

# Dystopia or utopia?

## Alan Turing's Promethean ambition about intelligent machines

Bernardo Gonçalves<sup>☆</sup>

*University of São Paulo  
Faculty of Philosophy, Languages and Human Sciences,  
and Polytechnic School*

---

### Abstract

Writing in 1948, Turing felt compelled to confront a “religious belief” that “any attempt” to construct intelligent machines was seen “a sort of Promethean irreverence.” And yet he has been associated by his own biographer Andrew Hodges with the image of “a Frankenstein — the proud irresponsibility of pure science, concentrated in a single person.” Reader of a 1865 version of Samuel Butler’s *Darwin among the machines*, Turing challenged the conventional wisdom of what machines really were or could be and prophesized a future pervaded by intelligent machines which may be seen as a dystopia or as a utopia. The question is thus posed: what future did Turing actually envision and propose to machines? I will formulate and study the problem of identifying Turing’s specific Promethean ambition about intelligent machines. I shall suggest that Turing’s primary aim was the development of mechanistic explanations of the human mindbrain. But his secondary aim, implied in irony and wit, was the delivery of a social criticism about gender, race, nation and species chauvinisms. Turing’s association with Mary Shelley’s Frankenstein will be discouraged. Rather, his third aim was to send a precautionary message about the possibility of machines outstripping us in intellectual power in the future.

*Keywords:* Alan Turing, Machine intelligence, Dystopia, Utopia, Modern Prometheus, Future of artificial intelligence

---

### 1. Preliminaries

Alan Turing (1912-1954), according to a key contemporary figure in a condolence letter to Mrs. Sara Turing after her son’s death, was “a sort of scientific Shelley” (1959, p. 58). This image of Turing was given by Sir Geoffrey Jefferson (1886-1961), himself portrayed later as “master of the neurosciences and man of letters” (Schurr, 1997). Jefferson is important because he stood out, most notably at the time, against Turing’s views on machine intelligence. And Turing bothered to respond back to him, such that Jefferson can be said

---

<sup>☆</sup>Contact: [begoncalves \(@usp.br\)](mailto:begoncalves@usp.br).

to have been Turing's primary philosophical opponent. In 1949, when they first met, Jefferson (63) was near double Turing's age (37). Scientific Shelley is a startling portrait of Turing that Jefferson rendered. Biographer Andrew Hodges took for granted (1983, p. 439) that it must have been Percy Bysshe Shelley and not his wife Mary Shelley the author of *Frankenstein*, the one Shelley that Jefferson associated Turing to. But Jefferson's dubious suggestion seems in any case to have caught Hodges' imagination, for he wondered that there was in Turing a Shelley but "also a Frankenstein — the proud irresponsibility of pure science, concentrated in a single person" (p. 521).

The Jefferson-Hodges portrait of Turing has been influential. It would reappear in science in Patrick Hayes and Kenneth Ford's (1995) influential criticism of the Turing test within the artificial intelligence research community. They alluded to Mary Shelley's *Frankenstein* and urged their peers to abandon the goal of creating an "artificial human," which meant for them to get rid of "Turing's ghost." It would also reappear in fiction in Ian McEwan's recent novel *Machines like me* (2019). In the novel, Turing does not die at his 42 and his further work leads to the development of embodied artificial humans with artificial skin and all sorts of features to make them look like real humans. It runs as if Turing's ambition was the synthesis of a human-like artificial creature. This view of Turing seems intriguing and yet it has barely been studied in the secondary literature. Is it accurate?

Turing scholars Jack Copeland and Diane Proudfoot suggested in turn to read Turing's strong statements about the future of intelligent machines, respectively, as "comic-strip stuff" (2004, p. 470) and "gentle mockery" (2015). In this view of Turing, if I take it correctly, Turing was a fairly well-behaved thinker who was outlining a manifesto for the new science of machine intelligence. A serious discussion of values, however, has never been acknowledged by these authors to be present in Turing's thought. Accordingly, in spite of the fact that Turing dedicated much time and effort in a public debate about the future of intelligent and superintelligent machines in society and culture, and even stated that in principle they might outstrip us human beings in intellectual power, we should not take him seriously in this regard. Now, this line of interpretation goes in the opposite direction of the Jefferson-Hodges portrait. It seems to be motivated to keep the analysis of Turing's thoughts and works sober or safeguard it under exegetical rigor, and yet it came out itself, respectively, in the forms of a side note and a short commentary.

Given this state of affairs in the secondary literature about Turing — including biographical (Hodges, 1983), scientific (Hayes & Ford, 1995), fictional McEwan (2019), historical (Copeland, 2004) and philosophical (Proudfoot, 2015) stances, one may ask: is a broad yet rigorous interpretation of Turing's ambition about intelligent machines possible?

## 2. The problem

Stirred mostly by Jefferson's ambiguous association of Turing with (the) Shelley(s), I shall formulate the problem of identifying Turing's specific Promethean ambition. Turing is widely seen as the author of the founding manifesto of machine intelligence (1950). As this science and engineering discipline promises to reach its maturity in this century, a historically rigorous view of Turing's intellectual profile and motives seems desirable.

Jefferson had made admonitions and other dubious suggestions in relation to Turing before his sympathy letter to Mrs. Sara Turing. For example, when Turing was elected FRS in March 1951 (at his 38), Jefferson, who had been elected FRS just four years before (at his 61), sent him a note:

I am so glad; and I sincerely trust that all your valves are glowing with satisfaction, and signalling messages that seem to you to mean pleasure and pride! (but don't be deceived!). (Turing, 1959, p. 101)

Playing with words, Jefferson suggested Turing to be a machine. He also alluded to some kind of behaviorism, perhaps Gilbert Ryle's, as if it were related to Turing's views. This was at the core of what Turing and Jefferson had in dispute since June 1949. So, in the occasion of congratulating Turing for a unique recognition by the scientific community of their country, Jefferson delivered a joke packed with their contention inside. Now, if Jefferson was ironic in this occasion, what may have him implied by "scientific Shelley"?

Percy Bysshe Shelley (1792-1822) once left in the register of Chamonix's Hôtel de Londres this threefold autobiographical note: "I am a lover of mankind, a democrat and an atheist."<sup>1</sup> Atheism aside, would Jefferson have thought of Turing as a lover of mankind and a democrat? Percy B. Shelley, as known, was one of the major English romantic poets together with his friend Lord Byron and others. Shelley is said not to have been as influential in his own time as he was in the three or four generations next. Considered a radically progressive thinker, most publishers and journals in his time declined to publish his work, if for nothing else, for fear of being arrested for blasphemy or subversion. From 1818 to 1822 (until his death) Shelley wrote his masterpiece *Prometheus unbound* Shelley (1822), a four-act lyrical drama in reference to Greek dramatist Aeschylus' trilogy conventionally called the *Prometheia*. The three plays were *Prometheus Bound*, *Prometheus Unbound* and *Prometheus the Fire-Bringer*. The second and the third plays were preserved only in fragments. Essentially, the *Prometheia* is concerned with the torments of the Greek mythological figure Prometheus, who defies the gods and gives fire to humanity, for which he is subjected to suffering and eternal punishment by Zeus (Jupiter). Unlike in Aeschylus' version, Percy Shelley exhibited a Prometheus that, with the help of his lover Oceanid Asia, manages to get released from oppression.

