STUDIA SEMIOTYCZNE (SEMIOTIC STUDIES), t. XXXIV, nr 1 (2020), s. 109–158
ISSN 0137-6608, e-ISSN 2544-073X
DOI: 10.26333/sts.xxxiv1.06
Article

JEFF BUECHNER *

USING KREISEL'S WAY OUT TO REFUTE LUCAS-PENROSE-PUTNAM ANTI-FUNCTIONALIST ARGUMENTS

SUMMARY: Georg Kreisel (1972) suggested various ways out of the Gödel incompleteness theorems. His remarks on ways out were somewhat parenthetical, and suggestive. He did not develop them in subsequent papers. One aim of this paper is not to develop those remarks, but to show how the basic idea that they express can be used to reason about the Lucas-Penrose-Putnam arguments that human minds are not (entirely) finitary computational machines. Another aim is to show how one of Putnam's two anti-functionalist arguments (that use the Gödel incompleteness theorems) avoids the logical error in the Lucas-Penrose arguments, extends those arguments, but succumbs to an absurdity. A third aim is to provide a categorization of the Lucas-Penrose-Putnam anti-functionalist arguments.

KEYWORDS: functionalism, Computational Liar, Gödel incompleteness theorems, finitary computational machine, mathematical certainty, finitary reasoning, epistemic refutation, metaphysical refutation, epistemic justification, recursively unsolvable, epistemic modality, finitary computational description.

1. Introduction

J. R. Lucas (1961) argued that for any finitary computational machine hypothesized to simulate full human mentality, there will be a Gödel sentence for that machine it cannot prove to be true, but which human beings can prove to be

* Rutgers University. The Saul Kripke Center, CUNY, The Graduate Center. E-mail: buechner@newark.rutgers.edu. ORCID: 0000-0002-6679-8209.

true. David Lewis (1969) responded that Lucas (and any other human being) can prove the Gödel sentence for that machine to be true if and only if they can also prove the theorems in Lucas arithmetic. But Lewis doubts a finitary human can do that, since Lucas arithmetic uses infinitary rules of inference—and so there might be infinitely many premises in a given proof. Lucas (1970), in turn, responded that Lewis failed to appreciate the dialectical character of Lucas' argument. Lewis (1979), in response, argued that even appreciating the dialectical character of the Lucas argument, Lucas cannot prove true the Gödel sentence of any finitary machine hypothesized to simulate full human mentality.

Roger Penrose (1989; 1994) improved upon Lucas' argument by proposing a neurobiological mechanism by which human beings might "see" the truth of the Gödel sentence of any finitary computational machine hypothesized to simulate full human mentality. Hilary Putnam argued (1995), famously, that Penrose commits a simple logical error. The finitary computational machine might have a program so long that no human being could physically survey it—and thus not be able to prove that it is consistent. If so, then even if full human mentality is not completely described by that finitary program, our failure to prove its consistency would not distinguish us from the finitary computational machine which (by the Gödel incompleteness theorems) fails to prove its own consistency. If so, the Gödel incompleteness theorems could not be used to arrive at a conclusion that functionalism as a theory of the human mind is a false theory, since it could not be demonstrated that there is an objective truth human minds can verify that no finitary computational machine can verify. The Penrose error is that even if human minds can "see" the truth of the Gödel sentence for the finitary computational machine that is hypothesized to describe human mentality, physically human beings are finite (in terms of time and space limitations). If the program of the finitary computational machine is so long that no human could survey it (such as read it) in their lifetime, then no human being could "see" that it is consistent (if it is). It is a logical error in Penrose's argument, since it is a possibility that, if true, undermines the argument by showing that the conclusion of the argument is false. The burden of proof is on Penrose's shoulders—to show that the possibility cannot be true. But this Penrose cannot do, since the ultimate finitary computational description of human mentality is yet to be written (if, in fact, there is one).

Putnam went on to construct an anti-functionalist argument using the Gödel incompleteness theorems (1988; 1994a; 1994b), applying it to both demonstrative and non-demonstrative reasoning. He does not apply the Gödel incompleteness theorems to a finitary computational program hypothesized to simulate full human mentality. Instead, he exploits the Kaplan-Montague paradox—the basic idea of which is the Computational Liar. The Computational Liar shows—if Putnam is right—that any attempt to formalize human reasoning must fail because any formal description of human reasoning can always be transcended by human reasoning. (Although Putnam does not make it, a distinction needs to be made between (i) prima facie, any formal system can be transcended by another formal system and (ii) any formal description of human reasoning can be transcended by human reasoning. It would be a mistake to reduce (ii) to (i)—that is not what Putnam claims.)

But his argument leads to a dilemma. If not all methods of inquiry are shown to be subject to the Gödel incompleteness theorems, one can take Kreisel's way out. But if all methods of inquiry are subject to the Gödel incompleteness theorems, there is an absurdity. I will provide (in section 7 of this paper) a categorization of the Lucas-Penrose-Putnam anti-functionalist arguments employing the Gödel incompleteness theorems.

What Putnam did not notice is that there is another way to show that human minds and any finitary computational machine hypothesized to simulate human minds are epistemically indistinguishable (even if they are de facto metaphysically distinguishable). What the Gödel incompleteness theorems show is that it is impossible to either prove the Gödel sentence of a formal system subject to the Gödel incompleteness theorems or to prove the consistency of that formal system using finitistic reasoning within that formal system (which delivers its theorems in the epistemic modality of mathematical certainty). Not even an infinitary mind can do that—an infinitary mind would use infinitary reasoning.

However, it is left open that either the Gödel sentence of a formal system subject to the Gödel incompleteness theorems or the consistency of that formal system can be proved with less than mathematical certainty or in some other epistemic modality. Both a human mind and a finitary computational machine might be able to do that. If so, both can prove the same thing, and no difference can be made between the two. This is the lesson from Kreisel's way out of the Gödel incompleteness theorems—and if taken, adds an interesting wrinkle to the Lucas-Penrose-Putnam anti-functionalist arguments. (Roger Penrose, in a preface to a reprinting of *The Emperor's New Mind* [Penrose, 1999], notes that one loophole to his argument is that "our capacity for [mathematical] understanding might be […] inaccurate, but only approximately correct". He says he will address this loophole to his argument in *Shadows of the Mind* [1994], but he does not.)

2. Kreisel's Way Out of the Gödel Incompleteness Theorems

Kreisel (1972) raises the question of whether there is non-mathematical evidence that can be used to establish the soundness of a formal system F (adequate for mathematical reasoning, and so subject to the Gödel incompleteness theorems). He observes that it does not logically follow from the fact that a formal system is subject to the second Gödel incompleteness theorems that there are absolutely no means available to prove its consistency. It only follows logically that its consistency cannot be mathematically demonstrated with mathematical certainty using finitistic reasoning. It is left open that its consistency can be proved by other means, viz., mathematically with less than mathematical certainty (typically by statistical reasoning) and non-mathematically, with less than mathematical certainty, by abstract philosophical reasoning (*a priori* reasoning that is not encodable into a formal system).

He believes that there are two different ways to realize the possibility of nonmathematical evidence to prove the soundness of F, both of which are left open by the Gödel incompleteness theorems. The first kind of nonmathematical evidence to prove the soundness of *F* is inductive evidence and the second is a metaphysical nonmathematical interpretation. Both kinds of evidence require substantial explanation—unfortunately, Kreisel's explanations are brief.

Nonmathematical inductive evidence is taken by Kreisel to be based on our experience with formal systems, such as our experience with Principia Mathematica. In one way of understanding what our experience of formal systems delivers, our confidence in the soundness of formal systems is acquired by various case studies of formal systems. Kreisel rejects this view—calling it a sham—for two distinct reasons. The first reason is that we have little or no experience of proving the soundness of a formal system by inductive methods. From this Kreisel thinks it follows that we have no good ideas about what are the appropriate statistical principles that would be used in evaluating the inductive evidence. Without statistical principles we have a data set, but no means by which to find in it the data which is necessary for establishing the soundness of some formal system. Whatever statistical principles we choose, one job which they must be able to do is to ascertain that the nonmathematical inductive evidence establishes that the entire formal system is sound, and not that only some subsystem of the formal system is sound.

The second reason Kreisel rejects the idea of nonmathematical inductive evidence for establishing the soundness of a formal system is that it is not done by using the experience we acquire from case studies of soundness proofs of formal systems. It is, instead, done by—at least in the case of Principia Mathematica—reflection on the intended meaning of the terms in the language of Principia Mathematica. However, what is interesting about Kreisel's point is that the act of reflecting upon what is the intended meaning of the terms in the language of a formal system may or may not be a computable procedure. There might not be a computational description of such acts. If there is no computational description of such acts, then there is some cognitive activity that humans can do which no machine can do. In which case, there would be a difference between humans and machines even if neither humans nor machines can prove the Gödel sentence of some formal system. Of course it would be a research project to show that acts of reflection upon the intended meanings of terms in some language (whether it is a formal language or not) have no computational description. (We shall see below that, using an ingenious Gödelian argument, Putnam attempts to close the door on both statistical methods and abstract philosophical methods for demonstrating CON(**PA**) by arguing that they are subject to the Gödel incompleteness theorems.)

The other way of proving the soundness of *F* is by an abstract but nonmathematical interpretation of *F*. Kreisel cites as an analogy the identification in intuitionistic mathematics of what is mathematical with what is intuitionistically acceptable. He notes that in intuitionism set-theoretic concepts are metaphysical and then claims that it might be possible to establish the soundness of some set-theoretic formal system using a metaphysical nonmathematical interpretation. Kreisel believes that this way of proving the soundness of *F* is more realistic than using inductive evidence to establish the soundness of *F*. I don't know what he means by "realistic" in this context. Perhaps he means that there is a wealth of mathematical and foundational work in intuitionism, and so we have a better understanding of what an abstract nonmathematical interpretation of *F* would look like than we do of statistical principles.

An interpretation is usually understood to be a map from syntactical objects (that is, symbols) to objects which need not be syntactical—perhaps mathematical objects. What, then, is a nonmathematical interpretation? Could it still be a map and yet be nonmathematical? And what does it mean to say it is metaphysical? Kreisel restricts the metaphysical nonmathematical interpretation to an abstract metaphysical nonmathematical interpretation. But if it is a map and it is abstract, it is not clear how it could not be mathematical.

Regardless of what Kreisel actually means by a metaphysical nonmathematical interpretation of *F*, using it to establish the soundness of *F* is different from proving the soundness of *F* within a classical formal system using finitary reasoning in the following respect: the proof of soundness of *F* within a classical formal system using finitary reasoning will be with mathematical certainty. (See below for a discussion of Church's view that the theorems of a given system of logic are proved with mathematical certainty.) On the other hand, the proof of the soundness of *F* using a metaphysical nonmathematical interpretation will perhaps not be with mathematical certainty. Kreisel's way out is the use of statistical proofs of consistency of **PA** with less than mathematical certainty or proofs in another epistemic modality such as (nonmathematical philosophical proofs). For more on the epistemic modality of a proof see 4.1 below.

3. Penrose on the Role of Trust in Mathematics

The key idea of Kreisel's way out is that one might be able to prove CON(**PA**) with less than mathematical certainty (using statistical methods) or in some other epistemic modality (such as metaphysical nonmathematical reasoning). Throughout the rest of this paper we will see how these possibilities enter into the Lucas-Penrose-Putnam anti-functionalism arguments. Recently Penrose has argued that trust plays an important role in mathematical proofs (2016). He claims that in order to trust a mathematical argument, we must trust that the rules of the formal system are sound. In cases where it cannot be established that the formal system is consistent because of the restriction imposed by the second Gödel incompleteness theorem, we need to trust that the formal system is consistent. If we do, then we can prove true the Gödel sentence and the consistency of that formal system by ascending to a stronger formal system—which we trust to be consistent.

We can view trust in the soundness of the rules of a formal system as an epistemic modality alternative to mathematical certainty delivered by proofs in a formal system. What Penrose fails to see is that if a finitary computational machine can meaningfully trust a formal system to be consistent, then there is no metaphysical difference between it and human minds. The move Penrose makes to show that human minds can determine the consistency of CON(**PA**) is one which defats his anti-functionalist argument, since it is open that finitary computational machines can do the same. The burden of proof is upon Penrose—to show that no finitary computational machine can exhibit the attitude of trust. (See Buechner, 2011, for an argument that finitary computational machines can engage in relations of trust with other finitary computational machines and with human beings.)

4. Two Uses of the Gödel Incompleteness Theorems in Refuting Functionalism

I introduce a distinction between two different uses of the Gödel incompleteness theorems in anti-functionalist arguments. This distinction has not been made in the literature—and it is important to make it because the conclusions of the arguments made under each use are significantly different. Perhaps the reader is puzzled: "Isn't there only one use of the Gödel theorems in refuting functionalism?" There are two different ways in which one can attempt to refute functionalism using the Gödel incompleteness theorems, and the conclusions about functionalism differ in each. Additionally, each method of refutation opens up different possibilities in the Lucas-Penrose-Putnam anti-functionalism arguments.

4.1. Metaphysical Uses of the Gödel Incompleteness Theorems in Refuting Functionalism

One way of using the Gödel incompleteness theorems in anti-functionalist arguments concludes that the human mind does not have the nature of a finitary computational machine, in which case, functionalism is false. This refutation establishes a metaphysical difference between human minds and finitary computational machines: human minds do not have the nature of such machines.

The metaphysical use of the Gödel incompleteness theorems in refuting functionalism is found in (Gödel, 1995; Lucas, 1961; Penrose, 1989): if it can be shown there is a mathematical truth that can be proved by a human mind, but that cannot be proved by a finitary computational machine (that, by hypothesis, finitely computationally models that human mind) then the human mind is not computationally modeled by that finitary computational machine. Whatever is the nature of the human mind, it does not have the nature of a finitary computational machine, since the human mind is different from the finitary computational machine in virtue of its causal powers, which enable it to prove a theorem that the latter cannot prove.

Another way of putting the same point: the human mind can prove that the program of the finitary computational device which purports to model it is correct, while the program cannot prove of itself that it is correct (assuming that there is no additional program embodied in the finitary computational device). So there is a cognitive power that the human mind possesses that is not possessed by the finitary computational machine. A human mind could justify the truth of the claim that the program that purports to describe it is correct, while the program itself cannot do that. But if the program, by hypothesis, describes all of the cognitive powers of the human mind, then it cannot be a complete finitary computational description of the human mind, since it lacks (at least) one cognitive power a human mind possesses.

This application of the Gödel incompleteness theorems shows functionalism is a false philosophical view by demonstrating that human minds are not identical with finitary computational machines. This non-identity claim is a metaphysical claim about the nature of the human mind: they do not have the nature of finitary computing machines. Functionalism is the view that human minds are identical with finitary computational machines (of some kind). The metaphysical argument (using the Gödel incompleteness theorems) demonstrates that human minds are not identical with finitary computing machines. Hence functionalism is false if the metaphysical argument is sound.

