

Human an Machine Logic: A Rejonder (*)

J.R. Lucas

Fellow of [Merton College](#), Oxford

Fellow of the [British Academy](#)

We can imagine a human operator playing a game of one-upmanship against a programmed computer. If the program is F_n , the human operator can print the theorem G_n , which the programmed computer, or, if you prefer, the program, would never print, if it is consistent. This is true for each whole number n , but the victory is a hollow one since a second computer, loaded with program C , could put the human operator out of a job.... It is useless for the 'mentalist' to argue that any given program can always be improved since the process for improving programs can presumably be programmed also; certainly this can be done if the mentalist describes how the improvement is to be made. If he does give such a description, then he has not made a case. (1)

Good, together with Turing and Benacerraf, (2) argues against 'mentalism' by denying that there could be any peculiarly mental powers, since if there were, they could be described, and if they could be described, a computer could be programmed to simulate them. I am not convinced that to be is to be describable in computer language, but I shall not discuss that here: for it seems to me that even if this step were valid, the whole argument would work against only one form of mentalism, and would completely fail to establish mechanism. For if mechanism is true, a complete specification of the mental mechanism of each human being can in principle be given. But once given, it proves inadequate, in that it cannot produce as true the Gödelian formula G , which a human being can see to be true. This is no 'hollow victory' for the human being, since *this* was the specification the mechanist put forward. It is of course true that a second computer, that is, another computer, could do as well as a human operator on *this* test: but that is not relevant, for it was not that computer that was supposed [156] to match the human being. And if the mechanist now makes it relevant, by shifting his ground and saying that it is the second computer that matches the human being, then there is *another* Gödelian formula which the second computer cannot produce as true but which a human being can. In my original paper I took the argument as far as ω . Good takes a further step to $\omega + 1$, and is all set to go on to somewhere between (ω to the power of ω to the power of ω) and τ (if it exists): but the principle remains the same; the human being can always 'trump' any particular machine, and the mechanist can always produce another machine that will not fail the test which stumped the first machine. If the game played between the mentalist and the mechanist is the game of one-upmanship, the result must be a draw. But this is not the game I proposed. I proposed a matching game, not a one-upman game: the question at issue was not superiority---is the mind better than the machine, or *vice versa*?---but equality---is this, specified, machine equal, in respect of intellectual powers, to this particular person? Good thinks the mentalist must be a sort of Balliol man, and adduces arguments to show that his effortless superiority is illusory. But mine is a different argument, less arrogant, depending on a subtle shift of the onus of proof. It is not a straightforward proof, starting from some acceptable

premises and leading to the conclusion that minds are better than machines, but rather it is a schema of refutation, showing how if the mechanist were to particularise enough to say what machine was equivalent to a named man, the mentalist can repute at least that equivalence. But since he can do it for any specified machine, he can do it for all. Not effortlessly superior, but incontrovertibly unequal. This is not enough to re-establish traditional dualism-which seems to me open to many philosophical objections: but it is enough to refute traditional materialism and its modern mechanist version.

Notes

(*) For the copyright of the papers of J.R. Lucas see <http://users.ox.ac.uk/~jrlucas/> [back](#)

(1) I.J. Good, [Human and Machine Logic](#), *This Journal*, **18**,1967, pp.145-146. [back](#)

(2) A.M.Turing, 'Computing Machinery and Intelligence', *Mind*, **59**, 1950, pp.444-445; reprinted in James R. Newman, *The World of Mathematics*, iv, New York, 1956, 2110; and in Alan Ross Anderson, *Minds and Machines*, Englewood Cliffs, N.J., 1964, pp. 15-16. Paul Benacerraf, [God, the Devil and Gödel](#), *The Monist*, **51**, 1967, pp.22-23. [back](#)