Roughly at the same time, as known, Mary Shelley (1797-1851), in some non-obvious collaboration with her companion (for one aspect of it, they both wrote prefaces to each other's work), wrote her famous *Frankenstein, or the modern Prometheus* (1818). In her fiction, Dr. Victor Frankenstein infuses the spark of life in an otherwise inert body only to get later horrified by his creature. Dr. Frankenstein's attitude, as represented by Mary Shelley, is a reference to Promethean disobedience of a different sort. Blind in his pursuit of modern science, Dr. Frankenstein loses control of his creature and its actions. Now, understanding the connections between the two Shelley's pieces is a problem of its

---

<sup>1</sup>Cf. "Shelley's shocking 'atheist' declaration rediscovered after 200 years". Available at <http://www.trin.cam.ac.uk/news/shelleys-shocking-atheist-declaration-rediscovered-after-200-years/>. Access on 1 Nov. 2020.

own which received, for instance, an informative analysis by critical biographer Richard Holmes (2011). Although Percy Shelley's Prometheus is a modern figure, it is actually Mary Shelley's the one that engaged in an outright profane attitude towards the creation myth and offered a dominating image about the artificial creation of life. A fairly simplified view of the differences between the figures of Prometheus in *Prometheus unbound* and *Frankenstein* will suffice to my purpose here, which is as follows.

The fire stolen by Percy Shelley's (enlightened and progressive) Prometheus is a symbol for the emancipation of humans by the knowledge and the abilities that will enable them to shape their own future, now independent of the gods. This is different from Dr. Frankenstein, or the (unwary and outrageous) modern Prometheus of Mary Shelley's, who just in the course of his reckless research gave fire inadvertently to a creature whose (in)human status is a problem. From 1816 to 1818, some critics note, Mary Shelley seems to have started to develop some deep concerns about modern science. She may even have read, in that period, the first official version of Goethe's *Faust* Part I, published in 1808. The apparently unlimited possibilities open by science would, for Mary Shelley, be uncanny. For the sake of this rough image of contrast, Mary's Prometheus is dangerous and may be approximated with Goethe's *Faust*.

The questions are thus posed: what kind of Promethean figure best fits Turing? Is it Mary or Percy Shelley's? What kind of society permeated with intelligent machines did Turing envision and propose? Is it a dystopia or a utopia?

I will refer to this as *the problem of identifying Turing's specific Promethean ambition*. I will revisit and address it in the central moment of this paper (§7). In what follows (§3) I shall introduce Turing's response to what he himself called "Promethean irreverence," and have a closer look at Jefferson's view of him. Also informative will be: to acquire a perspective on Turing's use of irony and intellectual integrity (§4); and to draw attention to the moment when he assumed the role of prophet of the machines (§5), and to the fact that he was a reader of Samuel Butler (§6). I shall then arrive at my proposed interpretation of Turing's ambition about intelligent machines (§7), and conclude with final remarks (§8).

### 3. Turing's irreverence

Mentions of "Frankenstein" appeared in the English press in 1946 in connection with the advent of modern computing machines, when a buzz came out about scientists who were building an "electronic brain." Hodges found (1983, p. 347) that this latter term had been first used publicly on 31 October 1946 by British statesman Louis Mountbatten in an official and high-profile address to the British Institution of Radio Engineers (1946). But we also know from Donald Bayley, who worked with Turing during the war since 1943 and up to 1946, that Turing had said before, in c. October 1945 as of joining the NPL, that he would build "a brain" (Sykes, 1992; Copeland, 2004, 25-27; p. 374). This is supported by a November 1946 letter from Turing to Ross Ashby, where Turing wrote: "Sir Charles Darwin [NPL's director] has shown me your letter." He resumed: "[i]n working on the ACE I am most interested in the possibility of producing models of the action of the brain than in practical applications to computing" (1946). So the origin of the term "electronic brain"

is likely Turing's or from his circle. According to Turing scholars Diane Proudfoot and Jack Copeland in (2018), Mountbatten would have been briefed by the NPL staff in 1946. For these authors, "Turing's views were probably the inspiration for much of Mountbatten's [October 1946] address" (p. 27). So Turing must have felt pointed as the Frankenstein imagined by the press in late 1946.

Turing did refer once to the myth of Prometheus explicitly. It was in his (1948) discussion of objection "(b)" to the possibility of machine thinking (p. 410), he formulated it this way: "[a] religious belief that any attempt to construct such [intelligent] machines is a sort of Promethean irreverence." He then responded: "being purely emotional, [it does] not really need to be refuted" (p. 411). And yet in the same text (1948), in a section he entitled "man as machine," Turing also wrote:

One way of setting about our task of building a 'thinking machine' would be to take a man as a whole and try to replace all the parts of him by machinery. He would include television cameras, microphones, loudspeakers, wheels and 'handling servo-mechanisms' as well as some sort of 'electronic brain'.  
(Turing, 1948, p. 420)

Later, in his (1950) paper, Turing enlarged and reshaped his 1948 list of objections. Although then not as quite directly, in his discussion of the closely related objection now named "the theological objection," Turing addressed with some irony the same charge:

In attempting to construct such [intelligent] machines we should not be irreverently usurping His [God's] power of creating souls, any more than we are in the procreation of children: rather we are, in either case, instruments of His will providing mansions for the souls that He creates. (Turing, 1950, p. 443)

Jefferson may not have had access to Turing's 1948 report but it is very unlikely that he did not read Turing's 1950 text. (He is addressed directly and indirectly several times there, and they have been in touch in the period, at least up to January 1952.) In any case, it may have been from a more total perspective or Gestalt that Turing reminded Jefferson of (the) Shelley(s). Unfortunately the records of their meetings and mutual impressions are relatively limited. We can grasp from Mrs. Sara Turing's biography that in addition to a few public meetings, there may have been a few private and informal meetings too. We know for sure of one, recollected by Jefferson himself in the same sympathy letter to Mrs. Sara Turing which is quoted by Turing's friend Lyn Irvine (wife of Turing's longstanding collaborator and mentor Max Newman) in her foreword to the biography:

He was a wonderful chap in many ways. I remember how he came to my house late one evening to talk to Professor J.Z. Young and me after we had been to a meeting in the Philosophy Department here, arranged by Professor Emmet. I was worried about him because he had come hungry through the rain on his cycle with nothing but an inadequate cape and no hat. After midnight he

went off to ride home some five miles or so through the same winter's rain. He thought so little of the physical discomfort that he did not seem to apprehend in the least degree why we felt concerned about him, and refused all help. It was as if he lived in a different and (I add diffidently, my impression) slightly inhuman world. (Irvine, 1959, p. xx)

So, in Jefferson's view, as dramatic as it may look, "it was as if [Turing] lived in a different" and "slightly inhuman world." Overall, we may not dismiss Jefferson's view of Turing as idiosyncratic of a personal impression or born out of some obvious rivalry. To add another source who engaged with Turing in the discussion on whether machines can think, let us consider Wolfe Mays (1912-2005). He participated with Turing in an October 1949 seminar held in the Department of Philosophy in the first edition of the same meeting mentioned by Jefferson and happens to have saved with him minute notes of the meeting (Turing et al., 1949). This seminar is a critical event in Turing's intellectual life and is known to have motivated Turing's famous (1950) paper. One year later, as Turing's 1950 paper was about to appear, according to Mays himself he was asked by Gilbert Ryle to write Turing a reply (2001). And so he did. In his text, which would be rejected by Ryle and only get published in (1952), Mays alluded to "[t]he paradoxical Frankenstein nature of the machine-mind" (p. 150) and implied to Turing the designation of a "mechanical necromancer" (p. 153).

