Pázmány Papers Vol. 3, Nr. 1 (2025)
ISSN 3004-1279
https://doi.org/10.69706/PP.2025.3.1.1

The Ethics of Logic

Gödel's Dichotomy and the Ethical Imperative

János V. Barcsák[1]

Abstract

The logic of ethical reasoning has been analysed in many different contexts in analytic philosophy. But is there such a thing as the ethics of logic? In other words, does logic, or the use of formal systems lead to certain commitments that can be considered ethical? In this paper I will explore these questions in the context of Kurt Gödel's "Gibbs Lecture" delivered at Brown University in 1951. In this address to the American Mathematical Society Gödel assesses the philosophical consequences of his incompleteness theorems in terms of what Solomon Feferman has called "Gödel's dichotomy," according to which either "*the human mind (even within the realm of pure mathematics) infinitely surpasses the powers of any finite machine, or else there exist absolutely unsolvable Diophantine problems*" [that is, basic arithmetical problems]. In my paper I will argue that both these options are thoroughly problematic in their epistemological implications. Gödel's discussion, however, leaves a third option open, as well. For he concedes that the mind (human reasoning) *can* be represented by a finite machine (that is, by a well-defined formal logical system) which does not understand its own functioning and does not know its own consistency. Although Gödel does not consider this to be a genuine third option, I will argue that this conception is perhaps the most fruitful, or least problematic, model of how human reason can contain knowledge. As such, however, this approach requires certain commitments which can best be described as ethical. In particular, it calls for (1) a commitment to the consistency of human reasoning and (2) a commitment to truth, as the truth of the undecidable proposition pertaining to the consistent system of human reasoning. I will argue that these ethical commitments are inevitable once we deploy formal-logical systems to produce knowledge about reality. To this extent, therefore, these commitments constitute the ethics of logic.

Keywords

Gödel, formal systems, Turing-machine, the mind, ethical commitment, consistency, undecidability

[1] Pázmány Péter Catholic University, barcsak.janos@btk.ppke.hu

In logic, there are no morals.
(Carnap 1937, 52)[2]

1. Introduction

The phrase "ethics of logic" seems to be a contradiction in terms. Logic is, after all, the use of formal deductive systems wherein we move from formalized premisses to conclusions, from axioms to theorems, in a fully controlled, mechanical way. There is nothing ambiguous in the process, there can be no two ways of going about it. How, then, can such a conception of logic lead to questions of right or wrong, of ought or ought not? How can it lead to questions of ethics? My contention in this paper is that, although logic in itself *can* indeed be thought of as a mere formal game, or, as Kurt Gödel (1995c, 319) puts it in a critical reference to Rudolf Carnap's conception of mathematics, as an "idle running of language" from premisses to conclusions, once we *use* logic to describe or make reference to some reality distinct from it (be it mathematical or empirical reality), the *purely logical* elements of the logical system – pace Rudolf Carnap – will make certain ethical commitments inevitable.

The ethical commitments that I will eventually arrive at are not new. Basically, I will show that the use of logic leads to a double commitment to immanent consistency and to truth (not in the sense of some essence but rather as the truth of the other). It is a double commitment, therefore, that is strongly reminiscent of the double commitment to a logic of "all or nothing" (Derrida 1988, 122) and to "quasi-transcendence" (Derrida 1988, 152) that Jacques Derrida advocates in his "Afterword" to *Limited Inc*, or of the double commitment that Paul de Man's analysis of the *Social Contract* reveals (de Man 1979), or of that which characterizes Alain Badiou's truth procedures (Badiou 2005), or of that which relating to the other requires in Derek Attridge's analysis (Attridge 1999, 2004). What I hope to achieve in this paper is to arrive at these familiar conclusions in a novel way, a way that might perhaps have the merit of giving a rigorous account of the logical necessity of those familiar conclusions, and that might thus discover a common logical foundation behind all these instances.

I will outline the ethical implications of formal logic in the context of Kurt Gödel's paper, "Some basic theorems on the foundations of mathematics and their implications" (1995c). Gödel read this paper on 26 December 1951 in front of

[2] Original italics.

members of the American Mathematical Society as the twenty-fifth Josiah Willard Gibbs lecture (Boolos 1995, 290), and for this reason it is often referred to simply as the "Gibbs lecture." It survives in manuscript form and was published in Volume III of the *Collected Works of Kurt Gödel* in 1995. It is in this text that Gödel formulated what he called his "disjunctive conclusion," and what Solomon Feferman (2006) has referred to as "Gödel's dichotomy." It is in terms of this dichotomy that I shall try to formulate the ethical imperatives that the use of formal logic requires.

2. The Dichotomy

In the Gibbs lecture Gödel presents his dichotomy as an inevitable consequence of what he refers to as the "basic fact, which might be called the incompletability or inexhaustibility of mathematics" (Gödel 1995c, 305). This "basic fact," as he explains further, is most clearly and most universally established by "certain very general theorems," by which he of course means his incompleteness theorems (Gödel 1995c, 308–309). It is, in particular, the second incompleteness theorem that leads him to the formulation of his "disjunctive conclusion," which he therefore refers to as a "mathematically established fact" (Gödel 1995c, 310).

Regardless of the specific trajectory by which Gödel arrives at his dichotomy, however, I think that he essentially relies on two premisses to formulate it: (1) the inevitable presence of undecidable propositions in formal systems, and (2) the exact delineation of a class of systems that can be considered entirely mechanical ("formal" in this sense). Although in the Gibbs lecture he does not actually name these as the premisses necessary for his "disjunctive conclusion," I think that to understand this conclusion, that is, Gödel's dichotomy, we need to grasp these and only these two conditions. So let us examine them one by one.

