Article
An Analysis of Turing's Criterion for 'Thinking'

Diane Proudfoot [ID]

Department of Philosophy, University of Canterbury, Christchurch 8140, New Zealand;
diane.proudfoot@canterbury.ac.nz

Abstract: In this paper I argue that Turing proposed a new approach to the concept of thinking, based on his claim that intelligence is an 'emotional concept'; and that the *response-dependence* interpretation of Turing's 'criterion for "thinking"' is a better fit with his writings than orthodox interpretations. The aim of this paper is to clarify the response-dependence interpretation, by addressing such questions as: What did Turing mean by the expression 'emotional'? Is Turing's criterion subjective? Are 'emotional' judgements decided by social consensus? Turing's take on these issues impacts current philosophical debates on response-dependent concepts and on the nature of artificial intelligence.

Keywords: Turing; Turing test; emotional concept; response-dependence; intelligence; thinking; objectivity

1. Introduction

In 'Computing machinery and intelligence', Turing famously said that his computer-imitates-human game gives us a 'criterion for "thinking"' [1] (p. 443). As a philosopher of mind, Turing is typically credited, implicitly or explicitly, with two proposals: a behaviourist criterion of intelligence (or thinking) in machines, based on the imitation game; and a computational theory of mind. I have argued against these two widely accepted interpretations of Turing and instead for a new interpretation of Turing's test: the response-dependence interpretation, based on Turing's claim that intelligence is an 'emotional concept' [2,3]. Here, I clarify this interpretation of Turing's 'criterion for "thinking"' and reply to challenges.

Sections 2–4 make the case against the orthodox behaviourist and computationalist interpretations of Turing, and present the response-dependence interpretation. In Section 5 I analyze Turing's different uses of the term 'emotional' and reply to claims that in discussing intelligence Turing rejected any role for 'emotional' concepts. Sections 6 and 7 address the question whether Turing's 'criterion for "thinking"' can be objective. Section 8 addresses the question whether his approach implies that judgements of intelligence are determined by social consensus. Last, I reply to the claim that scholars should not take seriously Turing's (purportedly) 'careless' philosophical remarks (Section 9).

One preliminary point: although people today often distinguish testing for *intelligence* from testing for *thinking*—and treat the latter as the philosophical goal of AI—there is no evidence that Turing made this distinction when presenting his 'criterion for "thinking"'. This paper follows Turing's practice.

2. Orthodox Interpretations of Turing and His Test

The orthodox behaviourist interpretation of Turing's test for intelligence in machines emerged in the early 1950s, after the publication of his famous 'Computing machinery and intelligence'. For example, Wolfe Mays said, in criticizing the test: 'Even if it were possible to construct a machine whose behaviour was indistinguishable from that of a human being, and even if we accept [Turing's] behaviourist criterion . . . ' [4] (p. 151). Mays was, like Turing, at Manchester University, and in fact built a 'logic machine' with the programmer Dietrich Prinz, who worked on the Manchester machine—the larger version of the world's first modern computer. 70 years later, the standard view of Turing's test

is still that it provides a behaviourist definition (or sufficient condition) of intelligence in machines. For example, Andrew Hodges, writing in the centennial version of his influential biography, said that 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"' [5] (p. 334).[1]

2.1. Flaws in the Behaviourist Interpretation

This interpretation of Turing's test faces considerable problems. The first is that the computer-imitates-human game focuses, not on the machine's behaviour, but rather on the *interrogator*. It is an experiment to see if the interrogator will make the correct identification. Having introduced 'Computing machinery and intelligence' by describing a man-imitates-woman game, Turing says:

> We now ask the question, 'What will happen when a machine takes the part of [the man in the man-imitates woman game]?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? [1] (p. 441)

A machine, it appears, does well in the computer-imitates-human game when the interrogator in that game is fooled no less frequently than the interrogator in the earlier game. A behaviourist will likely attempt to explain away Turing's focus on the interrogator by saying that testing the interrogator is merely an indirect way of testing the machine, but this would be a very poor way of testing the machine. Numerous objections to the Turing test (understood as a behaviourist test) make this clear: for example, the interrogator might be very gullible or the machine's programmer very lucky (for such objections, see, e.g., [8,9]).

The second reason why the orthodox interpretation is unconvincing is that it fails to explain the *design* of the imitation game. Why have two contestants, why hide them, and why require the machine to deceive? This makes little sense on the orthodox interpretation—after all, why not simply give the computer a set of cognitive tasks? Just because the design of the game does not fit with the behaviourist reading, commentators frequently redescribe the test as a 2-player rather than a 3-player game (for examples, see [10,11])—and others simply call the design of the game 'peculiar' or 'strange' [12] (pp. 466, 509). A more plausible response, however, is to say that the philosophical view underlying Turing's test of intelligence in machines is *not* behaviourism.

The third reason why the behaviourist interpretation fails is that it is inconsistent with Turing's own remarks on intelligence (see further Section 3). Before turning to his writings, however, there is a second widely-accepted interpretation of Turing's philosophy of mind.

2.2. Another Orthodox Interpretation

The second claim typically made about Turing's philosophy of mind is that he was an early proponent of a computational theory; according to Gordana Dodig-Crnkovic, for example, the current project of natural info-computationalism is 'a contemporary natural philosophy that builds on the legacy of Turing's computationalism' [13] (p. 1). The belief that Turing was a computationalist is widespread, amongst both computer scientists and philosophers. For example, Herbert Simon said:

> The materials of thought are symbols—patterns, which can be replicated in a great variety of materials (including neurons and chips), thereby enabling physical symbol systems fashioned of these materials to think. Turing (1950) was perhaps the first to have this insight in clear form. [14] (p. 24).

We find this view of Turing amongst both fans and critics of computationalism. For example, John Searle, well-known critic of the computational theory of mind, said:

> There is a story about the relation of human intelligence to computation that goes back at least to Turing's classic paper ... [Namely that] all we would need to account for mental processes would be computational processes between the syntactical elements in the head. [15] (pp. 202–203)

An equally famous fan of computationalism, Jerry Fodor, said: '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' [16] (pp. 3–4).

I will argue that this second orthodox interpretation of Turing is also mistaken (in Section 4.2).

3. What Turing Actually Said: Intelligence as an Emotional Concept

Turing told us how he understood the concept of intelligence or thinking in 'Intelligent Machinery' [17], a 1948 report for the National Physical Laboratory (NPL), written during a sabbatical in Cambridge before he moved to Max Newman's Royal Society Computing Machine Laboratory at Manchester University. The title of the last section is 'Intelligence as an emotional concept'. Turing made implicit references to his approach to intelligence elsewhere in his writings, but in this report he presents his view explicitly. His remarks are at odds with the orthodox behaviourist interpretation. 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. [17] (p. 431).