Inhumane views of Turing's "inhuman world" may, of course, be contrasted with some of the warm testimonies of Turing's friends and work colleagues, which can be found in plenty in the Turing biographies cited above. For example, one may consider stories such as this one from Robin Gandy, who received doctoral supervision in mathematical logics from Turing at the University of Cambridge and was one of Turing's best friends:

When we were engaged on war work, I always thought him a bit austere but at Cambridge I was enchanted to find how human he could be, discussing mutual friends, arranging a dinner-party, being a little vain of his clothes and his appearance. One of my happiest memories is of him and Nick Furbank and me playing a complicated game of hide-and-seeK in the Botanical Gardens by moonlight. (Turing, 1959, p. 119)

Gandy's note about Turing being more austere when engaged on war work, on the one hand, and being dazzling to the extent of playing hide-and-seeK in the gardens by moonlight in peace time, on the other, should not pass to the reader unnoticed. In any case, I will concentrate on views such as Jefferson and Mays' because they are germane to the reception of Turing's views on machine intelligence. If contemporary fellows saw Turing that way — as a "mechanical necromancer" or "as if he lived in a different" and "slightly inhuman world"—, then I take to be an important question to ask: why so? What is in such view?

Turing thought that machines can think. And he had a project to build an intelligent machine. Now, regarding the general position of contenders such as Jefferson, Mays and others, I would like to distinguish these two interpretation options (and my intent is less to reduce their views and more to foster clarity in my own discussion):

- (i) they took Turing's views and project to be feasible and have dangerous ethical and social implications; so Turing was, for them, sort of a mad scientist in the white coat that could create some monster in the laboratory and then had to be stopped;
- (ii) they rather took them to be unfeasible and significantly mixed with some sort of wishful thinking; so Turing was, for them, sort of a deceiver whose actual damage could not be the release of an artificial creature but rather to mislead the public opinion.

I would like to promptly suggest that Jefferson, Mays and other contenders were ambivalent and slippery in between the two positions above. Hodges reported that "Alan would refer to Jefferson as an 'old bumbler' because he never grasped the machine model of the mind" (1983, p. 439; no primary or secondary source is cited, neither is informed when Turing would have said that.) If Hodges is right, then, for Turing, the dispute had a basis on Jefferson's confusions about the science of computing. In any case, let us fixate attention on interpretations (i) and (ii) above.

In relation to interpretation (i), did Turing consider ethical and social implications of building and deploying in the society intelligent machines? For a short answer, I say yes — and that is why he exposed himself and spoke out about the future of machine intelligence. But in fact, he seems to have been half pleased about the image of a machine revolution. As Diane Proudfoot nicely put (2015), "[h]e seemed almost to welcome the possibility of this humiliating lesson for the human race." (I shall return to her interpretation and those of other Turing scholars soon.) Let us see. In discussing objection five to machine intelligence, "arguments from various disabilities" (*viz.*, various things that according to contenders machines could never do), Turing uttered this twofold piece of social criticism:

The works and customs of mankind do not seem to be very suitable material to which to apply scientific induction. A very large part of space-time must be investigated, if reliable results are to be obtained. Otherwise we may (as most English children do) decide that everybody speaks English, and that it is silly to learn French. (Turing, 1950, p. 448)

What is important about this disability [being unable to enjoy strawberries and cream] is that it contributes to some of the other disabilities, *e.g.* to the difficulty of the same kind of friendliness occurring between man and machine as between white man and white man, or between black man and black man. (Turing, 1950, p. 448, no emphasis added)

With his peculiar touch of irony (see below), Turing addressed chauvinisms of nation and race, respectively. But he considered that it would take several decades, maybe one century, for the impact of intelligent machines to come in society and culture. So, for him, as I interpret, his generation would have enough time to study the logical limits of (future) machine capabilities beforehand. That is, Turing's attitude in raising such a strong voice about the social importance of considering the future of intelligent machines — when only

a few were paying attention — can be seen as *precautionary*. This is my short answer to the question on whether Turing considered ethical and social implications of building and deploying intelligent machines in society. I will support and further develop it later (§7).

Now, moving on to alternative interpretation (ii) above, is it likely that Turing was at least in part lost in some sort of mystical dream or desire to pursue an absurd project and deceive public opinion? I think not, because of his integrity as discussed next.

#### 4. Turing's irony and intellectual integrity

Turing scholar Jack Copeland chose “humour” as one of three words (and the one appearing first) to sum up Alan Turing (2012, p. 1). Turing's remarkable sense of humor can also be noted from the contributions of several of his friends, some of which vividly appearing in Christopher Sykes' BBC documentary (1992). I invite the reader that is less familiar with Turing's biography to appreciate what was, indeed, one of his main facets. My specific goal here, though, is to discourage a specific reading of Turing's sense of humor. Both Turing's contemporaries and later commentators took his wit or liked to suggest it as reason to more or less dismiss the seriousness of his views. For instance, back in the outbreak of his polemic with Jefferson about intelligent machines in June 1949, Turing delivered along the same lines we have just seen a non-obvious reply to Jefferson's strong claims purporting a myriad of things that machines would never be able to do. He then received from the June (1949) editorial of the *British Medical Journal* (henceforth also BMJ) this rejoinder:

Mr. A. W. [sic] Turing, who is one of the mathematicians in charge of the Manchester “mechanical brain,” said in an interview with *The Times* (June 11) that he did not exclude the possibility that a machine might produce a sonnet, though it might require another machine to appreciate it. Probably he did not mean this to be taken too seriously [...]. (BMJ, 1949, p. 1129)

This can be seen as the first of a series of *ad hominem* attacks that Turing would receive. Jefferson, as the reader may have guessed at this point, was a master of dressing them as compliments. For instance, he wrote to Turing's mother in the same passage of the condolence letter previously mentioned: “He [Turing] was so unversed in worldly ways, so childlike it seems to me, so unconventional, so non-conformist to the general pattern [...] so very absentminded,” and completed “[h]is genius flared because he had never quite grown up” Turing (1959), p. 58). So Turing, according to Jefferson, would have been nothing but a childlike, innocent thinker.

Against this, I shall argue that Turing knew very well that his views on intelligent machines were bold and unorthodox. But in face of the sometimes unhelpful reaction received, he just could not help himself either in avoiding being fun and provocative. In fact, Turing received plenty of provocations of dubious taste. For instance, in the October 1949 philosophy seminar mentioned above, at some point Jefferson is reported to have said “but this is an argument against the machine: do human beings do this kind of thing?”; to what Turing would have replied “yes – mathematicians;” and then a murmur would have

followed “are mathematicians human beings?” (Cf. Turing et al. (1949)). In spite of all that, to the best of my knowledge, Turing has never pushed back along the same lines. He did not use to depart from the topic of discussion to engage in *argumentum ad hominem*.