In his famous 1931 article "On Formally Undecidable Propositions Of Principia Mathematica And Related Systems," Gödel (1992) developed an ingenious method of producing a proposition in the notation of a formal system (in this case it was Russell and Whitehead's *Principia Mathematica*) that is perfectly meaningful inside the given system, and yet cannot be decided by it. It cannot be decided because deriving either the proposition itself or its negation from the axioms of the system would inevitably lead to formal contradiction. The inevitable presence of such propositions constitutes the first condition of Gödel's dichotomy.

The second condition arises from the fact that Gödel's method of constructing such propositions, as it later became clear,[3] is applicable in the case of any "formal" system of a certain minimal level of complexity;[4] that is, in the case of any system that can be viewed as fully mechanical and thus humanly controllable in its entirety. The class of systems that can be considered "formal" in this sense, that is, fully mechanical and controllable, was very clearly delineated in the course of the twentieth-century development of formal logic. Gödel (1995c, 305) names Alan Turing as the one who provided the most precise formulation of what it means for a system to be fully mechanical, and for the formulation of his dichotomy Gödel considers the class of formal systems that can be considered mechanical in Turing's sense, that is, that can be unambiguously represented by certain abstract entities called Turing-machines (Turing 1936). It is in this sense that I will likewise use the term "mechanical": mechanical is that which can be represented by a Turing-machine.

The second condition Gödel actually needs for his dichotomy arises from the clear definability of this class of mechanical systems: it is the assumption that these systems comprise everything that the human mind can "effectively" calculate; that is, everything that can be mechanized in human thought.[5] Assuming this second condition and employing the first one – that is, that in any mechanical system we can always construct an undecidable proposition – we immediately arrive at Gödel's dichotomy:

> *either … the human mind (even within the realm of pure mathematics) infinitely surpasses the powers of any finite machine, or else there exist absolutely unsolvable Diophantine problems* (Gödel 1995c, 310)[6]

The Diophantine problems Gödel refers to here are essentially problems in arithmetic that can be reduced to the problem of finding solutions to a type of relatively simple equations.[7] The second alternative in his dichotomy is, therefore, that there are relatively simple and unambiguous mathematical problems that are undecidable even in principle.

[3] See on this for example (Zach 2007, 432).

[4] The level of complexity required is that the system should be able to represent basic arithmetic.

[5] This assumption is usually referred to as the Church-Turing thesis and Gödel's dichotomy is in an important sense his special interpretation of this thesis.

[6] Original italics.

[7] In an unpublished paper from the 1930s Gödel proved that undecidable propositions of such simple form existed in all formal systems of the family referred to above (Gödel 1995d). See on this also (Feferman 2006, 139).

3. … and What It Means

Gödel's dichotomy has been thoroughly analysed in the literature, especially by Stewart Shapiro (1998) and Solomon Feferman (2006). Following these presentations and simplifying a little, we can state that what the problem boils down to is whether or not all true mathematical propositions (sufficiently formalized) are provable in a strictly formal, mechanical way. If we designate the set of all true propositions by T, the question is whether this set is coextensive with the set of all provable propositions (K), where "provable" means demonstrable in one or another of the type of mechanically controllable systems mentioned above. The problem is, in other words, whether K = T (Shapiro 1998, 278).

If K = T – that is, if all true propositions are mechanically demonstrable – then it is impossible to capture mathematical demonstrability in a single well-defined formal system, since in any such system there will be undecidable propositions, which – assuming bivalence – are either themselves true or their negation is true. There will, therefore, be at least one truth that the given system cannot prove.[8] Thus, if all mathematical truths are mechanically provable, then there is no one well-defined mechanical system that can capture all of mathematical provability. In other words, the human mind – in its ability to prove mathematical propositions – is different from, and indeed exceeds the capabilities of, any machine.

If, on the other hand, the set of provable propositions does not coincide with the set of truths (K ≠ T), then – since all provable propositions are obviously true – there will have to be propositions that are true, but not provable. What is more, these propositions will be "*absolutely unsolvable*," since by definition they do not belong to the set of humanly provable propositions (K).

4. Can All Rationally Posed Problems Be Rationally Solved?

To bring the issue a little closer to the philosophical problems that it implies, we could approach it from the perspective of the philosophical conviction that apparently motivated Gödel to formulate his dichotomy. Both Shapiro (1998, 278–279; 290) and Feferman (2006, 145) as well as Boolos (1995, 294) note that in the background of this dichotomy is Gödel's belief that all rationally formulated problems must be rationally solvable. If this were not so, then – as he put it to Hao Wang (1974, 324) – "it would mean that human reason is utterly irrational by asking questions it cannot

[8] Gödel (1995c, 309) actually uses the statement of the given system's consistency (Con_s) to establish this.

answer, while asserting emphatically that only reason can answer them." We can therefore see Gödel's dichotomy in the context of the complications that arise from the question "Can all rationally posed problems be rationally solved?"

At first sight we would expect that the answer to this question must be yes. If, for example, we pose the simple arithmetical question "Does 2 +2 = 4?," then we expect that we can answer it within arithmetic. This expectation is only further strengthened with the introduction of formal systems. Such systems, after all, were developed precisely to provide an exact sense of what we mean by a *rationally* posed question; that is, to make such questions perfectly unambiguous and thus to make sure that they can just as unambiguously be resolved. A formalized system of arithmetic, for example, serves the purpose of assuring that all arithmetical problems can be clearly formulated and just as clearly solved. What is more, as Turing showed, all this can be represented as a purely mechanical operation where no contingent irrational influences can interfere with the process of reasoning, and where we can exert control without limitation.

