Stanford Encyclopedia of Philosophy

The Turing Test

First published Wed Apr 9, 2003; substantive revision Wed Jan 26, 2011

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.

The phrase "The Turing Test" is sometimes used more generally to refer to some kinds of behavioural tests for the presence of mind, or thought, or intelligence in putatively minded entities. So, for example, it is sometimes suggested that The Turing Test is prefigured in Descartes' *Discourse on the Method*. (Copeland (2000:527) finds an anticipation of the test in the 1668 writings of the Cartesian de Cordemoy. Gunderson (1964) provides an early instance of those who find that Turing's work is foreshadowed in the work of Descartes.) In the *Discourse*, Descartes says:

> If there were machines which bore a resemblance to our bodies and imitated our actions as closely as possible for all practical purposes, we should still have two very certain means of recognizing that they were not real men. The first is that they could never use words, or put together signs, as we do in order to declare our thoughts to others. For we can certainly conceive of a machine so constructed that it utters words, and even utters words that correspond to bodily actions causing a change in its organs. … But it is not conceivable that such a machine should produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence, as the dullest of men can do. Secondly, even though some machines might do some things as well as we do them, or perhaps even better, they would inevitably fail in others, which would reveal that they are acting not from understanding, but only from the disposition of their organs. For whereas reason is a universal instrument, which can be used in all kinds of situations, these organs need some particular action; hence it is for all practical purposes impossible for a machine to have enough different organs to make it act in all the contingencies of life in the way in which our reason makes us act. (Translation by Robert Stoothoff)

Although not everything about this passage is perfectly clear, it does seem that Descartes gives a negative answer to the question whether machines can think; and, moreover, it seems that his giving this negative answer is tied to his confidence that no mere machine could pass The Turing Test: no mere machine could talk and act in the way in which adult human beings do. Since Descartes explicitly says that there are "two very certain means" by which we can rule out that something is a machine—it is, according to Descartes, inconceivable that a mere machine could produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence; and it is for all practical purposes impossible for a machine to have enough different organs

to make it act in all the contingencies of life in the way in which our reason makes us act—it seems that he must agree with the further claim that nothing that can produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence can be a machine. Given the further assumption—which one suspects that Descartes would have been prepared to grant—that only things that think can produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in their presence, it seems to follow that Descartes would have agreed that the Turing Test would be a good test of his confident assumption that there cannot be thinking machines. Given the knowledge that something is indeed a machine, evidence that that thing can produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence is evidence that there can be thinking machines.

The phrase "The Turing Test" is also sometimes used to refer to certain kinds of purely behavioural allegedly logically sufficient conditions for the presence of mind, or thought, or intelligence, in putatively minded entities. So, for example, Ned Block's "Blockhead" thought experiment is often said to be a (putative) knockdown objection to The Turing Test. (Block (1981) contains a direct discussion of The Turing Test in this context.) Here, what a proponent of this view has in mind is the idea that it is *logically possible* for an entity to pass the kinds of tests that Descartes and (at least allegedly) Turing have in mind—to use words (and, perhaps, to act) in just the kind of way that human beings do—and yet to be entirely lacking in intelligence, not possessed of a mind, etc.

The subsequent discussion takes up the preceding ideas in the order in which they have been introduced. First, there is a discussion of Turing's paper (1950), and of the arguments contained therein. Second, there is a discussion of current assessments of various proposals that have been called "The Turing Test" (whether or not there is much merit in the application of this label to the proposals in question). Third, there is a brief discussion of some recent writings on The Turing Test, including some discussion of the question whether The Turing Test sets an appropriate goal for research into artificial intelligence. Finally, there is a very short discussion of Searle's Chinese Room argument, and, in particular, of the bearing of this argument on The Turing Test.

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 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. On the one hand, participants in the Loebner Prize Competition—an annual event in which computer programmes are submitted to the Turing Test—come nowhere near the standard that Turing envisaged. (A quick look at the transcripts of the participants for the past decade reveals that the entered programs are all easily detected by a range of not-very-subtle lines of questioning.) On the other hand, major players in the field often claim that the Loebner Prize Competition is an embarrassment precisely because we are 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). (The programs entered in the Loebner Prize Competition are 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.)

Even if Turing was very far out in assessment of how soon it will be before we have computer programs that can pass the Turing Test, 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 machinese—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.)

2.1 The Theological Objection

Substance dualists believe that thinking is a function of a non-material, separately existing, substance that somehow "combines" with the body to make a person. So—the argument might go—making a body can never be sufficient to guarantee the presence of thought: in themselves, digital computers are no different from any other merely material bodies in being utterly unable to think. Moreover—to introduce the "theological" element—it might be further added that, where a "soul" is suitably combined with a body, this is always the work of the divine creator of the universe: it is entirely up to God whether or not a particular kind of body is imbued with a thinking soul. (There is well known scriptural support for the proposition that human beings are "made in God's image". Perhaps there is also theological support for the claim that only God can make things in God's image.)