Turing emphasizes this point when he adds: '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' [17] (p. 431).

What is Turing claiming here? We might give three different answers. First, he might be making a very general point, namely that judgements of intelligence (or thinking) always involve a *judge*, who inevitably comes with his or her own cognitive resources.[2] On this view, judgements of intelligence are just like other non-technical judgements, all of which therefore involve 'emotional concepts'. But if so, why does Turing bother to stress a fairly mundane point—and why does he write as if the concept of intelligence, being 'emotional', is exceptional? Second, Turing might be suggesting that judgements of intelligence that are 'determined as much by our own state of mind and training' import *error* and that the 'temptation to imagine intelligence' leads us astray (on this see further Section 5). But if so, this is entirely at odds with the fact that in his famous imitation game, what matters is precisely that the human observer does imagine intelligence in the machine.

Third, and the answer I propose, is: for Turing, that we respond to an intelligent entity in a particular way is *part of the concept* of intelligence, rather than a mere indicator of the behavioural properties that (according to the behaviourist) determine judgements of intelligence. The observer's 'state of mind' is a crucial part of the conditions under which we properly apply the concept of intelligence (or thinking). It follows that whether an entity is in fact intelligent is not determined solely by its behaviour, capacities for behaviour, or dispositions to behaviour. Reading Turing in this third way fits with what comes next in the 'Intelligence as an emotional concept' section of his 1948 report, with his famous 1950 paper, and with his discussion of the problem of free will and determinism.

In the 'Intelligence as an emotional concept' section, Turing remarks, 'It is possible to do a little experiment on these lines, even at the present state of knowledge'. [17] (p. 431). This 'little experiment' is the first version of Turing's computer-imitates-human game (discussions of Turing's game almost always cite his 1950 paper, but there were three versions of the game). In its first format, the game is restricted to chess playing—after all, in the 1940s chess was regarded as a paradigmatic intellectual activity. Here, Turing described a three-player game, involving two human players and a 'paper machine' (a human being, with pencil, paper, and rules to follow, who acts the part of a programmable computer). One of the humans (that is, the judge) plays chess with both the other human and the paper machine. This was not solely a thought experiment; Turing said that it was 'a rather idealized form of an experiment I have actually done' [17] (p. 431). What was the outcome of the experiment? According to Turing, the judge 'may find it quite difficult to tell which he is playing' [17] (p. 431).

Turing's 1950 computer-imitates-human game is another experiment 'on these lines'—that is, it too is based on the assumption that the concept of intelligence is an 'emotional concept'.

The Response-Dependence Interpretation

That the concept of intelligence is in Turing's view an 'emotional concept' explains a puzzle in 'Computing machinery and intelligence'. In this paper, Turing begins with one question, 'Can machines think?', and substitutes another question, 'Are there imaginable digital computers which would do well in the imitation game?' [1] (p. 448). Turing calls the second question a 'variant' of the former, yet these two questions seem entirely dissimilar in meaning, at least at first sight [1] (p. 448). If, however, we remember that for Turing intelligence (or thinking) is an emotional concept, we can make sense of—that is, provide a philosophical justification for—the move from one question to another. Insofar as the second question specifies a sufficient condition for attributing thinking, the second question *is* a 'variant' of the first. So, the philosophical principle underlying the imitation game is precisely that intelligence is an emotional concept.

In modern philosophical terminology, Turing proposed a *response-dependent* 'criterion for "thinking"'. The suggestion is that thinking or intelligence is a response-dependent concept, just as, several thinkers claim, value concepts (e.g., good or beautiful) or secondary-quality concepts (e.g., colour) are response-dependent concepts. Beginning with a basic response-dependence theory—which standardly employs a biconditional—and applying it in the case of intelligence or thinking would give us something like this (without specifying whether the schema is a priori, or necessary, or reductive):

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

That Turing's 'criterion for "thinking"' is solely a *sufficient* condition, however, is an important feature of his account. He made it plain in the 1950 paper that an intelligent machine might fail his test [1] (p. 442).

The task for any response-dependence approach is to specify non-vacuous *normal subjects and conditions*, and we can find these in the words Turing used to describe his three versions of the imitation game (in 1948, 1950, and a 1952 radio symposium). We then arrive at something like the following, 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.

The difference between the 1948 chess-playing imitation game and the version in the famous 1950 paper is that the latter game is unrestricted; it covers almost any possible topic of conversation [1] (p. 442) and so is a much harder test for the computer.[3] In addition, the interrogator must be average, that is not expert about machines; the test involves participation in ordinary human conversation [18] (p. 495).

Turing was interested in real-world machines and in 'building a "thinking machine"' [17] (p. 420). He was also aware of the theoretical possibility of a humongous lookup table machine that would do well in the imitation game (he discussed this with Max Newman in the 1952 broadcast [18] (p. 503)). So, to incorporate his interest in real-world machines and exclude merely logically possible counterexamples to his test, the account so far should be *world-relativized*. Adding this, we arrive at Turing's response-dependent 'criterion for "thinking"', where x is a computer. I shall call this simply *Turing's criterion*:

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

In the rest of this paper, the expression 'Turing's criterion' (or TC) indicates this formulation.

Another consideration in favour of this interpretation (for a less-compressed version of the interpretation, see [2]) is the fact that we can see Turing implicitly using the notion of an emotional concept, so understood, in a different philosophical context—his discussion of the problem of free will and determinism.

4. A Criterion for Free Will

In the late 1940s and early 1950s, the question 'Can machines act freely?' was prominent in speculation about the new 'electronic brains'. Ada, Countess of Lovelace famously said a century earlier that Babbage's Analytical Engine 'has no pretensions whatever to *originate* anything. It can do whatever *we know how to order it* to perform' [19] (p. 722). In the 1940s, we see the use of the expressions 'originate' and 'self-originate' being used to gloss the problem of free will and determinism. Also in the 1940s, people around Turing frequently claimed that machines could *not* behave freely. For example, Douglas Rayner Hartree, computer pioneer and Plummer Professor of Mathematical Physics at Cambridge, said, 'These machines can only do precisely what they are instructed to do by the operators who set them up' [20]. Similarly, Geoffrey Jefferson, who held the first chair of neurosurgery in the UK, at the University of Manchester, argued that the electronic brain was not sufficiently like the biological brain: 'the machine', he said, 'can answer only problems given to it, and, furthermore, . . . the method it employs is one prearranged by its operator' [21] (p. 1109). Turing, in contrast, was confident that a computer could act freely, saying with respect to chess-playing, 'Once [the Manchester computer] has mastered the rules it will think out its own moves'.[4]

Turing generalized Lovelace's remark into the claim that machines in general cannot 'originate anything', and said: 'A variant of Lady Lovelace's objection states that . . . a machine can never "take us by surprise"' [1] (p. 455). If we take Lovelace as making *this* objection to the possibility of free will (or self-origination) in machines, then Turing replies: 'This statement . . . can be met directly. Machines take me by surprise with great frequency' [1] (p. 455). He admits, however: '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' [1] (p. 456). Moreover, Turing agrees that sometimes an experimenter's surprise is irrelevant to the question whether the machine is acting freely: for example, he says that he is often surprised by machines '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' [1] (pp. 455–456). Yet the fact that in some cases being surprised by a machine is irrelevant to whether the machine can act freely does not imply that in every case this is so; and so Turing says, 'This whole question will be considered again under the heading of learning machines' [1] (p. 455).

