Stanford Encyclopedia of Philosophy

The Turing Test

First published Wed Apr 9, 2003; substantive revision Mon Oct 4, 2021

The phrase "The Turing Test" is most properly used to refer to a proposal made by Turing (1950) as a way of dealing with the question whether machines can think. According to Turing, the question whether machines can think is itself "too meaningless" to deserve discussion (442). However, if we consider the more precise—and somehow related—question whether a digital computer can do well in a certain kind of game that Turing describes ("The Imitation Game"), then—at least in Turing's eyes—we do have a question that admits of precise discussion. Moreover, as we shall see, Turing himself thought that it would not be too long before we did have digital computers that could "do well" in the Imitation Game.

---

1. Turing (1950) and the Imitation Game

Turing (1950) describes the following kind of game. Suppose that we have a person, a machine, and an interrogator. The interrogator is in a room separated from the other person and the machine. The object of the game is for the interrogator to determine which of the other two is the person, and which is the machine. The interrogator knows the other person and the machine by the labels '*X*' and '*Y*'—but, at least at the beginning of the game, does not know which of the other person and the machine is '*X*'—and at the end of the game says either '*X* is the person and *Y* is the machine' or '*X* is the machine and *Y* is the person'. The interrogator is allowed to put questions to the person and the machine of the following kind: "Will *X* please tell me whether *X* plays chess?" Whichever of the machine and the other person is *X* must answer questions that are addressed to *X*. The object of the machine is to try to cause the interrogator to mistakenly conclude that the machine is the other person; the object of the other person is to try to help the interrogator to correctly identify the machine. About this game, Turing (1950) says:

> I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning. … I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.

There are at least two kinds of questions that can be raised about Turing's predictions concerning his Imitation Game. First, there are empirical questions, e.g., Is it true that we now—or will soon—have made computers that can play the imitation game so well that an average interrogator has no more than a 70 percent chance of making the right identification after five minutes of questioning? Second, there are conceptual questions, e.g., Is it true that, if an average interrogator had no more than a 70 percent chance of making the right identification after five minutes of questioning, we should conclude that the machine exhibits some level of thought, or intelligence, or mentality?

There is little doubt that Turing would have been disappointed by the state of play at the end of the twentieth century. Participants in the Loebner Prize Competition—an annual event in which computer programmes are submitted to the Turing Test— had come nowhere near the standard that Turing envisaged. A quick look at the transcripts of the participants for the preceding decade reveals that the entered programs were all easily detected by a range of not-very-subtle lines of questioning. Moreover, major players in the field regularly claimed that the Loebner Prize Competition was an embarrassment precisely because we were still so far from having a computer programme that could carry out a decent conversation for a period of five minutes—see, for example, Shieber (1994). It was widely conceded on all sides that the programs entered in the Loebner Prize Competition were designed solely with the aim of winning the minor prize of best competitor for the year, with no thought that the embodied strategies would actually yield something capable of passing the Turing Test.

At the end of the second decade of the twenty-first century, it is unclear how much has changed. On the one hand, there have been interesting developments in language generators. In particular, the release of Open AI's GPT-3 (Brown, et al. 2020, Other Internet Resources) has prompted a flurry of excitement. GPT-3 is quite good at generating fiction, poetry, press releases, code, music, jokes, technical manuals, and news articles. Perhaps, as Chalmers speculates (2020, Other Internet Resources), GPT-3 "suggests a potential mindless path to artificial general intelligence". But, of course, GPT-3 is not close to passing the Turing Test: GPT-3 neither perceives nor acts, and it is, at best, highly contentious whether it is a site of understanding. What remains to be seen is whether, within the next couple of generations of language generators – GPT-4 or GPT-5 – we have something that can be linked to perceptual inputs and behavioural outputs in a way that does produce something capable of passing the Turing Test. (For further discussion, see Floridi and Chiriatti (2020).)

On the other hand, as, for example, Floridi (2008) complains, there are other ways in which progress has been frustratingly slow. In 2014, claims emerged that, because the computer program *Eugene Goostman* had fooled 33% of judges in the Turing Test 2014 competition, it had "passed the Turing Test". But there have been other one-off competitions in which similar results have been achieved. Back in 1991, *PC Therapist* had 50% of judges fooled. And, in a 2011 demonstration, *Cleverbot* had an even higher success rate. In all three of these cases, the size of the trial was very small, and the result was not reliably projectible: in no case were there strong grounds for holding that an average interrogator had no more than a 70% chance of making the right determination about the relevant program after five minutes of questioning. Moreover—and much more importantly—we must distinguish between the test the Turing proposed, and the particular prediction that he made about how things would be by the end of the twentieth century. The percentage chance of making the correct identification, the time interval over which the test takes place, and the number of conversational exchanges required are all adjustable parameters in the Test, despite the fact that they are fixed in the particular prediction that Turing made. Even if Turing was very far out in the prediction that he made about how things would be by the end of the twentieth century, it remains possible that the test that he proposes is a good one. However, before one can endorse the suggestion that the Turing Test is good, there are various objections that ought to be addressed.