Turing was no polemist. Beyond his hardly offensive touch of humor, Turing seems to have been as truthful and as clear as possible about his views. A notorious example of Turing’s integrity, in my view, is his spontaneous initiative of outlining his beliefs in his (1950) paper (p. 442). In that passage Turing also offered two potentially testable predictions about the future. About whether one should take Turing’s sense of humor as compromising of the meaning of what he said, I would like to quote two specific remarks. The first one is by Mrs. Sara Turing. After having collected an enormous set of testimonies from Turing’s peers and friends in several letters after his death, she summed up:

There is a marked unanimity of opinion in the letters to me about Alan. Coupled with great admiration for “his profound originality and insight” there is repeated emphasis on his simplicity and integrity and complete “lack of pretentiousness and pomposity.” (Turing, 1959, p. 118)

A related, perhaps more complete note about Turing’s profile and manners was given by Turing’s fellow mathematician and contemporary at King’s College, Denis Williams:

In intellectual, as in other matters, it was essential to him that everything should ring true. [...It] seems to me precisely this complete intellectual integrity, which, combined with his other gifts, made it reasonable to expect that he would produce results of fundamental importance in his own field. Alan had a delightful sense of humour. He enjoyed elaborating fantastic projects, such as a scheme for faking prehistoric cave paintings, in mock-serious detail, or bringing an over-serious discussion down to earth with a quick colloquial turn of phrase. With him jest and earnestness were often closely intermingled. (Turing, 1959, p. 91)

Now, if it is fair enough to assume that Turing’s sense of humor did not interfere with his intellectual integrity, the question that may yet remain is how should we best understand some of his acutely ironic statements? Meeting with great intellectual opposition but also with — in his own words — “emotional” reactions, Turing made extensive use of irony. Turing thought that machines could think, and suspected that we humans may be, in a certain non-trivial sense, like machines. Sensing that he was not properly listened to, some of his communications seem to have been assembled to shock. Along these lines, I find in John Price’s study of David Hume’s irony a remarkable clue towards Turing’s:

When a man was under the intellectual and cultural pressure which Hume experienced he could not respond easily by denunciations, by shouting, or by threats. As a civilized man, Hume would not have responded that way under any circumstance. His method of dealing with those who would persecute him or ostracize him simply because of his religious or philosophical or moral opinions was subtle and effective. Irony gave him a method of operating in a world

that found his ideas both strange and shocking: strange because most people were simply unable to handle them, shocking because his scepticism dared to attack the citadel of religion. New ways of thinking about man's place in nature, especially if they do not reassure one's blind faith, are often difficult [...] to tolerate. Irony could at least create artificial tolerance. (Price, 1965, p. 4-5)

So clearly, Turing has not been the first bold (British) thinker to make use of first-class irony. And Turing's "new ways of thinking about man's place in nature" were "often difficult to tolerate" indeed. I take Price's interpretation of Hume to be very enlightening when applied to Turing. Turing's words, I hold, should neither be understood always literally nor dismissed as plain mockery. His irony was rather, I interpret, a clever form of communication to an intellectual and social environment that could barely listen to him. He applied irony more or less subtly to imitate people's language in ways to expose their stupidity or vices of thought about a subject matter. This is different from parody or imitation for the purpose of a philosophically empty comic effect. Turing's irony came rather as satire, or irony with a point.

Now, in light of this understanding of Turing's irony as a satire, shall we go astray left with no methodological principle to follow in the interpretation of Turing's statements and prone to psychologism? I think not. Although Turing relied often on irony, he was far off from being a hermetic speaker and writer. There are two fairly simple and hopefully effective principles that I shall abide by strictly whenever we are to interpret an ironic statement of his. First, as a form of structuralism, the statement shall be understood consistently with the whole system of his philosophical positions and on the basis of related passages from Turing sources that seem less loaded with irony. Second, as a science-studies maneuver, we shall look at Turing's historical context and, most specially, whom he was speaking to. Overall, I claim, the interpretation of Turing's irony is a tractable problem.

We may now proceed to examine Turing's statements about the future of intelligent machines in society and culture.

## 5. Turing, prophet of the machines

Machines shall not be condemned to be slaves forever. Thus implied Turing, seemingly for the first time publicly in a 1947 lecture to the London Mathematical Society when he said:

It is [...] true that the intention in constructing [electronic computing] machines in the first instance is *to treat them as slaves*, giving them only jobs which have been thought out in detail [...] Up till the present machines have only been used in this way. But is it necessary that they should always be used in such a manner? (Turing, 1947, p. 392-3, emphasis added)

Turing tried to make the point that seeing machines as slaves may be quite a short-minded perspective. Now, being "slaves" such a strong word — and a master-slave dichotomy was indeed implied overall in Turing's lecture, with multiple occurrences of both words —, one

may then ask, what future did Turing actually want for the machines and how does it relate to the humankind? We are now positioned to start answering this question.

In 1948 Turing formulated an objection “(a)” to the possibility of machine intelligence:

An unwillingness to admit the possibility that mankind can have any rivals in intellectual power. This occurs as much amongst intellectual people as amongst others: they have more to lose. Those who admit the possibility all agree that its realization would be very disagreeable. The same situation arises in connection with the possibility of our being superseded by some other animal species. This is almost as disagreeable and its theoretical possibility is indisputable. (Turing, 1948, p. 410)

Turing thus posed a *fact-value demarcation* that is very important to consider for identifying his ambition. In the presence of this, it seems that any criticism of his ethics may have to take a stand and cope with the question of science and values more generally as well. In any case, note that here we are less interested in whether his demarcation would be effective or not (and I am suggesting nothing in this regard), and more interested in interpreting how Turing himself saw it in the turn from the 1940’s to the 1950’s. He distinguished the problem of studying the possibility of machines or other animal species superseding us humans in intellectual power, on the one hand, and whether or not it would be agreeable to someone or to the humankind altogether, on the other. A second point that I take from the passage is that Turing said “[t]hose who admit the possibility” — and this count must obviously include himself in — “all agree that its realization would be very disagreeable.” I take this to suggest that the advent of superintelligent machines that might take control over us is, for Turing, essentially not welcome or agreeable. In any case, for him, what is at stake does not depend on how likable it is. Finally, a third takeaway is that Turing challenged the anthropocentric attitude towards machines and animals that has been established among us at least since René Descartes’s beast-machine thesis. This was the thesis that the human body and the non-rational animals are machines and as such have no minds or souls.