There are several different kinds of remarks to make here. First, there are many serious objections to substance dualism. Second, there are many serious objections to theism. Third, even if theism and substance dualism are both allowed to pass, it remains quite unclear why thinking machines are supposed to be ruled out by this combination of views. Given that God can unite souls with human

bodies, it is hard to see what reason there is for thinking that God could not unite souls with digital computers (or rocks, for that matter!). Perhaps, on this combination of views, there is no especially good reason why, amongst the things that we can make, certain kinds of digital computers turn out to be the only ones to which God gives souls—but it seems pretty clear that there is also no particularly good reason for ruling out the possibility that God would choose to give souls to certain kinds of digital computers. Evidence that God is dead set against the idea of giving souls to certain kinds of digital computers is not particularly thick on the ground.

2.2 The 'Heads in the Sand' Objection

If there were thinking machines, then various consequences would follow. First, we would lose the best reasons that we have for thinking that we are superior to everything else in the universe (since our cherished "reason" would no longer be something that we alone possess). Second, the possibility that we might be "supplanted" by machines would become a genuine worry: if there were thinking machines, then very likely there would be machines that could think much better than we can. Third, the possibility that we might be "dominated" by machines would also become a genuine worry: if there were thinking machines, who's to say that they would not take over the universe, and either enslave or exterminate us?

As it stands, what we have here is not an argument against the claim that machines can think; rather, we have the expression of various fears about what might follow if there were thinking machines. Someone who took these worries seriously—and who was persuaded that it is indeed possible for us to construct thinking machines—might well think that we have here reasons for giving up on the project of attempting to construct thinking machines. However, it would be a major task—which we do not intend to pursue here—to determine whether there really are any good reasons for taking these worries seriously.

2.3 The Mathematical Objection

Some people have supposed that certain fundamental results in mathematical logic that were discovered during the 1930s—by Gödel (first incompleteness theorem) and Turing (the halting problem)—have important consequences for questions about digital computation and intelligent thought. (See, for example, Lucas (1961) and Penrose (1989); see, too, Hodges (1983:414) who mentions Polanyi's discussions with Turing on this matter.) Essentially, these results show that within a formal system that is strong enough, there are a class of true statements that can be expressed but not proven within the system (see the entry on provability logic). Let us say that such a system is "subject to the Lucas-Penrose constraint" because it is constrained from being able to prove a class of true statements expressible within the system.

Turing (1950:444) himself observes that these results from mathematical logic might have implications for the Turing test:

> There are certain things that [any digital computer] cannot do. If it is rigged up to give answers to questions as in the imitation game, there will be some questions to which it will either give a wrong answer, or fail to give an answer at all however much time is allowed for a reply. (444)

So, in the context of the Turing test, "being subject to the Lucas-Penrose constraint" implies the existence of a class of "unanswerable" questions. However Turing noted that in the context of the Turing test, these "unanswerable" questions are only a concern if humans can answer them. His "short" reply was that it is not clear that humans are free from such a constraint themselves. Turing then goes on to add that he does not think that the argument can be dismissed "quite so lightly."

To make the argument more precise, we can write it as follows:

1. Let C be a digital computer.
2. Since C is subject to the Lucas-Penrose constraint, there is an "unanswerable" question q for C.
3. If an entity, E, is not subject to the Lucas-Penrose constraint, then there are no "unanswerable" questions for E.
4. The human intellect is not subject to the Lucas-Penrose constraint.
5. Thus, there are no "unanswerable" questions for the human intellect.
6. The question q is therefore "answerable" to the human intellect.
7. By asking question q, a human could determine if the responder is a computer or a human.
8. Thus C may fail the Turing test.

Once the argument is laid out as above, it becomes clear that premise (3) should be challenged. Putting that aside, we note that one interpretation of Turing's "short" reply is that claim (4) is merely asserted—without any kind of proof. The "short" reply then leads us to examine whether humans are free from the Lucas-Penrose constraint.

If humans are subject to the Lucas-Penrose constraint then the constraint does not provide any basis for distinguishing humans from digital computers. If humans are free from the Lucas-Penrose constraint, then (granting premise 3) it follows that digital computers may fail the Turing test and thus, it seems, cannot think.

However, there remains a question as to whether being free from the constraint is necessary for the capacity to think. It may be that the Turing test is too strict. Since, by hypothesis, we are free from the Lucas-Penrose constraint, we are, in some sense, too good at asking and answering questions. Suppose there is a thinking entity that is subject to the Lucas-Penrose constraint. By an argument analogous to the one above, it can fail the Turing test. Thus, an entity which can think would fail the Turing test.