Gödel's proof of the inevitability of undecidable propositions, however, complicates this situation, for it applies to all mechanical systems, including those formalizing arithmetic. He showed that there are arithmetic propositions as unambiguous as 2 +2 = 4 – propositions that are even reducible to the Diophantine problems mentioned above – which cannot be decided in any formalized, mechanical system of arithmetic on pain of contradiction.

It turns out, therefore, that the seemingly innocent question "Can all rationally posed problems be rationally solved?" proves to be a rather complicated one, and one that gives rise to a dichotomy. For if the answer to this question is yes – which is the first alternative in Gödel's dichotomy – then we must conclude that the rationality which decides all rationally posed questions (that is, the human mind) cannot be captured in a specific formal system, since any such system will have propositions that are perfectly meaningful and rational in terms of the system itself, yet cannot be decided within the given system. If, therefore, we maintain that all rationally posed questions can be answered rationally, then we also claim that rationality cannot be captured in any formal logical system: "*the human mind infinitely surpasses the powers of any finite machine*," or, to simplify further, the mind is not a machine.

One might, however, insist that there is strong empirical evidence to suggest that the mind is a Turing-machine. The brain, after all, works precisely like a digital

computer: neurons either fire or do not fire, there is no third option.[9] This view, however, leads directly to the second disjunct in Gödel's dichotomy, for it implies that there must be rationally posed questions that are *absolutely* unsolvable. These rationally posed questions, as Gödel showed, can in fact be very simple arithmetical questions, so simple that they can be reduced to the form of Diophantine problems. And yet they will, on this interpretation, be *absolutely* undecidable; that is, undecidable not only within the logical system in which they are formulated, but in any humanly devisable logical system. This is so because if the mind is indeed a machine in Turing's sense, then all that the mind can prove is captured by a given formal logical system. But, as Gödel proved, in all such systems there will be undecidable propositions. We will therefore always be able to ask the perfectly rational question of whether these propositions (or their negations) are solvable in the system that constitutes the mind, but this question will be impossible to answer by the mind, even in principle.

5. The First Disjunct and Gödelian Arguments

In the Gibbs lecture, Gödel is cautious not to take sides on the question of which of the two disjuncts is likelier. He merely states his dichotomy as a "mathematically established fact" (Gödel 1995c, 310). He is also careful to point out that "the case that both terms of the disjunction are true is not excluded, so that there are, strictly speaking, three alternatives" (1995c, 310). This concession, however, merely sharpens the dichotomous nature of his disjunctive conclusion, for if it can be true both that the mind is not a machine and that there are absolutely undecidable questions, then it follows that (1) even if there are absolutely undecidable questions, the mind does not necessarily have to be a machine, and (2) even if the mind is not a machine, there can be absolutely undecidable questions. What Gödel's dichotomy boils down to is, therefore, that (1) *if* there are no absolutely undecidable questions, then the mind is certainly not a machine, and (2) *if* the mind *is* a machine, then there must be absolutely undecidable questions.

[9] This of course is a much subtler issue than my presentation here suggests. It involves the philosophical problems pertaining to Turing's thesis (also known as the Church-Turing thesis), which maintains that all "effective" (that is, in the strictest sense rational) operations of the mind are captured by Turing-machines (see note 5 above). Gödel's troubled relation to this thesis (he endorsed it, while rejecting the conclusion that there is no rationality beyond what is captured by Turing machines) is discussed at length by Judson C. Webb in his note to Remark 3 of Gödel's "Some remarks on the undecidability results" (Feferman, Solovay and Webb 1990, 292–304; Gödel 1990). In the Gibbs lecture Gödel (1995c, 309n13) himself also considers the remote but conceivable possibility "that brain physiology would advance so far that it would be known with empirical certainty" that the brain is a machine in Turing's sense.

In spite of Gödel's cautious formulation, however, the reasoning that leads to his dichotomy can of course be used to make philosophical claims. We have seen, for example, that in his private communications Gödel rejected the option that there are rationally posed questions that are rationally unsolvable. In light of this, the argument that leads to the first disjunct could be used as a general argument for the non-mechanizability of the mind.

This is in fact the direction the most famous so-called "Gödelian arguments" adopt. The most influential (or at least best known) of these are the ones put forward by John R. Lucas (1961, 1996) and Roger Penrose (1989, 1994), both of whom use the incompleteness theorems to establish that the mind is not a machine, thus endorsing the first disjunct in Gödel's dichotomy – without, however, taking into consideration the second alternative. To be more precise, both Lucas and Penrose use the inevitable presence of undecidable propositions to prove that the human mind cannot be captured in a specific formal system or Turing-machine, because as soon as such a system or machine is specified, the mind will always be able to construct – using Gödel's method – an undecidable proposition inside that system and see that it is true, whereas the given system will be unable either to prove or to disprove this proposition. In this way, as Lucas (1961, 116) puts it, the mind "can always go one better than any formal, ossified, dead, system can. Thanks to Gödel's theorem, the mind always has the last word."

Shapiro (1998, 282–283) summarizes the standard criticism of this "Gödelian argument" (referencing Hilary Putnam (1960)) as follows:

> neither Lucas nor anyone else knows that G_S [the undecidable sentence of the system constructed by Gödel's method] is true. He only knows that *if* S [the formal system proposed as the model of the mind] *is consistent* then G_S is true. But the machine (or formal system) "knows" this conditional proposition as well, since
>
> $$\mathrm{Con}_S \rightarrow G_S$$
>
> is a theorem of S (as seen by the proof of the second incompleteness theorem). Lucas can claim to know G_S outright only if he can claim to know Con_S. But how does he establish this last premise?[10]

[10] Original italics. See on this also Boolos (1995, 294–295).