For Turing, the ability to learn is crucial to intelligence; as a route to artificial intelligence, he recommended building 'comparatively simple' machines that could be subjected to 'a suitable range of "experience"' [22] (p. 473). Some of these machines were neural networks, which he discussed in his 1948 report and regarded as analogous to the cortex of the human infant (on Turing's anticipation of connectionism, see [23]). In teaching a simple machine, the experimenter, Turing said, is to provide examples from which the machine can generalize. In this case, the machine might produce surprising behaviour that *is* relevant to the question whether it can behave freely. According to Turing:

> We should be pleased when the [learning] machine surprises us, in rather the same way as one is pleased when a [human] pupil does something which he had not been explicitly taught to do. [24] (p. 485)

Teaching the machine can cause it to produce interesting (rather than silly or bizarre) behaviour that surprises the experimenter:

> 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 [24] (p. 485)

In effect, Turing's proposal is that we treat the 'child machine' [1] (p. 460) no differently to the human child: if in these circumstances we would say that a human child is acting freely, we should say this too of a machine.

4.1. Response-Dependence Compatibilism

Turing's treatment of the question 'Can machines act freely?' is analogous to his treatment of the question 'Can machines think?'. He proposes a criterion for free will; and he begins by taking the question 'Can machines act freely?' and substituting for it the question 'Can learning machines take us by surprise?'. What justifies this substitution? For Turing, it is clear, attributions of free will are also determined in part by the responses of the surprised and interested educator. In short, the concept of free will is an emotional concept.

Turing's view is compatibilist. It is a response-dependence sort of compatibilism, quite distinct from other sorts of compatibilism that are based on, for example, second-order reflective mental states or the absence of impediments to action. We can set out an emotional-concept approach to free will in much the same way as for intelligence or thinking. A basic account might go like this (using the standard biconditional and without specifying whether the schema is a priori, or necessary, or reductive):

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

Turing's concern is free will in computers, and there is no reason to think that that he was proposing more than a sufficient condition of free will (as in the case of his 'criterion for "thinking"'). So, where x is a computer:

> 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.

One normal condition is the process of educating the machine; in this case, the normal subject is an experimenter who, Turing says, is 'largely ignorant' of what is going on inside the machine—just as in the case of a human teacher and human child [1] (p. 462). Given that Turing's focus is always on real-world machines, we can also world-relativize this criterion. Adding this, we arrive at Turing's response-dependent criterion for free will), where x is a computer. I shall call this *Turing's criterion for free will* (or TC$_F$):

> 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.

As in the case of intelligence, the criterion does not exclude other sufficient conditions of free will in computers. (For a less-compressed version of this interpretation of Turing on free will, see [25].)

4.2. The Flaw in the Computationalist Interpretation

As an example of computationalism with which to compare Turing's views, I shall take David Chalmer's notion of 'minimal computationalism' [26]. This is a maximally inclusive form of computationalism; the minimal computationalist does not take a position on the many debates around computation, such as whether it is essentially semantic or 'mechanistic', whether analog computation is a species of computation, and so on. According to (Chalmer's account of) the minimal computationalist: '[T]he right kind of computational structure suffices for the possession of a mind, and for the possession of a wide variety of mental properties' [26] (p. 324). It follows that 'a model that is computationally equivalent to a mind will itself be a mind' [26] (p. 344).

Treating the concepts of thinking and free will as emotional concepts is incompatible with minimal computationalism.[5] As we have seen, Turing said that '[t]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'. It follows that no (response-*in*dependent) properties of the machine or brain—including its computational structure or profile—in themselves suffice for thinking.[6] Moreover, if we regard free will as an essential property of mind (as many commentators in the 1940s and 1950s appeared to do), we arrive at the same result; since response-independent properties are insufficient for free will, they are insufficient for mind. So, Turing was not a

minimal computationalist about the mind—nor, if other versions of computionalism all entail minimal computationalism, any sort of computationalist.

This completes the case against the orthodox behaviourist and computationalist interpretations of Turing and his test. Next, I shall clarify and develop Turing's criterion, by addressing a series of questions. The first question relates to the expression 'emotional concept', which I interpret as a term of art for Turing and translatable in modern philosophical vocabulary as a response-dependent concept. But might Turing have intended a more everyday meaning?

5. What Did Turing Mean by 'Emotional'?

Some theorists understand Turing's use of the term 'emotional' to suggest, not a conceptual role for response, but rather the presence of bias or prejudice in judgements of intelligence in machines. This understanding fits with an everyday use of 'emotional' to indicate strong, and perhaps cognitively unjustified, feeling.

According to Peter Millican, for example, Turing 'need not be interpreted as saying that our judgements of intelligence *ought to be* made on an "emotional" basis, nor that intelligence is a *response-dependent* concept' [29] (p. 38). Millican claims instead that, in the section of Turing's report entitled 'Intelligence as an emotional concept', he should be read as concerned with the 'overcoming of *prejudice*' [29] (p. 38). For Millican, Turing's view is that 'judgements of intelligence are subjectively dependent on our own assumptions, and even "emotional" in character', and that we should 'guard against this' [29] (p. 38). This reading is very different from the response-dependence interpretation: according to the latter, Turing simply takes human response as central to our concept of thinking, with no negative implications for the cognitive or epistemic status of judgements of intelligence.

Bernardo Gonçalves' position is similar to Millican's. Gonçalves claims that Turing's view of intelligence as an emotional concept is 'negative in a certain sense' and that:

> Turing's concept of intelligence as emotional seems to have been conceived to prevent or block confirmation bias towards the conservative position that machines never thought, can't think and will never think. [30] (pp. 67–68)

According to Gonçalves, in talking of an 'emotional' response to machines, Turing was thinking only of negative reactions that some people in the 1940s and 1950s had to the idea of an 'electronic brain' (on popular reactions to the early computers, see [31]).

