Rethinking Turing’s test and the philosophical implications

Abstract: In the 70 years since Alan Turing’s ‘Computing Machinery and Intelligence’ appeared in *Mind*, there have been two widely-accepted interpretations of the Turing test: the canonical behaviourist interpretation and the rival inductive or epistemic interpretation. These readings are based on Turing’s *Mind* paper; few seem aware that Turing described two other versions of the imitation game. I have argued that both readings are inconsistent with Turing’s 1948 and 1952 statements about intelligence, and fail to explain the design of his game. I argue instead for a *response-dependence* interpretation (Proudfoot 2013). This interpretation has implications for Turing’s view of free will: I argue that Turing’s writings suggest a new form of free will compatibilism, which I call response-dependence compatibilism (Proudfoot 2017a). The philosophical implications of rethinking Turing’s test go yet further. It is assumed by numerous theorists that Turing anticipated the computational theory of mind. On the contrary, I argue, his remarks on intelligence and free will lead to a new objection to computationalism.

Keywords: Turing Turing test response-dependence free will computational theory of mind computationalism

Proudfoot D. (2020). Rethinking Turing’s Test and the Philosophical Implications. *Minds and Machines* 30: 487–512. http://dx.doi.org/10.1007/s11023-020-09534-7.

1. Introduction

This paper is in four parts. First, a reminder of the standard interpretations of the philosophical point of Turing's imitation game, and why in my view these are mistaken (Section 2). Second, my own account of Turing's approach to the concept of intelligence or thinking—as I read Turing, he does not differentiate between intelligence and thinking when talking about his 'test'—and how his conception of thinking plays out in the imitation game. I argue that Turing's view was in fact very different from the views that are standardly attributed to him (Sections 3-4). Third, one implication of rethinking Turing's test as I do is that it leads to a new reading of Turing on free will in machines—again, this is an approach that is very different to the view of free will that has been attributed to him, but which is consistent with his own words (Sections 5-8).

Finally, a broader implication of rethinking Turing's test is as follows: despite the many philosophers and computer scientists who assume that Turing was (in modern terminology) a proponent of the computational theory of mind, he was not (Sections 9-10). Turing had a more subtle approach to the mind—the one that we find expressed in his imitation game, once we read the 1948 and 1952 versions of the game in addition to his famous 1950 paper.

To elaborate Turing's views, I also set them in the philosophical debates of the time, often carried out in public.

2. Standard interpretations of Turing's test of intelligence in machines: behaviour or inner states

The canonical view of Turing's test, namely that it is a behaviourist test of thinking, can be seen in a well-known quotation from Ned Block. Block said that 'The Turing test conception of intelligence' is as follows: 'Intelligence … is the disposition to produce a sensible sequence of verbal responses to a sequence of verbal stimuli, whatever they may be' (1981, p. 11). This, Block said, is 'a clearly behaviorist formulation' (ibid., p. 11). On this familiar interpretation of Turing's test, the philosophical thesis underlying the test is, roughly speaking: if the computer *behaves* as if it is intelligent, then it *is* intelligent. If it walks like a duck and quacks like a duck, it just *is* a duck.

This interpretation has a long history: in the 1950s, Turing's colleague at Manchester University, Wolfe Mays, characterized Turing's imitation game as aiming 'to construct a machine whose behaviour was indistinguishable from that of a human being', and Turing's 'criterion' of thinking as 'behaviourist' (1952, p. 151). The behaviourist interpretation remains the standard reading of the imitation game. For example, Andrew Hodges reiterated this reading in the latest edition of his influential 1983 biography. According to Hodges, Turing 'favoured the idea of judging a machine's mental capacity simply by comparing its performance with that of a human. It was an *operational* definition of "thinking"' (2014, p. 334).

Certainly, behaviourism was popular in the 1940s and 1950s, and operationalizing a phenomenon in order to study it is common scientific practice. However, there are serious problems for the canonical behaviourist reading of Turing's imitation game. These problems include the following. First, the imitation game does *not* test machine behaviour. Instead, it tests the response of the *interrogator* or judge. The goal of the imitation game is that the interrogator be 'taken in by the pretence' and Turing tells us that a machine scores well in the computer-imitates-human game if the interrogator in that game is fooled no less frequently than the interrogator in Turing's man-imitates-woman game (Turing et al., 1952, p. 495; 1950, p. 441). Why would a behaviourist test the interrogator rather than the machine? The behaviourist must surely say: if the interrogator is fooled, we can infer that the computer's behaviour is appropriately human-like. However, this strategy makes the Turing test a test of machine behaviour only by making it unnecessarily circuitous. And the inference employed is invalid: as many critics have pointed out, we cannot infer from an interrogator's being fooled that the computer's behaviour is equivalent to that of a human being—the interrogator may simply be gullible or the programmer of the machine may be lucky.

The second problem for the behaviourist interpretation is that behaviourism does not explain the *design* of the game. Why would a behaviourist base a test on *deception* rather than merely give the machine a series of cognitive tasks? Moreover, even allowing this deception, why would a behaviourist include a human contestant rather than merely hide the machine?

The third problem for the canonical interpretation of Turing's test is that Turing's own words repudiate behaviourism. For example, in 1948, in 'Intelligent machinery', the paper in which Turing presented his first, chess-playing version of the imitation game, he said: 'The extent to which we regard something as behaving in an intelligent manner is determined as much by our own state of mind and training as by the properties of the object under consideration' (1948, p. 431). This has the consequence that mere behaviour—which is a property of the object under consideration in the imitation game—does *not* suffice for intelligence or thinking.

Moreover, in Turing's view, how we *respond* to the entity matters. For example, he went on to say: 'If we are able to explain and predict its behaviour or if there seems to be little underlying plan, we have little temptation to imagine intelligence' (ibid., p. 431). So much for the behaviourist interpretation.

The canonical interpretation of Turing's imitation game has a rival, which can be seen in a statement by James Moor: 'If a machine passed a rigorous Turing test, then we would have good inductive grounds for attributing intelligence or thinking to it', Moor claims (2001, p. 82). This is an inductive or epistemic interpretation of Turing’s test. It takes intelligence or thinking to consist in inner states or processes. According to Moor, the imitation game provides evidence about these inner states. He says:

> I believe that another human being thinks because his ability to think is part of a theory I have to explain his actions. The theory postulates a number of inner information processes …[T]here is no reason why my knowledge of computer thinking can not arise in the same way. (1976, p. 251)

On this interpretation of Turing’s test, Turing argued as follows: if a computer behaves like a human, then it is likely that it has functionally similar information-processing. If it walks like a duck and quacks like a duck, it’s likely that it has the *innards* of a duck.

This epistemic or inductive view of the test also goes back to the 1950s; at that time it was said, for example, that ‘Turing and others have tried to define thought processes in terms of formal machine states or logic’ (Ritchie and Mays 1957, p. 261). The epistemic view is also held by many today. For example, Tomasio Poggio and Ethan Meyers of the MIT Center for Brains, Minds and Machines, propose, they say, ‘a stronger version of the original Turing test … an open-ended set of Turing++ Questions that we are developing at the Center for Brains, Minds and Machines at MIT’ (2016, p. 73). Their test is intended to uncover information about the machine’s inner states:

> Each question roughly corresponds to a distinct neural nodule in the brain. … Our Turing++ Questions require more than a good imitation of human behavior … [They] check for human-like performance/behavior, human-like physiology, and human-like development. (Ibid., p. 75)

This goes beyond Moor’s hypothesis that the machine’s behaviour is evidence of inner information-processing, to include evidence of brain anatomy and development.

There are significant problems, however, for the epistemic or inductive reading of Turing’s test. Certainly, Turing talks about the ‘state of mind’ of the human computer and by analogy of the machine (1936, pp. 75-6). However, the ‘inner states’ interpretation is not consistent with Turing’s explicit statements about intelligence. Turing did not intend his imitation game as an indirect way of finding out *how* a machine’s behaviour is generated. In a 1952 radio symposium with Turing, Max Newman, Turing’s collaborator at Bletchley Park and Manchester, said, ‘[I]f I have understood Turing’s test properly, you are not allowed to go behind the scenes and criticise the method [by which the machine arrives at its answers], but must abide by the scoring on correct answers, found reasonably quickly’ (Turing et al. 1952, p. 496). Turing agreed with Newman. In fact, rather than focus on hypothesized brain structures generating behaviour, Turing said, ‘As soon as one can see the cause and effect working themselves out in the brain, one regards it as not being thinking, but a sort of unimaginative donkey-work' (ibid., p. 500). Not only is gaining evidence of how the machine produces human-like behaviour not the aim of the test, for Turing no inner state or process constitutes thinking. So much for the rival interpretation.

2.1. Sceptical interpretations of Turing's test

The question at issue is: what justifies Turing's claim that 'the question, "Can machines think?" should be replaced by "Are there imaginable digital computers which would do well in the imitation game?"' (Turing 1950, p. 448). Why should we replace the first question with the second question? The canonical and rival interpretations of the imitation game are attempts to answer this 'Why?' question, but they do so in ways that are inconsistent with Turing's explicit remarks on thinking or intelligence. We need a different account.

It might be suggested, however, that Turing did not have any philosophical principle in mind when devising his imitation game. For example, according to Daniel Dennett, Turing designed his test 'to be nothing more than a philosophical conversation-stopper'; Turing 'proposed—in the spirit of "Put up or shut up!"—a simple test for thinking that was *surely* strong enough to satisfy the sternest skeptic (or so he thought)', Dennett says (1985/2004, p. 296). Nonetheless, Dennett's reading of Turing's test assumes that the imitation game is a test for *thinking*—and therefore that Turing must have had some philosophical answer to the 'Why?' question, even if not a developed answer.

