The Legacy of Alan Turing, Volumes 1 and 2

Volume 1: Machines and Thought, edited by Peter Millican and Andy Clark. Oxford: Clarendon Press, 1996. Pp. x + 297. £30.

Volume 2: Connectionism, Concepts and Folk Psychology, edited by Andy Clark and Peter Millican. Oxford: Clarendon Press, 1996. Pp. ix + 281. £30.

These volumes contain the proceedings of a Colloquium held at the University of Sussex in 1990 in celebration of the fortieth anniversary of the publication of Turing's "Computing Machinery and Intelligence" (*Mind* 59 (1950), pp. 433-60). The quality of the contributions is very uneven. Eight of the twenty-seven papers have appeared previously (in whole or in part). Volume 1 is well focussed on Turing's work and legacy, and includes valuable contributions by two of Turing's closest associates, Donald Michie and the late Robin Gandy. These two papers excepted, there is little evidence of Turing scholarship. The date of Turing's death is wrongly shown (p. 59), the editors include no bibliography of Turing's relevant work, and (with the exception of Michie) the contributors pay no attention to Turing's less familiar papers. Nevertheless, the volume contains an occasional gem. Volume 2 concerns Turing's legacy in only the broadest of senses, addressing a heterogeneous collection of topics in cognitive science and the philosophy of mind. (Clark claims in his introduction only that the contributions "provide graphic proof of both the multiplicity and complexity of the issues raised by a computational perspective on mind" (p. 5).) Of the thirteen papers in the volume, three mention Turing only in passing, and nine make no reference to him whatsoever. Only one, by Hofstadter, goes any way to meet the expectations raised by the title of the set (and this only in a short epilogue to his paper).

Under pressure of space, we discuss only papers that bear a substantive relationship to Turing's work (and of these we omit the interesting papers by Clark Glymour and Iain Stewart). Page numbers and chapter numbers refer to Volume 1 unless labelled otherwise.

Chs 1 - 4

The Turing test is the subject of the papers by Robert French, Donald Michie, Ajit Narayanan and Blay Whitby. The mood is predominantly critical, and Whitby suggests that in consequence of this critique the Turing test will be "consigned to history" (p. 53).

Narayanan and Whitby both object to the practice of calling Turing's imitation game a "test", claiming that Turing himself did not do so. Whitby offers this claim as evidence for the "important suggestion that Turing's paper [has not been] interpreted as ... [Turing] himself intended" (p. 54), and both he and Narayanan use this claim in support of their respective exegeses. These authors are mistaken. Turing freely referred to the imitation game as a test:

> I would like to suggest a particular kind of test that one might apply to a machine. You might call it a test to see whether the machine thinks, but it would be better to avoid begging the question, and say that the machines that pass are (let's say) "Grade A" machines. The idea of the test is that the machine has to try and pretend to be a man, by answering questions ... [The question whether] machines really could pass the test ... [is] not the same as "Do machines think", but it seems near enough for our present purpose, and raises much the same difficulties. ("A Lecture and Two Radio Broadcasts on Machine Intelligence by Alan Turing" (ed. Copeland, B.J.), in Michie, D., Muggleton, S. (eds) *Machine Intelligence 15*, Oxford University Press (1998))

The Turing test is commonly said to provide an "operational definition" of intelligence (for example, by Andrew Hodges in *Alan Turing: The Enigma* (London: Vintage, 1992): "Turing offered an operational definition of 'thinking' or 'intelligence'" (p. 415, see also p. 266)). A moment's philosophical reflection should suffice to show that the Turing test cannot provide the claimed operational definition, for an intelligent machine (or alien) may fail just because its verbal behaviour is distinctively non-human, even when it is doing its best to pass itself off as one of us. Michie and French both aim to provide examples of ways in which an interrogator can identify the non-human player. Michie draws attention to the phenomenon of “superarticulacy” (pp. 41-3). One wants an expert system to be able to articulate how it reaches its judgements. Notoriously, human experts are often not good at articulating their expertise, and this has been an important obstacle in the attempt to transfer knowledge from human experts to artificial systems. In a side-by-side comparison with a human expert, an artificial system’s superarticulacy may unmask it.