Three Senses of 'Emotional'

In fact, Turing used the expression 'emotional' in three distinct ways—to discuss emotional *concepts*, emotional *arguments*, and emotional *communication* respectively. I have presented the first use in this paper, where an emotional concept is a response-dependent concept. Turing employed this sense of 'emotional' again when he said that 'the idea of "intelligence" is itself emotional rather than mathematical'. [17] (p. 411).

Turing's second use of 'emotional' was to describe certain objections to the possibility of machine intelligence. Turing distinguished between (what he described as) 'rational' and 'irrational' objections to the possibility of machine intelligence [24] (p. 485). In his view, some objections to AI are 'purely emotional' [17] (p. 411). Purely emotional objections, Turing said, 'do not really need to be refuted. If one feels it necessary to refute them there is little to be said that could hope to prevail' [17] (p. 411). In this sense of 'emotional', an emotional objection is not susceptible to reasoned argument. Those who propose such objections, Turing said, 'would probably not be interested in any criteria' for a machine's being intelligent [1] (p. 451).

Turing's third use of the expression 'emotional' was to refer to inputs and outputs to his P-type unorganised machines—these, he said, were 'pain' (or 'punishment') and 'pleasure' (or 'reward') inputs [17] (p. 425). A pain signal cancels a tentative entry in the P-type's machine table and a pleasure signal confirms it. Organizing the P-type requires additional inputs, since, Turing joked, if a child learns only by means of punishment and reward he or she 'would probably feel very sore indeed' [1] (p. 461). Turing called these

additional channels of communication 'sense stimuli': 'The sense stimuli are means by which the teacher communicates "unemotionally" to the machine, i.e., otherwise than by pleasure and pain stimuli', he said [17] (p. 426). By implication, pain and pleasure signals are the mechanism of *emotional* communication, which is to facilitate machine learning.

Turing was certainly aware of a bias against the possibility (and actual emergence) of thinking machines, on the part of pioneering computer scientists and the general public; indeed, he satirized sceptical and dystopian views of AI (on this see [32]). However, given Turing's multiplicity of uses of the expression 'emotional', there can be no argument merely from his saying that the concept of intelligence is an 'emotional' concept to conclude that he viewed all judgements of intelligence as irrational or partial.[7] Moreover, no other remark in the important last section of Turing's 1948 paper suggests that an emotional concept is linked to bias (or prejudice, or irrationality).

Nevertheless, even if judgements of an 'emotional' concept need not involve bias, it might be assumed that such judgements are inherently *subjective*—and as such no proper basis for judgements of intelligence.

6. Is Turing's Criterion Subjective?

Discussions of response-dependence approaches frequently use the term 'objective' in the sense that no judgement is objective unless it describes a *mind-independent* state of affairs (otherwise it is subjective). In this sense, Turing's criterion is subjective. However, there is a demand, growing in popularity, that an account of objectivity 'should not claim that we can have objective knowledge only of objects that exist independently of human beings', but instead should also apply to 'phenomena dependent on human conceptualizations' [33] (p. 1190). In effect, that is, we should separate the concepts of objectivity and realism. So, in this and Section 7 I consider only the question whether Turing's criterion is subjective—distinguishing this from the question whether on the response-dependence interpretation intelligence is *real*, even though mind-dependent (I address the latter question in a forthcoming paper [28]).

6.1. A Mathematically Empty Notion?

Gonçalves appears to take just from Turing's saying that intelligence is an 'emotional' concept that in his view judgements of intelligence are subjective; he describes Turing's account as involving a 'subjective element' and 'subjective observers' [30] (pp. 99, 245). Gonçalves is not alone in taking this implication from Turing's use of the term 'emotional'. For example, Gualtiero Piccinini remarks, 'Turing called intelligence an "emotional concept," meaning that there is no objective way to apply it.' [34] (p. 117). Teresa Numerico also says that, for Turing, '[a]s suggested by the title of the last paragraph of his paper, intelligence is an emotional and subjective concept' [35].

Further, according to Gonçalves, any 'subjective' account (or test) of intelligence, including Turing's criterion, is unsatisfactory. In his view, this applies to the response-dependence interpretation, since (he says) subjective 'emotional' judgements have no scientific value. An 'emotional concept', Gonçalves claims, is a 'mathematically empty notion'; with such a notion, 'neither Turing nor anyone could get anywhere near building an intelligent machine' [30] (p. 68). Gonçalves concludes that, since building an intelligent machine 'was the very reason why Turing eventually found himself debating whether machines can think', the response-dependence interpretation 'does not seem to make sense at all' [30] (p. 68).

However, much more needs to be said here, to make out a case against the response-dependence interpretation. What meaning is given to the term 'subjective' (or 'objective') in these various claims about Turing's approach? If TC is a 'subjective' criterion for 'thinking', exactly what notion and test of objectivity does it fail to satisfy?

6.2. Which Conception of Objectivity?

For Gonçalves, objectivity is restricted to *scientific* objectivity, which he understands in terms of (presumably standpoint-independent) *measurement*. He says, 'By an objective decision, I mean one depending solely on the properties of the object under consideration as measured by a scientific instrument' [30] (p. 65). With this notion of objectivity, Gonçalves declares that, on the response-dependence interpretation, judgements of intelligence are not objective: on this interpretation, he says, Turing's view is that 'intelligence is to be . . . decided inter-subjectively but never objectively' [30] (p. 65).[8]

Gonçalves' understanding of 'objective', however, is only one among various senses of this term. Even if we restrict the domain to science, his understanding is not a universally accepted conception of objectivity (on the variety of conceptions, see [36]). Many philosophers of science dispute both the possibility of a conception like Gonçalves' (claiming, for example, that measurement too is standpoint-relative) or its desirability (arguing, for example, that this conception stifles science, and should be replaced by epistemic pluralism).

In a response-dependence theory, the notions of 'normal' or 'ideal' subjects and conditions are intended precisely to accommodate the intuition that response is crucial to the concept concerned while avoiding individual biases that may distort judgements. This fits with an alternative conception of objectivity, according to which judgements are objective if and only if they are free from 'personal biases and idiosyncrasies' [36] (p. 31) This alternative conception allows the possibility that, using Turing's criterion, judgements of intelligence in machines *are* objective (or are *not* subjective).[9]

In addition to the response-dependence framework, specific features of Turing's test are intended to exclude bias. For example, the imitation game controls for what he called 'irrelevant disabilities'. Turing said that a machine should not be punished for disabilities, such as an 'inability to shine in beauty competitions', that are 'irrelevant' to whether the machine can think [1] (p. 442); to avoid this, the machine is hidden. Another example is that the imitation game is anthropomorphism-proofed. Turing was aware of an anthropomorphic bias towards the position that machines *can* think; in his 1948 report, he said that playing against a 'paper machine' gives 'a definite feeling that one is pitting one's wits against something alive' [17] (p. 412). Indeed, the tendency to anthropomorphize may be the biggest source of bias in judging intelligence in machines. Turing's 3-player computer-imitates-human game disincentivizes anthropomorphizing: we know from early Loebner contests that interrogators wish to avoid making the embarrassing mistake of misidentifying a machine as a human—they would rather misidentify a human as a machine—and they are extremely suspicious (see [2,38,39]). The three-player game also controls for anthropomorphizing: it can be assumed that that the parallel blind interrogations of the machine and human contestants are, across sufficiently many trials, affected equally by an interrogator's tendency to anthropomorphize (for greater detail on anthropomorphism-proofing, see [40]). These features of the imitation game tally with the conception of an 'objective' method as one 'that screens out the possibility of individual biases or idiosyncrasies distorting the results' [33] (p. 1198).