Aaron Sloman goes further than Dennett, saying:

> It is often claimed that Turing was proposing a test for intelligence. I think that assumption is mistaken … [T]here is an alternative, much more defensible, reading of his paper as making a technological prediction, whose main function was to provide a unifying framework for discussing and refuting some common arguments against the possibility of intelligent machines. (2013, p. 606)[1]

One of Sloman's reasons for saying that Turing did not actually propose a test at all is, Sloman says, that Turing was 'far too intelligent to propose a test with so many flaws' (ibid., p. 606). The latter claim (with which I agree) suggests only, however, that the canonical behaviourist interpretation of Turing's test is mistaken—it is on this interpretation that the test is susceptible to so many objections. The difficulty for Sloman's view is that Turing repeatedly called his imitation game a 'test', and even more significantly, a '"*criterion*" for thinking' (Turing 1950, p. 443; see

---

[1] Similarly, Marvin Minsky said that the Turing test is a 'joke' and that Turing 'never intended it as the way to decide whether a machine was really intelligent' (Minsky on Singularity 1 on 1: the Turing test is a joke! Interview with Nikola Danaylov, 2013, *www.youtube.com/watch?v=3PdxQbOvAlI*). Also, according to Drew McDermott, all that Turing wanted to do was to 'shake people's intuitions up' (McDermott 2010).

further Proudfoot 2011, p. 951). We should, I shall presume, begin an investigation into the philosophical principle underlying Turing's test by taking Turing at his word.

If Turing had a philosophical thesis to justify his test, but this was neither behaviourism nor the 'inner state' view of thinking, what was it?

3. The emergence of the 'electronic brain'

On the 31st of October 1946, Admiral the Viscount Mountbatten of Burma gave an address to the British Institution of Radio Engineers (as their President). Mountbatten was no mere figurehead; he had been the Mediterranean Fleet's Wireless Officer during World War II and had invented and patented gadgets. So far as I am aware, his 1946 speech was the first public manifesto of artificial intelligence—or machine intelligence, as it was called in the UK at the time (for more on this speech and the reaction to it, see Proudfoot and Copeland 2019). According to Mountbatten, who had been briefed by the team working on Turing's Automatic Computing Machine: 'It is now considered possible to evolve an electronic brain' (Mountbatten 1946, p. 223).[2] Machines, Mountbatten declared, 'are now being designed to exercise those hitherto human prerogatives of choice and judgement' (ibid., p. 224). In his view, 'we are really facing a new revolution … a revolution of the mind' (ibid., p. 224).

According to one newspaper reporting Mountbatten's speech, the 'whole world of science has been stirred by one revelation ... the electronic brain'.[3] The speech was recounted worldwide, in national, regional, and local newspapers—making the front page of several—and also on radio. It is salutary to note that 21st century dystopian fear of AI was common even in the 1940s and early 1950s.

In the UK, although some participants in this debate were confident that the 'new brain' could 'never take over completely' from the human brain, since it was created by the latter,[4] others feared that 'an aristocracy of super-minds' would reduce human beings to 'moronic button-pushers, lever-pullers, and dial-watchers'.[5] The risk was that humans 'will perish, victims of their own brain products'.[6] Outside the UK, Mountbatten's 'now-famous broadcast' (the *Palestine Post*[7]) generated similar reactions. In Switzerland, *La Sentinelle* said that the new machines gave humanity the

---

[2] In a letter to the *Times* on 13 November, Sir Charles Darwin, Director of the National Physical Laboratory, explained that Mountbatten had been 'fully informed' about the ACE (Darwin 1946).

[3] *The Mercury*, 3 January 1947, p. 5.

[4] *The Motherwell Times* 8 November 1946.

[5] 'The Brain Machine', *The Evening Telegraph and Post*, 2 November 1946, p. 4.

[6] *The Motherwell Times* 8 November 1946.

[7] 'Electronic "Brain"', *Palestine Post*, 7 November 1946.

possibility of ‘sensational progress’ and of solutions ‘in an instant’ to ‘international conflicts’.[8] However, in New Zealand the (Christchurch) *Press* focused on Mountbatten’s warning that scientists could not take up a ‘Pontius Pilate attitude’ to the new computing machines;[9] and the *Galveston Daily News* (in Texas) was alarmed by a future world ‘peopled in part by human beings of the old style and in part by mechanical monsters’.[10] Even in Switzerland, *Die Tat* foresaw ‘the end of the world’ and suggested that ‘man … [might] become a machine himself’.[11]

Several of the new computer scientists joined this public debate, and they did not always agree with Mountbatten. Only a few days after Mountbatten’s speech, on the 7th November, mathematician and computing pioneer Douglas Rayner Hartree wrote to *The Times*, saying that ‘the term “electronic brain” … is misleading in that it ascribes to the machine capabilities that it does not possess’ (Hartree 1946). This term ‘would suggest to the layman that equipment of this kind could “think for itself”, whereas this is just what it cannot do’, Hartree said (1949, p. 70).[12]

It is in the context of this public debate on the question ‘Are computers electronic brains?’ that Turing presented his first version of the imitation game, in 1948.

3.1. Intelligence as an emotional concept

Turing’s first version of the imitation game appeared in the final section of his paper ‘Intelligent Machinery’, under the heading ‘Intelligence as an emotional concept’ (Turing 1948, p. 431).[13] Here Turing wrote the words quoted in Section 2 above:

> The extent to which we regard something as behaving in an intelligent manner is determined as much by our own state of mind and training as by the properties of the object under consideration. (Ibid., p. 431)

And continued, as quoted in part above:

> If we are able to explain and predict its behaviour or if there seems to be little underlying plan, we have little temptation to imagine intelligence. With the same object therefore it is possible that one man would consider it as intelligent and another would not; the second man would have found out the rules of its behaviour. (Ibid., p. 431)

---

[8] ‘Le “cerveau electronique”’, *La Sentinelle* 8 November 1946; ‘Zwischen Gestern und Morgen’, *Die Tat* 3 November 1946.

[9] ‘An “Electronic Brain”: Reference by Lord Mountbatten’, *The Press* 4 November 1946.

[10] *Galveston Daily News* November 4, 1946.

[11] ‘Le “cerveau electronique”’, *La Sentinelle* 8 November 1946; ‘Zwischen Gestern und Morgen’, *Die Tat* 3 November 1946.

[12] Even worse, according to Hartree ‘The fashion which has sprung up in the last 20 years to decry human reason [in favour of machine reasoning] is a path which leads straight to Nazism’ (Hartree quoted in ‘“ACE” will speed jet flying’, *Daily Telegraph*, 8 November 1946).

[13] For Turing’s other uses of the expression ‘emotional’ in the case of machines, see Proudfoot 2014.

Turing's discussion of intelligence as an emotional concept was not purely theoretical. He said, 'It is possible to do a little experiment on these lines, even at the present state of knowledge. It is not difficult to devise a paper machine which will play a not very bad game of chess' (ibid., p. 431). At this time Turing was investigating computer chess routines. In 1948, though, a 'paper machine'—a human being acting the part of the computing machine, following a set of rules with pencil and paper—was the only sort of programmable computer that was easily available.

Next, Turing described the first version of his imitation game. He said:

> Now get three men as subjects for the experiment A, B, C. A and C are to be rather poor chess players, B is the operator who works the paper machine. (In order that he should be able to work it fairly fast, it is advisable that he be both mathematician and chess player.) Two rooms are used with some arrangement for communicating moves, and a game is played between C and either A or the paper machine. C may find it quite difficult to tell which he is playing. (Ibid., p. 431)

Turing added, 'This is a rather idealized form of an experiment I have actually done' (ibid., p. 431).

In his 1948 report Turing made it clear that the imitation game is an experiment to see how the human chessplayer *responds* to the computer. He said, 'Playing against such a machine gives a definite feeling that one is pitting one's wits against something alive' (ibid., p. 412; see Proudfoot 2011 on anthropomorphism and Turing's test).

This is the first version of Turing's imitation game. In the unrestricted 1950 imitation game, C would be replaced by a human conversationalist, who can communicate with A and B on 'almost any one of the fields of human endeavour' (Turing 1950, p. 442).

4. Rethinking Turing's test: a response-dependence approach to intelligence

An 'emotional concept', in current philosophical terminology, is a *response-dependent* concept. Mark Johnston provides the classic statement of a response-dependent concept as one that 'exhibit[s] a conceptual dependence on or interdependence with concepts of our responses in certain specified conditions' (1989, p. 145). Examples of response-dependence concepts are secondary-quality and value concepts: colour, beauty, goodness, and suchlike.

A response-dependence theory of colour might be stated as follows:

x is *red* if and only if, in normal (ideal) conditions, *x looks* red to normal (ideal) subjects. The work of the theory is in making out what will count as normal conditions and normal subjects, in addition to deciding whether this proposition is necessarily true, a priori, reductive, and so on.

Applying this formula in the case of intelligence (or thinking), a basic response-dependence theory of intelligence might be stated as follows:

> *x* is *intelligent* (or *thinks*) if and only if, in normal (ideal) conditions, *x* appears intelligent to normal (ideal) subjects.

Turing's focus was intelligence in computing machines, and his test employs solely a sufficient condition (see Turing 1950, p. 442). So, a response-dependence approach suited to Turing's words looks more like this (where *x* is a computer):

> *x* is *intelligent* (or *thinks*) if, in an unrestricted computer-imitates-human game, *x* appears intelligent to an average interrogator.[14]

If we assume the 1950 version of the imitation game, where the interrogator is allowed to ask questions on 'almost any one of the fields of human endeavour', the normal condition specified is an unrestricted computer-imitates-human game. Turing also specified the normal or ideal subject; he said that the interrogator must be 'average' (ibid., p. 449). AI experts are excluded.