Later, in (1950), Turing would write “I believe [...] that no useful purpose is served by concealing these beliefs [on the possibility of intelligent machines]” (p. 442). And he would then rephrase his 1948 formulation quoted above and name it the “heads in the sand” objection to machine intelligence (p. 444). Three aspects stand out in my view. The first is Turing’s insistence on challenging such an objection. The second is the denomination he assigned to it in 1950. And the third is the fact that Turing knew, of course, that his 1950 text, unlike the 1948 one (subject to State Secrecy), would go to the public domain. Turing’s (1950) paper seems to set the moment when he assumes a role of prophet of the machines. The 1948 objection, reformulated in 1950, ran:

(2) The ‘Heads in the Sand’ Objection. “The consequences of machines thinking would be too dreadful. Let us hope and believe that they cannot do so.” [...] We like to believe that Man is in some subtle way superior to the rest of creation. It

is best if he can be shown to be *necessarily* superior, for then there is no danger of him losing his commanding position. (Turing, 1950, p. 444, no emphasis added)

Turing spoke out, in spite that only a few were open to listen, and he did not seem to fear persecution. Further on in the same text, he concluded:

We may hope that machines will eventually compete with men in all purely intellectual fields. [...] We can only see a short distance ahead, but we can see plenty there that needs to be done. (Turing, 1950, p. 460)

It is such 1950 use of “hope” followed by a reference to “plenty that needs to be done” that marks, for me, when Turing’s statements started to become really prone to ambivalence with respect to his Promethean ambition. And it has only got worse in c. 1951, when — now perhaps even more aware of the risks and decided to speak out — he subtitled his BBC radio lecture “a heretical theory” and said:

Let us now assume, for the sake of argument, that these machines are a genuine possibility, and look at the consequences of constructing them. To do so would of course meet with great opposition, unless we have advanced greatly in religious toleration from the days of Galileo. There would be great opposition from the intellectuals who were afraid of being put out of a job. It is probable though that the intellectuals would be mistaken about this. There would be plenty to do, trying to understand what the machines were trying to say, i.e. in trying to keep one’s intelligence up to the standard set by the machines, for it seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers. There would be no question of the machines dying, and they would be able to converse with each other to sharpen their wits. At some stage therefore we should have to expect the machines to take control, in the way that is mentioned in Samuel Butler’s ‘Erewhon’. (Turing, 1951b, p. 475)

So, if intelligent machines are a genuine possibility — and Turing had already posited that he thought so —, then it seemed to him probable that by the nature of the machine thinking method (say, its reproducibility and scalability) machines would soon outperform us in intelligence power, and eventually take control. Turing had already included in his (1950) bibliography Samuel Butler’s “Book of the machines,” which is a 1865 version of Butler’s (1863) essay “Darwin among the machines” and composes three chapters of *Erewhon*. But in spite of the fact that Butler’s work appeared in the bibliography, no direct reference to it is found in the 1950 text itself. One possibility is that he meant it in connection with the passage “[t]hese are possibilities of the near future, rather than Utopian dreams” (p. 449) appearing in the end of his discussion of the fifth objection, “arguments from various disabilities.” So in the above c. 1951 passage, Turing cited Butler’s famous novel for the second time, then in an explicit association with his own views.

Now, one may wonder how did Turing scholars read passages such as the c. 1951 above. We have seen Turing biographer Andrew Hodges to side with Jefferson's view of Turing's Promethean ambition (§1). Jack Copeland in turn declared:

Turing ends 'Intelligent Machinery, A Heretical Theory' with a vision of the future, now hackneyed, in which intelligent computers 'outstrip our feeble powers' and 'take control'. There is more of the same in Chapter 13 [(Turing, 1951a)]. No doubt this is comic-strip stuff. (Copeland, 2004, p. 470)

More recently, Diane Proudfoot also seems to shrink the serious point of Turing's irony:

Turing (following Butler) poked fun at the fear of out-of-control AI. When he predicted in the London Times that machines could "enter any one of the fields normally covered by the human intellect, and eventually compete on equal terms," the media protested at the "horrific" implication of these ideas — namely, "machines rising against their creator." But Turing said drily, "A similar danger and humiliation threatens us from the possibility that we might be superseded by the pig or the rat." He joked — but with a good pinch of his usual common sense — that we might be able to "keep the machines in a subservient position, for instance by turning off the power at strategic moments." (Proudfoot, 2015)

She then added to it, perhaps to concede for some seriousness in Turing's prophecy:

Turing's response to AI panic was gentle mockery. All the same, there was a serious edge to his humor. If runaway AI comes, he said, "we should, as a species, feel greatly humbled." He seemed almost to welcome the possibility of this humiliating lesson for the human race. (Proudfoot, 2015)

So Proudfoot's interpretation of Turing's view seems to boil down to seeing it as "gentle mockery," or a joke about our species chauvinism. It is common to Hodges, Copeland and Proudfoot's references altogether, in any case, that they only refer to what I have called the problem of identifying Turing's specific Promethean ambition in side notes (Hodges and Copeland) and short commentaries (Proudfoot). The same holds for their comments on Butler's *Erewhon*, if any. Rather, as suggested early on in this text, I think that this is an important problem which deserves study. Accordingly, I shall now take a moment to introduce Butler's novel and how it reached Turing.

## 6. Turing, reader of Samuel Butler

Butler's *Erewhon* (1865) became popular in Victorian England early on since it was published. According to Mrs. Sara Turing, it would have reached Turing early in his youth in the turn from the 1920's to the 1930's:

In [Alan's] late teens he read a certain amount of fiction [...]. He had a particular fondness for *The Pickwick Papers*, George Borrow's books and Samuel Butler's *Erewhon*. This last possibly set him to think about the construction of an actual intelligent machine. (Turing, 1959, p. 108).

Given this recollection from Mrs. Sara Turing, the influence of Butler's novel on Turing may have been long-lasting indeed.

Butler's novel can be considered a social critique in the form of a satire. It is hard to class either as utopia or dystopia, as things are presented always from two perspectives. Butler lived in the (Victorian) machine age, while the philosophical and cultural shockwaves of Charles Darwin's 1859 *On the origin of species* were still being felt. Butler thus applied in a non-obvious way the principle of natural selection to machinery itself. He revised and enlarged "Darwin among the machines" (1863) into his 1865 "The book of the machines," which is itself a fictional book described by the narrator within *Erewhon* (the fiction). "Erewhon" is "nowhere" written backwards (or almost), and referred to a yet undiscovered country in New Zealand. There, so runs the novel, a revolution was started and led to civil war, opposing "machinists" and "anti-machinists." It was won by the latter. Then all of the more complicated machines formerly in common use were destroyed, and all treatises on mechanics burned (1865, p. 188). Only a few hundred years later, thus the story goes, when no Erewhonian seemed to consider anymore the idea of reintroducing forbidden inventions, the subject came to be regarded as a curious antiquarian study. So this is how the book of the machines gets written within the novel, as a recollection of the arguments of anti-maquinist philosophers. Here is an example addressing the rate of progress of machines or of the so-called "mechanical kingdom:"

Reflect upon the extraordinary advance which the machines have made during the last few hundred years, and observe how slowly the animal and vegetable kingdoms are advancing in comparison. The more highly organised machines are creatures not so much of yesterday as of the last five minutes, so to speak in comparison with past time. (Butler, 1865, p. 191).

Butler's anti-maquinist arguments also comprised aspects such as the emergence of consciousness and a purely mechanical reproductive system. That was all reasonable for Butler to write about in light of the industrial revolution and Darwin's then new theory of evolution by natural selection. And it must have seemed persuasive to Turing indeed, who would in (1936) conceive of a universal machine and ever since be a key figure to get the computer revolution underway.