The Gödel incompleteness theorems (in the context of this metaphysical argument) provide a mathematical proof that the human mind is not identical to a finite computing machine and thus does not have the nature of a finite computing machine. (This claim can be generalized: the Gödel incompleteness theorems provide a mathematical proof that the human mind is not identical to any kind of finite computing machine and thus does not have the nature of any kind of finite computing machine. It can be generalized because the Gödel sentence unprovable in finitary computing machine1 can be proved in a stronger finitary computing machine2. However, a new Gödel sentence can be expressed in finitary computing machine2 that cannot be proved in it. This is true for all finitary computing machines.) So we have a mathematical proof of a negative metaphysical claim about the human mind: it is not any kind of finitary computing machine. We will call this use of the Gödel theorems "MGF" ("Metaphysical claims that are consequences of using the Gödel theorems to refute functionalism".)

It would be a mistake to claim that the Gödel incompleteness theorems specify an exact bound on the extent of the metaphysical difference between human minds and a given finitary computing machine. For instance, given a finitary computing machine that cannot prove its program is consistent, the extent to which the human mind differs from it is that the human mind can prove the program is consistent. This is not informative, since it says nothing positive about the cognitive functions necessary for human minds to prove that the program describing their mentality is consistent. It does say something negative, though. It says that no human mind can prove the program is consistent by simulating a finitary computing machine.

What is not usually addressed in metaphysical refutations of functionalism that use the Gödel incompleteness theorems is the epistemic modality of the provability relation in the formal system in which the reasoning occurs. A (sound) proof in a formal system (whether or not it is subject to the Gödel incompleteness theorems) proves a theorem with mathematical certainty. Our justification for believing the theorem is true is that it has been proved with mathematical certainty. So the Gödel theorems need to be qualified: the second incompleteness theorem says that no formal system subject to the Gödel incompleteness theorems can prove its own consistency with mathematical certainty. Here the epistemic modality—the way in which we come to know the truth of the claim made in the proof—is mathematical certainty. But there are other ways than mathematical certainty by which we can come to know the truth of a claim made in a proof. As the epistemic modality of a proof changes, so does the nature of the proof.

It is left open by the Gödel theorems that the formal system can prove its consistency with less than mathematical certainty or in some other epistemic modality. A statistical proof that a formal system (that is subject to the Gödel incompleteness theorems) is consistent has less than mathematical certainty. (Probabilistic proofs have this feature; see Wigderson, 2019.) A nonmathematical philosophical proof that such a formal system can prove its consistency would be a proof in another epistemic modality than that of a proof in logic or in mathematics. A proof using diagrams or pictures would be a proof in an epistemic modality other than mathematical certainty because the nature of a picture proof differs from the nature of a proof in a system of logistic. Intuitionistic reasoning in Brouwer's version of intuitionism might also be an example. Only a proof using a symbol system found in the formal languages of logic or in classical mathematics would have mathematical certainty. (Understanding in what epistemic modalities other than mathematical certainty there can be proofs of mathematical truths is an important and open research topic.)

If the only means of achieving mathematical certainty that *S* is true is to prove S in a formal system by finitistic reasoning within that formal system, then if S is either a Gödel sentence for that formal system or a consistency claim about that formal system, it follows that no human being (whether finitary in its cognitive powers or infinitary in its cognitive powers) can prove S is true with mathematical certainty using finitary reasoning within that formal system. So no human mind can prove the master program for a finitary computing machine simulating human mentality is correct with mathematical certainty by engaging in finitistic reasoning described by that master program. If so, human minds are indistinguishable from the finitary computing machine. On the other hand, there is no prohibition on the human mind proving the correctness of the master program with either less than mathematical certainty or in some other epistemic modality. But neither is the finitary computing machine prohibited from this, either. (This is so, unless proof with less than mathematical certainty or in another epistemic modality is subject to the Gödel incompleteness theorems. In that case, it is ruled out for the finitary computing machine to do that. But then it is also ruled out for human beings to do so as well.) If human minds can perform infinitary reasoning, and can prove the correctness of the master program using infinitary reasoning, this would distinguish human minds from finitary computing machines (which, by definition, cannot perform infinitary reasoning). But since it is an open question whether human minds can perform infinitary reasoning, this line of argument cannot establish its conclusion.

If the MGF argument is sound, then we know, with mathematical certainty, that we are not finitary computing machines. What is the provenance of the qualifier "mathematical certainty"? The Gödel theorems show that any formal system subject to the them cannot prove its Gödel sentence nor its consistency sentence with mathematical certainty using finitistic formalizable reasoning within that formal system. Why mathematical certainty? Why not logical certainty? Because there are different systems of logic—such as relevance logic—what is provable with logical certainty in one kind of logic might not be provable in some other kind of logic. Since the finitary reasoning in classical first-order logic can be described mathematically, the theorems of that logic are said to be proved with mathematical certainty.

Where does the claim that proofs in a formal system of logic carry mathematical certainty come from? Alonzo Church (1956) uses the phrase "mathematical certainty" in his discussion of proofs in mathematics that are translated into first-order logic. For Church, the only way to achieve mathematical certainty is a proof system where the axioms are effectively specified and in where, for any line in the proof, there is an effective procedure by which one can tell that it is an authentic line in the proof. This finitary reasoning in first-order logic can be described mathematically. An auditor of a proof

> [M]ay fairly demand a proof, in any given case, that the sequence of formulae put forward is a proof; and until this supplementary proof is provided, he may refuse to be convinced that the alleged theorem is proved. This supplementary proof ought to be regarded […] as part of the whole proof of the theorem, and the primitive basis of the logistic system ought to be so modified as to provide this, or its equivalent. (Church, 1956, p. 53)

The only logistic systems for which Church's requirement is satisfied are those in which the axioms and the rules of inference are effectively specified—these are finitary proof systems in which there are only finitely many lines in a proof and the pedigree of each line in the proof can be effectively ascertained. Infinitary logistic systems are different, for rules of inference are not effectively specified. A mind that has infinitary capacities can effectively specify them, but the notion of "effectiveness" then belongs to alpha-recursion theory, a theory of effectivity for infinite minds. Church obviously assumed human minds are finitary in his discussion.

So if the MGF argument is sound, then we know, with mathematical certainty, that human minds are not identical with any kind of finitary computing machine. This is an extraordinarily strong claim. Compare it with the following claim: we know, with mathematical certainty, that B follows from A and $A \rightarrow B$. This claim is trivial. On the other hand, one does not know with mathematical certainty that one is (now) looking at a tree. The claim an MGF argument makes is strong, then, in the sense that the information it establishes about the nature of the human mind has important value. (I do not suggest, in using the phrase "extraordinarily strong", that the claim is thereby unlikely to be true.)

But the strength of the claim should make us suspicious of it. The assumption that underlies the metaphysical claim is that human minds can prove the correctness of the finitary computing machine's master program (for simulating human mentality). But we have seen that this assumption needs to be qualified: human minds can prove, with mathematical certainty using finitistic reasoning, the correctness of the computing machine's master program. This, though, is highly unlikely to be true. If a human mind has infinitary cognitive capacities, it might do so (for instance, by employing Turing's infinitary procedure; see Turing, 1939). But do we have infinitary cognitive capacities? Some philosophers and cognitive scientists believe we do not have infinitary cognitive capacities. Others believe that we do. So a stalemate is reached in the absence of evidence concluding one or the other position.

If the assumption underlying the MGF argument is changed by changing the qualification to "with less than mathematical certainty or in some other epistemic modality", then the MGF argument cannot establish its conclusion, since it is also available for a finitary computing machine to prove the correctness of its own master program with less than mathematical certainty or in some other epistemic modality. Thus the metaphysical claim is bankrupt and the refutation of functionalism using the Gödel incompleteness theorems is drained of its force. This is a significant philosophical result overlooked in the anti-functionalism debate. If it is true that human minds are not completely describable by a finitary computational machine and that human minds are able to verify the consistency of Peano arithmetic, i.e., CON(**PA**), how is it done? It cannot be done by employing a recursively axiomatized finite proof system to do it, since for any such proof system (strong enough to capture arithmetic), the Gödel incompleteness theorems apply. On the other hand, if we use a recursively axiomatized finite proof system which is too weak to be subject to the Gödel incompleteness theorems, then this will not distinguish us from any finitary computational machines, since finitary computational machines are also capable of proving theorems in such weak proof systems.

In such a finitary proof system, there is nothing human minds can prove which a finite computational machine (of the appropriate kind) cannot prove. How, then do we differ from the finite machine? We know from Gentzen's proof of CON(**PA**) by transfinite induction, that infinitely long derivations can secure CON(**PA**). We also know that within formalized systems of Peano arithmetic, proofs of transfinite induction for any ordinal up to, but not including the infinite ordinal epsilon$_0$, are available. However, we need transfinite induction along a well-ordered path of length epsilon$_0$ to prove CON(**PA**). The issue, then, is this:

if human minds know the truth of CON(**PA**) with mathematical certainty, is the only mathematical method by which we do it the use of infinitely long derivations? There cannot be a finitary method of reasoning that proves CON(**PA**) with mathematical certainty within the formal system for **PA**. One can find stronger formal systems in which CON(**PA**) can be proved by finitistic reasoning, but only if CON(stronger formal system) can be verified. If it is verified, then we do it this way only if we have infinitary cognitive capacities, and that is at present an open question.

4.2. Epistemic Uses of Gödel's Incompleteness Theorems in Refuting Functionalism

MGF arguments show the nature of the human mind differs from the nature of physical finitary computing machines. MGF arguments are philosophically satisfying, since they rule out one metaphysical possibility about the nature of the human mind—that our minds have the nature of finitary computational machines. Even though they do not have the resources to describe the true nature of the human mind, their importance lies in showing what the human mind is not. But MGF arguments are not the only use of the Gödel theorems in the functionalism debate. Even if we assume that human minds are finitary computing machines, we can still enlist the Gödel incompleteness theorems to make philosophically important claims about the human mind. Call these uses of the Gödel theorems "EGF" ("Epistemic claims that are consequences of using the Gödel incompleteness theorems to refute functionalism"). There are two different kinds of EGF arguments.

4.2.1. The first kind of EGF argument.

Assume that human minds are finitary computational in nature. (However, the argument is the same if human minds cannot be fully described by finitary computational machines.) Suppose human cognition is finitely computationally described by computer program *P*. If we assume human beings can prove truths of Peano arithmetic, *P* is subject to the Gödel incompleteness theorems (since *P* must be equipped with enough syntax to arithmetize metamathematics, which is necessary for the Gödel theorems to take root). CON(*P*) expresses the consistency (or correctness) of *P*. Since it is equivalent to *P*'s Gödel sentence, it follows that *P* can't prove it is consistent. Assuming we are correctly described by *P*, human beings cannot verify the consistency of *P*.

Since the project of cognitive science is to find *P*, then that project can never be epistemically justified (since it cannot be established that *P* is consistent). Any science of the human mind that views the human mind as a finitary computing machine will not be able to epistemically justify its claims, because we cannot verify that the correct program of the finitary computing machine is consistent. Human beings will not be able to prove, with mathematical certainty, *P* is consistent. Human beings cannot prove the consistency of *P* in the epistemic modality of mathematical certainty. To do so, our reasoning about *P* would have accord with that of a finitary computing machine, to which the notion of "proof with mathematical certainty" applies. This is a radical form of philosophical skepticism: we have a mathematical proof (of which we are mathematically certain) that we cannot know, with mathematical certainty, the correct computational theory of how our minds work.

EGF arguments do more than provide a new form of philosophical skepticism. They also address the competence/performance distinction essential for the viability of cognitive science. A critical distinction is made in cognitive science between how the human mind actually works and how it ought to work—between a performance level description and a competence level description of the human mind. Without such a distinction, the very idea of a psychological law is jeopardized. EGF arguments show three basic assumptions essential for cognitive science to be viable cannot consistently obtain: (i) that the human mind can be represented (at a level of computational description) by a computational device, (ii) that its cognitive capacities can be viewed as finitely computable functions and (iii) that there is a competence description of the human cognitive mind. The Gödel incompleteness theorems show the first two assumptions are incompatible with the third. If we take the first two to be part of Marr's (2010) implementation level and the third to be Marr's theory of the function (the what, i.e., the function, which is computed), Gödel's theorems reveal an incompatibility in Marr's foundational program for cognitive science. (For details, see Buechner, 2010.)

4.2.2. The second kind of EGF argument.

Assume that human minds are not finitary computational in nature (but that we do not know this fact). If so, any finitary computational machine conjectured to describe human mentality fails to do so—it either fails to describe all of human mentality or else it falsely describes parts of human mentality. Suppose it is conjectured human mentality is correctly described by computer program *P*, which is subject to the Gödel incompleteness theorems. Suppose, additionally, the length of *P* is infeasibly long for a human being to survey. In which case, no human being will be able to establish that *P* is consistent.

Since no human being will be able to verify that *P* is consistent (which is an epistemic claim), we cannot use the mathematical theory of computation or cognitive science to show that there is a metaphysical difference between human mentality and a finitary computational machine. Although this kind of EGF argument does not refute functionalism, it reveals a shortcoming in it—that we cannot use it to establish metaphysical claims about the human mind. Additionally, since cognitive science and functionalism might be false theories (if *P* is inconsistent), any psychological claims made within cognitive science and any philosophical claims made within functionalism might be false, and we could never fully justify those claims no matter how much evidence we had supporting them.

4.3. Correct and Incorrect Readings of the Gödel Theorems

In arguments that use the Gödel theorems to attempt to refute functionalism and in critical discussions of those arguments, an obvious point has been overlooked. What the Gödel incompleteness theorems show is that there is no mathematically certain finitistic mathematical proof of the Gödel sentence and the consistency sentence of any formal system susceptible to the Gödel theorems. We cannot fintistically prove, with mathematical certainty, the Gödel sentence and the consistency sentence of Gödelizable formal systems. What is overlooked is the epistemic modality of mathematical certainty that qualifies the proof relation. Perhaps it is overlooked since the method of proof within a system of logic is what delivers mathematical (or logical) certainty.

The standard reading is that we cannot prove CON(**PA**), period. By failing to qualify "prove", it appears the claim is that there is no proof of any kind of CON(**PA**). This is an incorrect reading of the Gödel incompleteness theorems. The correct reading is that we cannot prove CON(**PA**) with mathematical certainty by finitistic reasoning in a formal system for **PA**. (John von Neumann, in his tribute to Gödel, notes that "for no such system can its freedom from inner contradiction be demonstrated with the means of the system itself" [1969, p. x]. This is a correct reading of the Gödel incompleteness theorems.)

It does not follow, however, that we cannot prove CON(**PA**) with less than mathematical certainty or prove it in some other epistemic modality than mathematical certainty (as Kreisel rightly noted). (The claims of statistical proofs are with less than mathematical certainty. Epistemic modalities other than mathematical certainty might include pictorial proofs and nonmathematical philosophical reasoning.) The same remarks hold if we transpose the discussion of the Gödel incompleteness theorems to the context of what we know about CON(**PA**). If we substitute "know the truth of" for "prove", the same point applies. We cannot know the truth of CON(**PA**) with mathematical certainty. It is left open by the Gödel theorems that we can know the truth of CON(**PA**) with less than mathematical certainty and that we can know the truth of CON(**PA**) in some epistemic modality other than mathematical certainty.