This is not quite the whole story, since Turing was interested in *real-world* machines.

4.1. The threat of a humongous lookup table

Turing was aware of what has become the most famous objection to his test, the threat of a 'humongous lookup table'. In 1952 Max Newman said to Turing:

> I still feel that too much of our argument is about what hypothetical future machines will do. It is all very well to say that a machine could easily be made to do this or that, but, to take only one practical point, what about the time it would take to do it? (Turing et al. 1952, p. 503)

Newman continued

> It would only take an hour or so to make up a routine to make our Manchester machine analyse all possible variations of the game of chess right out, and find the best move that way—if you didn't mind its taking thousands of millions of years to run through the routine. (Ibid., p. 503)
>
> Turing agreed with Newman, saying:
>
> To my mind this time factor is the one question which will involve all the real technical difficulty. If one didn't know already that these things can be done by brains within a reasonable time one might think it hopeless to try with a machine. The fact that a brain *can* do it seems to suggest that the difficulties may not really be so bad as they now seem. (Ibid., p. 503)

---

[14] It would be anachronistic to ask whether Turing proposed a response-dependence understanding of the property, rather than the concept, of intelligence—in fact his remarks suggest both.

Turing's words suggest that we should qualify the response-dependence approach to intelligence, to take account of his focus on the real world.[15] (Response-dependence theorists of colour or goodness frequently relativize their theories to the real world; see e.g. Johnson 1989, Yates 2008.)

Strictly, then, the response-dependence account goes like this:

> *x* is *intelligent* (or *thinks*) if in the actual world, in an unrestricted computer-imitates-human game, *x* appears intelligent to an average interrogator.

The clause 'in the actual world' forestalls the use of exhaustive lookup table programs that could not exist in the real world—such as Block's imaginary Aunt Bubbles machine—as purported counterexamples to Turing's test. Block himself concedes that Aunt Bubbles 'is only logically possible, not physically possible'; the number of strings is 'too vast to exist', he says, and 'even if they could exist, they could never be accessed by any sort of a machine in anything like real time' (1995, p. 381).[16]

To conclude the first part of my argument, I suggest that we rethink Turing's test as follows (see Proudfoot 2013 for more detail). For Turing, 'the idea of "intelligence" is itself emotional rather than mathematical' (Turing 1948, p. 411). This explains why Turing replaced the question 'Can machines think?' with the question 'Are there imaginable digital computers which would do well in the imitation game?'. It also explains why the imitation game is an experiment to test the interrogator's response, rather than the machine's behaviour or inner processing.

This is not to suggest that Turing was other than profoundly interested in the behaviour and architecture of computing machines. However, if intelligence or thinking is a response-dependent concept (or property), a machine's behaviour or inner processing is not sufficient to make the machine *intelligent*—any more than, if colour is response-dependent, the electro-magnetic radiation emitted by the microphysical properties of an object suffices to make that object *red*.

If we rethink Turing's test in this way, what are the philosophical implications? The first relates to the question whether computers can act *freely*.

---

[15] See Copeland 2000b for an alternative way of interpreting Turing's emphasis on real-world machines, to avoid the humongous lookup table objection.

[16] What if Aunt Bubbles could be built in the actual world (because the calculation underlying the claim that the number of strings is 'too vast to exist' is mistaken—or, as is sometimes claimed, as a consequence of technological progress)? Drew McDermott argues that a humongous lookup table machine could be 'computationally equivalent to a very clever program indeed' (2014, p. 144)—if so, on both the behaviourist and inner-state readings of the Turing test, the humongous lookup table machine's passing the test would not count as a false positive.

5. Turing vs Jefferson on the early computers

Neurosurgeon Geoffrey Jefferson participated alongside Turing in the public debate about electronic brains. Jefferson held the first chair of neurosurgery in the UK, at the University of Manchester, and in 1949 the Royal College of Surgeons awarded him its highest honour, the Lister Medal. Jefferson's Lister Oration, on the 9th of June 1949, was titled 'No mind for mechanical man' (for more on this speech and the reaction to it, see Proudfoot and Copeland 2019). In his address, Jefferson said:

> [A]lthough electronic apparatus can probably parallel some of the simpler activities of nerve and spinal-cord, … it still does not take us over the blank wall that confronts us when we come to explore thinking, the ultimate in mind. (1949, p. 1110)

Jefferson acknowledged that computers might be electronic brains, but he thought it too early to say. In his view, 'the extreme variety, flexibility, and complexity of nervous mechanisms' was 'greatly underestimated' by the new computer scientists (ibid., p. 1110).

Jefferson proposed an extremely demanding test for a machine to pass if it were properly to be said to think. He declared:

> Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain—that is, not only write it but know that it had written it. (ibid., p. 1110)[17]

Turing's response to Jefferson was reported in *The Times* two days later. Turing said, 'I do not see why [the computer] should not enter any one of the fields normally covered by the human intellect, and eventually compete on equal terms'.[18]

Again, news of Jefferson's speech—and Turing's response—went viral. Many newspapers were horrified by Turing's claims, claiming in now familiar terms that the 'spectre of the machines rising against their creator … is horrific' and pointed to a future 'technological domination', the outcome of which 'no man can foresee'.[19] Overall, opinion in the papers came down on Jefferson's

---

[17] Jefferson's challenge to Turing is one source of Turing's treatment of (what he called) 'The Argument from Consciousness' (Turing 1950, p. 451). A frequent objection to Turing's test is that it fails as a criterion of thinking in machines just because it does not test for consciousness, and that alternative tests that can be used as a criterion of consciousness are required—for example, the 'AI Consciousness Test', which is presented as a 'zombie filter' (Schneider 2019, p. 56; see also Schneider and Turner 2017). This objection to Turing's test is, however, premised on the mistaken view that Turing's test offers a behaviourist criterion of thinking.

[18] Turing quoted in 'The Mechanical Brain: Answer Found to 300-Year-Old Sum', *Times* [London, England], 11 June 1949, p. 4. In this article, Turing added, 'I do not think you can even draw the line about sonnets, though the comparison is perhaps a little unfair because a sonnet written by a machine will be better appreciated by another machine'.

[19] *The Press and Journal* 23 June 1949. Overseas newspapers, again including local and regional papers, followed suit (see e.g. the *Palestine Post* for 24 June 1949). For example, the *Zanesville Times* Recorder headlined 'Electronic brain can't write poem' (*Zanesville Times Recorder* July 9, 1949); and the *Canton*

side. In a letter to the *Times* on 14 June 1949, the Catholic priest and theologian Illtyd Trethowan said that 'responsible' scientists would 'dissociate themselves' from Turing's views (Trethowan 1949). (On Turing's response to the public's fear of artificial intelligence, see Proudfoot 2015, 2018.)

An important element in the debate between Jefferson and Turing is their very different answers to the question: can computers act freely? According to Jefferson, although much in humans 'can be properly explained by conditioned reflexes and determinism … there is a fringe left over in which free will may act … a fringe that becomes larger and larger the more complex the nervous system' (1949, p. 1107). So, can a computing machine have free will? For Jefferson, the answer is likely *no*. He said, 'It can be urged, and it is cogent argument against the machine, that it can answer only problems given to it, and, furthermore, that the method it employs is one prearranged by its operator' (ibid., p. 1109).

Turing disagreed. In his view, 'Once [the computer] has mastered the rules it will think out its own moves'—the machine, that is, not the programmer.[20] Again, Turing's comments were greeted with shock in the newspapers. Dom Trethowan declared that those who believe that human beings are 'free persons'—which proposition, he claimed, is 'unintelligible' if humans have 'no unextended mind or soul, but only a brain'—must ask 'how far Mr. Turing's opinions are shared, or may come to be shared, by the rulers of our country' (Trethowan 1949).

What exactly was Turing's 'opinion' on free will in machines?

6. Turing on free will: early and late views

In 1927, when Turing was fifteen years old, Arthur Stanley Eddington had given his Gifford Lectures on 'The nature of the physical world', and Turing later took the best-selling corresponding book out of the Sherborne School Library several times, in total for nearly three months.[21] According to Eddington, physical determinism makes free will problematic, but quantum physics permits indeterminism—and so free will. He said, 'It is a consequence of the advent of the quantum that *physics is no longer pledged to a scheme of deterministic law*. … [S]cience thereby withdraws its moral opposition to freewill' (1928, pp. 294-95).

Young Turing took a similar line in an unpublished note titled 'Nature of Spirit'.[22] He said:

Herald reported Jefferson as declaring that electronic brains 'will never be able to bridge the gap between brain and mind' ('Electronic brain is able to solve many human acts', *Canton Herald* July 7, 1949).

[20] Turing quoted in 'Mechanical Brain Is Learning To Play Chess', *The Irish Times*, 13 June 1949, p. 7.

[21] I am grateful to the School Archivist, Rachel Hassall, for this information.

[22] 'Nature of Spirit' is a hand-written manuscript in the Turing Digital Archive, King's College Cambridge, catalogue reference AMT/C/29; all quotations in the text are from this manuscript. The

> [M]odern science … has come to the conclusion that when we are dealing with atoms & electrons we are quite unable to know the exact state of them … This means then that the theory which held that as eclipses etc were predestined so were all our actions, breaks down too.

And continued:

> We have a will which is able to determine the action of the atoms probably in a small portion of the brain, or possibly all over it. The rest of the body acts so as to amplify this.