For the Gödelian arguments to succeed, therefore, the consistency of the specific formal system proposed as the one comprising all of rationality (that is, the mind) must be established. Because of Gödel's second incompleteness theorem, however, this is impossible. To be precise, it is impossible to establish the given system's consistency *with mathematical certainty*,[11] and for this reason ultimately all "Gödelian arguments" go by the board.

This of course is a very schematic and superficial representation of the carefully laid out philosophical arguments on both sides, and I will not be able to do justice either to the arguments of Lucas and Penrose or to those of their opponents in this article.[12] One can, however, hardly doubt the validity of George Boolos's verdict on these Gödelian arguments. As he puts it, "It is fair to say that the arguments of these writers have as yet obtained little credence" (Boolos 1995, 295). Or to quote Hilary Putnam (2011, 332) on the same issue, "[t]hat Lucas and Penrose have failed to prove their claims about what the Gödel theorem 'shows about the human mind' is widely recognized."

Lucas and Penrose of course published their original Gödelian arguments long before the first publication of the Gibbs lecture in 1995, so they could not rely on Gödel's disjunctive conclusion. We might observe in passing, however, that Gödel's formulation of his dichotomy could not add much to the force of these arguments either.[13] What the first disjunct in his "mathematically established fact" ultimately amounts to is just that *if* we maintain that all rationally posed questions are rationally solvable, then we also inevitably claim that rationality in its entirety (that is, "the mind") cannot be captured in a single well-defined system or Turing-machine. It is unclear, however, how we could establish the antecedent in this conditional proposition. As George Boolos (1995, 294) points out, we might wonder why there should not be absolutely unsolvable questions even in the narrow field of arithmetic.[14] That there are questions which are practically unsolvable has been shown by Solomon Feferman and Robert Solovay (1990, 292),[15] and, as Boolos (1995, 294) observes,

[11] On the significance of *mathematical certainty* in Lucas's Gödelian argument see (Shapiro 1998, 281).

[12] For more recent versions of the arguments on either side see (Putnam 2011; Penrose 2011).

[13] This is probably why Gödel is cautious not to argue for the priority of either disjunct even if his private conviction was that the mind is not a machine. Boolos (1995, 294) even observes that "Gödel's disjunctive conclusion concerning the significance of his incompleteness theorems stands in contrast with the conclusion drawn by writers such as Ernest Nagel and James R. Newman (1958), J.R. Lucas (1961), and Roger Penrose (1989) …"

[14] Cf. on this also Feferman (2006, 147).

[15] Cf. also Feferman (2006, 149).

> there are many persons who, influenced by the picture of the mind as a Turing machine, find the falsity of the first and the truth of the second alternative [in Gödel's dichotomy] a pair of propositions they are quite willing to maintain. Others, while reserving judgment on the question whether (the mathematical abilities of) a mind can be (represented by) a Turing machine, simply find it extremely plausible that there are mathematical truths unprovable by any humanly comprehensible proof.

To this view we could of course respond – recalling Gödel's concession that both disjuncts may be true at the same time – that the existence of such questions does not prove the mechanizability of the mind, but this in itself is obviously no proof of the first disjunct either. It is no wonder, therefore, that Gödel himself does not use this line of argument to prove his convictions – expressed in private communications – that the mind is not a machine and that there are no rationally posed questions that are absolutely undecidable.

6. The Second Disjunct and Gödel's Platonism

While in the Gibbs lecture Gödel himself does not claim that his dichotomy can directly lead to any philosophical conclusions, he does consider the possible philosophical consequences of this "mathematically established fact." He summarizes these as follows:

> Corresponding to the disjunctive form of the main theorem about the incompletability of mathematics, the philosophical implications *prima facie* will be disjunctive too; however, under either alternative they are very decidedly opposed to materialistic philosophy. Namely, if the first alternative holds, this seems to imply that the working of the human mind cannot be reduced to the working of the brain, which to all appearances is a finite machine with a finite number of parts, namely, the neurons and their connections. So apparently one is driven to take some vitalistic viewpoint. On the other hand, the second alternative, where there exist absolutely undecidable mathematical propositions, seems to disprove the view that mathematics is only our own creation; for the creator necessarily knows all properties of his creatures, because they can't have any others except those he has given to them. So this alternative seems to imply that mathematical objects and facts (or at least *something* in them) exist objectively and independently of our mental acts and decisions, that is to say, [it seems to

> imply] some form or other of Platonism or "realism" as to the mathematical objects. (Gödel 1995c, 311–312)

While Gödel stresses that both disjuncts in his dichotomy question the validity of "materialistic philosophy," it becomes clear from the remaining part of his discussion that it is in fact the consequences of the second disjunct – that is, mathematical Platonism – that Gödel is most eager to establish. He uses the second disjunct to set this theme and devotes the remaining part of the Gibbs lecture to the defence of the Platonist interpretation of mathematics, which he characterizes at the end of the lecture as

> the view that mathematics describes a non-sensual reality, which exists independently both of the acts and [of] the dispositions of the human mind and is only perceived, and probably perceived very incompletely, by the human mind. (Gödel 1995c, 323)