However, even if we were to accept the 'subjective' label for Turing's criterion (solely to acknowledge that judgements of intelligence are in part determined by 'our own state of mind and training'), would it follow that it is unsatisfactory?

7. Does Turing's Criterion Have the Vices of Subjectivity?

Several philosophers argue that the notions of 'objective' and 'subjective' should be eliminated. For example, Ian Hacking famously exhorted, 'Let's not talk about objectivity' [41]. According to Hacking, the expression 'objective' (and 'subjective' also, it would follow) is an 'elevator word', and he advises us to 'stick to ground-level questions' [41] (p. 20). Elevator words, Hacking claims, generate 'grandiose important-sounding but idle controversies' [41] (p. 24). In his view, 'the invocation of objectivity gets you nowhere' and he advises: 'Instead, get some facts' [41] (p. 28). Which facts? Hacking holds that 'objectivity is not a virtue but the absence of various types of vice' [41] (p. 22); subjectivity,

then, is the presence of specific vices, such as ignoring evidence, proceeding by whim, or ignoring criticism. Investigating whether these vices are present is 'ground-level' work.

Gonçalves' characterization of Turing's criterion as 'subjective' is a good example of the repeated use of an elevator-word. He does also, however, attribute three vices to Turing's criterion. First, as quoted above, Gonçalves claims that TC fails to get us 'anywhere near' building an intelligent machine. Second, he says that, given TC, 'Turing's research program would better have been to build a machine not to learn in general, but simply to deceive' [30] (p. 68). And, third, since an interrogator's response is 'subjective', the engineer of the machine should simply opt for the easiest way to induce the appropriate response in the 'subjective observer'—regardless of whether the machine really is intelligent. The result of the last, Gonçalves claims, is that TC will 'fal[l] prey to mechanical parrots'— lookup table machines or chatbots like those entered in the annual Loebner Prize Contest in Artificial Intelligence will do well in Turing's imitation game [30] (p. 119). This, Gonçalves says, would go against Turing's aim for his test, since he excluded mechanical parrots as 'legitimate contestants' [30] (p. 120).

However, Gonçalves fails to demonstrate that Turing's criterion possesses these vices.

Not a Test for Mechanical Parrots

Gonçalves' reason for claiming that TC could not get us 'anywhere near' building an intelligent machine appears to be that the interrogator's response is (he says) 'mathematically empty': it tells you nothing about how to *build* a computer that would do well in the game. However, if we are to get anywhere near designing a human-level (or human-like) intelligent machine—or 'artificial general intelligence'—we need to know when we have achieved this.[10] We need a test, and moreover one that is anthropomorphism-proofed. TC provides this, and so in this respect is definitely not 'empty'. A test that supplies both a criterion of thinking and a means of developing candidate intelligent machines might be ideal, but there is no reason to think that anything less than this is scientifically vacuous. And in any case, TC *can* contribute to developing an intelligent machine, by means of a series of increasingly demanding tests; in this case the human player in the imitation game changes, for example from child to adult, to reflect increasing cognitive skills.

There is also no reason to think that, just because TC tests the interrogator's response, a machine should be built 'simply to deceive'. Or that one built 'simply to deceive' (for example, by using canned remarks triggered by the interrogator's inputs, as employed by chatbots in Loebner Prize Contests) would do better in the game than other machines.[11] Nor does it follow from using TC that a humongous 'mechanical parrot' would do well in the game. On the contrary, as Max Newman said, although the Manchester machine could be programmed to find the best move in chess by 'analys[ing] all possible variations of the game', the machine would take 'thousands of millions of years' to run the program ([18] (p. 503); on this, see [39]). And, as noted above, Turing's criterion is world-relativized precisely to exclude merely logically possible lookup-table machines (on this, see [2]).

Summing up Sections 6 and 7, there is scope to say that, on some plausible conception of objectivity, TC *can* yield objective judgements of intelligence in machines. Moreover, if labelling TC as 'subjective' is to be more than handwaving, specific 'vices' must be demonstrated; Gonçalves, at least, has *not* done this.

Some theorists regard inter-subjective agreement as sufficient for objectivity. This leads to another question about Turing's criterion.

8. Are 'Emotional' Judgements Decided by Society?

A recent account that answers *yes* is presented by Shlomo Danziger. For Turing, 'what matters', Danziger claims, is whether machines would be seen as intelligent by human society; 'If they would be—then they could be said to be "intelligent", and no further inquiry would be needed' [46] (p. 162). In Danziger's view, Turing took a 'social-relationist' as 'opposed to' (what Danziger calls) a 'realist' approach to the mind [46] (p. 162).

According to Danziger, his interpretation is 'close' to the response-dependence interpretation, but differs from it in 'small but crucial points' [46] (p. 169). Turing, Danziger says, regarded the connection between the imitation game and intelligence 'not as a *logical* connection, but as a *causal* one': machines' doing well in the game does not make it the case that they '*are* intelligent', but such success 'would eventually *cause* people to see machines' as intelligent [46] (p. 170). The criterion for a machine truly to *be* intelligent, according to Danziger, is 'sociolinguistic': the machine must be 'perceived by society as a potentially intelligent entity' [46] (p. 158).

Piccinini may offer a similar interpretation when (as quoted above) he says that, for Turing, 'there is no objective way to apply' the concept of intelligence. Piccinini adds, regarding Turing's 1948 chess-playing imitation game, that direct experience with a chess-playing machine 'could convince one to attribute intelligence to the machine'; and regarding the 1950 unrestricted imitation game, that Turing 'hoped that . . . people would modify their usage of terms like "intelligence" and "thinking", so that such terms apply to the machines themselves' [34] (p. 117).

On the social-relationist interpretation, Turing's imitation game is not a test of intelligence or thinking at all, since it does not test the view of 'human society'; and the response-dependence interpretation's analysis of Turing's 'criterion for "thinking"' is incorrect.[12] However, Danziger's interpretation is at odds with what Turing actually said.

