Etica & Politica / Ethics & Politics, 2003, 1
http://www.units.it/dipfilo/etica_e_politica/2003_1/9_monographica.htm

Lucas against Mechanism: a rejoinder (*)

J.R. Lucas

Fellow of Merton College, Oxford
Fellow of the British Academy

Professor Lewis (1) and Professor Coder (2) criticize my use of Gödel's theorem to refute Mechanism. (3) Their criticisms are valuable. In order to meet them I need to show more clearly both what the tactic of my argument is at one crucial point and the general aim of the whole manoeuvre.

Lewis argues that I have established that there is a certain *Lucas arithmetic* which is clearly true and cannot be the output of some Turing machine. If I could produce the whole of Lucas arithmetic, then I would certainly not be a Turing machine. But there is no reason to suppose that I am able in general to verify theoremhood iii Lueas arithmetic. Coder's first complaint touches on a similar point. Some men cannot follow Gödel's theorem, and my argument cannot show that *their* minds are not machines, and therefore has not shown that minds are essentially different from machines.

In order to refute Mechanism, however, I do not need to show that any mind can do all of Lucas arithmetic or that all minds can understand Gödel's theorem. Mechanism makes a universal claim. A single counter-instance is fatal. Once its defences are breached, we may follow up, and take over further positions we have not fought for. And this I do. But my first concern is the limited one of securing a single counter-instance. For that I do not need the whole of Lucas arithmetic: a single theorem will do, provided that I can see it to be true and the machine cannot. *The* machine: which machine? the argument at this stage becomes 'dialectical', dynamic, and finicky. (4) If I lay claim to less than the whole of Lucas arithmetic, I might still be a machine. Indeed, anything which I, a finite being living for only a finite time, actually do, could be copied by some machine. Yet the possibility of a machine's being able to simulate actual behaviour does not seem very important: to be a person is not merely to have actually done some things but to be able to do others as yet unspecified. A machine, to be equivalent to a mind, must not merely do everything the mind has done, but have the potentiality of doing anything the mind could do. A machine fails to be equivalent to a mind if there is anything it cannot do and a mind can. And for each particular machine there is some theorem in Lucas arithmetic which it cannot produce as true. In order therefore to refute Mechanism, what I have to do is to show that a mind can produce not the whole of Lueas arithmetic, but only a small, relevant part. And this I think I can show, thanks to Gödel's theorem. Gödel's theorem shows us how, making the relevant modifications for each particular case, we can find a formula which we can prove *to* the machine *(i.e.* in the machine's system) to be unprovable by the machine (provided it is not an inconsistent machine that can 'prove' everything), but which we can see to be a theorem in Lucas arithmetic. The argument depends essentially on the order of play. I can neither prove the whole of Lueas arithmetic nor some part of it that is outside the output of any machine. But since the Mechanist claims that every mind can be represented as a machine, I produce a particular mind and challenge him to specify which machine represents this mind, and when he has specified a particular machine, I show that there is a particular formula which that particular [149] machine cannot prove but which this particular mind can. The Mechanist's claim, as detailed by him, is refuted. Moreover, it is fairly easy for us rational

beings to see that however the Mechanist varies the specification of his detailed claim the upshot must be the same. And therefore there is and must be a counter-instance to the claim, and Mechanism must be false.

So much for the in-fighting. Lewis's reconstruction of my argument does, however, reveal the general line of advance. Lucas arithmetic (or something richer still), rather than Peano arithmetic, represents the sort of arithmetic that minds, in contrast to machines, can do. For minds can operate in accordance with Lewis's `infinitary rule of inference'.

R: If S is a set of sentences and P is a consistency sentence for S, infer P from S.

I am not myself happy with the term `infinitary rule of inference', as the whole point of conducting our inferences by rules is that the rule should be definite and finite. Many of our inferences, particularly many nonmathematical ones, cannot be formalised, and I believe this to be a fact of fundamental importance for philosophy. Even in mathematics not all can, and it is a distinguishing mark of a rational being that he can draw valid inferences even in the absence of a definite rule allowing him to. The pattern of inference expressed by Lewis's R is one such that a mind would accept, and his Lucas arithmetic typifies the difference between the sort of arithmetic that minds can do and the sort that machines can do, even though no mind can produce the whole of Lucas arithmetic or be able in general to verify theoremhood in Lucas arithmetic. This also reveals the 'essential' difference that worries Coder. He complains that I have shown only that a few smart mathematicians are not machines, and therefore have not established that minds in general are essentially unlike machines. But although the ability to understand Gödel's theorem is limited, the ability to think up new arguments is not. Many men could understand and accept Lewis's R even if they could not follow the proof of Gödel's theorem, and therefore their arithmetic would lie within the Lucas arithmetic (or something richer), but not within the narrow confines of Peano arithmetic. They, like the high-powered mathematicians, are not machines, because they share with them a power of reasoning which cannot be completely reduced to rule observance. This is what makes the essential difference between minds and machines. The peculiarity of minds that can follow Gödel's theorem is not that they are more mind-like than non-mathematical ones, but that their originality and power of reasoning without the rule-book is such that even the Mechanist must admit it. Mathematical ability is not more unmachinelike than other intellectual powers, but more easily compared with the powers of machines, and therefore more easily proved to be different.