Whether the second disjunct can be used to support this view is highly questionable. For first of all the antecedent in the conditional form of the dichotomy (that is, that the mind is a machine) is apparently impossible to establish.[16] Secondly, as Boolos (1995, 298) points out, even if we accept that mathematics is not entirely our own creation, it is not at all clear what we mean by the "objective existence" and the "independence" of mathematical objects. He concludes, therefore, that "it may be argued that we lack an interpretation of the key terms in this putative consequence [that mathematics is not our own creation but has objective content] under which it is true but not trivially true." (Boolos 1995, 298)

Pursuing these considerations, however, would land us in the deep water of Gödel's Platonism – a vast topic which has been thoroughly analysed and criticized in the literature, but which I will not attempt to engage with in the present article. For an evaluation of the philosophical relevance of Gödel's dichotomy, however, it is essential to note that Gödel fundamentally thinks of mathematics as being comprised of formalizable, completely mechanical axiomatic systems which describe or refer to an objectively existing (conceptual) reality. His Platonism is the view that the connection between these formalised systems and their referent is an essential

[16] In his very thorough study of the relevance of the incompleteness theorems to philosophical claims about the non-mechanizability of the mind, Stewart Shapiro (1998, 275) concludes that even the idea of the mind being a machine is too vague to provide a ground for such claims. As he puts it, "*there is no plausible mechanist thesis on offer that is sufficiently precise to be undermined by the incompleteness theorems*."

one. Meaning or reference is not arbitrarily assigned as a result of conventions,[17] but is there from the start, the axioms of the formal systems of mathematics being "correct mathematical propositions, and moreover, evident without proof" (Gödel, *1951 1995, 305). What this means is that for Gödel mathematics has objective content. The formalized propositions of arithmetic, for instance, are true or false by virtue of whether or not they correspond to mathematical objects and their relations that, according to Gödel, exist independently of the human mind. Gödel, in other words, insists that mathematics is ultimately the description of an objectively existing (conceptual) reality.

7. Philosophical Consequences: The Reference of Formal Systems

If we accept this view, however, then the two disjuncts in Gödel's dichotomy will both highlight serious epistemological problems; problems that can be seen as generally applying to all instances when a formal system is used to describe or refer to some independently existing reality.

The first one, according to which the mind is not a machine, is essentially Lucas and Penrose's position. As we have seen, this view has received many just criticisms. In my opinion, however, the main problem with this approach from a philosophical point of view is not so much what the standard criticisms point out, but rather that it inevitably leads to a loss of objectivity in knowledge, insofar as knowledge can be captured in formalized logical systems. Gödel's first disjunct and the Gödelian arguments maintain that human reason, the mind, is superior to any formal-logical system in that it can recognize a truth that the system itself cannot capture; namely, it can recognize the consistency of the system, or equivalently, the truth of the undecidable proposition, while – as Gödel showed – the system itself is incapable of proving either. But if there is indeed a truth pertaining to the system in the sense that it is true only *of* this system and is formulated solely *in terms of* this system, and if this objective truth can still only be recognized outside the system and not internally, then this questions the objectivity of the truth of the other theorems of the system. The theorems are those propositions that can be derived from self-evident truths (the axioms) through evidently truth-preserving operations (the rules) – this is after all what formal systems are all about. If we accept that the undecidable proposition or the proposition stating the consistency of the system can be recognized by the

[17] Much of the second half of the Gibbs lecture – as well as the six versions of a long unpublished paper called "Is mathematics syntax of language?" – is devoted to critiquing Rudolf Carnap's conventionalist account of mathematics.

mind as true, although it cannot be derived in the system, then this means that the mind intuits or recognizes more than what is self-evident in the axioms of the given system (since everything that can be mechanically preserved of the objective content of the axioms is contained in the theorems of the system). And why do we come to intuit or recognize this? Only because of the immanent necessities of our formal-logical system, only because – as Gödel showed – we can always mechanically produce an undecidable proposition in every mechanical system.

Can this, however, lead to an objective truth in Gödel's sense; that is, in the sense of a syntactic formula directly containing an objectively existing mathematical reality? If it could, it would mean that an objective fact would depend for its existence on nothing except that which belongs to the completely mechanical, fully controllable part of a formal logical system, that is, on nothing else but "the acts and dispositions of the human mind." Gödel insists that the axioms of what he terms "mathematics proper" are *evident* truths in the sense that they directly contain objective reality, immediately reflecting objective states of affairs.[18] Similarly, the rules that are used to manipulate the axioms and derive theorems from them are also *evidently* truth-preserving. But if there is a proposition that is true in the same sense as the evident axioms but that must be true regardless of how reality is, merely because of the syntactic construction of the formal system, that is, merely because of "our mental acts and decisions," then this would certainly compromise Gödel's criterion for the objective existence of the content of mathematics, which – he insists – exists "independently both of the acts and [of] the dispositions of the human mind." This in turn would question the ground on which the axioms can be considered *evidently* true,[19] and ultimately, it would undermine the very possibility of capturing something objective in the theorems of a formal-logical system. It would undermine the very possibility of knowledge – or at least the potential of formal-logical systems to contain knowledge.

The second alternative in Gödel's dichotomy, the one in which the mind is equivalent to a formal-logical system and there are absolutely undecidable

[18] At the beginning of the Gibbs lecture Gödel (1995c, 305) makes a distinction between "hypothetico-deductive" systems "such as geometry (where the mathematician can assert only the conditional truth of the theorems)" and "mathematics proper," "the body of those mathematical propositions which hold in an absolute sense, without any further hypothesis." For Gödel, therefore, "mathematics proper" is distinguished by being built on evidently true axioms, axioms that directly contain reality.