Not to Be Decided by Gallup Poll

Although Turing famously predicted that 'the use of words' would change so that by the end of the 20th century we would be able 'to speak of machines thinking without expecting to be contradicted' [1] (p. 449), it is highly unlikely that his hope or aim was to modify ordinary linguistic usage. Turing often mocked the general public's attitudes to machines; and he showed his disregard for society's view when he said that it is 'absurd' to try to answer the question 'Can machines think?' by means of 'a statistical survey such as a Gallup poll' [1] (p. 441).

In addition, Turing said that, when a machine is deemed to have an ability usually reserved for human beings, people claim that how the machine does this is 'really rather base'—they say, 'Well, yes, I see that a machine could do all that, but I wouldn't call it thinking' [18] (p. 500). This is bad news for Danziger's interpretation. If Turing himself thought—correctly[13]—that human beings would regard machines' success at supposedly intelligence-demanding tasks merely as evidence that the tasks did not after all require thought, he could hardly have intended the imitation game in the way Danziger claims, namely as a means to cause society to perceive machines as intelligent.

As we have seen, in describing his computer-imitates-human game, Turing referred to his 'criterion for "thinking"' and to the 'advantages' of the 'proposed criterion' [1] (p. 442). He treated successful performance in the game as sufficient for thinking, absent any mention of social acceptance. Notably, he described his question 'Are there imaginable digital computers which would do well in the imitation game?' as a 'variant' on the question 'Can machines think?'—whereas, on the social-relationist interpretation, these questions are not variants, since doing well in the game does *not* suffice for satisfying the 'sociolinguistic' criterion.

The social-relationist interpretation faces additional exegetical problems. Danziger explains Turing's claim that intelligence is an emotional concept as follows: 'Intelligence, so to speak, is in the eye of the beholder' [46] (p. 161). He also claims that Turing provides a 'descriptive' rather than a 'normative' account of intelligence [46] (p. 170). Here, Danziger is close to the view that judging intelligence is a matter of taste—too close, I suggest, to attribute this view to Turing, a natural scientist who wanted to build a thinking machine. In contrast, TC specifies idealized subjects in idealized conditions; the aim of the response-dependence interpretation is to recognize response as part of our concept of intelligence *without* conceding that intelligence is in the eye of the beholder. Turing's criterion gives us one way of *properly* judging intelligence in machines.[14]

Turing's words align more with TC than with the social-relationist interpretation. In a 1951 talk entitled 'Intelligent Machinery: a Heretical Theory', Turing said of the claim 'You cannot make a machine to think for you' that it was 'a commonplace that is usually accepted without question' [22] (p. 472). Nevertheless, he said, 'It will be the purpose of this paper to question it.' [22] (p. 472). Turing's attack on commonplace beliefs rebuts the claim that in his view 'what matters' is how society sees machines.

One last question, to fill in my interpretation of the historical Turing. In this paper, I base my approach to Turing's 'criterion for "thinking"' on his explicit remarks about intelligence. But could it be that Turing, in writing these words, was not entirely serious?

9. Philosophy or Propaganda?

Millican seems to suggest that we should simply ignore Turing's actual words where they conflict with the orthodox interpretations of Turing's test. He says:

> In some cases Turing's own position might not have been entirely clear, and sometimes he might simply have been careless in expressing it. In seeking to understand the Turing Test, therefore, literal reading of the text must be subject to the discretion of interpretative judgement. [29] (p. 39) [15]

Recall that Millican claimed that Turing 'need not be interpreted as saying that our judgements of intelligence *ought to be* made on an "emotional" basis, nor that intelligence is a *response-dependent* concept'. So, is the response-dependence interpretation based on 'careless' comments?

The strategy of privileging 'interpretative judgement' over Turing's text leaves Millican with questions to answer. On the response-dependence interpretation, it is the fact that for Turing intelligence is an emotional concept that explains his move from the question 'Can machines think?' to the question 'Are there imaginable digital computers which would do well in the imitation game?'. Having rejected the response-dependence interpretation, how can Millican explain Turing's move? He says only that Turing 'never explains or even discusses' why he replaces one question with the other [29] (p. 39). This absence of explanation is an example of what Millican describes as 'vagueness' and 'loose argument' in Turing's 1950 paper; the paper 'falls far short of the rigour' that we now expect of a paper in a journal such as *Mind*, he says [29] (p. 39).

Yet what is to explain this (supposed) vagueness and lack of rigour in a paper written by a mathematical logician of extraordinary ability, who had been thinking about artificial intelligence for several years? In answer, Millican quotes Robin Gandy's remark that Turing's 1950 paper was 'intended not so much as a penetrating contribution to philosophy but as propaganda', and that some discussions of the paper 'load it with more significance than it was intended to bear' [29] (p. 39). [16] In short, Turing was not serious.

The Cost of Rejecting Turing's Text

Claiming that Turing was 'careless' and 'not . . . entirely clear' in describing his 'criterion for "thinking"', just because Turing's words do not fit Millican's reading, is a high-risk strategy. To disregard Turing's own words, the explanatory pay-off would have to be high. However, Millican's account leaves much unexplained.

Millican describes the beginning of 'Computing machinery and intelligence', which depicts the man-imitates-woman game, as 'cavalier and rather confusing' [29] (p. 38)—another example of the 'vagueness' and 'loose argument' in Turing's paper. For Millican, the man-imitates-woman game is of 'little relevance' to Turing's test [29] (p. 42). However, even if we ignore Turing's words about this game (Section 2.1), this leaves the question: if the computer-imitates-human game is not benchmarked against the man-imitates-woman game, how *is* it scored?

Perhaps it might be claimed that this question too is simply one that Turing never answers, just because his aim in the 1950 paper was merely propaganda rather than philosophy. However, this would not explain Turing's related remarks about intelligence

in his 1948 report. Millican does not suggest that *this* report is vague or careless, or that it was intended non-seriously or as propaganda—after all, it was written for the Executive Committee of the NPL, so this is unlikely. Without a reason to reject Turing's account of intelligence in the section on 'Intelligence as an emotional concept', the attempt to explain away the related claims in his 1950 paper merely as 'careless' and 'not . . . entirely clear' cannot get off the ground.

Proponents of the orthodox behaviourist interpretation of Turing's test put the cart before the horse when they describe the test's design as 'peculiar' or 'strange' just because it does not fit their interpretation. Rejecting Turing's own words about his 'criterion for "thinking"' makes the same mistake.

10. Conclusions

On the response-dependence interpretation, Turing's 'criterion for "thinking"' is an original contribution to fundamental debates in the philosophy of artificial intelligence and to current discussions of response-dependence. It is not undermined by the challenges addressed in this paper.