French uses questions like the following to illustrate his claim that the Turing test is “culturally-oriented” and so “virtually useless as a real test for intelligence” (p. 12): on a scale of 0 (completely implausible) to 10 (completely plausible), rate “‘Flugbloggs’ as a name Kellogg’s would give to a new breakfast cereal”, rate “‘Flugly’ as the surname of a glamorous female movie star”, “rate banana splits as medicine” (pp. 18, 21). An artificial intelligence that is perfectly matched to its job—overseeing mining on the moon, say—may make a poor showing in these “rating games”. One might reply on behalf of the Turing test that the use of such questions will not necessarily assist the interrogator to identify the computer, since the computer will evidently maximise its chances of avoiding identification if it attempts to pass itself off as a foreigner of some sort. Conveniently, French claims to discern “an assumption ... tacit in Turing’s article”, namely that the computer must pass itself off as a member of the interrogator’s own culture (p. 15). He leaves it a mystery why Turing would have wished to impose a restriction which makes the test harder for the computer to pass and yet offers no conceptual gain. In fact, Turing stipulated that the machine “be permitted all sorts of tricks” in discharging its task of evading identification (“A Lecture and Two Radio Broadcasts on Machine Intelligence by Alan Turing”).

Turing was mistaken, French says, in his “claim ... that in the not-too-distant future it [will] in fact be possible actually to build ... a machine [that can] pass the Turing Test” (p. 11), since (so French argues) the interrogator will always be able to distinguish the computer by asking enough questions of the rating game type (pp. 12, 23). In fact, Turing’s only prediction of success in the test was this rather careful one: “an average interrogator will not have more than 70 per cent. chance of making the right identification after five minutes of questioning” (“Computing Machinery and Intelligence”, p. 442)—a view that clearly is consistent with the possibility of there being certain questions which, if asked, would easily unmask the computer, provided these are not ones likely to occur to an average interrogator.

French terms his rating game questions “subcognitive”, meaning that they probe the candidates’ “subconscious associative network ... that consists of highly overlapping activatable representations of experience” (p. 16). This description of the questions is driven by connectionist theory, of course. An AI researcher might say with some justice that in so far as French’s sample questions (“rate dry leaves as hiding places”, “rate pens as weapons”, “rate jackets as blankets”, and so forth (pp. 20-21)) have any one thing in common, it is that the majority of them probe the candidates on their “common sense knowledge” of the world. Viewed in this light, French’s rating games fail, for the most part, to provide any new challenge. Nor can he assume that only connectionist devices will perform satisfactorily in these games: it remains to be seen how high a score can be obtained by a conventional computer equipped with a massive store of common sense knowledge, such as Doug Lenat’s presently incomplete CYC.

Turing himself envisaged that the process of constructing the machine that is to “imitate an adult human mind” may involve subjecting a “child machine” to “an appropriate course of education”, which would possibly include allowing the machine “to roam the countryside” equipped with “the best sense organs that money can buy” (“Computing Machinery and Intelligence”, pp. 455, 456, 457, 460, and “Intelligent Machinery” (National Physical Laboratory Report, 1948), p. 13; page references to the 1948 report are to the reprinting in Meltzer, B., Michie, D. (eds) *Machine Intelligence 5*, Edinburgh University Press (1969)). Turing canvasses the possibility that the “child machine” should consist of an initially unorganised network of neuron-like elements (see below). One might reasonably conjecture that the resulting adult machine would do rather well in the Turing test. In the end, French’s case rests on his claim that unless a machine “resembled us *precisely* in *all* physical respects” its experiences of the world would differ from ours in a way “detectabl[e] by the Turing test” (pp. 22, 23; our italics). But French offers no argument whatsoever for this claim.

French also imagines the interrogator measuring the reaction times of contestants in the Turing test during semantically primed word/nonword decision tasks and comparing these to results from a sample population of humans. However, introducing such measurements is illegitimate. The specifications of the Turing test are clear: the interrogator is allowed only to put questions and there is no provision for the use of timing mechanisms to administer the decision task and to measure the contestants' reaction times, nor for the use of any other apparatus. One might as well allow measurement of the contestants' magnetic fields or energy dissipation.

Irrespective of the success or otherwise of French's ingenious efforts to assist the interrogator, the fundamental point at issue is correct and obvious: the Turing test suffers from anthropocentric bias and is thus of little use to AI engineers as a benchmark test of intelligence. But then Turing never claimed it would be. He certainly would not have agreed that the test offers an operational definition of "thinking" or "intelligence". He claims of the imitation game only that it "has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a man" ("Computing Machinery and Intelligence", p. 434), saying "machines [may] carry out something which ought to be described as thinking but which is very different from what a man does" (ibid., p. 435) and "I don't want to give a definition of thinking" ("A Lecture and Two Radio Broadcasts on Machine Intelligence by Alan Turing").

