
URL: *http://kryten.mm.rpi.edu/ka_sb_proofs_offprint.pdf*

Barwise, J. & Etchemendy, J. (1999), *Language, Proof, and Logic*, Seven Bridges, New York, NY.

Bellman, A., Bragg, S. & Handlin, W. (2012), *Algebra 2: Common Core*, Pearson, Upper Saddle River, NJ. Series Authors: Charles, R., Kennedy, D. & Hall, B. Consulting Authors: Murphy, S. & Wiggins, G.

Benacerraf, P. (1973), 'Mathematical Truth', *Journal of Philosophy* **70**, 661–679.

Bringsjord, S. (1992), *What Robots Can and Can't Be*, Kluwer, Dordrecht, The Netherlands.

Bringsjord, S. (2015*a*), 'A Vindication of Program Verification', *History and Philosophy of Logic* **36**(3), 262–277. This url goes to a preprint.
URL: *http://kryten.mm.rpi.edu/SB_progver_selfref_driver_final2_060215.pdf*

Bringsjord, S. (2015*b*), 'The Symbol Grounding Problem—Remains Unsolved', *Journal of Experimental & Theoretical Artificial Intelligence* **27**(1), 63–72. DOI: 10.1080/0952813X.2014.940139.

Bringsjord, S. & Govindarajulu, N. (2020), 'The Theory of Cognitive Consciousness, and Λ (Lambda)', *Journal of Artificial Intelligence and Consciousness* **7**(1), 155–181. The URL here goes to a preprint of the paper.
URL: *http://kryten.mm.rpi.edu/sb_nsg_lambda_jaic_april_6_2020_3_42_pm_NY.pdf*

Bringsjord, S. & Govindarajulu, N. S. (2017), Are Autonomous-and-Creative Machines Intrinsically Untrustworthy?, *in* H. Abbass, J. Scholz & D. Reid, eds, 'Foundations of Trusted Autonomy', Springer, Berlin, Germany, pp. 317–335.
URL: *https://doi.org/10.1007/978-3-319-64816-3*

Bringsjord, S. & Govindarajulu, N. S. (2018), Artificial Intelligence, *in* E. Zalta, ed., 'The Stanford Encyclopedia of Philosophy'.
URL: *https://plato.stanford.edu/entries/artificial-intelligence*

Bringsjord, S. & Sundar Govindarajulu, N. (2020), 'Rectifying the Mischaracterization of Logic by Mental Model Theorists', *Cognitive Science* **44**(12), e12898.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/cogs.12898*

Chisholm, R. (1978), 'Is There a Mind-Body Problem?', *Philosophic Exchange* **2**, 25–32.

Clarke-Doane, J. (2017), What is the Benacerraf Problem?, *in* F. Pataut, ed., 'New Perspectives on the Philosophy of Paul Benacerraf: Truth, Objects, Infinity', Springer Verlag, Cham, Switzerland, pp. 17–43.

Field, H. (1989), *Realism, Mathematics, and Modality*, Basel Blackwell, Oxford, UK.

Goldman, A. (1967), 'A Causal Theory of Knowing', *The Journal of Philosophy* **64**(12), 357–372.

Goldrei, D. (1996), *Classic Set Theory*, Chapman & Hall, Boca Raton, FL.

Goodman, N. (1955), *Fact, Fiction, and Forecast*, Harvard University Press, Cambridge, MA. The 4th edition of 1983 is also from HUP.

Hoare, C. A. R. (1961), 'Algorithm 64: Quicksort', *Communications of the ACM* **4**(7).

Lakoff, G. & Nuñez, R. (2000), *Where Mathematics Comes From: How The Embodied Mind Brings Mathematics Into Being*, Basic Books, New York, NY.

Lewis, C. I. (1929), *Mind and the World-Order: Outline of a Theory of Knowledge*, Charles Scribner's Sons, New York, NY.

Lewis, C. S. (1947), *Miracles*, Macmillan, New York, NY.

Luger, G. (2008), *Artificial Intelligence: Structures and Strategies for Complex Problem Solving (6th Edition)*, Pearson, London, UK.

Menuge, A. (2016), 'Knowledge of Abstracta: A Challenge to Materialism', *Philosophia Christi* **18**(1), 7–27.

Paris, J. & Vencovská, A. (2015), *Pure Inductive Logic*, Cambridge University Press, Cambridge, UK.

Pinker, S. (1997), *How the Mind Works*, Norton, New York, NY.

Pollock, J. (2006), *Thinking about Acting: Logical Foundations for Rational Decision Making*, Oxford University Press, Oxford, UK.

Reppert, V. (2003), *C.S. Lewis's Dangerous Idea*, InterVarsity Press, Downers Grove, IL.

Reppert, V. (2009), The Argument From Reason, *in* W. L. Craig & J. P. Moreland, eds, 'The Blackwell Companion to Natural Theology', Wiley-Blackwell, West Sussex, UK, pp. 344–390.

Ross, J. (1992), 'Immaterial Aspects of Thought', *The Journal of Philosophy* **89**(3), 136–150.

Russell, S. & Norvig, P. (2020), *Artificial Intelligence: A Modern Approach*, Pearson, New York, NY. Fourth edition.

Searle, J. (1980), 'Minds, Brains and Programs', *Behavioral and Brain Sciences* **3**, 417–424.

Wason, P. (1966), Reasoning, *in* 'New Horizons in Psychology I', Penguin, Harmondsworth, UK, pp. 135–151.