Notes

1.   A notable exception to the orthodox behaviourist interpretation is the inductive or epistemic interpretation, the canonical form of which is given by James Moor [6,7]. I argue that this interpretation too does not fit Turing's actual words in [2] (pp. 402–404).

2.   I am grateful to an anonymous reviewer for Philosophies for this first reading.

3.   Turing does not say how long the game is or how many occurrences are involved.

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

5.   For a less-compressed version of this argument against reading Turing as a computationalist, see [3].

6.   According to Michael Wheeler [27], my interpretation assumes one particular version of response-dependence theories; I respond to this objection in a forthcoming paper [28].

7.   Gonçalves conflates Turing's use of the two expressions, 'emotional' and 'purely emotional', when he refers to Turing's discussion of the 'heads in the sand' and 'theological' objections as the 'specific context' in which Turing 'stated that intelligence is an emotional concept' [30] (p. 66).

8.   I assume that the implicit premise here is that neither the imitation-game interrogator nor the whole game can qualify as a 'scientific instrument'.

9.   Not, however, if, 'mechanical objectivity' [37] is taken to be the ideal.

10.   This is again seen by many working in the field as a central and feasible goal. For example, John McCarthy said, 'Human-level AI will be achieved . . . I'd be inclined to bet on this 21st century' [42] (p. 1174); similarly, Ben Goertzel says, 'AGI [artificial general intelligence] should be the focus of a significant percentage of contemporary AI research . . . [and] dramatic progress on AGI in the near future is something that's reasonably likely' [43] (p. 1163). Brian Cantwell Smith, who thinks human-level AI is much further away, nevertheless argues that the 'ultimate' goal is 'genuine judgment' in machines—this is a 'form of dispassionate deliberative thought' that is 'the standard to which human thinking must ultimately aspire' [44] (pp. xvi, xv).

11.   There is a caveat to this. So as not to be caught out simply by (say) mathematical speed, the machine contestant in the game must 'deceive' the interrogator with respect to its speed and accuracy at typing or mathematics, and suchlike abilities. This is the source of the familiar 'artificial stupidity' objection to Turing's test [45].

12.   The view that Turing's test was not in fact intended as a test for intelligence or thinking has been put forward by several commentators (e.g., [47–49]).

13   This is exactly how, half a century later, critics of Deep Blue and Watson responded. For example, Douglas Hofstadter reacted to Deep Blue's defeat of Garry Kasparov in 1997, 'My God, I used to think chess required thought. Now, I realize it doesn't' (as quoted in B. Weber, 'Mean chess-playing computer tears at meaning of thought', New York Times (19 February 1996)).

14   To what degree attempts to defend the objectivity or normativity of response-dependent concepts are successful is a different question; for the argument in this paper, it is enough that there are serious debates around this.

15   According to Millican, Turing should have adopted a theory of intelligence that left aside the 'human-centred' notion of thinking, which Millican links with 'subjectivity' and 'consciousness' [29]. In Millican's view, we should reserve the word 'intelligence' for 'sophisticated information processing for some purpose' [29] (pp. 49, 24). Turing, he claims, 'should have been prepared to break away from the human-centred paradigm of "intelligence" that he had strategically highlighted in his famous Test'—that is, when he asked if machines can think [29] (p. 49). This is because Turing's 'own work gave ample reason to reinterpret intelligence' in this way [29] (p. 24). Millican concedes that this 'would involve some revision of our naïve conceptual scheme, away from a human-centred view of intelligence' [29] (p. 49). Nevertheless, Turing 'should have had the candour' (in the online version of Millican's paper, this is 'courage') to make this move [29] (p. 49).

16   The quotation is taken from [50] (p. 125). Daniel Dennett makes a remark that might seem to agree with Millican's and Gandy's positions. According to Dennett, Turing did not design his test as 'a useful tool in scientific psychology' but rather 'to be nothing more than a philosophical conversation-stopper' [51] (p. 122). Nevertheless, in Dennett's view, Turing's test *is* a test of intelligence; he continues, '[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).' [51] (p. 122).

References

1. Turing, A.M. Computing Machinery and Intelligence. Mind 1950, 59, 433–460. Reproduced in 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, pp. 441–464.
2. Proudfoot, D. Rethinking Turing's Test. *J. Philos.* **2013**, *110*, 391–411. [CrossRef]
3. Proudfoot, D. Rethinking Turing's Test and the Philosophical Implications. *Minds Mach. Dordr.* **2020**, *30*, 487–512. [CrossRef]
4. Mays, W. Can Machines Think? *Philosophy* **1952**, *27*, 148–162. [CrossRef]
5. Hodges, A. *Alan Turing: The Enigma, revised*; Vintage: London, 2014.
6. Moor, J.H. An Analysis of the Turing Test. *Philos. Stud.* **1976**, *30*, 249–257. [CrossRef]
7. Moor, J.H. Turing Test. In *Encyclopedia of Artificial Intelligence*; Shapiro, S., Ed.; Wiley: Hoboken, NJ, USA, 1987; Volume 2, pp. 1126–1130.
8. Block, N. Psychologism and Behaviorism. *Philos. Rev.* **1981**, *90*, 5–43. [CrossRef]
9. Ford, K.M.; Hayes, P.J. On Computational Wings: Rethinking the Goals of Artificial Intelligence. *Sci. Am. Presents* **1998**, *9*, 78–83.
10. French, R.M. The Turing Test: The First 50 Years. *Trends Cogn. Sci.* **2000**, *4*, 115–122. [CrossRef]
11. Shieber, S.M. The Turing Test as Interactive Proof. *Nous* **2007**, *41*, 686–713. [CrossRef]
12. Pinar Saygin, A.; Cicekli, I.; Akman, V. Turing Test: 50 Years Later. *Minds Mach.* **2000**, *10*, 463–518. [CrossRef]
13. Dodig-Crnkovic, G. Alan Turing's Legacy: Info-Computational Philosophy of Nature. *arXiv* **2012**, arXiv:1207.1033.
14. Simon, H.A. Machine as Mind. In *Android Epistemology*; Ford, K.M., Glymour, C., Hayes, P.J., Eds.; The MIT Press: Cambridge, MA, USA, 1995; pp. 23–40.
15. Searle, J.R. *The Rediscovery of the Mind*; MIT Press: Cambridge, MA, USA, 1994.
16. Fodor, J.A. *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*; Bradford Books: Cambridge, MA, USA, 2001.
17. Turing, A.M. Intelligent Machinery, 1948. Reproduced in Copeland Ed. 2004; pp. 410–432.
18. Turing, A.M.; Braithwaite, R.; Jefferson, G.; Newman, M. Can Automatic Calculating Machines Be Said to Think? Reproduced in Copeland Ed. 2004; pp. 494–506.
19. Lovelace, A.A. Notes by the Translator (Addenda to Her Translation of L.F. Menabrea, 'Sketch of The Analytical Engine invented by Charles Babbage'). In *Scientific Memoirs, Selected from the Transactions of Foreign Academies of Science and Learned Societies, and from Foreign Journals*; Taylor, R., Ed.; Richard and John E. Taylor: London, UK, 1843; Volume 3, pp. 691–731.
20. Hartree, D.R. The 'Electronic Brain. *The Times*, 7 November 1946; 5.
21. Jefferson, G. The Mind of Mechanical Man. *Br. Med. J.* **1949**, *1*, 1105–1110. [CrossRef] [PubMed]
22. Turing, A.M. Intelligent Machinery: A Heretical Theory, 1951. Reproduced in Copeland Ed. 2004; pp. 472–475.
23. Copeland, B.J.; Proudfoot, D. On Alan Turing's Anticipation of Connectionism. *Synthese* **1996**, *108*, 361–377. [CrossRef]
24. Turing, A.M. Can Digital Computers Think? *Reproduced in Copeland Ed.* 2004; pp. 482–486.
25. Proudfoot, D. Turing and Free Will: A New Take on An Old Debate. In *Philosophical Explorations of the Legacy of Alan Turing: Turing 100. Boston Studies in the Philosophy and History of Science*; Floyd, J., Bokulich, A., Eds.; Springer Verlag: Berlin, Germany, 2017; pp. 305–322.
26. Chalmers, D.J. A Computational Foundation for the Study of Cognition. *J. Cogn. Sci.* **2011**, *12*, 323–357.
27. Wheeler, M. Deceptive Appearances: The Turing Test, Response-Dependence, and Intelligence as an Emotional Concept. *Minds Mach. Dordr.* **2020**, *30*, 513–532. [CrossRef]

