Book Review

The Logic of Mind

R. J. Nelson

D. Reidel Publishing Company
Dordrecht, Holland, 1982
ISBN-90-277-1399-5
Pp. xvi, 392. $59.50

In 1936, Alan Turing published "On Computable Numbers, With An Application to the Entscheidungsproblem". In it he introduced the world to Turing machines. These he called simply computing machines, and amongst them he distinguished between "automatic machines" (deterministic Turing machines) and "choice machines" (nondeterministic Turing machines). For reasons that need not detain us, Turing focused on the computable real numbers, not on the nature and extent of the computable functions of natural numbers. In the last two sections of the paper (plus the appendix), Turing presents arguments for what we would now call Turing's Thesis (relativized to computable numbers). The real question at issue is "What are the possible processes which can be carried out in computing a number?" Turing continues:

> Computing is normally done by writing certain symbols on paper. We may suppose this paper is divided into squares like a child's arithmetic book. In elementary arithmetic the two dimensional character of the paper is sometimes used. But such a use is always avoidable, and I think that it will be agreed that the two-dimensional character of paper is no essential of computation. I assume then that the computation is carried out on one-dimensional paper, i.e., on a tape divided into squares.

Turing goes on to abstract other essentials of computation from *the case of a human being using pencil and paper to compute.* Thus, when he speaks (p. 136) of the "behaviour of the computer" being determined by "the symbols he is observing, and his 'state of mind' at that moment", Turing is using the personal pronouns nonmetaphorically. A computer is a person engaged in the act of computing. Indeed, after completing his analysis of the essentials of a computation by a computer, Turing notes (p. 137): "We may now construct a machine to do the work of this computer. To each state of mind of the computer corresponds an 'm-configuration' of the machine. The machine scans B squares corresponding to the B squares observed by the computer." And so on in the same vein. Turing never speaks of a computing machine as a computer; only people are computers.

Plus ca change, plus c'est la meme chose? We now do speak quite regularly of computing machines – even automatic computing machines – as computers. What of people? Ever since the birth of Artificial Intelligence (Turing's 1950 paper "Computing Machinery and Intelligence"; the Dartmouth Conference of 1956?), the idea has been abroad of conceiving of people on the model of what are now thought of as the paradigm cases of computers, that is, as computing machines. Indeed, Zenon Pylyshyn (1984) has recently urged us to recognize the "Computational Metaphor" for the Literal Truth it embodies: human beings, along with other intelligent, "really are cogitating creatures, really are computing machines."

In *The Logic of Mind,* R. J. Nelson, Professor (recently emeritus) of Philosophy at Case Western Reserve University, presents a sustained, systematic argument for this view, a view he calls "mechanism".

> This book presents a mechanist philosophy of mind. I hold that the human mind is a system of *computational* or *recursive* rules that are embodied in the nervous system; that the material presence of these rules accounts for perception, conception, speech, belief, desire, intentional acts, and other forms of intelligence. (p. xiii)

So much counts as almost conventional wisdom. What is striking about Nelson is his willingness to provide us with a systematic, specific version of mechanism, along with an equally ready ability to defend it. (With respect to his technical credentials, I should note that Prof. Nelson is the author (1968) of an excellent text on automata theory, *Introduction to Automata.* For Nelson, it is not enough to say people are computers and that mental processes are computational processes; one must also specify the kind of machine or automaton a person is. Nelson does just this.

First he must clear the ground of what he takes (as a rule, correctly) to be confusion and misinterpretation. This he does at the same time as he introduces the basic automata-theoretic notions out of which his account will be built. Thus, after telling us in his introduction what he means by mechanism (see the quotation above), he divides the domain of mental phenomena in two. In one part lie the so-called representational or contentful states and processes, together with their corresponding underlying capacities. These include, centrally, the propositional attitudes of belief and desire, together with the results of various peripheral pattern-recognition devices. Nelson calls such results perceptual takings, as in (say) seeing a hat to be red. It is to such "mental features", as he calls them, that his theory is directed. On the other side of the dichotomy lie "pure (raw) feels", as of pain. These, he claims, must be acknowledged to be outside the purview of mechanism.

In Chapters II and III, "What is a Rule of Mind" and "Behavior and Structure", Nelson lays out the fundamental concepts and distinctions he will work with. These two chapters constitute a solid, if sometimes opaquely written, introdution to and history of the theory of automata, both finite and infinite, as well as a detailed discussion of the associated notions of nerve networks, computer circuits, and grammars.

In Chapter IV, "Mechanism – Arguments Pro and Con", Nelson responds to several of the standard arguments against mechanism, including those (or is it just one) based on (as he characterizes it) "the now familiar misapplication of Godel's Theorem". He has in mind here an argument (or family of arguments) that attempt to use Godel's First and/or Second Incompleteness Theorems to show that humans can come to know truths of elementary arithmetic that no machine could ever know. Nelson's treatment of this particular morass is woefully brief and surprisingly unsophisticated, especially in light of the large philosophical and logical literature such arguments have both drawn upon and engendered. (See, especially, Webb 1980.) He does a much better job in arguing for the compatibility of (1) thoroughgoing Laplacian determinism and (2) people actually being nondeterministic automata. Even here, he might have made less heavy weather of things, and at the same time strengthened his argument, if he had noted that physicalism implies neither that mental types (his "mental features") are identical with physical types nor that mental tokens (individual mental events and states) are identical with physical events and states. Chapter IV begins with a reversal of the more or less customary hand-waving argument from analogy: "Digital computers are automata and share a large body of intelligent behaviors with humans; hence human beings are automata." That really is the argument – and in more or less just so many words.