[19] The drift of my argument here is similar to William Tait's main objection to what he terms "superrealism," a position into which – Tait insists – Gödel only occasionally lapses, but which is apparently the position Gödel adopts in the Gibbs lecture. Tait (2005, 98) points out that "the same grounds, whatever they are, upon which a proposition, undecided by our present axioms, is nevertheless really true or really false would seem to be grounds upon which the axioms themselves are really true or really false," and this questions the *evident* truth of the axioms and renders mathematics speculative (Tait 2005, 91).

propositions, is, as we have seen, generally considered by most commentators to be the more plausible option (Feferman 2006, 147; Boolos 1995, 294). I believe, however, that this alternative, too, is thoroughly problematic from an epistemological perspective. For there is a fundamental inconsistency in the very idea of an *absolutely* undecidable problem: as Gödel pointed out to Hao Wang (1974, 324), "it would mean that human reason is utterly irrational by asking questions it cannot answer, while asserting emphatically that only reason can answer them." The concept of an *absolutely* undecidable problem implies, in other words, that decidability goes beyond human thought, that a proposition formulated in the notation of a particular formal system has a specific meaning and truth value not just beyond the given system but beyond any humanly conceivable conceptual framework. It implies that there is a sense in which a proposition is either true or false outside any humanly confirmable way, outside whatever can be known. An absolutely undecidable proposition is not merely practically unknowable, like whether or not there is life on the planet Mars, or like what happened in the first microsecond after the Big Bang, or like whether or not the value of the digit of the $10^{10^{10^{10}}}$ th place of the decimal expansion of π–3 is equal to 0.[20] It is not even theoretically unknowable, like what happened in the first Planck-time after the Big Bang. It is *absolutely* unknowable, like what happened two minutes *before* the Big Bang. The problem with assuming the existence of such absolutely undecidable propositions is, therefore, that it disregards the context in which the concept of decidability is formulated and pretends "decidability" had an absolute sense beyond any system in which formulas can be decided.

One might object that this problem only occurs if we accept the antecedent in the conditional formulation of the second disjunct – namely, that the mind is a machine – which, as we have seen, is impossible to establish. In the next chapter I will briefly explain why this is still worth maintaining as a hypothesis, even if it is impossible to prove, and I will also outline the conditions on which this hypothesis can be maintained without contradiction. Even if we do not make this assumption, however, the notion of absolutely undecidable questions still remains problematic on a Platonist account such as Gödel's. It remains so because the objective content of such a proposition would have to be a (conceptual) reality that exists beyond any humanly conceivable framework, beyond anything accessible by the human

[20] Feferman suggests this as one of those problems that are "absolutely unsolvable from the standpoint of *practice*." As he explains, "this is an example of a mathematical 'yes/no' question, whose answer can be determined in principle by a mechanical check, but which, in all probability, cannot be settled by the human mind because it is beyond all remotely conceivable computational power on the one hand and there is no conceptual foothold to settle it by a proof on the other" (Feferman 2006, 149–150).

mind. The ultimate problem with assuming the existence of absolutely undecidable propositions is, therefore, that it implies a dogmatic presupposition of objective existence beyond knowledge, beyond anything humanly knowable.[21]

8. The Way Out: Ethics

I think, therefore, that Gödel's dichotomy in fact presents a genuine philosophical impasse. For the first disjunct requires compromising objectivity in knowledge, while the second entails a dogmatic assumption of objective existence beyond human knowability. Fortunately, however, Gödel himself shows the way out of this situation, for he concedes that

> it is not precluded that there should exist a finite rule producing all [the mind's] evident axioms. However, if such a rule exists, we with our human understanding could certainly never know it to be such, that is, we could never know with mathematical certainty that all propositions it produces are correct … If it were so, this would mean that the human mind (…) *is* equivalent to a finite machine that, however, is unable to understand completely its own functioning. (Gödel 1995, 309–310)[22]

Gödel himself does not consider this to provide a genuine third option.[23] In fact, he mentions this possibility before introducing his dichotomy and essentially identifies it with the second disjunct, where the mind is a machine and there are rationally posed questions that are absolutely unsolvable. The reason why this is not a genuine third option for Gödel is of course that he is after "a mathematically established fact." He first wants to obtain this fact – which he does in his "disjunctive conclusion" – and only after that does he proceed to draw philosophical conclusions from it. We have seen, however, that the philosophical inferences from Gödel's dichotomy are at best inconclusive and epistemologically highly problematic.

I think, therefore, that the best use we can make of Gödel's dichotomy (at least in a philosophical context) is to consider it as a reminder of the two major errors we can commit when trying to account for the potential of formal-logical systems to capture knowledge. If we want to maintain that such systems *can* impart knowledge

[21] This view comes close to what Paul Horwich (1982, 186) describes – in the context of the philosophy of science – as the metaphysical realist position, which he criticises precisely for the same reason, namely, because it "involves to an unacceptable, indeed fatal, degree the autonomy of facts."

[22] Original italics.

[23] For discussions of Gödel's concession see (Shapiro 1998, 281–282; Feferman 2006, 145–146).

to us about how the world is, that such systems can actually contain reality, then we must navigate between the Scylla of compromising objectivity in knowledge and the Charybdis of dogmatically assuming objective existence beyond knowledge.