If we accept a mathematical epistemology in which we can know mathematical propositions with less than mathematical certainty or in some other epistemic modality than mathematical certainty, new possibilities become available for the functionalism debate. For instance, if there are formal systems (in which the Gödel incompleteness theorems hold) in which CON(**PA**) is proved with less than mathematical certainty and the epistemic modality in which it is proved satisfies a reasonable notion of epistemic justification, then the limitations of the Gödel incompleteness theorems might be dramatically circumvented. Substitute "the correctness of its own computer program" for "CON(**PA**)" in the preceding sentence. If an anti-functionalist enlists the Gödel theorems to refute functionalism, she must show that the notion of justification under which a finite machine can prove the correctness of its own computer program with less than mathematical certainty is normatively bankrupt. Suppose that human beings are finitary computational machines. Define the goal of cognitive science to be discovery of the master computer program for the human mind. Assume the cognitive activities cognitive scientists engage in when they attempt to discover the master computer program are themselves described in that program. Suppose that in the future a cognitive scientist claims to have found the master computer program. Do we require that her belief that this is the correct master computer program must be mathematically certain in order to count as being epistemically justified? Whether that requirement does or does not appear to be too strong, it is clear that it is a question that must be addressed wherever the Gödel theorems are enlisted in the functionalism debate.

Even within mathematics there is evidence that this demand is negotiable. Mathematical proofs not formalized within a system of logic do not satisfy the stringent demands of mathematical certainty. Only proofs that are formalized in a formal system whose axioms, rules of inference and application of rules of inference are recursively specified can satisfy those stringent demands. Proofs in, for instance, algebraic topology do not meet them, though mathematicians do not feel that they need to translate those proofs into a formal system before they can be said to know (with adequate justification) the truths of algebraic topology.

The consequence is that no finitary being can prove CON(**PA**) finitistically with mathematical certainty. The reason this is so is obvious. If mathematical certainty is secured only in virtue of a finitistic proof within a system of logic, no finite being can prove CON(**PA**) with mathematical certainty unless they construct a finitistic proof of it within a system of logic. But the Gödel theorems forbid this. (A being with infinitary powers can construct a proof of CON(**PA**) with mathematical certainty only if constructions in a system of logic requiring infinitary operations confer mathematical certainty upon the theorems proved within that system. Church did not consider this matter in his discussion of mathematical certainty.)

When anti-functionalists, such as Penrose, claim that human beings can know CON(**PA**) they must qualify their claim. We cannot know CON(**PA**) with mathematical certainty. But if we can know it with less than mathematical certainty or in some epistemic modality than mathematical certainty, it is possible that a finitary computational machine can acquire that knowledge as well. If so, the Gödel incompleteness theorems cannot drive a wedge between what a human being can know and what a finitary computational machine can know.

5. Putnam's First Version of His Argument That Not All Methods of Inquiry Can Be Formalized

An early argument Putnam (1988) uses against the view that methods of inquiry can be formalized by a finitary computational machine is his Gödelian argument that that there can be no prescriptive competence description of human reasoning (including the reasoning in mathematical proofs). Suppose that there is a description *P* of human prescriptive mathematical competence. There will be many functions that are provably recursive according to *P*. List the index of each partial recursive function that *P* can prove to be total recursive. There will be infinitely many functions on this list—since a mathematician can (in principle) prove infinitely many functions are general recursive. This list of functions can be diagonalized, and the diagonal function will be total, since there are infinitely many functions on the list.

However, if it could be proved that *P* is a sound proof procedure, it could also be proved that the diagonal function is a total recursive function. Unfortunately, such a proof would also show that *P* is inconsistent. Why is that? Suppose that the proof is on the list—in which case, the diagonal function would be on the list. But by the definition of a diagonal function, if it is the j^{th} member on the list, then diagonal function (j) = diagonal function ($j + 1$). It follows that any formalization of human mathematical proof ability cannot both (i) be sound and (ii) can be proven to be sound using human mathematical proof abilities.

Putnam's conclusion needs to be emended: no formalization of human mathematical proof ability can both be sound and be such that it is part of human mathematical proof ability to finitarily prove that soundness, with mathematical certainty and from within *P*. We cannot prove with mathematical certainty and finitistic reasoning that *P* is correct. It follows that we cannot prove with mathematical certainty and finitistic reasoning that the competence theory for human mathematical proof ability is correct.

It is impossible for us—whether we are or are not subject to the Gödel incompleteness theorems—to finitarily prove with mathematical certainty from within *P* that the competence level description is true of us. If we were able to finitarily prove it is true of us, with mathematical certainty and from within P, we would have proven that the formal theory encapsulated by the competence description is consistent. But this is prohibited by Gödel's second incompleteness theorem. Notice we would have to ascend to a stronger computational system to finitarily prove, with mathematical certainty, the consistency of our competence description. If so, then the competence description that we finitarily prove to be correct, with mathematical certainty, in the stronger system is not our competence description. Since we ascended to a new computational system, the competence description of the weaker computational system is no longer true of us.

Suppose that human minds are not subject to the Gödel incompleteness theorems. The Gödel incompleteness theorems rule out the possibility that a finitary human mind can finitarily prove, with mathematical certainty, that a finitary computer program that simulates it is correct. What this means is that whether human minds are or are not subject to the Gödel incompleteness theorems, the human mind cannot finitarily prove with mathematical certainty that a program that simulates it is correct. Thus whether human minds are or are not subject to the Gödel incompleteness theorems, they cannot justify claims in cognitive science about its computational structure. EGF arguments do not need to show that there is something a human mind can do that any finitary computing machine cannot do in order to make philosophically interesting claims about the mind. In this case, the claim concerns the limits of cognitive science in providing a rigorous, scientific study of the human mind.

EGF argument (such as the one Putnam makes above) must (as we argued earlier) make a very strong assumption: that justifications of claims in cognitive science are mathematically certain. This follows from the use of the Gödel incompleteness theorems. We know, with mathematical certainty, that we cannot, with mathematical certainty, finitarily prove the correctness of the program, *P*, that describes our competence. If *P* is the master program for human cognition, we can't mathematically prove it is correct with mathematical certainty. Do any other scientific disciplines impose such stringent epistemic requirements upon the claims they make? I think it is too high a price to ask of cognitive science, and one that is incompatible with the epistemic demands other scientific disciplines impose upon their own claims. This is an important issue that deserves further attention.

Notice that statistical methods and proof methods in an epistemic modality other than that of mathematical certainty (we will call them 'weak methods') will be included in *P*. There's no absurdity or inconsistency in this inclusion, since they do not finitarily prove the correctness of *P* with mathematical certainty. Rather, they prove it with less than mathematical certainty or in some other epistemic modality. The central issue for EGF arguments is what we should take as the standard of epistemic justification of *P*. If we take the standard of epistemic justification to be mathematical certainty, then they refute computational functionalism. If the standard is less than mathematical certainty or some other epistemic modality, they lose all their potency in refuting functionalism.

This version of Putnam's anti-functionalist argument using the Gödel incompleteness theorems—that there can be no prescriptive competence description of human mathematical reasoning—succeeds only if the epistemic modality of the proof relation is that of mathematical certainty achieved by finitistic reasoning. Where that is not the case, the argument fails.

6. An Exposition of Putnam's Second Gödelian Argument Against Functionalism

Whether there is or is not a finitary computational description of total human mentality is an open question. However, if we cannot (now) know the ultimate finitary computational description of total human mentality—should there be one—then we cannot (now) know whether its program is (or is not) infeasibly long. This presents an irresolvable difficulty for any MGF or EGF arguments—such as the Lucas-Penrose arguments. To assume the program is feasibly long—and one which can be shown consistent by human minds—is a logical error. Putnam diagnosed this error in Penrose's argument. As we saw earlier, since is it possible the program is infeasibly long, it is therefore possible that even if human minds do not have a complete finitary computational description, they cannot be distinguished from finitary computational machines because they will not be able to prove the consistency of an infeasibly long program. (Even if we do have infinitary minds, our physical bodies in some of their aspects are finitarily restricted—and so we would not be able to read all of the lines in a program which is infeasibly long.) To neglect this possibility is a logical error. Yet Putnam makes a Gödelian argument against functionalism without making either logical error—he does not assume the program is feasibly long and he does not have to consider the possibility that it is infeasibly long. How is it done?

One way out of this difficulty for EGF and MGF arguments is to show that all epistemically justified methods that prove CON(*P*) with less than mathematical certainty or in some other epistemic modality (the weak methods) are subject to the Gödel incompleteness theorems. Putnam claims that all weak methods are subject to the Gödel incompleteness theorems. This argument appears in *Reflexive Reflections* (Putnam, 1994b). The argument employs Gödel's second incompleteness theorem. In what follows, I use the acronym "PGA" ("Putnam's use of the second Gödel incompleteness theorem in his argument that all weak methods are subject to the Gödel incompleteness theorems").

PGA claims that our prescriptive inductive competence is subject to the Gödel incompleteness theorems. Putnam cites his earlier work on Carnapian inductive logics and on computational learning theory, only to assert that it does not matter whether this work is taken into account in PGA, since PGA will assume there is some finitary computational description of our prescriptive inductive competence and that one does not need to know what that description looks to make the PGA argument. "P" denotes a finitary computational description of our inductive (or non-demonstrative) and demonstrative prescriptive competence.

Putnam uses an idea in the Montague-Kaplan *Paradox of the Knower* (Feferman, 1960) that is an application of self-reference. It is The Computational Liar (CL):

(CL) There is no evidence on which acceptance of the sentence CL is justified (Putnam, 1994b)

CL is arithmetizable, and its arithmetization is a sentence of arithmetic to which the Gödel diagonal lemma applies. The diagonal lemma tells us that for any predicate that is definable in the language of Peano arithmetic, there is some sentence that is true if and only if its Gödel number is false of that predicate. The diagonal lemma allows us to couple *P* with CL.

> It follows from Gödel's work that there is a sentence of mathematics which is true if and only if *P* does not accept that very sentence on any evidence, where *P* is any procedure itself definable in mathematics—not necessarily a recursive procedure. (Putnam, 1994b)

In an important caveat to CL, Putnam says that "[…] if the inductive logic *P* uses the notion of degree of confirmation rather than the notion of acceptance, then one replaces 'is justified' by 'has instance confirmation greater than .5', […]" (1994b, p. 426, note 5). This is significant, since the notions of a justified belief and of acceptance of a justified belief play critical roles in non-quantitative models of inductive reasoning, while "has instance confirmation greater than .5" and "degree of confirmation" play critical roles in both quantitative and logical models of inductive reasoning. This caveat gives us reason to think that Putnam takes *P* to be a computational description of any kind of inductive reasoning and not just logical models of inductive reasoning, such as those found in computational learning theory.

If there is evidence which justifies the acceptance of CL, it easily follows that CL is false, and it is a sentence of pure mathematics. Since *P* formalizes our prescriptive competence in demonstrative and non-demonstrative reasoning, our (fully justified) reasoning tells us to accept a mathematically false proposition.

The negation of CL is that there is evidence on which the acceptance of CL is justified. If there is evidence on which the acceptance of the negation of CL is justified, then we know from what was just established above that CL is a mathematically false sentence. (Putnam notes that it is an omega inconsistency.) It follows that should *P* converge on CL—that is gives an answer to CL—to which we are justified (by *P*), then that evidence for the answer licenses us to accept a mathematical falsehood. So it has been established that CL cannot be shown true or shown false using *P*, which is a computational description of our prescriptive competence in demonstrative and non-demonstrative (inductive) reasoning. (Gödel assumed that the formal system in which he worked is omega-consistent in order to show that proof of the negation of the Gödel sentence leads to contradiction, in this case, an omega-inconsistency. Omega-consistency is weaker than consistency. If a formal system is omega-consistent, it follows that it is consistent. Putnam makes the same assumption.)

Given that anyone is justified in believing that if *P* converges on CL, it licenses one to believe a sentence that is mathematically false, Putnam formulates a criterion of adequacy (CA) for accepting any formalization of human prescriptive demonstrative and non-demonstrative competence

(CA) The acceptance of a formal procedure *P* as a formalization of (part or all) of prescriptive inductive (demonstrative and non-demonstrative) competence is only justified if one is justified in believing that *P* does not converge on *P*'s own Gödel sentence (i.e., CL) as argument.

From CL and CA, it follows that no human being can demonstrate that *P* is prescriptive whenever our minds work in the exact way that *P* says they should work. When we believe CA and also believe that *P* is both complete and also correct in describing our prescriptive demonstrative and non-demonstrative competence, it easily follows that we will believe that *P* does not converge on CL. However, that is to believe CL. But notice that this belief is justified, and that (by assumption) all justification of beliefs can be formalized in *P*. Since we are committed to believing CL, we are in a contradiction. That is Putnam's ingenious PGA.

Notice that Putnam has not made any claims that there is something human minds can do that no finitary computing machine can do, nor has he assumed that *P* is feasibly long. (That is why Putnam does not commit the logical error that Penrose commits). He has, though, shown that *P* could not be justified within cognitive science without licensing us to believe a contradiction. One consequence of PGA is that any formal theory proposed in cognitive science of how we do inductive reasoning cannot be justified without also licensing us to believe a contradiction. (This is a disturbing and important result that has not caught the attention of cognitive scientists working on the problem of formally characterizing inductive reasoning.)

6.1. PGA and the Kaplan-Montague Paradox

Is it really the case that the key terms in CL can be arithmetized? If they cannot be arithmetized, then PGA fails. I contrast Putnam's Computational Liar with the version that Kaplan and Montague (1960) constructed in order to show the Gödel incompleteness theorems extend to the modal predicates "knowledge" and "necessity". Kaplan and Montague needed to find for the knowledge predicate suitable analogues of the Hilbert-Bernays derivability conditions for the provability predicate. Montague employed a weak epistemic system consisting of the four schemata:

(i.) $K\alpha \rightarrow \alpha$

(ii.) $K\alpha$, if α is an axiom of first-order logic

(iii.) $K(\alpha \rightarrow \mu) \rightarrow (K\alpha \rightarrow K\mu)$

(iv.) $K(K\alpha \rightarrow \alpha)$

Montague (1963) appreciated Tarski's insight (1983), in the latter's proof of the indefinability of truth in first-order logic, that two prima facie consistent theories cannot always be combined into a consistent theory. In Tarski's indefinability work, Robinson arithmetic relativized to ß cannot be combined with Tarski's schema for the language of Robinson arithmetic relativized to ß and extended with a truth predicate T. Montague saw that this insight can be generalized: two prima facie true theories, one a theory of its own syntax and the other a theory that has principles capturing the logic of concepts such as knowledge, belief or necessity, cannot be combined into a consistent theory. The tool necessary for the proof is the Gödel diagonal lemma:

Suppose T is an extension of Robinson arithmetic relativized to ß. Let α be a formula whose only free variable is v_0. Then there is a sentence ζ such that:

$$\vdash_T \zeta \text{ if and only if } \alpha(\boldsymbol{\zeta}/v_o), \text{ where,}$$

$$\text{if } n \text{ is the Gödel number of } \zeta, \boldsymbol{\zeta} \text{ is the } n^{th} \text{ numeral.}$$

The key to the Montague-Kaplan proof is the fact that knowledge is a property of "proposition-like" objects recursively built from atomic constituents. Given enough arithmetic, it is easy to associate with each "proposition-like" object a Gödel number. Then, structural properties and relations between "proposition-like" objects can be arithmetically simulated by explicitly defined arithmetical predicates of the Gödel numbers of the "proposition-like" objects.