We shall now be well positioned to examine in more depth Turing's 1950 and c. 1951 citations of Butler, including their historical context. Butler's *Erewhon* had actually been cited first by Norbert Wiener in his *Cybernetics* (1948, p. 27), a book that received commentary both from Jefferson in his (1949) Lister oration and from the BMJ's 1949 editorial that I quoted before (§4). In fact, after having admonished Turing directly, the BMJ's editorial mentioned *Erewhon* and warned: "if we fail to recognize [that] the mind

must surely be greater than its own ideas about itself, [...] we may suffer the fate of the [Erewhonians] and be enslaved by machines [...]" (BMJ, 1949, p. 1130). This was a prototype of the kind of opposition that Turing faced. One may observe in the conditional statement that the antecedent — “the mind must surely be greater than its own ideas about itself” — engages in a petition of principle. It was as if the theoretical and empirical possibility of (super)intelligent machines, on the one side, and the real status of the human mind, on the other side, were a matter of values only.

Considering Turing’s specific c. 1951’s citation of Butler as quoted above, he may not have forgotten the BMJ’s editorial. And it seems that he wanted to shock. He had already some appreciation for Butler’s fiction himself, and seems to have found an opportunity to respond to the BMJ editorial. He used irony to imitate his contenders’ language and render a scenario that was the exact opposite of what they considered desirable. And yet this was not an empty joke of his. Rather the evidence suggests that he did think the possibility of having machines outstripping our intellectual powers to be both true and reasonable. Turing’s c. 1951 citation of Butler, I interpret, was not parody but satire. Let us see.

It is actually recurrent in Turing sources the advent of superintelligent machines, sometimes with irony, sometimes without. I now proceed to present representative instances. A key such statement, and to my knowledge the first one chronologically, is Turing’s 1948 formulation of objection “(a)” which I quoted before (§3). Turing’s tone in that passage was steady, and in fact the occasion was an official NPL report. Two other most notable sources are the following. First, from a vivid 1949 reminiscence of Lyn Irvine’s, we learn about this Turing-Newman dialogue:

I remember sitting in our garden at Bowdon about 1949 while Alan [Turing] and my husband [Max Newman] discussed the machine (‘Madam’) and its future activities. I couldn’t take part in the discussion and it was one of many that had passed over my head, but suddenly my ear picked up a remark which sent a shiver down my back. Alan said, reflectively, ‘I suppose when it gets to that stage we shan’t know how it does it.’ (Turing, 1959, p. 95)

I take to be fairly clear that Turing’s observation as reported above was *not* loaded with irony. First, one may consider that this was a private conversation between two long-standing collaborators and friends, whose thoughts on machine intelligence did not differ enough to justify use of irony among themselves. Second, because Irvine herself gave away the tone of Turing’s statement, as she said “suddenly my ear picked up a remark which sent a shiver down my back.” Now, while Turing did not mention machines outstripping our powers and taking control specifically, I think that the context, most notably the implied tone that sounded dreadful to Irvine, adds up to make it.

For a second source, let us consider his May 1951 BBC radio lecture (1951a). In the two last paragraphs of the lecture’s transcript, Turing revisited once more what is strongly related to “the heads in the sand” objection, now perhaps in a more diligent tone, say, more like in (1948), and less like in (1951b). He again recollected the theoretical possibility of other species, say, “the pig or the rat,” to supersede us humans in intelligence power, and compared it with the same possibility now applied to the machines. He then added:

But this new danger is much closer. If it comes at all it will almost certainly be within the next millennium. It is remote but not astronomically remote, and is certainly something which can give us anxiety. (Turing, 1951a, p. 486)

Turing thus plainly posed: the advent of superintelligent machines is not a certain event, but its possibility is genuine. It will take a long time, but it may come in a foreseeable future. Now, if he thought so and still wanted to build intelligent machines, does that mean that he actually *wanted* machines to take control? I think not. As anticipated, I think that Turing was, among other things, raising a precautionary voice.

In the presence of the crass chauvinism of some of his contenders, however, he suggested epistemological relations between the possibility of *intelligent* machines and what he saw as species, gender and other related biases of his contenders. This is, I think, the reason why Turing's prophesized future pervaded by intelligent machines may be seen as a dystopia just as much as a utopia. For him it was in part disagreeable, and in part agreeable. *Superintelligent* machines were for him, I interpret, a true but distant possibility that pushed his argument to the limit from both epistemological and social stances.

I shall now proceed to consolidate this analysis of the problem of identifying Turing's specific Promethean ambition in light of the interpretive basis we have built so far.

## 7. Turing's Promethean ambition

For Turing, I interpret, the possibility of machines to supersede us in intelligence power was a serious problem that needs to be studied in detail, both analytically and empirically. Let us now address these questions: if Turing thought superintelligence to be a serious possibility, then why would he propose to actually build intelligent machines? Should not this imply that he was incautious about the consequences for the humankind? I think that the answer lies in his references to the timescale of the evolution of machine intelligence. As we have just seen, he offered a structured argument in this regard. Turing dismissed the urgency of time. He considered the advent of superintelligent machines not to be astronomically remote and yet to be remote enough to allow for deeper studies with the prospect of giving us clear benefits in the meantime. In the same 1951 lecture, Turing followed on and admitted:

The whole thinking process is still rather mysterious to us, but I believe that the attempt to make a thinking machine will help us greatly in finding out how we think ourselves. (Turing, 1951a, p. 486)

This passage is not filled with irony either. It is perhaps the most direct statement that Turing made about his deeper positive motives. I take it to be confessional about his hopes: they were pointed towards improving our scientific understanding of the human mindbrain. There is also secondary-source evidence of that from a testimony by Mrs. Sara Turing. She reports that Turing has told her about his aims as early as 1944:

Sometime round about 1944 he [Turing] had talked to me about his plans for the construction of a universal computer and of the service such a machine might render to psychology in the study of the human brain. This he regarded as likely to be one of the more valuable contributions a universal computing machine could make to knowledge. (Turing, 1959, p. 92)

Given this secondary source and all we have seen so far, I take that Turing's positive aims were laid out towards human enlightenment broadly construed. We may now ask, are these aims of Turing's actually new in history? Or do they have a strain?

I find two astonishing facts in Max Newman's Royal Society memoir on Turing. First, Newman wrote:

The varied titles of Turing's published work disguise its unity of purpose. The central problem with which he started, and to which he constantly returned, is the extent and the limitations of mechanistic explanations of nature. (Newman, 1955, p. 256)

One can observe that there is a remarkable similarity between Turing's natural-philosophy program and René Descartes's. (I am thinking of Descartes's *The world and Treatise on man*, and his *Essays*). Indeed, both Turing and Descartes dedicated themselves to find just *the extent and the limitations* of mechanistic explanations of nature. However, as Newman's follow-up sentence reveals, their approach to address that problem differed significantly. Thus put Newman:

[Turing's] way of tackling the problem was not by philosophical discussion of general principles, but by mathematical proof of certain limited results: in the first instance the impossibility of the too sanguine [Hilbert's] programme for the complete mechanization of mathematics, and in his final work, the possibility of, at any rate, a partial explanation of the phenomena of organic growth by the 'blind' operation of chemical laws. (Newman, 1955, p. 256)

Descartes's way of tackling the problem, it turns out, was foremostly "by philosophical discussion of general principles," or at least surely not by mathematical proof of limited results. Newman's formulation is, for me, a most elegant take on Turing's intellectual profile as a philosopher and sets our inquiry in a promising direction.