To steer clear of these two philosophical errors, however, we need to take a step back and sacrifice our desire for "mathematically established facts." And the best way to do this is, I think, to consider Gödel's concession that the mind *might be* "equivalent to a finite machine that, however, is unable to understand completely its own functioning" as a genuine third option with which we can avoid the philosophical errors pointed out by his dichotomy. Gödel does not consider this as a third option because he takes it for granted that if the mind is a well-defined system, then we can explicitly know its axioms and rules and thus we can mechanically construct propositions that are undecidable inside the system. This assumption is necessary for him because this is how he can deploy his incompleteness theorems to establish his dichotomy as a "mathematically established fact." For us to gain a genuine third alternative, however, we must waive the requirement of being able to specify exactly the axioms and rules of the system that the mind is. We must, in other words, merely *assume* that the mind is such a well-defined system, without being able to spell out all its axioms and rules explicitly. In fact, once we assume that the mind is such a well-defined system, we also accept that every reasoning, every knowing can only take place within this system, and hence, as a rule, we cannot know all its axioms, or even the mind's consistency. All that we can know is that *if* the mind is consistent, then it will have its own appropriate undecidable proposition.

These are of course rather severe limitations on knowability, yet I believe that they are still worth countenancing, because it is only in this way – that is, by assuming that the mind is a well-defined system that does not know its own operations – that we can avoid the epistemological problems entailed by the two alternatives in Gödel's dichotomy. In particular, this seems to me to be the only way in which we can salvage objectivity in knowledge without dogmatically presupposing an objective reality. This approach, however, can only succeed if we are willing to endorse certain commitments that can best be described as ethical. They are ethical in the sense that they do not derive from how things are in objective reality, but merely from what we *must do* if we want to give a reliable account in a formal logical system of how the world is objectively.

So, what *are* the ethical commitments that we must embrace to account for the possibility of describing reality in formal-logical systems? We must first assume that we are always inside the whole of this system we call the mind. In other words,

we must commit ourselves to a stance of *radical immanence*. This is not the kind of immanence that Quine's seamless "web of beliefs" involves (Quine 1951, 39–42), for Quine's approach is admittedly empirical. The web of beliefs, on his account, can be impacted from the outside by objective reality, which implies that we can have an external, transcendent, holistic view of this web. The problems with this account and its variants are discussed at length by Shapiro (Shapiro 1998, 294–300) and I will not reproduce the whole argument here. What it boils down to is that any such admittance of being dependent on an assumed empirical reality inevitably chips away from the formal, fully mechanical character of the system that we have assumed is identical with the mind. It compromises what is no doubt the most important aspect of formal systems in an attempt to account for knowledge: the guarantee that anything contained in such a system is entirely controllable by the human mind. We must, therefore, embrace a more radical immanence than this, one that resembles Derrida's "There is nothing outside the text" (Derrida 1997, 158), where of course "the text" is replaced by "formal systems."

From within, however, we must also commit ourselves to the *consistency* of the system that the mind is. This is of course a necessary correlate of assuming that the mind is equivalent to a formal system, and yet it must be highlighted because, with the assumption of radical immanence, this is what actually constitutes the fundamental ethical commitment. By Gödel's second incompleteness theorem, we cannot know the system's consistency (once we assume that we are always within this system). Consistency, therefore, cannot be mastered, controlled, or known with absolute certainty. And yet, if we want to maintain that formal systems can contain reality, we *must* commit ourselves to the consistency of the formal system that the mind is.[24]

Secondly, we must commit ourselves to truth. But emphatically not to the truth of any presumed objective state of affairs. Assuming objective existence dogmatically always leads to philosophical error. In particular, it leads to the philosophical error that Derrida dubs "empiricism" and that he analyses in great depth in his reading of Lévinas in "Violence and Metaphysics." To Derrida's incisive analysis I can add in the context of the present argument that one of the main problems with empiricism is that – as we have seen – it compromises any objectivity a well-defined system can ever hope to capture. The commitment to truth must, therefore, never be a commitment to an objective truth that transcends the system in which we formulate it. It can only be a commitment to the truth of the undecidable proposition of the

[24] I have argued elsewhere that consistency is in fact the maximal presupposition that we can maintain without jettisoning radical immanence.

system. It must, in other words, be a commitment to a truth that can never be mastered without giving up consistency, but that *immanently* transcends anything that the system in which we always already are can master.

Committing ourselves to this truth, assuming responsibility for it, and the concomitant fidelity to radical immanence and consistency are, therefore, the ethical imperatives that the use of logic demands. What these commitments dictate is, as I have anticipated, not something wholly new. It is what Derrida (1997, 61) refers to as a "pathway," what Badiou calls a "truth procedure," and what Attridge describes as relating to the other. What I have tried to show is that all these approaches can be seen as arising from the same logical necessity and thus that they all exhibit what we may call the ethics of logic.

References

Attridge, Derek. 1999. "Innovation, Literature, Ethics: Relating to the Other." *PMLA* 114 (1): 20–31. https://doi.org/10.2307/463424

—. 2004. *The Singularity of Literature*. London: Routledge.

Badiou, Alain. 2005. *Being and Event*. Translated by Oliver Feltham. London: Continuum.

Boolos, George. 1995. "Introductory note to *1951." In *Collected Works. Volume III. Unpublished Essays and Lectures*, by Kurt Gödel, edited by Solomon Feferman, John W. Dawson, Warren Goldfarb, Charles Parsons and Robert M. Solovay, 290–304. New York; Oxford: Oxford University Press.

Carnap, Rudolf. 1937. *The Logical Syntax of Language*. Translated by Amethe, Countess von Zepplin Smeaton. London: Routledge and Kegan Paul.

de Man, Paul. 1979. "Promises (Social Contract)." In *Allegories of Reading: Figural Language in Rousseau, Nietzsche, Rilke, and Proust*, by Paul de Man, 246–277. New Haven and London: Yale University Press.