Recall Putnam's Computational Liar:

CL There is no evidence on which acceptance of the sentence CL is justified.

We need to arithmetize the properties and relations in CL in order to use Gödel's diagonal lemma. Can "evidence", "acceptance", and "justified" be arithmetized? It is not obvious that they can. Consider the ramified type theory in Russell and Whitehead's *Principia Mathematica*. No one has succeeded in showing it is subject to the Gödel incompleteness theorems, for there is no general theory of the intensional provability relation. It will do no good to simply assert that consistency cannot be proved within any sufficiently strong system because Gödel's second incompleteness theorem tells us this. Richmond Thomason (1980; 1989) has pointed out in this connection that "it has never been possible to state the [second incompleteness] theorem at this level of generality with a degree of precision that will support a mathematical proof" (1989, p. 54).

Intensional provability relations link arithmetical theories to a given set of propositions when the arithmetical theory is able to prove each of the propositions in the set. That there cannot be a general theory of the kind Thomason specifies follows from an interesting result on the peculiarities of the intensional proof relation. It is a result of Feferman (1960) that Gödel's arithmetical formalization of the proposition that Peano arithmetic is consistent can be proved, under substitution of different linguistic expressions for the same classes of numbers in that arithmetical formalization.

PGA requires that "evidence", "acceptability", and "justified" can be arithmetized. We can formalize the evidence relation and the property of acceptance within computable learning theory, but this raises the question of whether that formalization captures all of the uses of these terms in inductive reasoning and if the terms can be arithmetized. What of the property of being justified? How would we axiomatize its basic features in the way that Kaplan and Montague axiomatized the basic features of knowledge? What happens to PGA if the notion of being justified is omitted? Without it, we cannot say that *P* tells us that we are prescriptively justified in believing an arithmetically false sentence. Thus we will not be able to show that an absurdity results if *P* converges upon either CL or the negation of CL. In which case, we cannot even express the condition of adequacy that is necessary for obtaining the contradiction.

Objection: It is true that omitting the notion of "justifies" in PGA blocks deriving the contradiction. But that is not a problem for the anti-functionalist end to which PGA is applied. You succumb to a dilemma if you argue there is no obvious arithmetization of "is justified". The first horn is that if there is an arithmetization of "is justified", then the contradiction is secured. For the second horn, suppose it cannot be arithmetized. If so, then it cannot be part of cognitive science. Thus, either way, cognitive science is in jeopardy. On the first horn, cognitive science cannot prove that it is correct and on the second horn, inductive reasoning can't be computationally described. On either horn, the anti-functionalist wins.

Response: The first horn of the dilemma is that if "is justified" is arithmetizable, then PGA is secured. Below we argue that even if PGA is sound, it cannot be used to secure the claim that human minds are not finitary computing machines or the claim that cognitive science cannot be justified. The second horn is easily dismissed, though. That "X" is not arithmetizable does not logically imply "X" is not formalizable. Why think any property or relation whatsoever, even though formalizable, can be arithmetized? Certainly, Gödel numbers can be assigned to formalized sentences and to formalized properties. But it does not follow from that fact that any formalized property is arithmetizable. The example of Principia ramified type theory, discussed above, illustrates the point. The burden of proof is upon Putnam, to show that the epistemic property of being justified, under a suitable formalization, can be arithmetized. (Artemov-Fitting logics of justification are not a method of reasoning to achieve justification, but a method for reasoning about justifications. An open question is whether a Montague-Kaplan type paradox could be constructed using their justification predicate.)

6.2 Strengthened PGA Leads to an Absurdity

One problem with PGA is that if not all inductive methods or, more broadly, methods of inquiry into the world, are subject to the Gödel incompleteness theorems, then it is possible that in using methods that are subject to the Gödel incompleteness theorems, we can employ weak inductive methods that are not subject to the Gödel incompleteness theorems to prove CON(method subject to the Gödel incompleteness theorems) or the Gödel sentence (of a method subject to the Gödel incompleteness theorems) in another epistemic modality or with mathematical certainty less than the degree of mathematical certainty of the proof procedure of the formal system in which the methods are formalized. Both human minds (that have or do not have a finitary computational description) and finitary computing machines that are subject to the Gödel incompleteness theorems can use weak methods that are not subject to the Gödel incompleteness theorems. Any EGF or MGF argument that ignores this possibility commits a logical error no less serious than the logical error Penrose commits in his anti-functionalist argument. On the other hand, if the above possibility is taken seriously, then EGF and MGF arguments can fail. What can be done? One suggestion is to show all methods of inquiry into the world are subject to the Gödel incompleteness theorems.

Suppose we strengthen PGA in the following way: all methods of inquiry into the world are subject to the Gödel incompleteness theorems. (Putnam appears to say this is how he wants his argument to be interpreted; see Putnam, 1988.) Such methods include all inductive methods, all demonstrative methods and all methods to which Putnam calls attention in (1988): rational interpretation, reasonable reasoning and general intelligence. Although he makes the strengthened PGA argument in (1994a), he alludes to it in:

> This is analogous to saying the true nature of r a t i o n a l i t y—or at least of human rationality—is given by some "functional organization", or computational description […]. But if the description is a formalization of our powers to reason rationally *in toto*—a description of a l l our means of reasoning—then inability to know something by the "methods formalized by the description" is inability to know that something i n p r i n c i p l e. (Putnam, 1988)

Strengthened PGA claims all inductive methods, all notions of epistemic justification, all methods of inquiry into the nature of the world are subject to the Gödel incompleteness theorems. The truth of (x) CON(method of inquiry$_x$) is essential to the soundness of PGA. If we can't prove (x) CON(method of inquiry$_x$), then we cannot show that strengthened PGA is sound. Why is that? If we can't prove (x) CON(method of inquiry$_x$), method of inquiry$_x$ might be inconsistent, in which case anything is provable. If so, we can't prove that the epistemic notions of "acceptance" and "justifies" are subject to the Gödel incompleteness theorems. Even if we can prove CON(method of inquiry$_i$) using method of inquiry$_j$ (a stronger extension of method of inquiry$_i$), if CON(method of inquiry$_j$) can't be proved, then it's possible that both CON(method of inquiry$_i$) and NOT-CON(method of inquiry$_i$) can be proved within method of inquiry$_j$. If each method of inquiry is subject to the Gödel incompleteness theorems, then no method of inquiry can be proved consistent. If no method of inquiry can be proved consistent, it is possible no method of inquiry is consistent.

I will now argue that strengthened PGA engenders an absurdity. Suppose that all methods of inquiry (such as statistical methods and methods that deliver proofs in another epistemic modality) are subject to the Gödel incompleteness theorems. That supposition would have as a consequence that all of our reasoning (in whatever method of inquiry that reasoning occurs) about the Gödel incompleteness theorems is subject to the Gödel incompleteness theorems. In which case, that reasoning might not be correct, and so that reasoning could not be epistemically justified. Why is that? For any method of reasoning, its' consistency cannot be proved. Thus it is left open that any method of reasoning might be inconsistent. Consider the following: for any chain of reasoning that establishes proposition *p*, it is possible there is another chain of reasoning that establishes not-*p*. This is so because it is possible that all methods of inquiry are inconsistent. If so, one could validly reason to p and one could validly reason to not-*p*, for any inconsistent method of inquiry. Thus, for all *p*, *p* cannot be epistemically justified, since for each *p*, one might validly infer not-*p* and *p*. This is an absurdity. Take this absurdity to be a reductio of the argument that all forms of reasoning are subject to the Gödel incompleteness theorems.

Given this absurdity, the most natural explanation of it is that one assumed that all methods of inquiry are subject to the Gödel incompleteness theorems. Give up that assumption, and the absurdity is removed. But giving up that assumption means there must exist some methods of inquiry that are not subject to the Gödel incompleteness theorems. If so, it is possible that any such method can prove CON(*P*) or CON(method of inquiry subject to the Gödel incompleteness theorems) with less than mathematical certainty or in some other epistemic modality. And if that is the case, then any finitary computational machine could also make such inferences. No cognitive difference would be registered between human minds and any finitary computational machine. (There might be significant cognitive differences between human minds and finitary computational machines which can compute functions that human minds cannot compute, owing to resource limitations, such as the length of time allowed for computing values of the function.)

7. A Fundamental Logical Problem for EGF and MGF

We now introduce a logical difficulty that arises in MGF and EGF arguments, how anti-functionalists might respond to it and whether Putnam can satisfactorily respond to it. We remark that a difficulty noticed by George Boolos (1986) will not be considered here. Boolos argued the Gödel disjunction (Gödel, 1995) is not derivable from the Gödel incompleteness theorems without first clarifying what it means for a human mind to be equivalent to a finite computing machine. What does it mean to assert that the human mind is equivalent to a Turing machine? We do not consider it here, because Nathan Salmon (2001) has convincingly argued the Gödel disjunction can be used to make philosophically interesting claims about the limitations of the human mind even if we do not have a precise description of what it is for human minds to be equivalent to Turing machines.

7.1. The Logical Problem Confronting EGF and MGF Arguments Is Recursively Unsolvable

The possibilities Kreisel (1972) notes for finitistically proving CON(**PA**) with less than mathematical certainty or in some other epistemic modality must be taken seriously by anti-functionalists who offer EGF or MGF arguments. Failure to take them into account is a logical error in EGF or MGF arguments. Why is that? Where a human agent can finitistically prove CON(P) with less than mathematical certainty or in some other epistemic modality—and that is how such human agents prove CON(**PA**) and the Gödel sentence for **PA**, so also might a finitary computational machine hypothesized to provide a computational description of human mentality. If so, neither MGF nor EGF arguments can distinguish human mentality from finitary computational machines. Failure to consider this possibility is a logical error in Lucas-Penrose-Putnam arguments. However, taking this possibility into account is a recursively unsolvable problem. The anti-functionalist is then faced with a dilemma: either the anti-functionalist fails to take into account Kreisel's way out, in which case they commit a logical error in their argument or else they do take it into account, in which case they must solve a recursively unsolvable problem.

The anti-functionalist might voice the following objection to the claim that they commit a logical error by failing to take into account Kreisel's way out: "The functionalist must find a specific method of inquiry or program that proves CON(P) with less than mathematical certainty or in another epistemic modality. The anti-functionalist is not required to find such a method. No logical error is committed by failing to consider the possibility of such a way out". This objection can be easily dismissed. The anti-functionalist makes the claim that a human mind not fully characterized by any finitary computational machine can determine the truth of CON(P). But it is possible that there is a method of inquiry or a program that can determine CON(P) with less than mathematical certainty or in some other epistemic modality. It is up to the anti-functionalist to dismiss that possibility. To dismiss it, the anti-functionalist must prove a negative existential claim: there is no such method of inquiry or program. It will be shown below that dismissing this possibility is a recursively unsolvable task.

Recall that Putnam's objection to Penrose's argument is that the program P might be so large that it cannot be humanly surveyed, and so no human could establish CON(P). Putnam only needs to cite the possibility that the program P is so large that no human could survey it. Since it is a possibility which, if true, would undermine Penrose's argument, Penrose must respond to it. It is not a legitimate argumentative move for Penrose to reply that Putnam must provide an actual P which cannot be humanly surveyed. The burden of proof is on Penrose—to show that the actual P can be humanly surveyed. Of course, P has yet to be written, since we do not now have a complete finitary computational description of human mentality (should there be one), so Penrose cannot counter Putnam. That is why Putnam's critique of Penrose's argument is so devastating.

Anti-functionalists who wish to avoid that logical error by taking these possibilities into account confront a computationally daunting task. Call that task "DISJUNCTION". It is the following: The anti-functionalist must show that either: (i) each method or program for mathematically or non-mathematically finitistically proving, with less than mathematical certainty or in some other epistemic modality, the consistency of *P* (the ultimate computer program that completely describes the human cognitive mind) is subject to the Gödel incompleteness theorems or (ii) if that cannot be done, because such a method or program is not subject to the Gödel incompleteness theorems, show that the proofs delivered by those methods or programs are not epistemically justified.

From DISJUNCTION there is a dilemma for anti-functionalists using EGF or MGF arguments:

First horn: The anti-functionalist must show, for each possible method or program capable of finitistically demonstrating the consistency of *P* with less than mathematical certainty or in some other epistemic modality, that it is either subject to the Gödel incompleteness theorems or that, where it is not subject to the Gödel incompleteness theorems, it is epistemologically inadequate.

Second horn: If the anti-functionalist does not enumerate all of these possibilities, a logical error is committed in their EGF or MGF argument.

DISJUNCTION has logical complexity Π(1,2). Suppose an anti-functionalist offers an MGF argument. In virtue of DISJUNCTION, they must be able to perform an infinitary computational task. If they have infinitely many resources, they will be able to complete the task. If not, then they will not. But if they do not complete the task, then they commit a logical error in MGF. Thus the anti-functionalist who uses an MGF argument must either have the capacity to make infinitary computations or else commits a logical error. But it is not known whether human beings do or do not have infinitary computational capacities.

The anti-functionalist must show that human beings can prove CON(*P*), but the machine for which *P* is its program cannot prove CON(*P*). Neither the human nor the machine can finitistically prove CON(*P*) with mathematical certainty in the program for *P*. So the anti-functionalist must finitistically prove CON(*P*) with less than mathematical certainty or in some other epistemic modality that is not available to the machine. To show these methods are not available to the machine, she must (according to DISJUNCTION) be able to make infinitary computations to canvass all of the possibilities for doing just that or else commit a logical error. But, once again, it is not known whether human beings do or do not have infinitary computational capacities.

7.2. DISJUNCTION is Π(1,2) in the analytic hierarchy

(a) The first disjunct in DISJUNCTION.

How many methods of reasoning are there for finitistically proving CON(*P*) with less than mathematical certainty or in some other epistemic modality? Since formal systems such as **PM**, **FOL**, and sentential logic prove their truths with mathematical certainty, and since the Gödel theorems tell us that we cannot finitistically establish CON(*P*) with mathematical certainty, those formal systems cannot be used. But probabilistic formal systems can deliver their truths with less than mathematical certainty.

For instance, assume we use a statistical method based on a Carnapian measure function to finitistically prove CON(*P*) with less than mathematical certainty. Then there are infinitely many possible methods that can be used, since Carnapian inductive logics employ a caution parameter that has infinitely many values and which differentiates different logics (Carnap, 1952; 1962). (This establishes an existence proof that there are infinitely many inductive methods. For recent work on new probabilistic proof methods in randomness and computation, see Wigderson, 2019) How many different systems of formal inductive reasoning are there? How many probabilistic logics are there? How many hybrid modal probabilistic logics? Thus far we have the following computational problem: (i) Look at each method for finitistically proving CON(*P*) with less than mathematical certainty or in some other epistemic modality. (ii) Show it is subject to the Gödel incompleteness theorems.