28. Proudfoot, D. Intelligence Naturalized, Turing-Style. Forthcoming.
29. Millican, P.J.R. Alan Turing and Human-Like Intelligence. In *Human-Like Machine Intelligence*; Muggleton, S., Chater, N., Eds.; Oxford University Press: Oxford, UK, 2021; pp. 28–51.
30. Gonçalves, B. Machines Will Think: Structure and Interpretation of Alan Turing's Imitation Game. Ph.D. Thesis, Tese Apresentada ao Programa de Pós-graduação em Filosofia do Departamento de Filosofia da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo,, Sao Paulo, Brazil, 2020.
31. Proudfoot, D.; Copeland, B.J. Turing and the First Electronic Brains: What the Papers Said. In *The Routledge Handbook of the Computational Mind*; Sprevak, M., Colombo, M., Eds.; Routledge: London, 2019; pp. 23–37.
32. Proudfoot, D. Mocking AI Panic. *Spectrum* **2015**, *52*, 46–47.
33. Koskinen, I. Defending a Risk Account of Scientific Objectivity. *Br. J. Philos. Sci.* **2020**, *71*, 1187–1207. [CrossRef]
34. Piccinini, G. Turing's Rules for the Imitation Game. *Minds Mach.* **2000**, *10*, 573–582. [CrossRef]
35. Numerico, T. Google AI and Turing's Social Definition of Intelligence. In SIS Summit Vienna 2015—The Information Society at the Crossroads session Conference Stream ICTS 2015, Vienna Austria, 3–7 June 2015. Available online: https://sciforum.net/manuscripts/2996/manuscript.pdf (accessed on 10 October 2022).
36. Reiss, J.; Sprenger, J. *Scientific Objectivity. The Stanford Encyclopedia of Philosophy (Winter 2020 Edition)*; Zalta, E.N., Ed.; Stanford: Stanford, CA, USA, 2020; Available online: https://plato.stanford.edu/archives/win2020/entries/scientific-objectivity/ (accessed on 10 October 2022).
37. Daston, L.; Galison, P. The Image of Objectivity. *Representations* **1992**, *40*, 81–128. [CrossRef]
38. Moor, J.H. The Status and Future of the Turing Test. *Minds Mach.* **2001**, *11*, 77–93. [CrossRef]
39. Copeland, B.J. The Turing Test. *Minds Mach.* **2000**, *10*, 519–539. [CrossRef]
40. Proudfoot, D. Anthropomorphism and AI: Turing's Much Misunderstood Imitation Game. *Artif. Intell.* **2011**, *175*, 950–957. [CrossRef]
41. Hacking, I. Let's Not Talk About Objectivity. In *Objectivity in Science*; Padovani, F., Richardson, A., Tsou, J.Y., Eds.; Boston Studies in the Philosophy and History of Science; Springer International Publishing: Cham, Switzerland, 2015; Volume 310, pp. 19–33. [CrossRef]
42. McCarthy, J. From Here to Human-Level AI. *Artif. Intell.* **2007**, *171*, 1174–1182. [CrossRef]
43. Goertzel, B. Human-Level Artificial General Intelligence and the Possibility of a Technological Singularity. *Artif. Intell.* **2007**, *171*, 1161–1173. [CrossRef]
44. Smith, B.C. *The Promise of Artificial Intelligence: Reckoning and Judgment*; MIT Press: Cambridge, 2019.
45. Artificial Stupidity. *The Economist*, August. 1992; 14.
46. Danziger, S. Where Intelligence Lies: Externalist and Sociolinguistic Perspectives on the Turing Test and AI. In *Philosophy and Theory of Artificial Intelligence 2017*; Springer International Publishing: Cham, Switzerland, 2018; pp. 158–174. [CrossRef]
47. Whitby, B. The Turing Test: AI's Biggest Blind Alley. In *The Legacy of Alan Turing*; Millican, P.J.R., Clark, A., Eds.; Oxford University Press: Oxford, UK, 1996; Volume I, pp. 53–62.
48. Narayanan, A. The Intentional Stance and the Imitation Game. In *The Legacy of Alan Turing*; Millican, P.J.R., Clark, A., Eds.; Oxford University Press: Oxford, UK, 1996; Volume I, pp. 63–80.
49. Sloman, A. The Mythical Turing Test. In *Alan Turing: His Work and Impact*; Cooper, S.B., van Leeuwen, J., Eds.; Elsevier: Amsterdam, The Netherlands, 2013; pp. 606–611.
50. Gandy, R. Human versus Mechanical Intelligence. In *The Legacy of Alan Turing*; Millican, P.J.R., Clark, A., Eds.; Clarendon Press: Oxford, UK, 1996; pp. 125–136.
51. Dennett, D.C. Can Machines Think. In *How We Know: Nobel Conference XX*; Shafto, M.G., Teuscher, C., Eds.; Harper & Row: San Francisco, CA, USA, 1985; pp. 295–316.