Ch. 5

"Machine as Mind" is classic Herb Simon ("computers ... have been thinking ... for forty years" (p. 100)). Simon points out that much of the putative evidence against his and Newell's now unfashionable hypothesis that the human mind is fundamentally a serial symbol-processing machine is misdirected.

Chs 8 - 10

Antony Galton's title is "The Church-Turing Thesis: Its Nature and Status". In fact Galton conflates two distinct theses, which we will refer to as "the Church-Turing thesis properly so called" and "the maximality thesis". A computer, in the sense of the term that was current when Church and Turing (independently) advanced their theses, was an obedient human calculator who worked in a manner demanding neither insight nor ingenuity, in accordance with a table of instructions devised by a numerical analyst. (Prior to the advent of automatic calculating machines in the 1940s, such work was the lot of many thousands of people in commerce, government, and research establishments.) The Church-Turing thesis properly so called is the assertion that every table of instructions that can be carried out by such a computer can also be carried out by a Turing machine. (It seems that it was Kleene who first coined the name "Church-Turing thesis" for this assertion (see his *Mathematical Logic*, New York: Wiley (1967), p. 232).) Thus the Church-Turing thesis properly so called relates the abilities of Turing machines and human computers; in and of itself it says nothing concerning the relationship between Turing machines and the class of possible computing machines. It is this latter relationship that is addressed by the maximality thesis, which asserts that anything computable by any means whatsoever is computable by Turing machine: the universal Turing machine is *maximal* among computing machines. (As Galton puts the thesis, the universal Turing machine "defin[es] the limits to computation in general" (p. 142).) When Galton asks whether the Church-Turing thesis is true, he sometimes has in mind the Church-Turing thesis properly so called and sometimes the maximality thesis. The first of these questions has been discussed previously by (among others) Péter and Kalmár, whose work Galton reviews. The second question is a good one, for the maximality thesis is a dogma of our time, and Galton deserves credit for seeing that there *is* a question to be asked. His conclusion on the matter is a cautious one: if we deny the maximality thesis

> we are postulating the existence of some [machine-] operation that Turing missed ... It does not seem very likely, but on the other hand we cannot entirely rule out the possibility. (p. 151)

This remark of Galton's is unfortunate. Turing was in fact the first to carry out a mathematical investigation of abstract machines able to perform operations of the sort that Galton suggests Turing missed. In his PhD thesis Turing introduced the idea of machines able to solve mathematical problems of a specific sort that cannot be solved by (what we now call) Turing machines ("Systems of Logic Based on Ordinals", published in *Proceedings of the London Mathematical Society* 45 (1939), pp. 161-228). Turing described these as "a new kind of machine" and called them "O-machines" (op. cit., p. 173). (In a letter written shortly after Turing's death, Gandy records that "Alan considered [this paper] had never received the attention it deserved (he wouldn't admit that it was a stinker to read)".)

O-machines generate digital output from digital input by means of a step-by-step procedure consisting of a small, fixed number of primitive operations; the procedure unfolds under the control of a finite program of instructions which is stored internally in the form of data on the machine's tape. At least one of the primitive operations cannot be simulated by a universal Turing machine. (Further discussion of O-machines may be found in Copeland's "Turing's O-machines, Searle, Penrose, and the Brain", *Analysis* 58 (April 1998).) The expression "super-Turing-machine" is gaining some currency as a general term for machines able to solve problems that cannot be solved by Turing machine. At present such machines are purely notional, but the idea that it may be physically possible to build one is gathering momentum. It may be physically possible, that is, to construct a machine which, under the idealisation of unbounded storage and unbounded computing time, can compute functions that a universal Turing machine is unable to compute. (Speculation that there may be physical processes—and so, potentially, machine-operations—whose behaviour cannot be simulated by Turing machine stretches back over at least four decades; see the references in "Turing's O-machines, Searle, Penrose, and the Brain".)

Aaron Sloman, too, broaches the possibility of machines whose behaviour does not "map onto a ... Turing machine"; in Sloman's phrase, such machines are not "Turing equivalent" (pp. 181, 183). His discussion of this possibility is less conservative than Galton's, and he raises the important issue of whether such mechanisms could play a role in human intelligence (p. 182). However, like Galton, he writes without reference to the literature on the topic and overlooks Turing's own contribution.