What of proving some proposition in an epistemic modality other than that of mathematical certainty? For instance, philosophical nonmathematical reasoning that cannot be translated into first-order logic might be an example. One problem, though, is that Hilbert's thesis that any argument can be translated into first-order logic makes it difficult to claim that there is reasoning in a natural language that cannot be captured in first-order logic.

There are infinitely many applicable methods of reasoning with less than mathematical certainty or in another epistemic modality. Each of them must be enumerated and checked for being subject to the Gödel incompleteness theorems. And there is an additional regress-like wrinkle. It is the following. Suppose a program P^* proves CON(*P*) with less than mathematical certainty or in some other epistemic modality. The anti-functionalist needs to verify that P^* is subject to the Gödel incompleteness theorems. (If not, then neither an MGF nor an EGF argument can be deployed.)

The wrinkle is that even if P^* is subject to the Gödel incompleteness theorems, there might be a program P^{**} that can be used to mathematically and finitistically prove CON(P^*) with less than mathematical certainty or in some other epistemic modality. Suppose that P^{**} is shown to be subject to the Gödel incompleteness theorems. If so, there is a possibility there is a program P^{***} that can be used to mathematically prove CON(P^{**}) with less than mathematical certainty or in some other epistemic modality. So we have the possibility of an infinite regress for each program or method for proving CON(P^*) and its star relatives that we have shown to be subject to the Gödel incompleteness theorems.

The procedure then, is the following. Look at each method$_{1,i}$ for proving CON(P) with less than mathematical certainty or in some other epistemic modality. Show it is subject to the Gödel incompleteness theorems. If it is, look at each method$_{2,j}$ for proving CON("method$_{1,i}$") with less than mathematical certainty or in some other epistemic modality. Show it is subject to the Gödel incompleteness theorems. If it is, look at each method$_{3,k}$ for proving CON("method$_{2,j}$") with less than mathematical certainty or in some other epistemic modality. Show it is subject to the Gödel incompleteness theorems. Continue in this way ad infinitum.

Let's consider an objection the anti-functionalist might raise to the specter of the infinite regress. She tells us that there will be no infinite regress, because of her dialectical situation in EGF or MGF arguments. Whenever computational functionalists propose a method *M*, all she has to do is to show *M* is subject to the Gödel incompleteness theorems. She plays a waiting game. She waits for the computational functionalist to propose a method, and only then does she need to show that the proposed method is subject to the Gödel incompleteness theorems (Lewis, 1969; 1979; Lucas, 1961; 1970).

This objection fails, for two reasons. The first is that methods of proof that prove a theorem with less than mathematical certainty or in some other epistemic modality are methods of proof that will be used to prove the consistency of the methods for proving CON(P) that are susceptible to the Gödel incompleteness theorems. So we are still considering a specific machine *M* and not any other machine, *M*'. The anti-functionalist does not, contra J. R. Lucas, play a wait and see game with the computational functionalist.

Second, all MGF and EGF arguments are responsible to certain epistemic standards: if there are any relevant possibilities that undermine the arguments, they must be examined. If it is possible there is a method or program *P* not subject to the Gödel incompleteness theorems that finitistically proves CON(M) with less than mathematical certainty or in some other epistemic modality, then that undermining possibility must be discharged.

The anti-functionalist implicitly makes a negative existence claim in EGF and MGF arguments: there is no method or program subject to the Gödel incompleteness theorems by which CON(P) can be finitistically shown correct with less than mathematical certainty or in some other epistemic modality. Since there are infinitely many possibilities for finitistically proving CON(P) with less than mathematical certainty or in some other epistemic modality, each of them must be taken into account. If not then the negative existence claim fails.

7.3. How Program Length Contributes to the Complexity of DISJUNCTION

Suppose that *P* is so long it can't be surveyed by any human agent, whether they are finitistically computationally describable or not. If that is the case, we will not know if there are any programs or methods that can be used to prove CON(P) with less than mathematical certainty or in some other epistemic modality. But there might be ways of compressing the length of P so that we can then determine if there are methods that can be used to prove CON(P). One way of doing this is to reduce P to some program P^* that is humanly surveyable. (One then looks at methods for proving CON(P^*) with less than mathematical certainty or in some other epistemic modality.) There are three ways in which this can be done. One method is by a relative interpretation of P in P^*, another is by a translation of P into P^* and the third is a reduction of P to P^*. There are logical differences between interpretations, translations and reductions, which are the subject of reductive proof theory. What is common to all three is that the map from P into P^* is recursive and preserves negation. The latter condition ensures that logical consistency is preserved under the map.

The maps between P and P^* preserve consistency, provided P^* is consistent. Since the assumption is that P is consistent, we need to find a short and consistent P^*. Suppose P^* is not feasibly short. It is possible there is a P^{**} that is consistent and feasibly short to which P^* can be reduced or translated or into which it can be interpreted. At each level of reduction for which there is a consistent and infeasibly long P^{n*}, it is possible that in a reduction to the next level, by either a translation, reduction or interpretation, there is a consistent and feasibly short $P^{(n+1)*}$.

To avoid in EGF and MGF arguments the logical error committed by Penrose, we have to consider the possibility that P is infeasibly long and then to consider how it might be compressed. The possibility of an infinite chain of reductions of length omega is a prospect that cannot be *a priori* ruled out. (The chain length could be omega, since a reduction might not decrease the length of P^{n*}.) There are also other methods that can compress P. For instance, P could be translated into another programming language in which compression devices called MACROS are available or other higher-order programming constructs that facilitate program compression. There are infinitely many different programming systems, so there are that many possibilities that might need examination in the search for a feasibly short P. There are also speed-up theorems in the theory of computability that tell us there's no recursive bound on the speed-up of some programs (over the initial program for which there is speed-up).

The anti-functionalist can object to the preceding infinite regress generated by program compression considerations in the same way she objected to the first infinite regress above: "The computationalist must first present to me a feasibly short P. Once that is done, we can then see if there are methods or programs that finitistically prove CON(P) with less than mathematical certainty or in some other epistemic modality". Once again, the anti-functionalist misconceives of her epistemic situation in the anti-functionalism dialectic. If it is possible that there is a feasibly short P, then she must examine the possibilities under which it can be obtained. Many of these possibilities (such as relative interpretability) might be dead-ends, might generate infinite regresses or might create trade-off problems.

7.4. The First Disjunct of DISJUNCTION is Π(1,2) in the Analytic Hierarchy

We first noted that there might be infinitely many distinct methods for finitistically proving CON(P) with less than mathematical certainty or in some other epistemic modality. For each such method, the anti-functionalist must show either that it is subject to the Gödel incompleteness theorems or that it is not epistemically justified. We then noted that for each method M that proves CON(P) and is shown subject to the Gödel incompleteness theorems, there might be a method M^* that proves CON(method M) with less than mathematical certainty or in some other epistemic modality. If so, the anti-functionalist must show method M^* is subject to the Gödel incompleteness theorems. In general, for each M that is shown subject to the Gödel incompleteness theorems, there might be an M^* that proves its correctness for which it must be shown it is subject to the Gödel incompleteness theorems. After that, we saw that if P (or any of the methods or any of the M^*'s) is infeasibly long, we need to see if we can compress it to obtain a feasibly short P (or short M^*, etc.) Each of these feasibly short M's must then be shown to be subject to the Gödel incompleteness theorems.

There are infinitely many methods of reasoning that might finitistically prove CON(P) with less than mathematical certainty or in some other epistemic modality. For each method M_i subject to the Gödel incompleteness theorems, it is possible there is a method or program that finitistically proves CON(M_i) with less than mathematical certainty or in some other epistemic modality. Let M_j be the method that finitistically proves CON(M_i), where $i \neq j$. If M_i is subject to the Gödel incompleteness theorems, then there might be an M_k ($i \neq j \neq k$) that finitistically proves CON(M_i) and which must then be shown by the anti-functionalist to be subject to the Gödel incompleteness theorems. For each of the infinitely many M_i's, there are infinitely many M_i^{n*}'s. Finally, for every M_i and M_i^{n*}, it is possible it is infeasibly long and thus we need to look for a compression of it into a feasibly short program. But for each M_i and M_i^{n*}, there might be an infinite sequence of compression reductions R_i.

Each method or procedure can be considered to be a function from the natural numbers to natural numbers. Determining that a method or procedure is or is not subject to the Gödel incompleteness theorems is a recursive predicate. The predicate is applied to each method or procedure, of which there are infinitely many. So there is a quantifier over the set of methods and procedures—it is a function quantifier. For all such methods or procedures, it is possible there exists a method or procedure not subject to the Gödel incompleteness theorems which verifies its consistency with less than mathematical certainty or in some other epistemic modality. (M_x) ($\exists M_y$) (M_y is not subject to the Gödel incompleteness theorems AND M_y proves CON(M_x) with less than mathematical certainty or in some other epistemic modality). In the analytic hierarchy, this sentence has logical complexity Π(1,2).

(b) The second disjunct in DISJUNCTION

Recall the second disjunct in DISJUNCTION: If a method or program for proving CON(*P*) with less than mathematical certainty or in some other epistemic modality is not subject to the Gödel incompleteness theorems, then show that the proofs delivered by that method or program are not epistemically justified. The anti-functionalist must show that for each method or program examined by the procedure described in the first disjunct of DISJUNCTION that is not subject to the Gödel incompleteness theorems, it is not epistemically justified. This must be done to save any EGF or any MGF argument. Suppose the anti-functionalist argument is an EGF argument. The claim is: *P* cannot be proved correct because it is subject to the Gödel incompleteness theorems (and thus cognitive science cannot be justified). But there might be other ways to prove CON(*P*) with less than mathematical certainty or in another epistemic modality. If those ways are subject to the Gödel incompleteness theorems, the claim remains intact. If any of those ways are not subject to the Gödel incompleteness theorems, they prima facie refute the claim. The only way to save the claim is to show that the methods or programs not subject to the Gödel incompleteness theorems are not epistemically justified. That is, proofs delivered by those methods or programs are not epistemically warranted.

Since any method or any procedure might not be subject to the Gödel incompleteness theorems, then every subset of the infinite methods tree might need to be tested for epistemic adequacy—that it is epistemically justified. Of course, no point in the infinite methods tree might need to be tested, if every point represents a method or program that is subject to the Gödel incompleteness theorems.

How we can show that a method or program is not epistemically justified? If what is proved by a method has a 50% chance of being true, we can conclude the method is not justified. However, what do we say when the probability of being true is greater than ½? What is the cut-off point? What if we do not have sufficient statistics for showing the likelihood of what a method proves? What epistemological theory do we employ in assessing epistemic justification of a method? Even if we are guided by statistical methods used in the sciences, those methods still make philosophical presuppositions about the nature of probabilities.

Suppose that a method uses nonmathematical philosophical reasoning (Kreisel, 1972) that contains no quantitative information necessary for obtaining probabilities. How do we assess these methods for epistemic justification? Is the epistemic justification of a quantitative method different in kind from the epistemic justification of a non-quantitative method? What does it mean to say we search the space of epistemologies for various construals of epistemic justification (Audi, 1988; Lehrer, 1990)? Given that EGF and MGF arguments are philosophical arguments claiming to refute a philosophical position in the philosophy of mind, any elucidation of the notion "epistemic justification of *P* (for any *P*)" must be philosophically respectable. If the philosophical construal of "epistemic justification of *P* (for any *P*)" is not philosophically respectable, the anti-functionalist will not be able to satisfy the second disjunct of DISJUNCTION.

These issues concerning epistemic justification are critical problems for the anti-functionalist. Without establishing that methods or procedures not subject to the Gödel incompleteness theorems are not epistemically justified, EGF and MGF arguments fail. The anti-functionalist must be prepared to decide what counts as epistemic justification of the correctness of *P* (for any *P*), and so what counts as the epistemic justification of cognitive science. Being able to assess the epistemic justification of methods that prove CON(*P*) with less than mathematical certainty or in another epistemic modality is a necessary condition for the success of EGF and MGF arguments. An important philosophical project, then, is elucidation of the notion "epistemic justification of proofs of CON(*P*) with less than mathematical certainty or in another epistemic modality".

7.5. Chains and Tangled Chains in the Methods Tree Exhibiting Defeater Relations

Suppose that a method or procedure is not subject to the Gödel incompleteness theorems and that it is not epistemically justified. Does it follow it can be dismissed by the anti-functionalist? No, for this method might epistemically justify CON(M^*), where method M^* is not subject to the Gödel incompleteness theorems and epistemically justifies CON(*P*). This may happen if we allow relative interpretations, translations and reductions between *P*, the method and M^*. But it can happen even if these relations do not occur. There might be chains in the methods tree, of arbitrary length, in which a method that does not epistemically justify *P* epistemically justifies a method which epistemically justifies *P*. Such chains can be of arbitrary length. Each of these chains must be examined by the anti-functionalist. It is well-known is epistemology that justification of a proposition can be defeated and can be restored after defeat, given the appropriate conditions (Pollock, 1999). The same can happen with methods for proving CON(*P*).

For example, suppose we have a chain in the methods tree of length 1,000 in which the 1,000th element in the chain is not subject to the Gödel incompleteness theorems. It is a method that does proves CON(*P*) with less than mathematical certainty or in some other epistemic modality, but is not epistemically justified when considered in isolation from all of the other methods in the chain. However, the 529th method in the chain epistemically justifies the 530th method in the chain, which, in turn, epistemically justifies the 531st method in the chain. This continues, until the 1000th element in the chain is epistemically justified.

Even if the n^{th} method in a chain is not epistemically justified by the n-1st method in that chain (where the two methods are consider in isolation from all other methods), it does not follow the anti-functionalist can dismiss it, since there might be chains, of arbitrary length starting with the n-k^{th} method, between the n-1st and n-k^{th} methods, which transmit epistemic justification in such a way that the n-1st method is epistemically justified, and in consequence of this, is able to epistemically justify the n^{th} method. Additionally, one method in a chain might defeat epistemic justification of another method in the chain. If a chain of methods is finitely long, the power set of that chain consists of all subsets of methods which might need to be considered by the anti-functionalist. If a chain is infinitely long (because there are infinitely many methods or programs), then all possible chains that can be built with those methods or programs will have the power set of that infinitely long chain, and need to be considered by the anti-functionalist.

An additional complication in building such chains is the existence of methods or programs that defeat epistemic justification of CON(P) or of CON(M_i). Moreover, those methods or programs might not be formalized or even formalizable—suppose they are instances of what Kreisel means by "nonmathematical philosophical reasoning". Justification can be achieved by many different forms of reasoning. If your aim is to show that some proposition is not justified, then you must consider all of the ways in which it could be justified.