Derrida, Jacques. 1988. "Afterword: Toward an Ethic of Discussion." In *Limited Inc*, by Jacques Derrida, 111–160. Evanston IL: Northwestern University Press.

—. 1997. *Of Grammatology*. Translated by Gayatri Chakravorty Spivak. Baltimore: Johns Hopkins University Press.

Feferman, Solomon, Robert M. Solovay, and Judson C. Webb. 1990. "Introductory Note to 1972a." In *Collected Works. Vol. II. Publications 1938–1974*, by Kurt Gödel, edited by Solomon Feferman, John W. Jr. Dawson, Stephen C. Kleene, Gregory

H. Moore, Robert M. Solovay and Jean van Heijenoort, 281–304. New York; Oxford: Oxford University Press.

Feferman, Solomon. 2006a. "Are There Absolutely Unsolvable Problems? Gödel's Dichotomy." *Philosophia Mathematica* 14 (2): 134–152. https://doi.org/10.1093/philmat/nkj003

—. 2006b. "The Nature and Significance of Gödel's Incompleteness Theorems." *mathematics.stanford.edu*. Princeton Institute for Advanced Study and Gödel Centenary Program. 17 November. Accessed March 17, 2021. https://math.stanford.edu/~feferman/papers/Godel-IAS.pdf

Gödel, Kurt. 1990. "Some Remarks on the Undecidability Results." In *Collected Works. Volume II. Publications 1938–1972*, by Kurt Gödel, edited by Solomon, et al. Feferman, 305–306. New York; Oxford: Oxford University Press.

—. 1992. *On Formally Undecidable Propositions Of Principia Mathematica And Related Systems*. New York: Dover Publications.

—. 1995a. "Is Mathematics Syntax of Language? – III." In *Collected Works. Volume III. Unpublished Essays and Lectures*, by Kurt Gödel, edited by Solomon Feferman, John W. Dawson, Warren Goldfarb, Charles Parsons and Robert M. Solovay, 334–356. New York, Oxford: Oxford University Press.

—. 1995b. "Is Mathematics Syntax of Language? – V." In *Collected Works. Volume III. Unpublished Essays and Lectures*, by Kurt Gödel, edited by Solomon Feferman, John W. Dawson, Warren Goldfarb, Charles Parsons and Robert M. Solovay, 356–362. New York, Oxford: Oxford University Press.

—. 1995c. "Some Basic Theorems on the Foundations of Mathematics and their Implications." In *Collected Works. Volume III. Unpublished Essays and Lectures*, by Kurt Gödel, edited by Solomon Feferman, John W. Dawson, Warren Goldfarb, Charles Parsons and Robert M. Solovay, 304–323. New York, Oxford: Oxford University Press.

—. 1995d. *[Undecidable Diophantine Propositions]*. Vol. III, in *Collected Works. Volume III. Unpublished Essays and Lectures*, by Kurt Gödel, edited by Solomon, et al. Feferman, 164–175. New York; Oxford: Oxford University Press.

Horwich, Paul. 1982. "Three Forms of Realism." *Synthese* 51 (2): 181–201. https://doi.org/10.1007/bf00413827

Lucas, John R. 1961. "Minds, Machines, and Gödel." *Philosophy* 36: 112–127. https://doi.org/10.1017/S0031819100057983

—. 1996. *Minds, Machines, and Gödel: A Retrospect*. Vol. 1, in *Machines and Thought. The Legacy of Alan Turing*, edited by Peter Millican and Andy Clark, 103–124. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198235934.003.0007

Penrose, Roger. 1989. *The Emperor's New Mind*. Oxford: Oxford University Press.

—. 1994. *Shadows of the Mind*. Oxford: Oxford University Press.

—. 2011. "Gödel, the Mind, and the Laws of Physics." In *Kurt Gödel and the Foundations of Mathematics*, edited by Matthias Baaz, Christos H. Papadimitriou, Hilary W. Putnam, Dana S. Scott and Charles L. Jr. Harper, 339–358. Cambridge: Cambridge University Press.

Putnam, Hilary W. 1960. "Minds and Machines." In *Dimensions of Mind: A Symposium*, edited by Sidney Hood, 138–164. New York: New York University Press.

—. 2011. "The Gödel Theorem and Human Nature." In *Kurt Gödel and the Foundations of Mathematics. Horizons of Truth*, edited by Matthias, et al. Baaz, 325–337. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511974236.018

Quine, Willard Van Orman. 1951. "Main Trends in Recent Philosophy: Two Dogmas of Empiricism." *The Philosophical Review* 60 (1): 20–43. https://doi.org/10.2307/2181906

Shapiro, Stewart. 1998. "Incompleteness, Mechanism, and Optimism." *The Bulletin of Symbolic Logic* 4 (3): 273–302. https://doi.org/10.2307/421032

Tait, William. 2005. "Beyond the Axioms: The Question of Objectivity in Mathematics." In *The Provenance of Pure Reason*, by William Tait, 89–104. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780195141924.003.0005

Turing, Alan M. 1936. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* s2–42 (1): 230–265. https://doi.org/10.1112/plms/s2-42.1.230

Wang, Hao. 1974. *From Mathematics to Philosophy*. London: Routledge and Kegan Paul.

Zach, Richard. 2007. "Hilbert's Program Then and Now." In *Philosophy of Logic*, edited by Dale Jacquette, 411–447. Amsterdam: Elsevier. https://doi.org/10.1016/B978-044451541-4/50014-2