Chapter V, "Functionalism, Behaviorism, and Rationalism", contrasts mechanism favorably with various close neighbors. There's something almost quaint about the respect with which Nelson treats behaviorism. Behaviorism is a movement now almost wholly out of fashion; but attention must be paid and Nelson's discussion is careful and illuminating. The last part of the chapter is devoted to a discussion of Chomsky's (neo)rationalism; the main point is well taken and even fairly well argued – namely, that no plausible story about linguistic competence and universal grammar requires tacit *propositional* knowledge. The most important and interesting (albeit least clear) part of this survey of the current scene comes at the beginning. Nelson takes great pains to make it clear that mechanism, or at any rate his version of it, is not functionalism; indeed, is not even a version of functionalism. (He does recognize that there are mechanistic versions of functionalism. For more on this, see Block and Fodor (1972).

What then is functionalism? By functionalism, I understand any of a family of views whose central tenet (if such there be) is that representational mental states are to be identified with – or are uniquely determined by – their [causal] roles in the lives of [normally] functioning organisms. In particular, this identification – or determination – is spelled out in terms of the relations holding among input (behaviorally or physiologically characterized), output (behaviorally or physiologically characterized) and other representational mental states. In fact, Nelson does not convince me that his version of mechanism is not actually a version of functionalism – though it most certainly is not the particular mechanistic functionalism (the Functional State Identity Theory) that he discusses and rejects. Most of his arguments aren't directed against functionalism; their target is rather one or another ill-considered mechanistic version of functionalism. For instance, he makes much of the fact that mental features like belief and desire are best thought of as automata, *not* as automata states. Within an automata-theoretic framework, this seems just right. Correct also is his claim that many "mechanistic functionalists" have spoken of beliefs and desires as states. Wrong, though, is the [implied] claim that the mentioned misidentification is an essential feature of, or intrinsic to, functionalism. Similar points could be made about many of his animadversions regarding functionalism; but so, too, should it be said that his caveats against particular theses espoused by functionalists are well-founded.

While all this "clearing of the ground" has been in progress, Nelson has smuggled in a crucial demurrer from current orthodoxy. He rejects the identification of humans or human minds with universal (deterministic or nondeterministic) Turing machines, for human memory is obviously finite. He rejects as well the view that "systems of man-environment pairs" are universal Turing machines. This latter idea might *seem* to be Turing's, modulo the idealization of unbounded amounts of paper in the environment. It is crucial to keep in mind, however, that Turing wasn't (primarily) interested in the nature of mind in general. For the special idealized case that did interest him, it was possible to mark off clearly that part of the environment that corresponds to the tape from the rest of the environment as well as from the "computer" himself. For, in principle, Turing really had in mind a one-dimensional tape, actually sectioned off into squares, with tokens of some specified character set actually written in the squares. Nor need there be any doubt as to what kinds of behavior correspond to moving left or right one square or n squares, etc. Quite apart from concern over problems of infinitude, none of these things can be said for the various ill-considered accounts to the effect that human-environment systems are universal Turing machines.

Nelson opts unambivalently for finitude; the minds of humans are non-deterministic finite automata. But this is not specific enough for Nelson. The heart of his account is to be found in Chapters VI-VIII, "The Logic of Acceptance", "Perception", "Belief and Desire." Nelson offers us detailed automata-theoretic analyses of certain basic mental capacities and propositional attitudes.

These include perceptually taking some object to be such-and-such, expectation, perceptual belief, belief more generally, and desire. Chapters IX, "Reference and Truth", and X, "Toward Meaning", attempt (cautiously and sketchily) to extend the analyses to the phenomenon of language. In Chapter XI, Nelson returns to raw feels, conscious experiences, and the relationship between mind and body. (I fear that Chapter XI sounds infinitely more interesting than it actually is.)

The basic theoretical concept is in Nelson's account is that of a **superautomaton.** A superautomaton consists of a collection of pattern recognition devices, modeled as (nondeterministic) finite state acceptors, cascaded together with a central executive, modeled as a [modified] "Moore machine". A [modified] Moore machine is a septuple T = [S, Q, O, q0, M, N Q(F)]. Here, S is a finite nonempty set of inputs; Q a finite nonempty set of internal states; O a finite nonempty set of outputs; q0 (an initial state) a distinguished member of Q; M a transition function (Q X S $\rightarrow$ Q); N an output function (defined only on Q; so N: Q $\rightarrow$ O); Q(F), the final states, are a finite nonempty subset of Q. Of course, in the case of nondeterministic Moore machines, M and N are relations, not functions.

I shall not attempt here to impart even the "flavor" of Nelson's analyses; indeed, I went to the trouble of specifying the major novel automata-theoretic feature of Nelson's superautomaton precisely as a surrogate for any such venture. The striking fact about Nelson's book is simply that someone has finally put forth a detailed and theoretically sophisticated version of mechanism. Nelson has had the courage of his (and others') convictions. In particular, he has had the courage – and the requisite technical mastery – to offer an account against which detailed objections *could* be made. Nobody else can make that claim.

David Israel
Bolt Beranek and Newman Inc.
SRI International

References

Block, N. and Fodor, J.A. 1972 What Psychological States Are Not. *Philosophical Review* 81: 159-181.

Nelson, R.J. 1968 *Introduction to Automata.* John Wiley & Sons, Inc., New York. New York

Pylyshyn, Z. 1984 *Computation and Cognition.* The MIT Press, Cambridge, Massachusetts.

Turing, A.M. 1936 On Computable Number with Applications to the Entscheidungsproblem. *Proceedings of the London Mathematical Society* 42: 230-265.

Turing, A.M. 1950 Computing Machinery and Intelligence. *LIXMind, LIX:* p. 433-460.

Webb, J.C. 1980 *Mechanism, Mentalism, and Metamathematics.* D. Reidel Publishing Co., Dordrecht, Holland.