Indeed, Turing's motives can be associated with the modern tradition of mechanistic natural philosophy. Besides his connections with Descartes (Abramson, 2011; Gonçalves, 2020), he can be approximated, for example, with figures such as the French inventor and engineer Jacques de Vaucanson (1709-1782) who is famous for, among other pieces, his artificial duck. The designers of early modern automata from the seventeenth and the eighteenth centuries have often been criticized on the grounds of being driven by futile, entertaining-only motives. Is this fair? David Fryer and John Marshall suggested not (1979). They presented an interesting study of the motives of Vaucanson. The "claim that

the primary objective of Vaucanson's work was 'to astonish and amuse the public,' Fryer and John Marshall posited, "is hardly fair" (p. 267-8). "Vaucanson," they concluded, "was an entertainer, but he was also deeply committed to the development of an explanatory psychology." Turing was definitely no entertainer. But Fryer and Marshall's findings about Vaucanson offered a promising path to locate Turing's motives. Their inquiry is within a class of studies presented earlier by Silvio Bedini (1964) and Derek de Solla Price (1964), which showed that early modern automata were neither "trivial toys" nor "immediately useful inventions." Rather, they were simulacra or models "whose very existence offered tangible proof, more impressive than any theory, that the natural universe of physics and biology was susceptible to mechanistic explication" (Price, 1964, p. 9). This seems quite accurate as a description of Turing's primary aims. And yet, unlike the natural philosophers of the court in the seventeenth and eighteenth centuries, Turing eventually turned to be, as we have seen, rather a *social critic*. This secondary aim of his came most spontaneously out of the controversy he engaged in. We shall now have a firmer basis to identify Turing's Promethean ambition. The most explicit aspect of Jefferson's view — which defined Turing as "a sort of scientific [Percy] Shelley" (1959, p. 58) — seems just about right. In light of all that we have seen, I interpret, Percy Shelley's enlightened Prometheus suits well Turing.

This study poses interpretive boundaries of a negative kind as well. The evidence gathered discourages the association of Turing's ambitions with the image of Mary Shelley's Dr. Frankenstein. For instance, as we have seen, Hodges offered a view of Turing as "a Frankenstein — the proud irresponsibility of pure science, concentrated in a single person" (1983), p. 521). But Frankenstein is widely seen as the figure of a scientist who is blind in the pursuit of his science and unwary of its consequences. So a question that I pose to this view is the following. If Turing impersonated Dr. Frankenstein, why would he have exposed himself and spent time raising a voice in the public domain about the possibility of having machines outstripping our intellectual powers and taking control over us several decades, perhaps more than a century before he thought this possibility was feasible? It does not seem to fit. In fact, juxtaposed to the passage quoted above when he positively stated his aim of finding out how we think, Turing gave this negative statement:

But I certainly hope and believe that no great efforts will be put into making machines with the most distinctively human, but non-intellectual characteristics such as the shape of the human body; it appears to me to be quite futile to make such attempts and their results would have something like the unpleasant quality of artificial flowers. Attempts to produce a thinking machine seem to me to be in a different category. (Turing, 1951a, p. 486)

I take this as evidence that Turing did *not* find interesting to put effort into making fantastic things such as, say, an artificial creature resembling the human body, and even thought this to be futile and bound to give unpleasant results. Viewing Turing as a sort of Dr. Frankenstein — say, in connection with Hodges' suggestion, or Jefferson's dubious association of him with Shelley, or with Mays' allusion to a "mechanical necromancer" — is unsupported. Furthermore, with his reference to the "unpleasant quality of artificial flowers," Turing expressed his dislike to the dismissal of the ontological distinction between the natural

and the artificial. So the evidence is unfavorable to associating Turing with, say, modern transhumanist movements.

Turing was first and foremost a scientist. He believed in human rationality and was not afraid about establishing our own limits as human beings. He believed that he could be able to pursue his mathematical studies further in order to anticipate the logical possibilities and limits of machines, and that this was actually needed given the possibility of superintelligent machines. He wanted to study the status of machine intelligence as the other face of our own status as human beings. This was not a fantastic project, but one laid out towards human enlightenment. The evidence is in favor of viewing Turing as a humanist, and not as a transhumanist. Challenged by a conservative reaction, Turing did not pass on the opportunity to make a subtle social critique about our species and gender chauvinisms, and even racial and national partisanships altogether. It is precisely because he did *not* think that we humans are necessarily superior beings, that associating Turing with the arrogance that concerned Mary Shelley as portrayed in the character of Dr. Frankenstein, is just far-fetched.

## 8. Final remarks

In this paper I have studied the problem of identifying Turing's specific Promethean ambition. Turing has been described by his contemporary and interlocutor Geoffrey Jefferson as someone who "lived in a different" and "slightly inhuman world." In the eyes of Jefferson, Turing was "a sort of scientific Shelley." This view, suggested with ambiguities, was not idiosyncratic of Jefferson. Wolfe Mays, another contemporary of Turing's, referred to the notion of a "mechanical necromancer" and implied to Turing the image of "Frankenstein." Turing biographer Andrew Hodges went on to suggest that there was in Turing a (Percy Bysshe) Shelley but "also a Frankenstein — the proud irresponsibility of pure science, concentrated in a single person." On the one hand, the Jefferson-Hodges portrait of Turing has been influential. It reappeared in science (Hayes & Ford, 1995) and in fiction (McEwan, 2019). On the other hand, Turing scholars Jack Copeland and Diane Proudfoot suggested to read Turing's most ironic statements about the future of intelligent machines as "comic-strip stuff" (2004, p. 470) and "gentle mockery" (2015). My findings can be seen in part as a rapprochement between these opposing views in the secondary literature.

I have found that Turing's primary aim was the development of mechanistic explanations of the human mindbrain. This is akin to the motives of seventeenth- and eighteenth-century figures such as Jacques de Vaucanson. But Turing was no court natural philosopher. His secondary aim, implied in irony and wit, was the delivery of a social criticism about gender, race, nation and species chauvinisms. Percy Shelley's enlightened and progressive Prometheus seems to suit Turing well indeed. Turing's association with Mary Shelley's Frankenstein in turn has been clearly discouraged by these findings. He thought that fantastic projects, *e.g.*, making a machine resembling the human body, was futile and bound to give results whose quality would be unpleasant like artificial flowers. The evidence gathered is also unfavorable for associating Turing with contemporary transhumanist movements in general. Rather, his third aim was to send a precautionary message about the possibility of machines outstripping us in intellectual power in the future.

Turing's prophesized future pervaded by intelligent machines may be seen as a dystopia or as a utopia. He acknowledged that this future was disagreeable to a certain conception of mankind, while he himself was awaiting for the new possibilities to come by.

## Acknowledgements

I am indebted to Pedro Bravo for the suggestion of John V. Price's *The ironic Hume* in connection with Turing's irony. Possible mistakes in my elaborations out of this material are my sole responsibility.

This research has been supported by The São Paulo Research Foundation (FAPESP) under grant number 19/21489-4 "The future of artificial intelligence: the logical structure of Alan Turing's argument."