Sloman's discussion of machines that are not "Turing equivalent" forms one prong of his attack on "the standard concept of computation" (p. 181): the concept is "too narrow" (p. 184). Another prong dwells on the difficulty of avoiding the position that "*all* processes are computational" (p. 183). Sloman points out that the attempt by fellow contributor Chris Fields to provide an account of "how physical processes might be characterised as computations" (p. 165) has the consequence that "measurable states of *any* observable physical process can be interpreted as the execution of an algorithm that could be run on a universal Turing machine" (p. 184, our italics). (Whether or not Fields would agree that his account has this consequence is not clear.) If any and every physical process can be characterised as computation, then the view that cognition is computation is vacuous. The truth of the theory that the brain is a computer is secured, but at the cost of making rocks, pails of water, and the solar system computers. This is the *trivialisation problem*. It is important: some take it to refute the computational view of mind (for example, Searle in *The Rediscovery of the Mind*, Cambridge, Mass.: MIT Press (1992), ch. 9). Sloman is pessimistic about the prospects of solving the problem: "It is not clear that we can find [a] general definition [of computation] that covers all the interesting cases and avoids ... triviality" (p. 184). (The interesting cases include Turing machines, analogue computers, neural networks, and super-Turing-machines.) His pessimism is unwarranted. The trivialisation problem has recently been tackled by Chalmers ("Does a Rock Implement Every Finite-State Automaton", *Synthese* 108 (1996), pp. 309-333) and Copeland ("What is Computation?", ibid., pp. 335-359). The difficulty is not peculiar to the notion of computation but is a more general one: formal concepts typically admit of nonstandard models or interpretations (Skolem's paradox is a spectacular illustration of this). In the case of computation, simple and natural conditions on the admissibility of interpretations turn out to exclude Searle's "Wordstar wall", arbitrary pails of water, and the like, from the class of computers.

Chs 6 - 7

J.R. Lucas provides a welcome retrospect of his famous "Gödel argument" (originally presented in "Minds, Machines and Gödel", *Philosophy* 36 (1961), pp. 112-127). He states his argument in some detail and deals skilfully with numerous objections to it. Lucas's target, now as then, is "mechanism"—the view that "minds [can] be explained as machines" (op. cit., p. 112). More fully, mechanism is the claim that "we can understand the operation of the [mind] in terms of the operations of its parts, and the operation of each part either shall be determined by its initial state and the construction of the machine, or shall be a random choice between a determinate number of determinate operations" (op. cit., p. 126). "Mechanism is false", says Lucas (op. cit., p. 112). He believes that he has established this by using the Gödel argument "to show that minds [are] not Turing machines" (p. 103). However, this latter conclusion is consistent with the mind's being a super-Turing-machine, and if the mind is a super-Turing-machine then mechanism is true. Lucas's overall argument contains a lacuna. A _narrow_ mechanist holds that the mind is a machine equivalent to a Turing machine; a _wide_ mechanist holds that the mind is a machine but countenances the possibility of information-processing machines that cannot be mimicked by a universal Turing machine, and allows in particular that the mind may be such a machine. In effect, Lucas moves straight from the claim that narrow mechanism is false to the claim that mechanism is false, apparently believing that mechanism entails narrow mechanism. The latter fallacy is common among mechanists also, and appears to be rooted in a popular misconception of Turing's work (see Copeland "The Broad Conception of Computation" _American Behavioral Scientist_ 40 (1997), pp. 690-716, esp. sect. 4).

In fact, the Gödel argument can be generalised so as to threaten wide mechanist conceptions of the mind also. Roger Penrose admits with some discomfort that the argument applies not only to the view that the mind is equivalent to a Turing machine but "much more generally", saying "No doubt there are readers who believe that the last vestige of credibility of my [version of the Gödel] argument has disappeared at this stage! I certainly should not blame any reader for feeling this way" ("Beyond the Doubting of a Shadow", _Psyche_ 2:23 (1996), sect. 3.10, and _Shadows of the Mind_, Oxford University Press (1994), p. 381). To date, Penrose has not made it clear what scientific conception of the mind can remain for one who endorses the Gödel argument, remarking only that, since the crucial step of the argument "can be applied in very general circumstances indeed", the mind is "something very mysterious" ("Beyond the Doubting of a Shadow", sect. 13.2). Mechanists will think it more credible that the Gödel argument contains some subtle error than that—as Lucas claims—"no scientific enquiry can ever exhaust the ... human mind" (op. cit., p. 127).