Strictly speaking, the position expressed here—physical behaviour is determined by the will, which is part of what Turing called the 'spirit'—seems closer to substance determinism than to indeterminism. Turing said:

> I think that spirit is really eternally connected with matter … [T]he body by reason of being a living body can 'attract' & hold on to a 'spirit', whilst the body is alive and awake the two are firmly connected, & when the body is asleep I cannot guess what happens ... (Turing, 'Nature of Spirit')

The young Turing had, it seems, an understanding of free will similar to that of Dom Trethowan, who was so shocked by the adult Turing's views on free will in computing machines (see end of last section).

So, did Turing's views on free will change between 'Nature of Spirit' and his 1949 comments in the *Times*? Not according to several theorists who write about Turing's mature views on free will, and who appear heavily influenced by the young Turing's ideas. According to Hodges, for example, during the war Turing was 'not now very interested in philosophical, as opposed to scientific, discussions of the problem of free will' (2014, pp. 173-74). In his 'last years', however, Hodges claims, Turing 'gave new attention to questions in *physics*'; he was 'trying to formulate "a new quantum mechanics"' (2007, pp. 8, 14). In Hodges' view, Turing was 'drawn to the physics of the brain and the question of its quantum-mechanical basis'—it is this, on Hodges' reading, that underlies Turing's mature view of free will (ibid., p. 17).

Scott Aaronson takes a similar line, claiming that 'one of Turing's obsessions throughout his life [was] the question of physics and free will' (Aaronson 2013). Aaronson cites Turing's remarks on a 1954 postcard, headed 'Messages from the Unseen World' and seemingly spoofing Eddington's (1929) *Science and the Unseen World*. The remarks are as follows: 'The universe is the interior of the light-cone of the Creation. Science is a differential equation. Religion is a

---

manuscript is also transcribed in Hodges 2014, pp. 82-3. It is not known exactly when Turing wrote 'Nature of Spirit'. Andrew Hodges suggests April 1932 (2014, pp. 82, 683); the King's College Archive has this date, probably supplied by Hodges (personal communication from the KCC Archivist, Patricia McGuire).

boundary condition'.[23] Aaronson remarks, 'I think I now understand what Turing could have meant by these remarks', and he proposes 'a viewpoint, suggested … by Turing himself, that tries to find scope for "freedom" in the universe's boundary conditions' (2013). This viewpoint is to solve the problem of free will in a deterministic universe by invoking 'a certain kind of in-principle physical unpredictability' (Aaronson 2013). In attaching this idea to Turing, Aaronson too claims that Turings's solution to the free will problem is based in physics.

I shall argue, however, that Turing's mature understanding of free will does not make the possibility of free will a matter of physics; rather, it has more in common with his approach to intelligence.

7. Self-origination and learning machines

In the 1940s and 1950s, free will was understood by many thinkers as *self-origination* (either rather than, or in addition to, the ability *to do otherwise*). For example, Gardner Williams wrote, 'True freedom is self-determinism or self-determination'; some theorists hold, Williams said, that 'if determinism were true man's state would be hopeless since he would lack all power of spontaneity and creativity' (1941, pp. 709, 707). Similarly, on the subject of free will, Raymond Bradley said that 'the question that really concerns us is whether we are ever entitled to report … that we have a power of *absolute spontaneity* or *self-origination*?' (1958, p. 44).

Following this usage, to ask if computers can act freely is to ask if computers can self-originate. Jefferson, as we have seen, held that the answer to this is *no*, because, he said, a machine's behaviour is 'prearranged'. Hartree too doubted that machines could possess free will, and for the same reason, saying 'These machines can only do precisely what they are instructed to do by the operators who set them up' (1946). This claim goes back, of course, to Ada, Countess Lovelace, who famously wrote:

> It is desirable to guard against the possibility of exaggerated ideas that might arise as to the powers of the Analytical Engine. … The Analytical Engine has no pretensions whatever to *originate* anything. It can do whatever *we know how to order it* to perform. (1843, p. 722)

In his *Mind* paper, Turing considered Lovelace's and Hartree's assertion that a machine cannot self-originate, under the title 'Lady Lovelace's objection' (1950, p. 455). Turing's

---

[23] AMT/D/4 image 14, The Turing Digital Archive. This postcard, sent to Robin Gandy, is dated 8 March 1954 (AMT/D/4 image 13). In a letter to Max Newman, Gandy wrote, 'During this spring [Turing] spent some time inventing a new quantum mechanics; it was not intended to be taken very seriously (almost in the "for amusement only" class)'. According to Gandy, this work showed Turing 'at his most lively and inventive' (letter from Robin Gandy to M.H.A. Newman, n.d. [1955], AMT/A/8 image 1d, Turing Digital Archive). Even so, Gandy remarked, 'no doubt [Turing] hoped that something might turn up in it which could be taken seriously' (ibid.).

discussion of this claim is striking. Just as he replaced the question ‘Can machines think?’ with the question ‘Can imaginable digital computers do well in the imitation game?’, so he proposed a substitute for the question ‘Can machines self-originate?’—a substitution having the result that his answer to the question is *yes*.

Turing said, first, ‘A variant of Lady Lovelace’s objection states that … a machine can never “take us by surprise”.’ (ibid., p. 455). If this were all that Lovelace’s assertion amounted to, Turing already knew that the assertion was false. He said, ‘This statement … can be met directly. Machines take me by surprise with great frequency’ (ibid., p. 455).

By replacing the original question about the possibility of self-origination in machines with a question about the possibility of surprise in the observer of machines, is Turing not, however, merely changing the subject? Turing recognized that Lovelace and others were unlikely to be persuaded by his substitution—he said, ‘I do not expect this reply to silence my critic. He will probably say that such surprises are due to some creative mental act on my part, and reflect no credit on the machine’ (ibid., p. 456). And indeed, Turing said, a computing machine could surprise him just because he had not thought sufficiently about how the machine would behave: he was often surprised by machines, he said, ‘because I do not do sufficient calculation to decide what to expect them to do, or rather because, although I do a calculation, I do it in a hurried, slipshod fashion, taking risks’ (ibid., p. 456). In these cases, Turing acknowledged, his surprise is of no philosophical consequence.

Yet could there be circumstances in which an observer’s being surprised by a computer *is* a criterion of the machine’s self-originating? Turing said that this question ‘will be considered again under the heading of learning machines’ (ibid., p. 455).

By 1950, Turing had been considering the topic of learning for several years; the pioneering computer scientist Donald Michie described himself, Turing, and Jack Good as (at Bletchley Park during World War II) forming ‘a sort of discussion club focused around Turing’s astonishing “child machine” concept’ (Michie 2002).[24] A child-machine is a simple unorganised machine that is to

---

[24] Later Michie said, ‘[The child-machine concept gripped me. I resolved to make machine intelligence my life as soon as such an enterprise became feasible’ (2002). Like Turing, Michie thought that the ‘hallmark’ of intelligence is ‘the ability to learn’ (1989, p. 118). According to Michie, like a ‘newborn baby’, a computer’s possibilities ‘depend upon the education which is fed into it’ (1966). Michie too emphasized the experimenter’s response to the machine, saying, with respect to ‘the child-machine concept, how much of value and use could a school teacher impart to a child with whom rapport was impossible?’ (2001, p. 17).

In the 1960s, Michie built famous early learning machines. His MENACE machine (Matchbox Educable Noughts-And-Crosses Engine) could be trained to improve its game. The FREDERICK robots (Friendly Robot for Education, Discussion and Entertainment, the Retrieval of Information, and the Collation of Knowledge, usually known as FREDDY), built in Michie’s lab at the University of learn and develop as human infants do (see further Proudfoot 2017b). According to Turing, one way to build intelligence in a machine is to begin with ‘a comparatively simple machine, and, by subjecting it to a suitable range of “experience” transform it into one … able to deal with a far greater range of contingencies’ (c. 1951, p. 473). This ‘might be called “education”’, he said (ibid., p. 473). The result of such a process is that ‘[b]it by bit one would be able to allow the machine to make more and more “choices” or “decisions”. [Eventually] interference would no longer be necessary, and the machine would have “grown up”’ (Turing 1948, p. 430).

In such cases, Turing claimed, surprise at the machine’s behaviour *is* a criterion of the machine’s self-originating. He said:

> We should be pleased when the [learning] machine surprises us, in rather the same way as one is pleased when a pupil does something which he had not been explicitly taught to do. (1951, p. 485)

We should, in short, respond to the child-machine as we do to a human infant. And so, Turing concluded:

> If we give the machine a programme which results in its doing something interesting which we had not anticipated I should be inclined to say that the machine *had* originated something, rather than to claim that its behaviour was implicit in the programme. (ibid., p. 485)

This is Turing’s reply to ‘Lady Lovelace’s objection’. In his view, machines can act freely, not in virtue of a quantum indeterminism—in fact Turing explicitly doubted that indeterminism could supply freedom[25]—but in virtue of how we respond to them in relevant circumstances. This suggests a novel compatibilist approach, built on the concept of the observer’s response, to what Turing called ‘an age-old controversy, that of “free will and determinism”’ (1951, p. 484).

8. Response-dependence compatibilism

A basic response-dependence theory of free will, analogous to the response-dependence approach to intelligence or thinking (and to response-dependence theories of colour), would be framed as follows:

> *x* is the *ultimate origin* of *x*’s actions (or *possesses free will*) if and only if, in normal (ideal) conditions, *x* appears to be so to normal (ideal) subjects.

Edinburgh, learned to manipulate various objects, including how to put differently shaped blocks together in order to create a toy. (MENACE was reported in Michie 1961; the first FREDDY robot was built in 1969 and reported in Barrow and Salter 1969.)

[25] According to Turing, a ‘digital computer with a random element’ is sometimes ‘described as having free will’ (1950, p. 445). Turing said, however: ‘I would not use this phrase myself’(ibid.).