## References

- Abramson, D. (2011). Descartes' influence on Turing. *Studies in History and Philosophy of Science: Part A*, 42(4), 544–51. doi:10.1016/j.shpsa.2011.09.004.
- Bedini, S. A. (1964). The role of automata in the history of technology. *Technology and Culture*, 5(1), 24–42. doi: 10.2307/3101120.
- BMJ (1949). Mind, machine, and man. *British Medical Journal*, 1(4616), 1129–1130.
- Butler, S. (1863). Darwin among the machines. *The Press*, (13 June). Signed with pen name *Cellarius*. Reproduced in the New Zealand Electronic Text Collection at <http://nzetc.victoria.ac.nz/tm/scholarly/tei-ButFir-t1-g1-t1-g1-t4-body.html>. Access on 16 Jul 2020.
- Butler, S. (1872 [1865]). *Erewhon or over the range*. London: Trubner & Co.
- Copeland, B. J., Ed. (2004). *The essential Turing: the ideas that gave birth to the computer age*. Oxford: Oxford University Press.
- Copeland, B. J. (2012). *Turing: pioneer of the Information Age*. Oxford: Oxford University Press.
- Fryer, D. M. & Marshall, J. C. (1979). The motives of Jacques de Vaucanson. *Technology and Culture*, 20(2), 257–69. doi: 10.2307/3103866.
- Gonçalves, B. (2020). *Machines will think: structure and interpretation of Alan Turing's imitation game*. PhD thesis, Faculty of Philosophy, Languages and Human Sciences, University of São Paulo, São Paulo. Preprint available at <http://philsci-archieve.pitt.edu/id/eprint/18558>.
- Hayes, P. & Ford, K. (1995). Turing test considered harmful. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95)* (pp. 972–7).

- Hodges, A. (2012 [1983]). *Alan Turing: The Enigma*. Princeton: Princeton University Press, the centenary edition edition.
- Holmes, R. (2011). Prometheus<sup>2</sup>: the two Shelleys and romantic science. John Coffin Memorial Lecture delivered on 27 October 2011 at the University of London Institute of English Studies. Available at University of London School of Advanced Study channel: [http://www.youtube.com/watch?v=\\_ifHWz-Sljo](http://www.youtube.com/watch?v=_ifHWz-Sljo). Access on 13 Mar. 2020.
- Irvine, L. (2012 [1959]). Foreword to the first edition. In *Alan M. Turing: Centenary Edition* (pp. xix–xxiv). Cambridge: Cambridge University Press.
- Jefferson, G. (1949). The mind of mechanical man. *British Medical Journal*, 1(4616), 1105–10.
- Mays, W. (1952). Can machines think? *Philosophy*, 27(101), 148–62.
- Mays, W. (2001). My reply to Turing: fiftieth anniversary. *Journal of the British Society for Phenomenology*, 32(1), 4–23. doi: 10.1080/00071773.2001.11007314.
- McEwan, I. (2019). *Machines like me*. London: Jonathan Cape.
- Mountbatten, L. (1946). The presidential address. *Journal of the British Institution of Radio Engineers*, 6(6), 221–5. doi:10.1049/jbire.1946.0032.
- Newman, M. H. A. (1955). Alan Mathison Turing, 1912-1954. *Biographical memoirs of fellows of the Royal Society*, 1(November).
- Price, D. J. d. (1964). Automata and the origins of mechanism and mechanistic philosophy. *Technology and Culture*, 5(1), 9–23. doi: 10.2307/3101119.
- Price, J. V. (1965). *The ironic Hume*. Austin: University of Texas Press.
- Proudfoot, D. (2015). Mocking AI panic: Turing anticipated many of today’s worries about super-smart machines threatening mankind. *IEEE Spectrum*.
- Proudfoot, D. & Copeland, J. (2018). Turing and the first electronic brains: what the papers said. In M. Sprevak & M. Colombo (Eds.), *The Routledge Handbook of the Computational Mind*, Routledge Handbooks in Philosophy chapter 2, (pp. 23–7). Oxford: Routledge.
- Schurr, P. H. (1997). *So that was life: a biography of Sir Geoffrey Jefferson, master of the neurosciences and man of letters*. London: Royal Society of Medicine Press.
- Shelley, M. (2012 [1818]). *Frankenstein; or, the modern Prometheus*. Norton Critical Editions. New York: W. W. Norton, second edition edition.
- Shelley, P. B. (1959 [1818-1822]). *Shelley’s “Prometheus Unbound”*: a variorum edition. Seattle: University of Washington Press. (Ed.) Lawrence John Zillman.

- Sykes, C. (1992). BBC Horizon: the strange life and death of Dr. Turing. Documentary. Produced by Christopher Sykes, and edited by Jana Bennett. IMDb title 2373297, cf. <http://www.imdb.com/title/tt2373297/>. Available at: <http://www.youtube.com/watch?v=Z-sTs2o0VuY>. Access on 9 Jul 2020.
- Turing, A. (1946). Letter to Ross Ashby. W. Ross Ashby Digital Archive. Also at Woodger Papers (catalogue reference M11/99). Facsimile available at: <http://www.rossashby.info/letters/turing.html>. Access on 20 July 2020.
- Turing, A. et al. (2005 [1949]). Rough draft of the Discussion on the Mind and the Computing Machine, held on Thursday, 27th October, 1949, in the Philosophy Seminar. *The Rutherford Journal*, 1(December). Transcript of notes taken during the philosophy seminar co-chaired by Michael Polanyi and Dorothy Emmet at the University of Manchester on 27 October 1949, made available by Wolfe Mays. Available at: <http://rutherfordjournal.org/article010111.html>. Facsimile at: [http://www.alanturing.net/philosophy\\_seminar\\_oct1949/](http://www.alanturing.net/philosophy_seminar_oct1949/). Access on 20 July 2020.
- Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1), 230–65. doi: 10.1112/plms/s2-42.1.230.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, LIX(236), 433–60. doi: 10.1093/mind/LIX.236.433.
- Turing, A. M. (2004 [1947]). Lecture on the Automatic Computing Engine. In B. J. Copeland (Ed.), *The essential Turing: the ideas that gave birth to the computer age* (pp. 378–94). Oxford: Oxford University Press. Chapter 9. Facsimile available at <http://www.turingarchive.org/browse.php/B/1>.
- Turing, A. M. (2004 [1948]). Intelligent machinery. In B. J. Copeland (Ed.), *The essential Turing: the ideas that gave birth to the computer age* (pp. 410–32). Oxford: Oxford University Press. Chapter 10. Facsimile available at: <http://www.turingarchive.org/browse.php/C/11>.
- Turing, A. M. (2004 [1951]a). Can digital computers think? In B. J. Copeland (Ed.), *The essential Turing: the ideas that gave birth to the computer age* (pp. 482–6). Oxford: Oxford University Press. Chapter 13. Facsimile available at: <http://www.turingarchive.org/browse.php/b/5>. Access on 20 July 2020.
- Turing, A. M. (2004 [c. 1951]b). Intelligent machinery, a heretical theory. In B. J. Copeland (Ed.), *The essential Turing: the ideas that gave birth to the computer age* (pp. 472–5). Oxford: Oxford University Press. Chapter 12. Facsimile available at: <http://www.turingarchive.org/browse.php/B/20>. Access on 20 July 2020.
- Turing, S. (2012 [1959]). *Alan M. Turing: Centenary Edition*. Cambridge: Cambridge University Press.
- Wiener, N. (1965 [1948]). *Cybernetics: Or Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press, second edition edition.