Or not so subtle. Turing for one believed that the Gödel argument is easily defeated:

> Recently the theorem of Gödel and related results have shown that if one tries to use machines for such purposes as determining the truth or falsity of mathematical theorems and one is not willing to tolerate an occasional wrong result, then any given [Turing] machine will in some cases be unable to give an answer at all. ... It has for instance been shown that with certain logical systems there can be no [Turing] machine which will distinguish provable formulae of the system from unprovable ... Thus if a [Turing] machine is made for this purpose it must in some

> cases fail to give an answer. On the other hand if a mathematician is confronted with such a problem he would search around and find new methods of proof, so that he ought eventually to be able to reach a decision about any given formula. This would be the argument. Against it I would say that fair play must be given to the machine. Instead of it sometimes giving no answer we could arrange that it gives occasional wrong answers. ... The argument from Gödel's and other theorems ... rests essentially on the condition that the machine must not make mistakes. But this is not a requirement for intelligence. ... [T]hese theorems say nothing about how much intelligence may be displayed if a machine makes no pretence at infallibility. ("Intelligent Machinery", p. 4, and "Lecture to the London Mathematical Society on 20 February 1947", pp. 123-4; the latter is in Carpenter, B.E., Doran, R.W. (eds) *A.M. Turing's ACE Report of 1946 and Other Papers*, Cambridge, Mass.: MIT Press (1986))

Elsewhere Turing asserts that the "danger of the mathematician making mistakes is an unavoidable corollary of his power of sometimes hitting upon an entirely new method" ("A Lecture and Two Radio Broadcasts on Machine Intelligence by Alan Turing"). Robin Gandy's paper is a spirited endorsement of Turing's short way with the Gödel argument.

[leave a double paragraph space]

As to Volume 2, a sad omission from a book that purports to deal with Turing's legacy in connectionism is any mention of Turing's own early contribution to this field. Indeed, Christopher Peacocke falsely claims that Turing's "conception of what mechanism has to involve is now superseded [since] the explanations appropriate for a connectionist system are still 'mechanical'" (Vol. 2, p. 115). (Equally unfortunate is the failure of Joseph Ford's paper "Chaos: Its Past, Its Present, but mostly Its Future" (Vol. 1)—which does not mention Turing at all—to draw attention to the fact that Turing was seemingly the first person to engage in the computer-assisted exploration of nonlinear dynamical systems.) Connectionists look back on Donald Hebb and Frank Rosenblatt as the founding fathers of their approach, but in fact both were preceded by Turing, who anticipated much of modern connectionism in his 1948 paper "Intelligent Machinery" (see our "On Alan Turing's Anticipation of

Connectionism”, <u>*Synthese*</u> 108 (1996), pp. 361-377, and “Turing, Wittgenstein and the Science of the Mind”, <u>*Australasian Journal of Philosophy*</u> 72 (1994), pp. 497-519). Here Turing introduces what he calls “unorganised machines”, giving as examples networks of neuron-like boolean elements connected together in a largely random manner (we shall call these “Turing nets”). He describes a certain form of Turing net as “the simplest model of a nervous system” and he hypothesises that “the cortex of the infant is an unorganised machine, which can be organised by suitable interfering training” (“Intelligent Machinery”, pp. 10, 16). The idea that an initially unorganised neural network can be organised by means of “interfering training” is undoubtedly the most significant aspect of Turing’s discussion. In Turing’s model, the training process renders certain neural pathways effective and others ineffective. He anticipated the modern procedure of simulating neural networks, and also the process of training them, in an ordinary digital computer, saying

> quite definite “teaching policies” ... could also be programmed into the [computer]. One would then allow the whole system to run for an appreciable period, and then break in as a kind of “inspector of schools” and see what progress had been made. (ibid., p. 21)

Turing claimed a proof (now lost) of the proposition that an initially unorganised Turing net with sufficient neurons can be organised to become a universal Turing machine with a given storage capacity (ibid., p. 15). This proof first opened up the possibility, noted by Turing (ibid., p. 16), that the human cognitive system is a universal symbol-processor implemented in a neural network.

Both volumes are handsomely produced and contain few typographical errors.

Research on which this article draws was supported in part by University of Canterbury Research Grant no. U6271.

The Turing Project
Philosophy Department
University of Canterbury

B. JACK COPELAND
DIANE PROUDFOOT

Christchurch
New Zealand
Fax +64-3-3642889
j.copeland@phil.canterbury.ac.nz
d.proudfoot@phil.canterbury.ac.nz