As before, most of the work of a response-dependence approach is in specifying normal conditions and subjects. In his discussion of child machines, Turing suggests both.

The normal condition is the ‘education’ of the child machine. As we have seen, the experimenter is to begin with:

> a comparatively simple machine, and, by subjecting it to a suitable range of ‘experience’ transform it into one which was more elaborate, and was able to deal with a far greater range of contingencies. (Turing c. 1951, p. 473)

Normal subjects are also specified. According to Turing:

> [T]he education of the machine should be entrusted to some highly competent schoolmaster who is interested in the project but who is forbidden any detailed knowledge of the inner workings of the machine (Ibid., p. 473)

Similarly, Turing said:

> An important feature of a learning machine is that its teacher will often be very largely ignorant of quite what is going on inside, although he may still be able to some extent to predict his pupil’s behaviour. (1950, p. 462)

Putting these elements together, we have Turing’s version of response-dependence compatibilism, where the entity concerned is a computer (see further Proudfoot 2017a). Again, the basic account goes like this, where the adjective ‘average’—Turing’s—indicates only that the teacher, though ‘highly competent’ is ‘forbidden any detailed knowledge’ of the machine’s architecture:

> *x* is the *ultimate origin* of *x*’s actions (or *possesses free will*) if, when teaching *x* to generalize from experience, *x* appears to be so to average educators.

Since there may well be other normal contexts, this can only be a sufficient condition of free will. In addition, as we have already seen, Turing is interested in real-world machines. So the just-stated formula should be modified to reflect this, as follows:

> *x* is the *ultimate origin* of *x*’s actions (or *possesses free will*) if in the actual world, when teaching *x* to generalize from experience, *x* appears to be so to average educators.

That Turing intended his account of the education of child machines to constitute a reply to ‘Lady Lovelace’s objection’ is clear from his remarking that ‘[t]his whole question’ would be considered again ‘under the heading of learning machines’ (ibid., p. 455). After describing the ‘largely ignorant’ teacher of the child machine, Turing emphasised the significance of this discussion for ‘Lady Lovelace’s objection’, saying: ‘The view that “the machine can only do what we know how to order it to do”’—this is the view of Lovelace, Hartree, and Jefferson—‘appears strange in face of this’ (ibid., p. 462).

In sum, Turing proceeds in exactly the same fashion with respect to intelligence (or thinking) and to free will. In the case of intelligence or thinking, he took the question 'Can machines think?' and replaced it with the question 'Can imaginable digital computers do well in the imitation game?'. Turing's reason for this replacement is that intelligence (or thinking) is a response-dependent concept.[26] Analogously, he took the question 'Can machines act freely?' and replaced it with the question 'Can learning machines take us by surprise?'. Response-dependence compatibilism justifies this move. For Turing, free will too is an emotional concept.[27]

9. Turing on computationalism—the standard story

Mountbatten's 1946 prediction of 'a revolution of the mind' was far-sighted. Today many people answer the question 'Are computers electronic brains?' with a resounding *yes*. The very different question facing us now is: Are brains (or minds) biological computers? Or, more plausibly: Do biological brains perform computations? In today's philosophy of mind and cognitive science, the standard answer to this last question is also *yes*. We are mostly computationalists now, it seems—even if some modern scientists reject computationalism, often echoing Jefferson's caution in think-pieces in the press.[28] For example, Rodney Brooks claims that 'the computational metaphor leads researchers to ask questions today that will one day seem quaint, at best' (2015, p. 296). According to Matthew Cobb, the brain-as-computer metaphor 'may be reaching its limits' (2020a, pp. 7; see also Cobb 2020b). Yet even Cobb agrees that for now this metaphor 'retains its dominance' (2020a, p. 377).

Perhaps the most famous assertion of the dominance of computationalism is Jerry Fodor's claim that the 'only psychological models of cognitive processes that seem even remotely plausible represent such processes as computational' (1975, p. 27). Many recent philosophers endorse Fodor's sentiment. Gualtiero Piccinini puts the philosophical state of play as follows:

> [C]omputationalism says that cognition is computation. Computationalism has been mainstream in philosophy of mind—as well as psychology and neuroscience—for several decades. … [It is] the most serious, and ambitious mechanistic explanation of cognition and intelligent behavior to date. (2009, pp. 515, 518)

---

[26] Again, it would be anachronistic to ask whether Turing proposed a response-dependence approach to the property, rather than the concept, of freewill.

[27] Even if the concept of free will is response-dependent, it may be that in fact only entities with certain observer-independent properties—for example, being equipped with a 'random element'—generate the appropriate response in observers.

[28] Cobb calls for 'modesty' and 'realism' with respect to 'the difficulties of drawing parallels between brains and artificial systems' (2020a, p. 379).

The question seemingly facing us today is which flavour of computationalism (or of the computational theory of mind[29]) to endorse—including, among recent theories, 'info-computationalism' (Dodig-Crnkovic 2012), 'ego-cognitive computationalism' (Magnani 2018), 'enactive computationalism' (Estrada 2018), and so on. Computationalism is a broad church, with a diverse membership.

There are, of course, deep questions as to the nature of computation and in addition major philosophical debates around computationalism. The latter includes arguments over whether computation is essentially semantic (e.g. Shagrir 2006, 2018; Sprevak 2010; Villalobos and Dewhurst 2017), 'iterative' (Copeland 1996), 'mechanistic' (Copeland 2000a; Piccinini 2015), and/or 'contextual' (Shagrir and Harbecke 2019). In addition, there are debates over whether analog computation is a species of computation (e.g. Shagrir 2010, Maley 2018), and if we should endorse 'ontological' (or 'realist') computationalism rather than 'instrumentalist' computationalism (see Villalobos and Dewhurst 2017). Even given this range of views, however, it can plausibly be said that 'computationalism is not only here to stay, it becomes stronger every year' (Milkowski 2018a, p. 537).

Was *Turing* a computationalist?

9.1. Turing's legacy?

Computationalism is frequently credited to Turing. In 1995 Herbert Simon claimed that Turing 'was perhaps the first to have this insight'—that the 'materials of thought are symbols'—'in clear form' (1995, p. 24).[30] Today, natural info-computationalism, for example, claims to build on 'the legacy of Turing's computationalism' (Dodig-Crnkovic 2012, p. 1). It is widely assumed, both in philosophy and in theoretical computer science, that Turing proposed a computational theory of mind.[31] In this section I describe three examples.

According to Fodor:

---

[29] For the argument that 'computationalism' about the mind differs from the 'computational theory of mind', see e.g. Miłkowski 2018. Likewise, inclusive views of computationalism sometimes refer to the 'computational approach' or 'computational paradigm', and even to computationalism as a research 'programme' or 'tradition', rather than as a theory.

[30] According to Piccinini, 'Contrary to a popular belief, modern computationalism is not due to Alan Turing but to Warren McCulloch and Walter Pitts' (2009, p. 517). In Piccinini's view, McCulloch and Pitts' famous 1943 paper proposes the 'the first modern computational theory of mind and brain' (2004, p. 176). Similarly, Morgan and Piccinini claim that 'the first rigorous computational theory of the mindbrain [was] proposed by McCulloch and Pitts' (2018, p. 123).

[31] Sometimes the attribution is implicit. For example, Carrie Figdor writes: 'until Turing, we lacked an empirically plausible model of how the mind could be material' (2018, p. 283).

> The cognitive science that started fifty years or so ago more or less explicitly had as its defining project to examine a theory, largely owing to Turing, that cognitive mental processes are operations defined on syntactically structured mental representations (2001, pp. 3-4)

Fodor explicitly attributes to Turing a computational theory of mind. He refers to: ‘Turing’s idea that mental architecture is computational in the proprietary syntactic sense’, ‘Turing’s kind of computational psychology’, and ‘Turing’s theory that thinking is computation’ (ibid., pp. 19, 20, 5).

Fodor, of course, was a fan of computationalism. Critics of computationalism, however, also credit Turing with giving us computationalism. For example, John Searle, a famous opponent of computationalism and of Turing’s test, criticizes what he calls the ‘primal story’ concerning the relation of human intelligence to computation—this story, he claims, generates ‘a whole iceberg of problems’ (1994, pp. 202, 207). The story, Searle says, goes as follows. Some human mental abilities (e.g. long-division) are algorithmic; in such a case, Searle says, ‘as described by Turing (1950), both I, the human computer, and the mechanical computer are implementing the same algorithm. I am doing it consciously, the mechanical computer nonconsciously’ (ibid., pp. 202-3). The story continues:

> Now it seems reasonable to suppose that there might be a whole lot of other mental processes going on in my brain nonconsciously that are also computational. And if so, we could find out how the brain works by simulating these very processes on a digital computer. … we could get a computer simulation of the processes for understanding language, visual perception, categorization, etc. … Now suppose that mental contents in the head are expressed syntactically in the head, then all we would need to account for mental processes would be computational processes between the syntactical elements in the head. (Ibid., pp. 202-3)

This story—a traditional form of the computational theory of mind—‘goes back at least to Turing’s classic paper’, Searle claims (ibid., p. 202).

Theoretical computer scientists as well as philosophers attribute computationalism to Turing. Eric Baum, for example, claims that (what he calls) ‘Turing’s “analysis of thought”’ defines ‘what is sometimes known as strong Artificial Intelligence—the thesis that the mind is equivalent to a computer program’ (2006, p. 6) According to Baum, the ‘computer science viewpoint on mind was basically formulated by Turing’ (ibid., p. 69). What is this viewpoint? Baum says:

> For every mental state you have, there is an exact corresponding state that the Turing machine tape passes through. … [F]rom the point of view of computation, the machine and you are in one-to-one correspondence (Ibid., p. 70)

And so, Baum concludes, ‘there is no reasonable sense in which the computer could be said not to be as “conscious” as you are’ (ibid., p. 70).