Suppose that M_k defeats justification of M_i, and M_i can prove correct CON(P) with less than mathematical certainty or in another epistemic modality. However, there might be a method or program M_{k-i} that defeats justification of M_k, thus restoring M_i so that it can prove correct CON(P). Call this a tangled chain of methods or programs. Notice this problem is similar to the logical problem facing defeater epistemologies (Pollock, 1999). There might be chains of defeaters, of arbitrary length, in which the 999^{th} member of the chain defeats the 347^{th} member of the chain, while the 876^{th} member of the chain defeats the 999^{th}. Simply enumerating and individually assessing each element in the chain is not enough. Each element in the chain must be evaluated for justificatory relations with every other sequence of elements in the chain.

Although formalizable methods or procedures can be considered to be functions over the natural numbers, I am less confident about methods or procedures for, say, nonmathematical philosophical reasoning. Perhaps they can be formalized and considered to be functions over the natural numbers. But the relation of one method justifying another might not be recursive, and might not even be formalizable. So it might be that no logical complexity measure can be assigned to the second disjunct of DISJUNCTION.

We have the following results:

(i) It is possible there are epistemically justified methods or programs which prove, with less than mathematical certainty or in some other epistemic modality, CON(P). EGF arguments must show there are no methods which can do that. If not, the conclusion of the EGF argument—that cognitive science cannot be demonstrated to be a correct theory—fails. EGF arguments assume human minds have a finitary computational description. Showing there are no epistemically justified methods or programs which can prove CON(P) with less than mathematical certainty or in some other epistemic modality is recursively unsolvable. Finitary human minds that have a finitary computational description cannot complete this task. If human minds have a metarecursive computational structure, they might be able to complete the task. But we do not know if human minds have a metarecursive computational structure.

(ii) To save the MGF conclusion that there is a cognitive task human minds can do that finitary computing machines can't, it must be shown either (a) that human minds can prove CON(*P*) with mathematical certainty or (b) that there is no epistemically justified method or program by which CON(*P*) can be proved, with less than mathematical certainty or in some other epistemic modality. Since only an infinitary mind can prove CON(*P*) with mathematical certainty (and only if mathematical certainty can be defined for an infinitistic system of reasoning), (a) has no empirical basis in cognitive science. There is no empirical evidence that human minds can perform infinitary tasks, such as constructing infinite proof trees. EGF arguments must establish (b), and we saw they cannot do so, because it is a recursively unsolvable task. It is a mystery how a human mind, even one that has no finitary computational description, could complete the task (unless it has a metarecursive computational structure, but we do not know whether this is so.)

8. A Categorization of Anti-Functionalist Arguments Using the Gödel Incompleteness Theorems Into Sixteen Cases

There are sixteen cases that are determined by partitioning anti-functionalist arguments into (i) epistemic and metaphysical uses of the Gödel incompleteness theorems—that is, EGF and MGF arguments, (ii) Penrose error cases (infeasibly long programs), and (iii) showing some, but not all weak inductive methods, are subject to the Gödel incompleteness theorems (PGA) and showing that all methods of inquiry into the world (i.e., all inductive methods) are subject to the Gödel incompleteness theorems (strengthened PGA).

There are eight cases when PGA or strengthened PGA succeeds. There are an additional eight cases when PGA or strengthened PGA fails. (We contend they both fail.) What is surprising is that even if PGA or strengthened PGA succeeds, the anti-functionalist acquires virtually no advantage over the computational functionalist in anti-functionalism arguments. It's important to note that in all MGF cases it is not assumed that human minds are finitary, nor is it assumed that they are infinitary. If human minds are infinitary and have a metarecursive structure, should we consider them to have a computational description analogous to finite minds with a computational structure? If human minds are infinitary and do not have a metarecursive structure, we should not consider them to have a computational description. But it is unknown whether human minds are or not infinitary. Similarly, although some cognitive scientists and philosophers believe human minds are finitary and can be described computationally, it is not known whether they are finitary.

In the first kind of EGF argument, it is assumed human minds are finite. Not so for the second kind of EGF argument (see Section 4.2.2 above). However, the second kind of EGF argument shows that metaphysical claims established by

MGF arguments are epistemically justified. Thus MGF arguments need to be categorized—the second kind of EGF argument does not. The phrase “EGF arguments” below refers to the first kind of EGF argument.

8.1. The First Eight Cases: PGA and Strengthened PGA Succeed

A successful PGA shows that some, though not all, weak methods of inquiry are subject to the Gödel incompleteness theorems. The first four cases cover a successful PGA. There are two cases for an EGF refutation of functionalism and two cases for an MGF refutation of functionalism. The two cases for each are when the computational description *P* is feasibly short and when it is infeasibly long.

Case (i): Recall that EGF arguments assume human minds are fully characterized by a finitary computational description. Suppose *P* is feasibly short. Since not all weak inductive methods have been shown to be subject to the Gödel incompleteness theorems, there may be weak methods that prove CON(*P*) with less than mathematical certainty or in another epistemic modality. If so, an EGF argument fails, since it is the point of an EGF argument to show that human minds cannot justify the finitary computational description *P* of themselves. That is, there isn’t a proof of CON(*P*) that is epistemically justified. But a weak method might provide such a proof.

Case (ii): Suppose an EGF argument and that *P* is infeasibly long. Since not all weak methods have been shown to be subject to the Gödel incompleteness theorems, use weak methods to perform a statistical analysis to recover the full size of *P* from the fragments available. Then use weak methods to establish CON(*P*), with less than mathematical certainty. The EGF argument fails, for the same reasons in case (i).

Case (iii): Assume an MGF argument. Recall that MGF arguments show human minds do not have a finitary computational description, and argue that human minds are metaphysically different from finitary computing machines, since there are cognitive activities we can perform, that finitary computing machines cannot perform. Suppose that *P* is feasibly short. Even if human minds do not have a finitary computational description, we cannot use weak inductive methods or programs subject to the Gödel incompleteness theorems to establish CON(*P*), in the epistemic modality of the proof procedures of the weak methods. We can only use weak methods or programs not subject to the Gödel incompleteness theorems to establish CON(*P*) with less than mathematical certainty or in another epistemic modality. However, finitary computing machines can do the same thing, so we can’t establish a metaphysical difference between them and human minds. The MGF argument fails.

Case (iv): Assume an MGF argument and that *P* is infeasibly long. Even if human minds do not have a finitary computational description, we cannot use weak inductive methods subject to the Gödel incompleteness theorems to do a statistical analysis of the fragments of *P* and recover *P* from that analysis and then prove CON(*P*). We can only use weak methods or programs not subject to the Gödel incompleteness theorems to do this. However, so can finitary computing machines. Once again, there is no metaphysical difference which we can establish between them and finitary human minds. The MGF argument fails.

Now we look at the four cases in which strengthened PGA succeeds. Recall that strengthened PGA shows that all methods of inquiry into the structure of the world (i.e., all inductive methods) are subject to the Gödel incompleteness theorems. The four cases are analogous to the four cases for PGA.

Case (v): Assume an EGF argument and that *P* is feasibly short. If so, then there are no weak methods or programs that can be used to show CON(*P*). In that case, the EGF argument succeeds, since we have shown that a human mind with a computational description *P* cannot justify *P*.

Case (vi): Assume an EGF argument and that *P* is infeasibly long. Since there are no weak methods or programs available for a statistical analysis of fragments of *P* to recover *P*, nor for showing CON(*P*), it follows that the EGF refutation succeeds. We have shown that a human mind with a finitary computational description *P* cannot justify *P*.

Case (vii): Assume an MGF argument and that *P* is feasibly short. There are no weak methods that can be used to show CON(*P*). In which case, even human minds with no finitary computational description will not be able to justify *P*. However, finitary computing machines cannot do this either. In which case, there is no discernible metaphysical difference (concerning computability) between human minds with no finitary computational description and finitary computing machines. Hence, the MGF argument fails.

Case (viii): Assume an MGF argument and that *P* is infeasibly long. There are no weak methods or programs that can be used to perform a statistical analysis on a fragment of *P* and recover *P*, nor to show CON(*P*). In which case, even human minds with no finitary computational description will not be able to justify *P*. However, finitary computing machines cannot do this either. In which case, there is no discernible metaphysical difference (concerning computability) between human minds with no finitary computational description and finitary computing machines. Hence, the MGF argument fails.

These analyses reveal an interesting truth. It is that all MGF arguments fail, even though either PGA or strengthened PGA succeeds. On the other hand, though EGF arguments fail even where PGA succeeds, EGF arguments succeed where strengthened PGA succeeds. Thus there is a critical philosophical difference between MGF and EGF arguments.

Note that if it is demonstrated that human minds are able to construct infinite proof trees and do not have a metarecursive structure that allows for a computational description analogous to a computational description of finitary minds, then all MGF arguments will succeed wherever PGA and strengthened PGA succeed. Using the Gödel theorems to refute functionalism by an MGF argument can only succeed if it is a fact (and known to us) that human minds can construct infinite proof trees, but have no metarecursive structure that allows for a computational description analogous to a computational description of finitary minds. If that cannot be demonstrated, then even though PGA or strengthened PGA succeeds, no MGF argument can succeed.

8.2. The Second Set of Cases: PGA and Strengthened PGA Fail

We now look at the same kinds of cases, under the assumption that PGA and strengthened PGA fail (in the way in which I have argued they fail). Cases ix–xii will happen when PGA fails. That is, PGA fails to show that some weak inductive methods are subject to the Gödel incompleteness theorems.

Case (ix): Suppose an EGF argument and that *P* is feasibly short. Since it has not been shown that any weak methods are subject to the Gödel incompleteness theorems, all weak methods are available for proving CON(*P*), with less than mathematical certainty. So a human mind that has a finitary computational description can prove *P* is correct (i.e., justify *P*). Since there are more weak methods available for proving CON(*P*) with less than mathematical certainty, and in other epistemic modalities, than there are when PGA succeeds, EGF arguments fail more often when PGA fails than they do when PGA succeeds.

Case (x): Suppose an EGF argument and *P* is infeasibly long. Since it has not been shown that any weak inductive methods are subject to the Gödel incompleteness theorems, all weak inductive methods are available for statistically recovering P and proving CON(*P*). So a human mind that has a finitary computational description can epistemically justify *P*. Since there are more weak methods available for recovery of *P* and proof of CON(*P*), and in other epistemic modalities, than there are when PGA succeeds, EGF arguments fail more often when PGA fails than they do when PGA succeeds.

Case (xi): Suppose an MGF argument and that *P* is feasibly short. Although all weak inductive methods are available for proving CON(*P*) with less than mathematical certainty, all of these methods are also available to finitary computing machines. In which case, there is no means of discerning a metaphysical difference (concerning computability) between human minds with no finitary computational description and finitary computing machines. MGF arguments fail when strengthened PGA fails, but no worse (or no better) than they failed when PGA succeeded.

Case (xii): Suppose an MGF argument and *P* is infeasibly long. Although all weak inductive methods are available for statistically recovering *P* and for proving CON(*P*) with less than mathematical certainty, all of these methods are available to the finitary computing machine. In which case, there are no means of discerning a metaphysical difference (concerning computability) between human minds with no finitary computational description and finitary computing machines. MGF arguments fail when strengthened PGA fails, but no worse (or no better) than they did when PGA succeeded.

Now we examine the four cases when strengthened PGA fails, because of the absurdity to which it succumbs. Recall the absurdity: *P* encompasses all of the epistemically adequate weak methods *M* of inquiry into the world that could prove CON(*P*) with less than mathematical certainty or in another epistemic modality. Suppose that all methods of inquiry are subject to the Gödel incompleteness theorems. For each method *M*, we cannot prove that it is consistent. So it is possible that each method *M* is inconsistent. For any chain of reasoning that establishes proposition *p*, it is possible there is another chain of reasoning that establishes not-*p*. One could validly reason to p and one could validly reason to not-*p*, for any inconsistent method of inquiry. Thus, for all *p*, *p* cannot be epistemically justified, since for each *p*, one might validly infer not-*p* and validly infer *p*. This is an absurdity. Take this absurdity to be a reductio of the argument that all forms of reasoning are subject to the Gödel incompleteness theorems.

Case (xiii): Suppose an EGF argument and *P* is feasibly short. The reasoning is exactly the same as it is for case (ix). All weak methods are available for proving CON(*P*) with less than mathematical certainty or in another epistemic modality. So a human mind that has a finitary computational description can epistemically justify *P*. Since there are more weak methods available for proving CON(*P*) with less than mathematical certainty, and in other epistemic modalities, than there are when PGA succeeds, EGF arguments fail more often when strengthened PGA fails than they do when strengthened PGA succeeds.

Case (xiv): Suppose an EGF argument and *P* is infeasibly long. The reasoning is exactly the same as it is for case (x). All weak methods are available for statistically recovering and proving CON(*P*) with less than mathematical certainty. So a human mind that has a finitary computational description can epistemically justify *P*. Since there are more weak methods available for recovery of *P* and proof of CON(*P*), and in other epistemic modalities, than there are when PGA succeeds, EGF arguments fail more often when strengthened PGA fails than they do when strengthened PGA succeeds.

Case (xv): Suppose a MGF argument and *P* is feasibly short. The reasoning is exactly the same as it is for case (xi). All weak methods are available for proving CON(*P*) with less than mathematical certainty or in some other epistemic modality to finitary computing machines and human minds with no finitary computational description. In which case, there is no means of discerning a metaphysical difference (concerning computability) between human minds with no computational description and finitary computing machines. MGF arguments fail when strengthened PGA fails, but no worse (or no better) than they did when strengthened PGA succeeded.

Case (xvi): Suppose a MGF argument and *P* is infeasibly long. The reasoning is exactly the same as it is for case (xii). All weak methods are available for statistically recovering *P* and for proving CON(*P*), with less than mathematical certainty or in another epistemic modality, to finitary computing machines and human minds with no finitary computational description. In which case, there is no means of discerning a metaphysical difference (concerning computability) between human minds with no finitary computational description and finitary computing machines. MGF arguments fail when strengthened PGA fails, but no worse (or no better) than they did when strengthened PGA succeeded.

That concludes the categorization of cases under PGA and strengthened PGA, where they succeed and where they fail. Do we have any reason to believe that *P* will be infeasibly long? Now, we have no such reason. We do not know what ultimate cognitive science will look like, so we do not know, now, whether in ultimate cognitive science the ultimate program *P* will be infeasibly long. We do not have a theory of feasible computability that will tell us whether programs that have outputs of certain kinds are feasibly short. We do not know if human minds can be completely described computationally. We do not know if there is an ultimate cognitive science.

9. Twelve Objections to the Absurdity Engendered by Strengthened PGA

There are several anti-functionalist objections to the absurdity that threatens to destroy PGA and strengthened PGA and thus threatens to destroy EGF and MGF arguments. I enumerate and respond to them below.