Coder also complains that my Gödelian argument is unnecessary, because minds are so obviously unlike machines that nobody could ever suppose them to be the same. Even the most moron-like mathematician behaves quite unlike a Turing machine, and it is only because, in Coder's view, I have misunderstood the term 'mechanical' that I think there is a problem which my argument can resolve. But Coder himself misunderstands my argument here, and imputes to me a sufficient condition where I only claim a necessary one. I do not think 'of being unable to calculate except according to mechanical procedures as a sufficient condition for being a machine', although I do argue that it is necessary condition, so [150] much so that if an apparent machine, with wires and wheels and valves, could reason cogently but not mechanically, I should be prepared to say that it was not, despite its appearances, really a machine but had a mind of its own after all. The fact that men may calculate only according to an algorithm does not prove them to be machines: but if a man's reasoning cannot be reduced to a particular algorithm, then it does follow that he cannot be really that sort of machine which can infer only in accordance with that algorithm.

Of course, there are other reasons too for thinking that men are not machines, and Coder, along with many others, is entitled to feel that my argument is not so much invalid as unnecessary. Perhaps it is. If we base ourselves on ordinary facts, it is quite clear that the activity of thinking is fluid, dynamic, tentative, spontaneous, sometimes creative, and not rule-bound, rigid, static, mechanical, formalised and ossified: but philosophers are very resistant to facts, particularly the so-called empiricists of the last generation, who, together with some cyberneticians of the

present age, have been utterly convinced that in the last resort the workings of the human mind cats be understood in purely mechanical terms. Often some metaphysical doctrine of materialism underlies this conviction. But in any case it is itself a metaphysical tenet, and cannot be shaken by appeal to the facts, because the *in the last resort* clause always offers an escape from the facts as they seem to be at present. The Mechanist can concede that human behaviour does not look very machine-like at present, while maintaining that ultimately he will be able to explain it in mechanical terms. Against such a position facts are of no avail. However much we point out ways in which ratiocination is unlike anything we expect of a machine, it always could be, so far as these facts go, that it is only complexity that differentiates minds from machines, and that there is no essential difference of kind between them, but only one of degree. If a tough-minded philosopher is prepared to brush all other arguments aside, my Gödelian argument will act as a long-stop. It is a *reductio ad absurdum.* Perhaps it is an absurd thesis anyhow. But not every philosopher would agree, and many of the arguments demonstrating its absurdity appeal too much to common sense to be telling with those who boast their lack of it. The merit of the Gödelian argument is that the *reductio* is itself by means of inescapable logical arguments to a position which even a tough-minded philosopher must admit to be absurd. I am using a steam-hammer to crack a nut; but a nut so hard as to be almost impenetrable to reason.

Notes

(*) For the copyright of the papers of J.R. Lucas see http://users.ox.ac.uk/~jrlucas/ back
(1) David Lewis, Lucas against Mechanism, *Philosophy,* XLIV, July 1969, pp. 231-233. back
(2) David Coder, Goedel's Theorem and Mechanism, *Philosophy,* XLIV, July 1969, pp.234-237. back
(3) J.R. Lucas, ``Minds, Machines and Gödel", *Philosophy,* XXXVI, 1961, pp.112-127; reprinted in Kenneth M. Sayre and Frederick J. Crosson, eds., *The Modelling of Mind,* Notre Dame, 1963, pp. 255-271; and in Alan Ross Anderson, ed., *Minds and Machines,* Englewood Cliffs, 1964, pp. 43-59. back
(4) For a fuller exposition of the dialectical nature of the argument see my Satan Stultified, *The Monist,* **52,** 1968, pp. 154-7. or *The Freedom of the Will,* Oxford, 1970, §27, pp.139-45. back