The view that computationalism goes back to Turing is sufficiently widespread that we can justifiably regard it as the standard account of Turing’s view of the mind (although, for doubts and difficulties facing this account, see Copeland and Shagrir 2013, Copeland 2000a). But can this story be correct, once we rethink Turing’s test?

10. Computationalism vs emotional-concept theory

To examine whether the standard account is true, I shall take as my focal case of computationalism what David Chalmers calls ‘minimal computationalism’ (Chalmers 2011). According to Chalmers, minimal computationalism is ‘compatible with a very wide variety of empirical approaches to the mind’, and thus computation can serve as ‘a true foundation for cognitive science’ (ibid., p. 324).[32] Minimal computationalism is, so to speak, a maximally inclusive computationalism; for example, it does not assume that computation is essentially semantic, and nor does it exclude analog computation. It would be difficult indeed to maintain that Turing is a computationalist if his views about the mind are inconsistent with minimal computationalism.

Minimal computationalism, Chalmers says, ‘is defined by the twin theses of computational sufficiency and computational explanation’ (ibid., p. 354). According to the first thesis, the ‘right kind of computational structure suffices for the possession of a mind, and for the possession of a wide variety of mental properties’. According to the second, computation provides ‘a general framework for the explanation of cognitive processes and of behavior’ (ibid., p. 324). The first thesis makes the grander claim, philosophically speaking. It leads to the declaration that ‘a model that is computationally equivalent to a mind will itself be a mind’ (ibid., p. 344). In Chalmers’ view, AI and computational cognitive science are ‘committed’ to minimal computationalism (ibid., p. 349).

As I shall argue, however, the thesis of computational sufficiency is inconsistent with a response-dependence approach to intelligence or thinking.

10.1. Upsetting the standard story

The minimal computationalist says, as we have just seen, that the ‘right kind’ of computational structure ‘suffices’ for the possession of a ‘wide variety of mental properties’. Turing, however, as

---

[32] The view that cognitive science depends on computationalism is widespread; for example, ‘Cognitive science was founded on a computational theory of mind’ (Morgan and Piccinini 2018, p. 123).

we have also seen, said, ‘As soon as one can see the cause and effect working themselves out in the brain, one regards it as not being thinking, but a sort of unimaginative donkey-work’. In the case of intelligence or thinking, response matters, not solely internal processing. It follows that computational structure, of whatever kind, is insufficient for intelligence or thinking. I shall call this the *response-dependence* objection to computationalism. In view of its origin, it can also be called the emotional-concept objection to computationalism.

The dispute between the computationalist and Turing here is analogous to the dispute between the sort of physicalist who claims that colour just is electromagnetic radiation and the response-dependence theorist who insists that response is part of the very concept of colour, even if radiation is part of the explanation of why humans in fact respond as they do. If thinking is an emotional concept, internal goings-on are insufficient for thinking, even if (in the thinker or the observer) they are part of the explanation why humans in fact respond as they do.

The minimal computationalist also claims that a model that is ‘computationally equivalent’ to a mind will itself be a ‘mind’. But again Turing said, ‘The extent to which we regard something as behaving in an intelligent manner is determined as much by our own state of mind and training as by the properties of the object under consideration.’ If thinking is a necessary property of mind, then it is false that a model computationally equivalent to a mind will itself be a mind *just in virtue of that computational equivalence*. That is again because, according to Turing, computational structure does not suffice for intelligence or thinking.

An analogous argument can be framed in terms of free will. It follows from Turing’s approach to free will that the ‘right kind’ of computational structure does not suffice for free will: the right kind of response matters (we find the machine ‘doing something interesting which we had not anticipated’). If—as Jefferson and many of Turing’s critics appear to believe—free will is an essential property of mind, then it is false that a model computationally equivalent to a mind will itself be a mind (solely in virtue of that computational equivalence). If this argument is correct, internal goings-on are not sufficient for free will, and hence not sufficient for mind.

I conclude that Turing’s criterion for thinking and his approach to free will are both at odds with minimal computationalism.

10.2. Differentiating the response-dependence objection

Computationalists (and others) argue that extant objections to computationalism are inadequate. For example, in Marcus’s view, ‘all the standard arguments about why the brain might not be a computer are pretty weak' (2015).[33] According to Piccinini, the 'periodic reports' of the 'demise' of computationalism are 'greatly exaggerated', and none of the objections to computationalism are 'conclusive' (2009, pp. 515, 524). Milkowski claims that most criticisms of computationalism 'fail' and that subscribing to these objections is 'probably a matter of ideology rather than rational debate' (2018b, p. 71).

Certainly, many objections to computationalism do not apply to minimal computationalism; they attack less inclusive, more specific forms of computationalism—for example, the claim that the mind is a serial machine, that its operations are digital, that computation is essentially semantic, that the brain is modular or localized, and so on. In this respect, Milkowski's (2018b, p. 58) claim that a number of objections to computationalism are 'red herrings' is accurate; many objections target elements of only some computationalist accounts. For this reason, Milkowski says that the 'majority' of objections to computationalism 'make computationalism a straw man' (2018b, p. 58). The response-dependence objection, however, does not attack such a straw man (2018b, p. 58). If intelligence or thinking is an emotional concept, as Turing claimed, then no internal processing—serial or parallel, semantic or mechanistic, symbolic or nonsymbolic, localized or distributed, discrete or continuous, biological or nonbiological—suffices for intelligence or thinking.

Moreover, the response-dependence objection is very different from other much-discussed criticisms of computationalism. For example, there are objections based on Gödel's theorem (e.g. Lucas 1961; Penrose 1991),[34] objections pertaining to consciousness (e.g. Searle 1990; Chalmers 1995),[35] and objections arguing for the 'triviality' of the computational theory of mind (e.g. Putnam 1988; Searle 1994). Computationalists' replies to those objections do not undermine the response-dependence objection.

There are other recent objections to computationalism that might appear to have some commonality with the response-dependence objection. For example, enactivists (and other proponents of 'E-approaches') claim that cognition is a matter of 'interaction' with the environment (see e.g. Kiverstein and Miller 2015; Raja and Anderson 2019; Hutto and Myin 2017, Hutto et al. 2019). According to a canonical enactivist, the mind is 'relational'—'a way of being in the world'—and for this reason computationalism is 'confused' (Evan Thompson in Heuman 2014)?[36] However,

---

[33] 'If the brain is not a serial algorithm-crunching machine … what is it?', Marcus asks (2015).

[34] For Turing's reply to such arguments, see his treatment of the 'Mathematical Objection' (Turing 1950, p. 450); see Copeland and Shagrir 2013 for a detailed analysis and assessment of Turing's reply.

[35] For Turing's response to considerations pertaining to consciousness, see his treatment of the 'Argument from Consciousness'.

[36] Shaun Gallagher makes the analogous claim that free will is relational: in his view, 'it is better to conceive of autonomy as relational, rather than as a pre-established character of human nature' (2019, p. 805). Gallagher's target appears to be a localized account of free will—'Understood enactively, freely

this superficial similarity obscures key differences between the enactivists' view and Turing's claim that intelligence is an emotional concept. First, enactivists typically oppose *specific* forms of computationalism about the mind—for example, (as in Kiverstein and Miller 2015) the hypothesis that psychological functions are localized in the brain, or (as in Hutto and Myin 2013) the hypothesis that cognition necessarily involves semantic content.[37] Second, enactivists typically conceive of cognition in terms of 'extensive and dynamically loopy processes that are responsive to information in the form of environmental variables' (Hutto and Myin 2017, p. 9). This posits a *causal* interaction—a feedback loop between processes in an individual's brain, body, and environment. For Turing, in contrast, whether the 'enactive, embodied activity'[38] of an entity is indeed intelligent or thinking is a yet further matter, determined in part by the observer.

Another example of an objection to computationalism that might seem akin to the response-dependence objection is Searle's claim that computation is 'observer relative', and so the theory that the brain is a digital computer lacks 'a clear sense' and does not even reach 'the level of falsehood' (Searle 1994, p. 225). Searle's view, however, is very different from Turing's. For Searle, what is determined in part by an observer is whether a machine performs computations; for Turing, it is whether a machine that is performing computations is intelligent (or thinks). Indeed, Turing's view is opposed to Searle's: according to the latter, only 'intrinsic' intentionality—i.e., intentionality that exists 'independently of any observer'—is 'genuinely mental' (ibid., pp. xiii, 156).[39] If thinking is an emotional concept, however, there is no such thing as 'intrinsic' thinking.[40]

11. **Conclusion**

Turing's claim that intelligence is an emotional concept is plainly inconsistent with the thesis of computational sufficiency. How, then, can this be reconciled with Turing's reference to the human brain as a 'digital computing machine', his claim that the 'infant human cortex' is an 'unorganised machine' (a network of logic gates), or his remark that in designing the Automatic Computing

---

willed action is not something that occurs in the head', he says (Gallagher 2017, p. 148). On his alternative approach, what matters for free will is that the individual possesses a certain property—'a specific type of consciousness … *embedded* or *situated* in the particular context' (ibid., p. 145). This is a very different view from the response-dependent account of free will that I have outlined in this paper.

[37] According to Kiverstein and Miller 2015, the function of any brain region is determined by 'its interactions with the other elements to which it is connected in a network' (p. 6). According to Hutto and Myin 2013, we can understand cognition 'as involving a complex series of systematic—but not contentfully mediated—interactions between well-tuned mechanisms' (p. 71).

[38] To use the enactivists' vocabulary (Hutto and Myin 2017, p. 10).