Objection 1: Even if *P* is infeasibly long, human minds can epistemically justify *P*, though no finite computing machine (which *P* formally characterizes) can. Since all epistemically adequate weak methods of inquiry into the world—including any that confer empirical justification upon CON(*P*)—are, by PGA, subject to the Gödel incompleteness theorems, no finite computing machine formally characterized by *P* can employ those methods to prove, with less than mathematical certainty or in another epistemic modality, CON(*P*). However, human minds can do that, since statistical methods fall under the weak methods subsumed by *P* and statistical methods are employed where human minds face resource limitations or do not have all the facts. The burden of proof is on the shoulders of the functionalist, to show that for programs greater than length L no statistical method subsumed under *P* can empirically justify CON(*P*).

Response: It is true that no finitary computing machine formally characterized by *P* can use the statistical methods subsumable under *P*, provided that strengthened PGA succeeds. But human minds, even under the assumption they have no finitary computational description, are similarly forbidden. If all formalized statistical methods are shown by strengthened PGA to be subject to the Gödel incompleteness theorems, then no finitary human mind can use them to recover *P* and then prove CON(*P*).

Objection 2: Finitary human mind can empirically justify CON(*P*) by reducing its consistency problem to a consistency problem for a formal system that does not subsume any of the weak methods of inquiry into the world that are subsumed by *P*. We then use weak methods to prove, with less than mathematical certainty, CON(REDUCING FORMAL SYSTEM) and use the reduction to conclude CON(*P*).

Response: If *P* subsumes all methods of inquiry into the world, then any formal system that does not subsume them is probably not a formal system to which *P* can be reduced. Suppose that, for the sake of argument, it is. Reductive proof theory requires there is a recursive function that maps every proof in the reduced system to a proof in the reducing system. Moreover, this mapping must itself be provable in a formal system that is, in general, included in the reducing system. When these conditions are satisfied, the reducing system will be a conservative extension of the reduced system. There is nothing in the reduced system that cannot be proved in the reducing system and, more importantly, there is nothing in the language of the reduced system that can be proved in the reducing system, though not proved in the reduced system. In other words, for any proof in PGA that any epistemically adequate weak method in *P* is subject to the Gödel incompleteness theorems, there will be a corresponding proof in the reducing system that whatever is the analogue of the weak method in *P* is subject to the Gödel incompleteness theorems.

Objection 3: If *P* is infeasibly long, it fails as an explanatory theory in cognitive science. Any finitary computational description we can't follow is one that can't be explanatory for us. Thus, under the assumption human beings have no finitary computational description that characterizes their complete mentality, an infeasibly long *P* secures for anti-functionalists the conclusion that cognitive science is not justified. If a scientific theory has no explanatory value, it loses epistemic justification.

Response: This objection does not advance the anti-functionalist even one square forward in the functionalism debate. If it turns out that *P* is infeasibly long, then human beings might never discover it. What we do discover will be an approximation to *P* which we do find explanatory and that is not infeasibly long. The objection which the anti-functionalist just voiced is really a skeptical objection, and it is one which could be voiced in any scientific discipline whatsoever. The anti-physicalists can say that the ultimate theory of physics is super-long and thus has no explanatory value. The same response to the anti-functionalist holds here as well. Yes—it is a worry, but no—it is not a worry that gives the anti-functionalist any advantage, for it is a general skeptical worry.

Objection 4: Let's try to refine the preceding objection. Genuine warranted assertibility and epistemic justification have no finitary computational description. These methods, because they are not formalizable, are not subject to the Gödel incompleteness theorems. Finitary human minds—under the assumption they have no finitary computational description—can use these methods to produce a proof of CON(*P*). So there is something a finitary human mind can do that no finitary computing machine can do.

Response: This is a confused objection. How can methods resisting formalization be used to prove the correctness of a formal system? Strengthened PGA shows that all epistemically adequate weak methods are subject to the Gödel incompleteness theorems. Thus it shows that all epistemically adequate weak methods have no complete finitary computational description. But if strengthened PGA fails, it is left open that there are formalizable epistemically adequate weak methods that can prove, with less than mathematical certainty or in another epistemic modality, CON(*P*). If strengthened PGA fails, then the anti-functionalist must compute the solution to a recursively unsolvable problem, in order to show that there are no epistemically adequate weak methods that are not subject to the Gödel incompleteness theorems. The point is that the only way we have of showing that there is no complete finitary computational description of X is by using a Gödelian argument. Strengthened PGA is such a Gödelian argument, but it fails.

Objection 5: The absurdity is a travesty of mathematical reasoning. If you are right, then you have shown that the Gödel theorems in their original context—proving the incompleteness of Peano arithmetic and the unprovability of CON(**PA**)—fail to work. One can run your absurdity argument on the provability predicate and easily reach the absurd conclusion that there is no unprovable sentence in Peano arithmetic. You would have shown that Gödel is wrong. Since that is too absurd to consider, we must conclude that you are wrong!

Response: That is an important objection However, you did not think very clearly about the matter at hand. The provability predicate is not defined by Peano arithmetic. We have independent reasons for believing in its cogency and we could construct it even if Peano arithmetic did not exist. What we are able to do in Peano arithmetic is to arithmetize it and then employ the diagonal lemma to secure the incompleteness theorems.

The situation is quite different when it comes to program *P*—the computational description of our methods of inquiry into the world. Recall that in PGA the analogue of the notion of "proof" for Peano arithmetic is the notion of "justifies". However, *P* defines the notion of justification. If there were no *P*, there would be no notion of justification. If it turns out that the notion of justification cannot itself be justified—and that is exactly what PGA attempts to show—then we have no coherent notion of justification. If there are truths about justification we are forbidden from justifying, the notion is incoherent. In which case, we can't appeal to the Montague-Kaplan-Thomason axioms for axiomatizing "justifies" so that it can meaningfully satisfy the Gödel diagonal lemma, since we have no reason to think that these axioms applied to "justifies" are true. On the other hand, we do have independent reasons for thinking that the Hilbert derivability conditions for the provability predicate are true, independently of the question of the consistency of Peano arithmetic.

Objection 6: You cannot be serious that human minds with no finitary computational description have no epistemic advantages over finitary computing machines. Can't a human mind with no finitary computational description survey an infeasibly long *P*? If not, then what could possibly be the difference between the human minds and finitary computing machines? Are you proposing that they are identical?

Response: No, we are not. But just because a human mind has no finitary computational description does not entail it is able to construct infinite proof trees or that it has the computational resources to survey an infeasibly long *P*. Human minds that have no finitary computational description might not have any epistemic advantages over finitary computing machines. Even infinitary agents cannot prove the consistency of Peano arithmetic using a finitary and effective proof, since finitary and effective proofs of it are prohibited by Gödel's incompleteness theorems. If all weak methods for proving CON(**PA**) are subject to the Gödel incompleteness theorems, then an agent with an infinitary mind can only employ an infinitary method to prove CON(**PA**). In which case, the anti-functionalist must demonstrate that human minds are infinitary or give up the view that there is an epistemic difference between human minds that have no finitary computational description and finitary computing machines governed by *P*.

If human minds, under the assumption they have no finitary computational description, prove CON(*P*) with less than mathematical certainty or in another epistemic modality, by weak methods not subject to the Gödel incompleteness theorems, they are not distinguishable from finitary computing machines that can similarly employ those weak methods to prove CON(*P*). If those weak methods are subject to the Gödel incompleteness theorems, then neither the human mind that has no finitary computational description nor the finitary computing machine can prove CON(*P*) in the characteristic epistemic modality of the proof procedures of the formal systems that formalize the weak methods.

The anti-functionalist wants to prove that all weak methods which could, under some standard of epistemic adequacy, prove CON(*P*), with less than mathematical certainty or in some other epistemic modality, are subject to the Gödel incompleteness theorems. Yet this task is just what engenders the absurdity. If all weak methods which could, under some standard of epistemic adequacy, prove CON(*P*) with less than mathematical certainty or in some other epistemic modality, are subject to the Gödel incompleteness theorems, then they cannot be used to prove CON(*P*), even by minds that have no finitary computational description. This is so, because whatever the epistemic modality of the proof of CON(*P*) no agent, no matter what its computational structure (whether finitary or metarecursive), can prove CON(*P*) in that epistemic modality.

Objection 7: An epistemic use of the Gödel theorems does, in fact, render a metaphysical conclusion. It shows that the cognitive structure of the human mind is subject to the Gödel incompleteness theorems. That, in turn, shows that we cannot be metaphysically distinguished from finitary computing machines.

Response: However, that is a moot conclusion, since the anti-functionalist who employs an EGF argument proceeds from the assumption that the human mind has a finitary computational description. That is, she proceeds from the adoption of the metaphysical picture of the human kind as a finitary computing machine. The Gödel theorems tell us about the limitations faced by such finitary computational descriptions, but the basic metaphysics is already in place. EGF arguments don't conclude to a metaphysical conclusion, as is done in MGF arguments.

Objection 8: That the anti-functionalist falls into an absurdity in escaping from the simple logical error of Penrose is a clever observation, but it is false. We do not say that an absurdity arises out of the fact that Peano arithmetic is subject to the Gödel incompleteness theorems. A formal system strong enough to carry out (minimally) Robinson arithmetic is one for which we cannot, with mathematical certainty, employing a finitistic and effective proof procedure, prove its consistency. However, that we cannot is not license for us to infer that we can reasonably doubt that Peano arithmetic is subject to the Gödel incompleteness theorems. That is absurd. It is too easy a move. Certainly, we would have encountered someone in mathematics making it long ago. But no one did, because it is nothing short of being numbingly stupid.

Response: You are quite right about Peano arithmetic. No absurdity—of the kind we have specified—arises, and it would be numbingly stupid to claim one does. It is the assumption that all forms of reasoning are subject to the Gödel incompleteness theorems which produces the absurdity. The absurdity shows the assumption is false. Gödel did not prove that all formal systems are incomplete—only those that are string enough for Peano arithmetic.

Additionally, the epistemic situations with respect to Peano arithmetic and with respect to *P* are quite different. There is probably not a single mathematician who genuinely doubts the consistency of Peano arithmetic. There are infinitary proofs of it—Gentzen discovered one in the mid-thirties and Ackermann polished it five years later. We have good reason to believe that the Gentzen proof works. We have, then, no reason to believe that the Gödel incompleteness theorems fail to hold of a formal system that encompasses Peano arithmetic. There is no absurdity, even though we cannot prove, with mathematical certainty, using a finitistic and effective proof procedure, the consistency of Peano arithmetic. From that we do not conclude that Peano arithmetic might be inconsistent.

The epistemic situation is much different with respect to *P*, which is a finitary computational description based on a cognitive theory, an ultimate one at that. We do not have the same intuitions about its consistency that we have about the consistency of Peano arithmetic, for we do not even have the cognitive theory that underlies *P*. It is a suppositional device to carry out the anti-functionalist argument. Nor, for the same reasons, do we have an infinitary proof of CON(*P*). If *P* encompasses all finitary methods of inquiry into the world, and we show that all of these methods are subject to the Gödel incompleteness theorems, then we have no methods of inquiry left with which to carry out the consistency proof of *P*, other than infinitary ones. We cannot, however, say that we have good reason to believe that *P* is consistent, since we have no idea what it will look like and, even if we did, it is still based on a cognitive theory which has to be tested. If we cannot test it, because all our procedures for testing it are subject to the Gödel incompleteness theorems, we are in an epistemic situation of maximal ignorance. We have no good reason to believe it is consistent and no good reason to believe it is inconsistent. In that epistemic situation, we cannot accept the result that all epistemic methods of inquiry are subject to the Gödel incompleteness theorems. The absurdity cannot be dismissed by comparing it with the disanalogous epistemic situation in Peano arithmetic. It is, then, a genuine epistemic problem for the anti-functionalist.

Objection 9: You mistakenly think that since PGA and strengthened PGA incur an absurdity, it is left open for finitary human minds and finitary computing machines to use any weak methods of empirical inquiry into the structure of the world. The absurdity does not entitle the agent to use all weak methods. Given there is an absurdity, how would you determine the weak methods which escape being subject to the Gödel incompleteness theorems because of the absurdity? You cannot stipulate there are weak methods that can be used by a human agent. Just as a paradoxical sentence (such as the Liar sentence) can't be assumed true, agents can't conclude from the absurdity of strengthened PGA that there are weak methods that are not subject to the Gödel incompleteness theorems and that are thereby legitimate to use.

Response: That is a perceptive point, but it is misguided. The analogy with Liar sentences is not acceptable. Once we show a Liar sentence is paradoxical, we cannot assume it is true, nor can we assume that it is false. In some truth theories, we withhold assignment of a truth-value to it, in which case it has a null functional status in our discourses.

On the other hand, the assumption in strengthened PGA that led to the absurdity is that all methods of inquiry into the world are subject to the Gödel incompleteness theorems. The absurdity shows that assumption is false—not all methods of inquiry into the world are subject to the Gödel incompleteness theorems. So it is left open that there are weak methods which are not subject to the Gödel incompleteness theorems.

Objection 10: Any intuitions about the consistency of *P* must be seen as evidence for the claim that we have infinitary capacities. We would not have those intuitions unless there is some infinitary reasoning process, below the threshold of conscious perception, which accounts for them. The best explanation of why we have these intuitions is that there is some infinitary reasoning mechanism in us which causes us to have those intuitions. Thus, even though there is an absurdity for the anti-functionalist who wants to show all weak methods are subject to the Gödel incompleteness theorems, the intuitions we would (since *P* does not exist—it is merely a hypothetical construct) have about the correctness of *P* are reliable indicators of our infinitary capacities. The absurdity is no hindrance to the anti-functionalist, since human minds are infinitary and we do not even need PGA.

Response: If we do have intuitions that *P* is consistent, and we set a probability level for the reliability of those intuitions higher than the reliability we would—in probabilistic terms—rate the weak methods for showing *P* is consistent, with less than mathematical certainty or in some other epistemic modality, and we know that there are no other weak methods available and that only infinitary methods can prove the correctness of *P* with mathematical certainty, what can we reasonably conclude about the nature of our cognitive capacities? We can't reasonably conclude that we have infinitary cognitive capacities. It would be the case that the best explanation of our intuitions is that an infinitary reasoning mechanism causes us to have them if we had no alternative explanations of them. But we have alternative explanations of how we could have such intuitions, and these explanations do not posit infinitary reasoning processes. For instance, we have experiences with cognitive theories of inductive reasoning, and we see an analogy between them and *P*. If they are known to be consistent, we conclude that it is highly likely *P* is consistent as well. We might, also, be simply mistaken. Our probabilistic intuitions are notoriously shaky, a fact well-known to cognitive psychologists. In that case, the best explanation for our intuitions is that we have made errors in probabilistic reasoning. If we had independent evidence the human mind performs infinitary operations, then the explanation of our intuitions about the correctness of P in terms of infinitary operations would be superior to the two alternatives we have just cited. But, in the absence of that evidence, the two alternatives are not inferior to it, since they are sensitive to established work in cognitive psychology, while there is no established work that shows we have infinitary reasoning powers.

Objection 11: We can use the Gödel incompleteness theorems to show that there are capacities which human minds have that finitary computing machines do not have. Let the formal system characterizing the capacities of a finitary computing machine be P. Suppose P is subject to the Gödel incompleteness theorems. Then the finitary computing machine can't prove CON(P) and can't prove its own Gödel sentence. However, a human mind can prove CON(P) and the Gödel sentence in P by ascending to a more powerful formal system, P^*, that contains P. The finitary computing machine characterized by P, however, cannot ascend to P^*.