Some people have suggested that the Turing Test is chauvinistic: it only recognizes intelligence in things that are able to sustain a conversation with us. Why couldn't it be the case that there are intelligent things that are unable to carry on a conversation, or, at any rate, unable to carry on a conversation with creatures like us? (See, for example, French (1990).) Perhaps the intuition behind this question can be granted;

perhaps it is unduly chauvinistic to insist that anything that is intelligent has to be capable of sustaining a conversation with us. (On the other hand, one might think that, given the availability of suitably qualified translators, it ought to be possible for any two intelligent agents that speak different languages to carry on some kind of conversation.) But, in any case, the charge of chauvinism is completely beside the point. What Turing claims is only that, if something can carry out a conversation with us, then we have good grounds to suppose that that thing has intelligence of the kind that we possess; he does not claim that only something that can carry out a conversation with us can possess the kind of intelligence that we have.

Other people have thought that the Turing Test is not sufficiently demanding: we already have anecdotal evidence that quite unintelligent programs (e.g., ELIZA—for details of which, see Weizenbaum (1966)) can seem to ordinary observers to be loci of intelligence for quite extended periods of time. Moreover, over a short period of time—such as the five minutes that Turing mentions in his prediction about how things will be in the year 2000—it might well be the case that almost all human observers could be taken in by cunningly designed but quite unintelligent programs. However, it is important to recall that, in order to pass Turing's Test, it is not enough for the computer program to fool "ordinary observers" in circumstances other than those in which the test is supposed to take place. What the computer program has to be able to do is to survive interrogation by someone who knows that one of the other two participants in the conversation is a machine. Moreover, the computer program has to be able to survive such interrogation with a high degree of success over a repeated number of trials. (Turing says nothing about how many trials he would require. However, we can safely assume that, in order to get decent evidence that there is no more than a 70% chance that a machine will be correctly identified as a machine after five minutes of conversation, there will have to be a reasonably large number of trials.) If a computer program could do this quite demanding thing, then it does seem plausible to claim that we would have at least *prima facie* reason for thinking that we are in the presence of intelligence. (Perhaps it is worth emphasizing again that there might be all kinds of intelligent things—including intelligent machines—that would not pass this test. It is conceivable, for example, that there might be machines that, as a result of moral considerations, refused to lie or to engage in pretence. Since the human participant is supposed to do everything that he or she can to help the interrogator, the question "Are you a machine?" would quickly allow the interrogator to sort such (pathological?) truth-telling machines from humans.)

Another contentious aspect of Turing's paper (1950) concerns his restriction of the discussion to the case of "digital computers." On the one hand, it seems clear that this restriction is really only significant for the prediction that Turing makes about how things will be in the year 2000, and not for the details of the test itself. (Indeed, it seems that if the test that Turing proposes is a good one, then it will be a good test for any kinds of entities, including, for example, animals, aliens, and analog computers. That is: if animals, aliens, analog computers, or any other kinds of things, pass the test that Turing proposes, then there will be as much reason to think that these things exhibit intelligence as there is reason to think that digital computers that pass the test exhibit intelligence.) On the other hand, it is actually a highly controversial question whether "thinking machines" would have to be digital computers; and it is also a controversial question whether Turing himself assumed that this would be the case. In particular, it is worth noting that the seventh of the objections that Turing (1950) considers addresses the possibility of continuous state machines, which Turing explicitly acknowledges to be different from discrete state machines. Turing appears to claim that, even if we are continuous state machines, a discrete state machine would be able to imitate us sufficiently well for the purposes of the Imitation Game. However, it seems doubtful that the considerations that he gives are sufficient to establish that, if there are continuous state machines that pass the Turing Test, then it is possible to make discrete state machines that pass the test as well. (Turing himself was keen to point out that some limits had to be set on the notion of "machine" in order to make the question about "thinking machines" interesting:

> It is natural that we should wish to permit every kind of engineering technique to be used in our machine. We also wish to allow the possibility that an engineer or team of engineers may construct a machine which works, but whose manner of operation cannot be satisfactorily described by its constructors because they have applied a method which is largely

experimental. Finally, we wish to exclude from the machines men born in the usual manner. It is difficult to frame the definitions so as to satisfy these three conditions. One might for instance insist that the team of engineers should all be of one sex, but this would not really be satisfactory, for it is probably possible to rear a complete individual from a single cell of the skin (say) of a man. To do so would be a feat of biological technique deserving of the very highest praise, but we would not be inclined to regard it as a case of 'constructing a thinking machine'. (435/6)

But, of course, as Turing himself recognized, there is a large class of possible "machines" that are neither digital nor biotechnological.) More generally, the crucial point seems to be that, while Turing recognized that the class of machines is potentially much larger than the class of discrete state machines, he was himself *very* confident that properly engineered discrete state machines could succeed in the Imitation Game (and, moreover, at the time that he was writing, there were certain discrete state machines —"electronic computers"—that loomed very large in the public imagination).

2. Turing (1950) and Responses to Objections

Although Turing (1950) is pretty informal, and, in some ways rather idiosyncratic, there is much to be gained by considering the discussion that Turing gives of potential objections to his claim that machines—and, in particular, digital computers—can "think". Turing gives the following labels to the objections that he considers: (1) The Theological Objection; (2) The "Heads in the Sand" Objection; (3) The Mathematical Objection; (4) The Argument from Consciousness; (5) Arguments from Various Disabilities; (6) Lady Lovelace's Objection; (7) Argument from Continuity of the Nervous System; (8) The Argument from Informality of Behavior; and (9) The Argument from Extra-Sensory Perception. We shall consider these objections in the corresponding subsections below. (In some—but not all—cases, the counter-arguments to these objections that we discuss are also provided by Turing.)

2.6 Lady Lovelace's Objection

One of the most popular objections to the claim that there can be thinking machines is suggested by a remark made by Lady Lovelace in her memoir on Babbage's Analytical Engine:

> The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform (cited by Hartree, p. 70)

The key idea is that machines can *only* do what we know how to order them to do (or that machines can never do anything really new, or anything that would take us by surprise). As Turing says, one way to respond to these challenges is to ask whether we can ever do anything "really new." Suppose, for instance, that the world is deterministic, so that everything that we do is fully determined by the laws of nature and the boundary conditions of the universe. There is a sense in which nothing "really new" happens in a deterministic universe—though, of course, the universe's being deterministic would be entirely compatible with our being surprised by events that occur within it. Moreover—as Turing goes on to point out—there are many ways in which even digital computers do things that take us by surprise; more needs to be said to make clear exactly what the nature of this suggestion is. (Yes, we might suppose, digital computers are "constrained" by their programs: they can't do anything that is not permitted by the programs that they have. But human beings are "constrained" by their biology and their genetic inheritance in what might be argued to be just the same kind of way: they can't do anything that is not permitted by the biology and genetic inheritance that they have. If a program were sufficiently complex—and if the processor(s) on which it ran were sufficiently fast—then it is not easy to say whether the kinds of "constraints" that would remain would necessarily differ in kind from the kinds of constraints that are imposed by biology and genetic inheritance.)

Bringsjord et al. (2001) claim that Turing's response to the Lovelace Objection is "mysterious" at best, and "incompetent" at worst (p.4). In their view, Turing's claim that "computers do take us by surprise" is only true when "surprise" is given a very superficial interpretation. For, while it is true that computers do things that we don't intend them to do—because we're not smart enough, or because we're not careful enough, or because there are rare hardware errors, or whatever—it isn't true that there are any cases in which we should want to say that a computer has *originated* something. Whatever merit might be found in this objection, it seems worth pointing out that, in the relevant sense of *origination*, human beings "originate something" on more or less every occasion in which they engage in conversation: they produce new sentences of natural language that it is appropriate for them to produce in the circumstances in which they find themselves. Thus, on the one hand—for all that Bringsjord et al. have argued—The Turing Test is a perfectly good test for the presence of "origination" (or "creativity," or whatever). Moreover, on the other hand, for all that Bringsjord et al. have argued, it remains an open question whether a digital computing device is capable of "origination" in this sense (i.e. capable of producing new sentences that are appropriate to the circumstances in which the computer finds itself). So we are not overly inclined to think that Turing's response to the Lovelace Objection is poor; and we are even less inclined to think that Turing lacked the resources to provide a satisfactory response on this point.