[39] For Searle, attributions of intentionality to computers are merely 'metaphorical or as-if' (1994, p. 156).

[40] Further, if there is no such thing as 'intrinsic' thinking, and if thinking is a necessary property of mind, then there is nothing 'genuinely mental' that is observer-independent.

Engine his interest was in 'producing models of the action of the brain' (Turing 1947, p. 382; 1948, p. 432; c. 1946)? The solution is simple. According to Turing, the brain is a computing machine, but whether it possesses certain properties—such as being intelligent, thinking, or possessing free will—depends upon the observer's response to the computing machine. Applying Turing's approach to another 'age-old controversy', the mind-body problem, we obtain a bold and interesting take on that problem: the brain is computational, but *not* the mind.[41]

The difficulties set out in this paper for the standard claims about Turing—that he was a behaviourist, an indeterminist with respect to free will, and a proponent (even the first) of the computational theory of mind—demonstrate the ongoing need to re-examine and rethink Turing's philosophical writings. On the response-dependence interpretation, Turing suggested a subtle and novel approach to the concepts of intelligence or thinking and free will, and ultimately to the concept of mind. Developing these ideas is of more than exegetical or scholarly importance. The ideas offer the possibility of philosophical progress, in contrast to the tired old debates that have dominated seventy years' discussion of Turing's test.

Acknowledgements: I am grateful to Jack Copeland and to three anonymous reviewers for *Minds and Machines* for their helpful comments on an earlier draft of this paper.

---

[41] Computationalism appears to be formulated chiefly as a thesis about the mind (e.g. 'Computationalism in the philosophy of mind is the view that mental processes, including perceptual processes, are computational' (Nico 2019)). On the other hand, some still say that computationalism is 'the view that the brain is some kind of computer' (Maley 2018, p. 78). Other theorists assume that there is no need for a distinction; thus Morgan and Piccinnini attribute 'the first rigorous computational theory of the mindbrain' to McCulloch and Pitts (1943)' (2018, p. 123).

References

Aaronson, S. (2013) The Ghost in the Quantum Turing Machine. arXiv:1306.0159v2 [quant-ph] 11 Oct 2013.

Barrow, H.G. and Salter, S.H. (1969). Design of low-cost equipment for cognitive robot research, *Machine Intelligence 5* (ed. B. Meltzer and D. Michie), Edinburgh University Press, pp. 555-66.

Baum, T. (2006). *What is thought?* Revised edition. Bradford Books, MIT Press: Cambridge, Mass.

Block, N. (1981) Psychologism and Behaviorism. *Philosophical Review* 90(1): 5-43.

Block, N. (1995) The Mind as the Software of the Brain. In E.E. Smith and D.N. Osherson (eds) *An Invitation to Cognitive Science, 2nd edition, Volume 3*. Cambridge, Mass.: MIT Press, pp. 377-425.

Bradley, R.D. (1958). Free will: Problem or pseudo-problem? *Australasian Journal of Philosophy* 36(1): 33-45.

Brooks, R.A. (2015). The computational metaphor. In J. Brockman (ed.), *This Idea Must Die: Scientific Theories That Are Blocking Progress*. New York: Harper Perennial, pp. 295-8.

Chalmers, D. Facing up to the problem of consciousness. *Journal of Consciousness Studies* 2(3): 200-219.

Chalmers, D. J. (2011). A Computational Foundation for the Study of Cognition. *Journal of Cognitive Science* 12(4): 323-57.

Cobb, M. (2020a). *The idea of the brain: the past and future of neuroscience*. Basic Books.

Cobb, M. (2020b). Why your brain is not a computer. *The Guardian* 27 February 2020. https://www.theguardian.com/science/2020/feb/27/why-your-brain-is-not-a-computer-neuroscience-neural-networks-consciousness

Copeland, B.J. (1996). What is computation? *Synthese* 108: 335-359.

Copeland, B.J. (2000a). Narrow Versus Wide Mechanism. *Journal of Philosophy* 96(1): 5-32.

Copeland, B.J. (2000b). The Turing test. *Minds and Machines* 10: 519–539.

Copeland, B.J. (ed.) (2004) *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, Plus The Secrets of Enigma*. Oxford: Oxford University Press.

Copeland, B.J. (2013) Turing and the Physics of the Mind. In S.B. Cooper and J. van Leeuwen (eds), *Alan Turing: His Work and Impact* (Amsterdam: Elsevier), pp. 651-666.

Copeland, B.J. and Proudfoot, D. (1996) On Alan Turing's Anticipation of Connectionism. *Synthese* 108 (1996): 361-7. Reprinted in R. Chrisley (ed.), *Artificial Intelligence: Critical Concepts in Cognitive Science, Volume 2: Symbolic AI*. London: Routledge, 2000.

Copeland, B.J. and Proudfoot, D. (2017) Connectionism: Computing with Neurons. In Jack Copeland et al., *The Turing Guide*. Oxford University Press, pp. 309-314.

Copeland, B.J. and Shagrir, O. (2013) Turing versus Gödel on Computability and the Mind. In Copeland, B.J., Posy, C.J., and Shagrir, O. (eds) *Computability: Turing, Gödel, Church, and Beyond*. Cambridge, MA: MIT Press, pp. 1-33.

Copeland, J. et al. (2017). *The Turing Guide*. Oxford: Oxford University Press.

Darwin, C.G. (1946) The 'Electronic Brain'. *The Times* [London, England] 13 November 1946, p. 7.

Dennett, D.C. (1985/2004). Can machines think? In Michael G. Shafto (ed.), *How We Know: Nobel Conference XX*, Harper & Row, San Francisco, CA, 1985. Reprinted in C. Teuscher (ed.), *Alan Turing: Life and Legacy of a Great Thinker*. Springer-Verlag Berlin Heidelberg, pp. 295-316. Page references are to the 2004 reprint.

Dodig-Crnkovic, G. (2012). Alan Turing's Legacy: Info-Computational Philosophy of Nature. arXiv:1207.1033 [cs.GL] (2012).

Eddington, A.S. (1928) *The Nature of the Physical World*. Cambridge: Cambridge University Press. Reprinted in 1948.

Eddington, A.S. (1929) *Science and the Unseen World*. Swarthmore Lecture 1929. New York: Macmillan.

Estrada, D. (2018). Conscious Enactive Computation. arXiv:1812.02578 [cs.AI]

Figdor, C. (2018). The Rise of Cognitive Science in the 20th Century. In A. Kind (ed.), *Philosophy of Mind in the Twentieth and Twenty-First Centuries*. Abingdon, Oxon. and New York: Routledge, pp. 280-302.

Fodor, J.A. (2001). *The Mind Doesn't Work that Way: The Scope and Limits of Computational Psychology*. Bradford Books: Cambridge, MA.

Gallagher, S. (2017). *Enactivist Interventions: Rethinking the Mind*. Oxford University Press.

Gallagher, S. (2019). Precis: Enactivist Interventions. *Philosophical Studies* 176: 803-806.

Hartree, D.R. (1946). The 'Electronic Brain'. *The Times* [London, England] 7 November 1946, p. 5.

Hartree, D.R. (1949) *Calculating Instruments and Machines*. Urbana: University of Illinois Press.

Heuman, L. (2014). The Embodied Mind: An interview with philosopher Evan Thompson. *Tricycle: the Buddhist Review* 24(1). https://tricycle.org/magazine/embodied-mind

Hodges, A. (2007). Alan Turing, the logical and physical basis of computing. British Computer Society. https://pdfs.semanticscholar.org/5496/25087e8dea2c478925b8f98c8af6b9ea0839.pdf?_ga=2.68059476.410227184.1583899374-1397298405.1583713945

Hodges, A. (2014) *Alan Turing: The Enigma*, revised edition. London: Vintage.

Hutto, D.E. and Myin, E. (2013). *Radicalizing Enactivism: Basic Minds Without Content*. Cambridge, MA: MIT Press.

Hutto, D.E. and Myin, E. (2017). *Evolving Enactivism: Basic Minds Meet Content*. Cambridge, MA: MIT Press.

Hutto, D.D., Myin, E., Peeters, A. and Zahnoun, F. (2019). The cognitive basis of computation Putting computation in its place. In Sprevak M., Colombo M. (eds), *The Routledge Handbook of the Computational Mind*. Routledge, London, pp. 272-282.

Jefferson, G. (1949). The Mind of Mechanical Man. *British Medical Journal* 1(4616): 1105-10.

Johnston, M. (1989) Dispositional Theories of Value. *Proceedings of the Aristotelian Society, Suppl. Vol.* 63: 139-174.

Johnston, M. (1989). Dispositional Theories of Value, *Proceedings of the Aristotelian Society*, Supplementary Volumes 63(1): 139–74.

Kiverstein, J. and Miller, M. (2015). The embodied brain: towards a radical embodied cognitive neuroscience. *Frontiers in Human Neuroscience* 9, 237. doi.org/10.3389/fnhum.2015.00237.

Lloyd, S. (2012) A Turing test for free will. *Philosophical Transactions of the Royal Society A* 370: 3597-3610.

Lovelace, A. A. (1843) Notes by the Translator (addenda to her translation of L.F. Menabrea, 'Sketch of The Analytical Engine invented by Charles Babbage'). In R. Taylor (ed.) *Scientific Memoirs, Selected from the Transactions of Foreign Academies of Science and Learned Societies, and from Foreign Journals, Volume 3*, pp. 691-731. London: Richard and John E. Taylor.

Lowe, E.J. (2008) *Personal Agency: The Metaphysics of Mind and Action*. Oxford: Oxford University Press.

Lucas, J.R. (1961). Minds, Machines, and Gödel. *Philosophy* 36: 112-27.