Response: That point is well-known in the functionalism debate. Perhaps ascent to P^* may prove futile, since P^* may be so long that finitary human minds cannot survey it and thus cannot prove that it is consistent. That is the Penrose error.

However, even if we discount the Penrose error, there is still a problem. Recall that what the second Gödel incompleteness theorem rules out is the possibility of finitistically proving, with mathematical certainty, and within the system P, CON(P). If one ascends to P^*, then CON(P) can be proved finitistically with mathematical certainty, period. However, this is true only if one can finitistically prove, with mathematical certainty, that P^* is consistent. But now the Gödel theorems take root in P^*. It is impossible to finitistically prove CON(P^*) with mathematical certainty, within P^*. That means that the ascent to P^* is futile unless P^* can be proved consistent. But that cannot be done within P^*. It can only be done by ascending to a stronger system P^{**} that contains both P and P^*. Within P^{**}, one can finitistically prove CON(P) and CON(P^*) with mathematical certainty, but only if P^{**} is consistent.

Notice the epistemic pattern which emerges. For any n less than omega, one can finitistically prove with mathematical certainty CON(P^n) in the formal system P^{n+1} only if one can finitistically prove, with mathematical certainty, CON(P^{n+1}). However, for any n less than omega, it is impossible to finitistically prove CON(P^n) with mathematical certainty within P^n.

The anti-functionalist will have to ascend infinitely high to the infinitary formal system Pomega, in order to finitistically prove, with mathematical certainty, CON(P). That is just to say that the anti-mechanist will have to possess the cognitive capacity to construct an infinite proof tree in order to finitistically prove, with mathematical certainty, CON(*P*). Indeed, this is true for any P^n, where *n* is less than omega.

It easily follows from these considerations that the anti-functionalist has no advantage over the functionalist in showing that there are cognitive capacities which finitary human minds possess, but which a finitary computing machine lacks. If human minds possess an infinitary cognitive capacity, there is something we possess that finitary computing machines lack. But there is no conclusive evidence that we possess an infinitary cognitive capacity. It is open to us to prove CON(*P*) with less than mathematical certainty or in another epistemic modality, but it is open to finitary computing machines to do the same as well. If the methods for proving CON(*P*) with less than mathematical certainty or in some other epistemic modality are subject to the Gödel incompleteness theorems, then the very same considerations expressed above will apply to this case also. In which case, the anti-functionalist has no advantage over the functionalist in demonstrating there is a cognitive capacity which human minds possesses that finitary computing machines lack.

Objection 12: The anti-functionalist using an MGF argument has an avenue of escape. Although there cannot be a finitistic proof within *P* that establishes, with mathematical certainty, CON(*P*), it is possible for a human mind (not susceptible to the Gödel incompleteness theorems) to prove CON(*P*), with mathematical certainty, by using mathematical reasoning that is not subject to the Gödel incompleteness theorems.

Response: That is a good objection, but it might not work. If the mathematical reasoning in question is captured by a formal system that is not subject to the Gödel incompleteness theorems, it might be too weak to finitistically prove CON(*P*) with mathematical certainty. Perhaps CON(*P*) could be finitistically proved with mathematical certainty in the ramified type theory of *Principia Mathematica*. But since there is no adequate theory of its intensional proof predicate (which is why it is not subject to the Gödel incompleteness theorems), it is not known whether such a proof will have mathematical certainty.

On the other hand, if there is a system of mathematical reasoning which is not subject to the Gödel incompleteness theorems only because it cannot be formalized (justified perhaps on philosophical grounds), such as Brouwer's view of intuitionism, it is not known whether such reasoning can establish its conclusions with mathematical certainty and it is not known whether such reasoning is (or is not) finitary.

There are systems of mathematical reasoning that are captured only by infinitary formal systems (such as the system in Turing's completeness theorem), that are not subject to the Gödel incompleteness theorems. But there is no conclusive evidence human agents can engage in infinitary reasoning, where proper infinitary reasoning implies the ability of the reasoner to construct infinitary proof trees. This will not help the anti-functionalist who uses an MGF argument.

The moral, then, is that the anti-functionalist can dream of a system of finitary mathematical reasoning which can finitistically prove CON(P) with mathematical certainty, and which is not subject to the Gödel incompleteness theorems But we have no reason to believe such a system of mathematical reasoning exists, nor that it is logically possible.

10. The Epistemology of Mathematical Certainty: A New Project for Philosophy of Mind

Proving that an arbitrary mathematical sentence is true is beyond the pale of a mechanical proof procedure, since the set of mathematical truths is not recursive, not recursively enumerable, and not definable in arithmetic. This is another reason why mechanical proof procedures that verify a proof of a theorem in mathematics must be mechanical. If we attempt to show that each line in a proof preserves truth by showing that each line in the proof is true in and of itself and without examining how it was obtained, there is no guarantee we will be able to complete the job of verifying the proof of the theorem (even if we have the time and resources). On the other hand, if the proof verification procedure is mechanical, then we do not check that each line of the proof is true. Rather, we check that it has the requisite syntactical form. The relation "p is a proof of α" is recursive, where "α" is a sentence in some language and "p" is a proof of that sentence. It follows that all of the theorems in that language are recursively enumerable. There is a fundamental dichotomy between proof and truth arising from these considerations. Mathematical truth is not recursively enumerable, while mathematical provability is recursively enumerable. One way of describing the Gödelian incompleteness phenomena is that they witness this dichotomy.

If we relax the standards of mathematical proof, we might not have assurance that intersubjective agreement can be reached as to whether a derivation is a legitimate proof of its conclusion. In which case, we cannot be assured we will be mathematically certain of the truth of the theorem derived. It is the epistemological requirement in mathematics that a proof establish with mathematical certainty the truth of its conclusion that allows the anti-functionalist to capitalize on the Gödel incompleteness theorems in EGF and MGF arguments. Relaxing this requirement in mathematics is relevant to the philosophy of mind. We must ask: what is the epistemic goodness of weak mathematical methods—those which do not confer mathematical certainty on what they establish?

An area in philosophy of mathematics that connects with philosophy of mind is mathematical intuitionism. Can intuitionistic reasoning as originally envisaged by Brouwer deliver mathematical certainty? Is it infinitistic? If so, does it have a metarecursive computational structure? Work needs to be done to explore Kreisel's musing:

> There is the old and familiar idea, or: idealization, which regards a t h o u g h t and, in particular, a p r o o f of a general proposition as an infinite object. [I]nfinite ob-

> jects are better *representations* of proofs than the words we use to communicate proofs… (1967, p. 203)

and Brouwer:

> These *mental* mathematical proofs that in general contain infinitely many terms must not be confused with their linguistic accompaniments, which are finite and necessarily inadequate, hence do not belong to mathematics. (1967, p. 460, note 8)

A virtue of Lucas-Penrose-Putnam anti-functionalist arguments is that they connect mathematical logic with the philosophy of mind and might cast light on issues in the foundations of mathematics.

Note 1: An anonymous reviewer of this paper made several important remarks: that Kreisel (1965; 1967) and Gödel (in his *Dialectica* paper; see Gödel, 1990) perhaps hold the view that human minds are capable of infinitary mental proofs, that Gödel (1995) perhaps believes mathematics is empirical (and so statistical methods would be an appropriate means of proving theorems), and that there is an interesting problem in Kripke's Schema (formalizing Brouwer's creating subject)—namely, the assumptions in his argument using the schema are incompatible with infinitary mental proofs. Van Atten (2018) provides an excellent discussion of this matter. If human minds are capable of infinitary mental proofs, the question of whether such mental acts have a metarecursive computational structure is raised and with it, whether such a computational structure can be accommodated within functionalism. I thank the anonymous reviewer for these remarks and other useful suggestions.

Note 2: This paper revises and expands two earlier versions (Buechner, 2007; 2010). The most prominent changes are the nature of the problem that I contend arises for Putnam's use of the Gödel incompleteness theorems to refute functionalism and the nature of the problem that arises for functionalists whose burden of proof is to show there are no ways (that avoid the incompleteness theorems) of establishing the consistency of first-order arithmetic with less than mathematical certainty or in some other epistemic modality than that of mathematical certainty. The most significant overlap is in the categorization of the Lucas-Penrose-Putnam anti-functionalist arguments. Although there are changes of emphasis in that categorization in this paper, I still believe it is a significant contribution to the role of the Gödel incompleteness theorems in the functionalism debate.

REFERENCES

Audi, R. (1988). *Belief, Justification, and Knowledge*. Belmont, CA: Wadsworth Publishing Company.

Boolos, G. (2001). Introductory Note to 'Some Basic Theorems on the Foundations of Mathematics and Their Implications'. In: S. Feferman et al. (Eds.), *Collected Works, Volume 3: Unpublished Essays and Lectures* (pp. 290–307). New York: Oxford University Press.

Brouwer, L. E. J. (1967). On the Domains of Definition of Functions. In J. van Heijenoort (Ed.), *From Frege to Gödel* (pp. 446–463). Cambridge, Mass.: Harvard University Press.

Buechner, J. (2007). *Gödel, Putnam, and Functionalism: A New Reading of 'Representation and Reality'*. Cambridge: MIT Press

Buechner, J. (2010). Are the Gödel Incompleteness Theorems Limitative Results for the Neurosciences? *J Biol Phys*, *36*, 23–44.

Buechner, J., Tavani, H. (2011). Trust and Multi-Agent Systems: Applying the "Diffuse, Default Model" of Trust to Experiments Involving Artificial Agents. *Ethics and Information Technology*, *13*, 39–51.

Carnap, R. (1952). *The Continuum of Inductive Methods*. Chicago: University of Chicago Press.

Carnap, R. (1962). *Logical Foundations of Probability* (2nd ed.). Chicago: University of Chicago Press.

Church, A. (1956). *Introduction to Mathematical Logic* (vol. 1). Princeton: Princeton University Press.

Feferman, S. (1960). Axiomatization of Metamathematics in a General Setting. *Fundamenta Mathematicae*, *49*, 35–92.

Gödel, K. (1990). On a Hitherto Unutilized Extension of the Finitary Standpoint. In: S. Feferman et al. (Eds.), *Collected Works, Volume II: Publications 1938–1974* (pp. 241–252). New York: Oxford University Press.

Gödel, K. (1995). Some Basic Theorems on the Foundations of Mathematics and Their Implications. In: S. Feferman et al. (Eds.), *Collected Works, Volume III: Unpublished Essays and Lectures* (pp. 304–323). New York: Oxford University Press.

Kaplan, D., Montague, R. (1960). A Paradox Regained? *Notre Dame Journal of Formal Logic*, *1*, 79–90.

Kreisel, G., Sacks, G. (1965). Metarecursive Sets. *Journal of Symbolic Logic*, *30*, 318–338.

Kreisel, G. (1965). Mathematical Logic. In: T. L. Saaty (Ed.), *Lectures on Modern Mathematics* (vol. 3, pp. 95–195). New York: Wiley.

Kreisel, G. (1967). Mathematical Logic: What Has It Done for the Philosophy of Mathematics? In: R. Schoenmann (Ed.), *Bertrand Russell: Philosopher of the Century* (pp. 201–272). London: George Allen & Unwin.

Kreisel, G. (1970). Church's Thesis: A Kind of Reducibility Axiom for Constructive Mathematics. In: A. Kino, J. Myhill, R. Vesley, (Eds.), *Intuitionism and Proof Theory* (pp. 121–150). Amsterdam: North-Holland.

Kreisel, G. (1972). Which Number Theoretic Problems Can Be Solved in Recursive Progressions on Π(1,1) Paths Through O? *Journal of Symbolic Logic*, *37*, 311–334.

Lehrer, K. (1990). *Theory of Knowledge*. Boulder, Colorado: Westview Publishing Company.

Lewis, D. (1969). Lucas Against Mechanism. *Philosophy*, *44*, 231–233.
Lewis, D. (1979). Lucas Against Mechanism II. *Canadian Journal of Philosophy*, *4*, 373–376.
Lucas, J. (1961). Minds, Machines and Gödel. *Philosophy*, *36*, 112–127.
Lucas, J. (1970). Mechanism: A Rejoinder. *Philosophy*, *45*, 149–151.
Marr, D. (2010). *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. Cambridge, MA: MIT Press.
Montague, R. (1963). Syntactical Treatments of Modality, with Corollaries on Reflection Principles and Finite Axiomatizabiity. *Acta Philosophica Fennica*, *16*, 153–167.
Penrose, R. (1999). *The Emperor's New Mind, with a New Preface by the Author*. New York: Oxford University Press.
Penrose, R. (1994). *Shadows of the Mind: A Search for the Missing Science of Consciousness*. New York: Oxford University Press.
Penrose, R. (2016). On Attempting to Model the Mathematical Mind. In: S. B. Cooper (Ed.), *The Once and Future Turing* (pp. 361–378). New York: Cambridge University Press.
Pollock, J. Cruz, J. (1999). *Contemporary Theories of Knowledge* (2nd edition). Lanham, Maryland: Rowman and Littlefield.
Putnam, H. (1988). *Representation and Reality*. Cambridge, MA: MIT Press.
Putnam, H. (1994a). *Words and Life*. Cambridge, MA: Harvard University Press.
Putnam, H. (1994b). Reflexive Reflections. In: H. Putnam, *Words and Life* (pp. 416–427). Cambridge, MA: Harvard University Press.
Putnam, H. (1995). Review of Roger Penrose's *Shadows of the Mind*. *Bulletin of the American Mathematical Society*, *32*, 370–373.
Salmon, N. (2001). The Limits of Human Mathematics. *Philosophical Perspectives*, *15*, 93–117.
Tarski, A. (1983). The Concept of Truth in Formalized Languages. In: A. Tarski, *Logic, Semantics, Metamathematics* (2nd edition, pp. 152–278). Indianapolis, Indiana: Hackett Publishing Company.
Thomason, R. (1980). A Note on Syntactical Treatments of Modality. *Synthese*, *44*, 391–395.
Thomason, R. (1989). Motivating Ramified Type Theory. In G. Chierchia, B. H. Partee, R. Turner (Eds.), *Properties, Types and Meaning, Volume 1: Foundational Issues* (pp. 47–62). Norwell, MA: Kluwer.
Turing, A. (1939). Systems of Logic Based on Ordinals. *Proceedings of the London Mathematical Society*, *45*, 161–228.
van Atten, M. (2018). The Creating Subject, the Brouwer-Kripke Schema, and Infinite Proofs. *Indagationes Mathematicae*, 29, 1565–1636.
von Neumann, J. (1969). Tribute to Dr. Gödel. In: J. J. Bulloff, T. C. Holyoke, S. W. Hahn (Eds.), *Foundations of Mathematics: Symposium Papers Commemorating the Sixtieth Birthday of Kurt Gödel* (pp. ix–x). Berlin: Springer-Verlag.
Wigderson, A. (2019). *Mathematics + Computation*. Princeton, NJ: Princeton University Press.