WILLIAM G. LYCAN

THE MORAL OF THE NEW LILLIPUTIAN ARGUMENT

(Received 10 June, 1982)

Despite its preternatural clarity and force, my 'New Lilliputian Argument' against Machine Functionalism has reaped some considerable confusion and misunderstanding. I am grateful to Professor Elugardo for providing a correct restatement of the Argument and yet an opportunity for me to sharpen its moral and to make some further clarificatory comments.

As I have used the term, 'Machine Functionalism' is the view that a correct metaphysical explication of a particular mental state (-type) would take the form. "To be in mental state M is to realize or instantiate machine program P and to be in functional state S relative to P", where a physical organism or system O realizes or instantiates P just in case a one-one mapping holds between some set or other of O's (possible) physical stimuli, inner states and responses and, respectively, the abstract input symbols, computational states and output symbols tabulated in P (cf. the more formal definition Elugardo quotes on p. 000). The intent of the Argument (hereafter 'NLA') was to show that any such explication of a mental state-type would be far too liberal, since if we are thus allowed to select any subset of physical states of O we like, organism-program correlations of the sort just mentioned are trivially easy to come by. In particular, if we attend to what Elugardo calls "certain small and uninteresting parts of Oscar's body", viz, those parts which Oscar shares with Harold, we can see that such a correlation holds trivially and parasitically between Oscar and the program P approriate for thinking that index cards ought to be painted phosphorescently.

Elugardo's own handling of the 'Hoky Stipulation' objection (p. 000) seems to me exactly right (I have made the same reply to Ned Block in conversation), so I shall pass directly on to his criticism of (Def.), which I think is confused as it stands. He questions whether "(Def.) does in fact specify a one—one mapping" (p. 000) and doubts that "(Def.) does in fact show ... that Oscar realizes a Turing Machine program in the way that [I describe] a Turing Machine realization" (p. 000). Since the NLA shows precisely that a one—one
mapping holds between some of Oscar's physical states and the relevant constituents of $P$, I think Elugardo must be conflating this issue with the other, quite different question he raises in the same breath, that of whether 'Machine Functionalism' as I have defined it has ever been held as such by any actual philosopher of mind. He complains that

the physical correlate of the input-symbol that Oscar is said to receive is not a physical input that Oscar himself characteristically receives; but rather, it is the physical input that Harold characteristically receives.... Surely this is not what the MFist has in mind when he speaks of machine realization... [W]hat counts as a physical input (output, state) of Harold need not count as a physical input (output, state) of Oscar even if the former input (output, state) occurs inside Oscar....

A general requirement emerges [:]... The machine characterization must ... describe a correlation of the person's own characteristic inputs, outputs, and states with the relevant abstract symbols in question (p. 000; italics mine).

In effect, Elugardo is proposing a restricted use of the technical term 'realize', according to which Oscar does not 'realize' program $P$ even though he does realize $P$ in the sense I originally defined and attributed to the 'Machine Functionalist'. The restriction is marked by the words I have italicized in the foregoing quotation; Elugardo's idea is to pick out a privileged subclass of Oscar's physical stimuli, states and responses and to count Oscar as 'genuinely realizing' $P$ only if a one-one mapping holds between this privileged subclass — Oscar's 'own characteristic' inputs, outputs and states — and the abstract symbols tabulated in $P$. Not just any subclass will do.

This restriction is intuitively appealing and does succeed in blocking the NLA. Moreover, it seems to block it in the right way; in my view, what the Argument calls for is precisely a tightening of the notion of 'realization'. But recall, as Elugardo does briefly on p. 000, that I deliberately defined 'Machine Functionalism' in terms of the old, unrestricted, ultraliberal notion of realization, and so what the NLA shows is that Machine Functionalism (thus defined) is false and needs revision. Anticipating this rejoinder, Elugardo suggests that Machine Functionalism (thus defined) may well be a straw man. And so it may; early writers on functionalism were not very explicit about what they meant by 'realize' (though Elugardo's historical point about Putnam is well-taken). Why did I not give them the benefit of the doubt? As I remarked in my original paper (pp. 285–286), the reason I focussed on the unrestricted notion of realization is that it is clear. Any attempt to tighten it up introduces obscurities and/or unexplained primitives. Elugardo concedes in closing that his own term 'characteristic' is a case in point. He does not tell us in any principled way which inputs etc. are Oscar's 'characteristic' ones (of