Feferman on Gödel and Free Will

Abstract

Feferman surveys the work of Gödel and Turing, and, although not himself holding a mechanist view of the mind as a whole, raises objections to the anti-mechanist argument put forward by Gödel, Penrose and me, chiefly to our assumption that any mechanism plausibly representing a human mind must be consistent. These objections are countered. In his final section Feferman puts forward his **Formalist-Mechanist Thesis II** that insofar as human mathematical thought is concerned, mind is mechanical in that it is completely governed by some open-ended schematic formal system. But it is questionable whether an open-ended schematic formal system can be accounted mechanist.

—o0o—

Feferman is right to dismiss logical determinism perfunctorily, although it puzzled Aristotle and the mediaeval Schoolmen and many people still. The inference from "There is going to be either a sea battle tomorrow or not" to "Either there is going to be a sea battle tomorrow or there is not going to be a sea battle tomorrow" is invalid for the same reason as the inference from "I know he is either in his room or in the library" to "Either I know he in his room or I know he is in the library"; 'is going to be' is a modal operator, like 'I know that', and as in modal logic generally,

$$\Box(p \vee q) \nvdash (\Box p \vee \Box q),$$

on pain of collapse into modal vacuity. Like Ryle, I spent time working through the argument in detail and delving into tense logic, an interesting topic that need not detain us here.[1]

Feferman gives a careful account of the much-criticized Gödelian argument against mechanism put forward by Gödel, Penrose and me. Like many other critics he highlights our assumption that any plausible mechanical model of the mind must be consistent. Is that assumption justified? By what right do we make out that our understanding of mathematics is consistent? Frege thought his was, and was wrong. How can we be confident that Zermelo-Fraenkel Set Theory (ZF) does not contain a hidden contradiction that may one day emerge?

So far as our understanding of mathematics is concerned, the answer is simple: if ZF turned out to be inconsistent, we should reformulate Set Theory, modifying or abandoning some axiom required for proving the contradiction. There is no guarantee that our current formalisations of mathematics are correct, or consistent. But that is not the point. What is in issue is not the consistency of our current formalisations of mathematics, but the consistency of a purported representation the whole of a person's mental activity, and hence that his mathematical output is a subset of the set of provable sentences of some formal system. The mechanist might argue for this on empirical grounds—the success of neuro-physiology and, until recently, an over-arching Laplacian determinism. If the empiricist argument succeeded, then the whole of a person's mental activity could be represented by a mechanism, and hence the whole of his mathematical work. But the Gödelian argument shows that that cannot be the case, and so not only Feferman's Formalist-Mechanist Thesis I is shown to be false, but the general mechanist thesis too.

Feferman complains that we work with highly idealized views of the nature of mind and of mechanism, and that the empirical support for the mechanist thesis cannot be expected to supply a complete perfect description of the workings of the brain any more than of the digestive system. But the mechanist's "empirical defense" claims that there *is* such a system, even though it cannot offer a complete perfect description of it. And *this* system, although not completely or perfectly described by the empirical scientist, must be completely and perfectly describable in principle. If this idealized representation of a particular human mind is consistent, then according to the claim it should not be able to do something that he can do. And if the idealized representation is not consistent, then

[1] J.R.Lucas, *The Future*, Oxford, 1989.

it is not selective, and does not distinguish between truth and falsity, which a human mind at least sometimes can do. So either way the Formalist-Mechanist Thesis I fails.

Many critics fail to distinguish this informal argument from a formal proof of consistency within the idealized representation. "Merely to find from a given machine M a statement S from which it can be proved that M, if consistent, cannot prove S is not to *prove* S—even if M *is* consistent."[2] But the word 'prove' is being used equivocally, the first two times meaning formally proved by M, the third meaning established as true by cogent argument not necessarily confined to the formal proof-procedures of M. The contention that if M were inconsistent, it would not be selective, and hence not a realistic model of a human mind is a cogent argument, though not a formal proof within M, and thus enables a human mind to prove beyond reasonable doubt, though not by a formal proof-sequence generated by M, that if M is made out to be a realistic model of his mind, there is a truth which he can establish and it cannot.

But perhaps, it might be suggested, the inconsistency is buried very deep, and only emerges after a long time, if at all. A Nietzschean might maintain that philosophers who persist in thinking too long and too hard about tricky concepts, such as the Self and God, would end up by going mad. It is an intelligible hypothesis. I leave it to others to develop and demolish it.

In his last section, **Mechanism and partial freedom of the will**, Feferman puts forward a mediating position between hard-line mechanism and the anti-mechanist position taken by Gödel, Penrose and me. It depends on an "open-ended schematic axiomatic system", which is open-ended "in the sense that its basic vocabulary may be expanded to any wider conceptual context in which its notions and axioms may be appropriately applied". This fits with mathematical practice—witness how the concept of number has been widened from the natural numbers to the integers, the rationals, the reals and the complex numbers—but is difficult to construe within the framework of formal logic. Standardly in logic the range of possible substitution instances is laid down in the initial formulation. The thus circumscribed possibilities are too restricted for many purposes. Feferman's opening up the range of admissible substitution extends these possibilities in an interesting way, and may yield valuable insights into the nature of mathematical thinking—which does, indeed, progress as much by the recognition of concepts as by the discovery of derivations. But even if its axiom schemata and rules of inference are fixed, a system with open-ended rules of substitution is not going to give the hard-line mechanist what he wants, and would be compatible with an open-ended and creative view of the human mind. The hard-line mechanist is offering ab explanation of the "Why Necessarily?" type, in contrast to others, including some connectionists, who merely seek to explain How Possibly the brain could work.[3] The latter may be called mechanists, but their conclusions, though illuminating, do not have determinist consequences, any more than evolutionists who explain how reptiles evolved into birds are committed to holding that the evolutionary path actually taken was one that had to go in that pre-determined way.

[2] George Boolos, Introductory note to *1951, *Kurt Gödel, Collected Works* III, Oxford, 1995, p.295; reprinted in George Boolos, *Logic, Logic, and Logic*, Cambridge, Mass., 1998, p.110.

[3] I owe this crucial distinction to W.H.Dray. See his "Explanatory Narrative in History", *The Philosophical Quarterly*, **4**, 1954, pp.15-27; or his *Laws and Explanation in History*, Oxford, 1957, ch.6.