Magnani, L. (2018). Eco-Cognitive Computationalism: From Mimetic Minds to Morphology-Based Enhancement of Mimetic Bodies. *Entropy* 20(6), 430.

Maley, C.J. (2018). Toward Analog Neural Computation. *Minds and Machines* 28(1): 77-91.

Marcus, G. (2015). Face It: Your Brain Is a Computer. *The New York Times* June 27, 2015.

Mays, W. (1952). Can Machines Think? *Philosophy* 27(101): 148–62.

McCulloch, W. S. and Pitts, W. H. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity, *Bulletin of Mathematical Biophysics* 7: 115–133.

McDermott, D. (2010). Don't improve the Turing test, abandon it. Address to the *Towards a Comprehensive Turing Test (TCIT)* workshop, 2010 AISB Convention in Leicester, UK, *www.youtube.com/watch?v=zXEB9ctWJp8*.

McDermott, D. (2014). On the Claim that a Table-Lookup Program Could Pass the Turing Test. *Minds and Machines* 24: 143-188.

Michie, D. (1961). Trial and error, *Science Survey 1961:* Part 2 (ed. S. A. Barnett and A. McLaren), Penguin (1961), 129–45; reprinted in D. Michie, *On Machine Intelligence*, 2nd edition. Chichester: Ellis Horwood (1986), pp. 11–23.

Michie, D. (1966). Edinburgh will set the pace, *The Scotsman* (17 February 1966), p. 11.

Michie, D. (1989). The Turing Institute: an experiment in cooperation', *Interdisciplinary Science Reviews* 14(2): 117–19.

Michie, D. (2001). Return of the imitation game [revised version]. *Linköping Electronic Articles in Computer and Information Science*, 6(28): 1–17.

Michie, D. (2002). Recollections of early AI in Britain: 1942–1965. Transcript of talk given by Donald Michie at the Computer Conservation Society's Artificial Intelligence: Recollections of the Pioneers seminar on 11 October 2002, The Artificial Intelligence Applications Institute (http://www.aiai.ed.ac. uk/events/ccs2002/CCS-early-british-ai-dmichie.pdf);

Miłkowski, M. (2018a). From Computer Metaphor to Computational Modeling: The Evolution of Computationalism. *Minds and Machines* 28: 515–541.

Miłkowski, M. (2018b). Objections to Computationalism: A survey. *Roczniki Filozoficzne/Annales de Philosophie/Annals of Philosophy* 66(3): 57-76.

Moor, J.H. (1976) An Analysis of the Turing Test. *Philosophical Studies* 30(4): 249–57.

Moor, J.H. (2001). The Status and Future of the Turing Test. *Minds and Machines* 11: 77–93.

Morgan, A. and Piccinini, G. (2018). Towards a Cognitive Neuroscience of Intentionality. *Minds and Machines* 28(1): 119-139.

Mountbatten, L. (1946) The Presidential Address, *Journal of the British Institution of Radio Engineers* 6(6): 221-225.

Nico, O. (2019). Perception without computation? In Sprevak M., Colombo M. (eds), *The Routledge Handbook of the Computational Mind*, pp. 410-423.

Penrose, R. (1991) The Emperor's New Mind. *RSA Journal* 139(5420): 506-514.

Pettit, P. (1991) Realism and Response-Dependence. *Mind* 100(4): 587-626.

Piccinini, G. (2004).The first computational theory of mind and brain: a close look at McCulloch and Pitts's 'Logical calculus of ideas immanent in nervous activity'. *Synthese* 141: 175–215.

Piccinini, G. (2009). Computationalism in the Philosophy of Mind. *Philosophy Compass* 4(3): 515–53.

Piccinini, G. (2015). *Physical Computation: A Mechanistic Account*. Oxford: Oxford University Press.

Poggio, T. and Meyers, E. (2016). Turing++ Questions: A Test for the Science of (Human) Intelligence. *AI Magazine* 37(1): 73-77.

Proudfoot, D. (2011). Anthropomorphism and AI: Turing's much misunderstood imitation game. *Artificial Intelligence* 175(5-6): 950-7.

Proudfoot, D. (2013). Rethinking Turing's Test. *Journal of Philosophy* 110(7): 391-411.

Proudfoot, D. (2014). Turing's Three Senses of 'Emotional'. *International Journal of Synthetic Emotions* 5(2): 7-20.

Proudfoot, D. (2015) Mocking AI Panic. *Spectrum* 52(7): 46-7, 30 June 2015.

Proudfoot, D. (2017a). Turing and Free Will: A New Take on An Old Debate. In J. Floyd and A. Bokulich (eds), *Philosophical Explorations of the Legacy of Alan Turing: Turing 100*. Boston Studies in the Philosophy and History of Science. Berlin: Springer Verlag, pp. 305-322.

Proudfoot, D. (2017b). Turing's Child Machines. In Jack Copeland et al., *The Turing Guide*. Oxford University Press, pp. 315-25.

Proudfoot, D. (2018) Alan Turing and evil Artificial Intelligence. *OUP Blog*, 19 January 2018. https://blog.oup.com/2018/01/alan-turing-evil-artificial-intelligence/

Proudfoot, D. and Copeland, J. (2019) Turing and the first electronic brains: what the papers said. In Sprevak M., Colombo M. (eds), *The Routledge Handbook of the Computational Mind*. Routledge, London, pp. 23-37.

Putnam, H. (1988) *Representation and Reality*. Cambridge, MA: MIT Press.

Raja, V. and Anderson, M.L. (2019). Radical Embodied Cognitive Neuroscience. *Ecological Psychology* 31(3): 166-181.

Ritchie, A.D. and Mays, W. (1957). Thinking and machines. *Philosophy* 32(122): 261.

Schneider, S. (2019). *Artificial You: AI and the Future of Your Mind*. Princeton, NJ: Princeton University Press

Schneider, S. and Turner, E. (2017). Is Anyone Home? A Way to Find Out If AI Has Become Self-Aware. *Scientific American* July 19, 2017. https://blogs.scientificamerican.com/observations/is-anyone-home-a-way-to-find-out-if-ai-has-become-self-aware/

Searle, J.R. (1990). Is the Brain's Mind a Computer Program? *Scientific American* 262(1): 25-31.

Searle, J.R. (1994). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

Shagrir, O. (2006). Why We View the Brain as a Computer. *Synthese* 153: 393–416.

Shagrir, O. (2010). Brains as analog-model computers. *Studies in History and Philosophy of Science* 41(3): 271–279.

Shagrir, O. (2018). In defense of the semantic view of computation. *Synthese* October 2018, DOI:10.1007/s11229-018-01921-z.

Shagrir and Harbecke (2019). The Role of the Environment in Computational Explanations. *European Journal for Philosophy of Science* 9, 37. DOI.org/10.1007/s13194-019-0263-7.

Brains as analog-model computers. *Studies in History and Philosophy of Science* 41(3): 271–279.

Simon. H.A. (1995). Machine as mind. In K.M. Ford, C. Glymour, and P.J. Hayes (eds), *Android epistemology*. Cambridge, MA: The MIT Press, pp. 23-40.

Sloman, A. (2013). Aaron Sloman absolves Turing of—the mythical Turing test. In Cooper, S.B. and van Leeuwen, J. (eds). *Alan Turing: His Work and Impact*. Elsevier Science: Amsterdam, pp. 601-11.

Smilansky, S. (2001) Free Will: From Nature to Illusion. *Proceedings of the Aristotelian Society* 101: 71-95.

Sprevak, M. (2010). Computation, individuation, and the received view on representation. *Studies in History and Philosophy of Science* (2010) 41: 260–270.

Trethowan, I. (1949). The Mechanical Brain. *The Times* [London, England], 14 June 1949, p. 5.

Turing, A.M. (1936) On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, Series 2 42: 230-65. Reprinted in Copeland ed. 2004, pp. 58-90; page references are to Copeland 2004.

Turing, A.M. (c. 1946) Letter from Turing to W. Ross Ashby. W. Ross Ashby: the correspondence of W. Ross Ashby, the British Library. Facsimile available in *The Turing Archive for the History of Computing*, http://www.alanturing.net/turing_archive/archive/l/l23/L23-001.html

Turing, A.M. (1947) Lecture to the London Mathematical Society, 20 February 1947. Reprinted in Copeland ed. 2004, pp. 378-94; page references are to Copeland 2004.

Turing, A.M. (1948) Intelligent Machinery. Reproduced in Copeland ed. 2004, pp. 410-32; page references are to Copeland 2004.

Turing, A.M. (1950) Computing Machinery and Intelligence. *Mind* 59: 433-460. Reproduced in Copeland 2004, pp. 441-64; page references are to Copeland ed. 2004.

Turing, A.M. (1951) Can Digital Computers Think? Reproduced in Copeland ed. 2004, pp. 482-6; page references are to Copeland 2004.

Turing, A.M. (c. 1951) Intelligent Machinery: a Heretical Theory. Reproduced in Copeland ed. 2004, pp. 472-5; page references are to Copeland 2004.

Turing, A.M., Braithwaite, R., Jefferson, G., and Newman, M. (1952) Can Automatic Calculating Machines Be Said to Think? Reproduced in Copeland ed. 2004, pp. 494-506; page references are to Copeland 2004.

Villalobos, M. and Dewhurst, J. (2017). Why post-cognitivism does not (necessarily) entail anticomputationalism. *Adaptive Behavior* 25(3) 117–128.

Williams, G. (1941). Free-Will and Determinism. *The Journal of Philosophy* 38(26): 701-712.

David Yates, D. (2008). Response-Dependence. *Philosophical Books* 49(4): 344–54.