We can respond to this concern by noting that the construction of questions suggested by the results from mathematical logic—Gödel, Turing, etc.—are extremely complicated, and require extremely detailed information about the language and internal programming of the digital computer (which, of course, is not available to the interrogators in the Imitation Game). At the very least, much more argument is required to overthrow the view that the Turing Test could remain a very high quality statistical test for the presence of mind and intelligence even if digital computers differ from human beings in being subject to the Lucas-Penrose constraint. (See Bowie 1982, Dietrich 1994, Feferman 1996, and Abramson 2008, for further discussion.)

2.4 The Argument from Consciousness

Turing cites Professor Jefferson's *Lister Oration* for 1949 as a source for the kind of objection that he takes to fall under this label:

> Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain—that is, not only write it but know that it had written it. No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants. (445/6)

There are several different ideas that are being run together here, and that it is profitable to disentangle. One idea—the one upon which Turing first focuses—is the idea that the only way in which one could be certain that a machine thinks is to be the machine, and to feel oneself thinking. A second idea, perhaps, is that the presence of mind requires the presence of a certain kind of self-consciousness ("not only write it but know that it had written it"). A third idea is that it is a mistake to take a narrow view of the mind, i.e. to suppose that there could be a believing intellect divorced from

the kinds of desires and emotions that play such a central role in the generation of human behavior ("no mechanism could feel …").

Against the solipsistic line of thought, Turing makes the effective reply that he would be satisfied if he could secure agreement on the claim that we might each have just as much reason to suppose that machines think as we have reason to suppose that *other* people think. (The point isn't that Turing thinks that solipsism is a serious option; rather, the point is that following this line of argument isn't going to lead to the conclusion that there are respects in which digital computers could not be our intellectual equals or superiors.)

Against the other lines of thought, Turing provides a little "*viva voce*" that is intended to illustrate the kind of evidence that he supposes one might have that a machine is intelligent. Given the right kinds of responses from the machine, we *would* naturally interpret its utterances as evidence of pleasure, grief, warmth, misery, anger, depression, etc. Perhaps—though Turing doesn't say this—the only way to make a machine of this kind would be to equip it with sensors, affective states, etc., i.e., in effect, to make an artificial *person*. However, the important point is that if the claims about self-consciousness, desires, emotions, etc. are right, then Turing can accept these claims with equanimity: *his* claim is then that a machine with a digital computing "brain" can have the full range of mental states that can be enjoyed by adult human beings.

2.5 Arguments from Various Disabilities

Turing considers a list of things that some people have claimed machines will never be able to do: (1) be kind; (2) be resourceful; (3) be beautiful; (4) be friendly; (5) have initiative; (6) have a sense of humor; (7) tell right from wrong; (8) make mistakes; (9) fall in love; (10) enjoy strawberries and cream; (11) make someone fall in love with one; (12) learn from experience; (13) use words properly; (14) be the subject of one's own thoughts; (15) have as much diversity of behavior as a man; (16) do something really new.

An interesting question to ask, before we address these claims directly, is whether we should suppose that intelligent creatures from some other part of the universe would necessarily be able to do these things. Why, for example, should we suppose that there must be something deficient about a creature that does not enjoy—or that is not able to enjoy—strawberries and cream? True enough, we might suppose that an intelligent creature ought to have the capacity to enjoy some kinds of things—but it seems unduly chauvinistic to insist that intelligent creatures must be able to enjoy just the kinds of things that we do. (No doubt, similar considerations apply to the claim that an intelligent creature must be the kind of thing that can make a human being fall in love with it. Yes, perhaps, an intelligent creature should be the kind of thing that can love and be loved; but what is so special about us?)

Setting aside those tasks that we deem to be unduly chauvinistic, we should then ask what grounds there are for supposing that no digital computing machine *could* do the other things on the list. Turing suggests that the most likely ground lies in our prior acquaintance with machines of all kinds: none of the machines that any of us has hitherto encountered has been able to do these things. In particular, the digital computers with which we are now familiar cannot do these things. (Except perhaps for make mistakes: after all, even digital computers are subject to "errors of functioning." But this might be set aside as an irrelevant case.) However, given the limitations of storage capacity and processing speed of even the most recent digital computers, there are obvious reasons for being cautious in assessing the merits of this inductive argument.

(A different question worth asking concerns the progress that has been made until now in constructing machines that can do the kinds of things that appear on Turing's list. There is at least room for debate about the extent to which current computers can: make mistakes, use words properly, learn from experience, be beautiful, etc. Moreover, there is also room for debate about the extent to which recent advances in other areas may be expected to lead to further advancements in overcoming these alleged

disabilities. Perhaps, for example, recent advances in work on artificial sensors may one day contribute to the production of machines that can enjoy strawberries and cream. Of course, if the intended objection is to the notion that machines can experience any kind of feeling of enjoyment, then it is not clear that work on particular kinds of artificial sensors is to the point.)

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.

2.7 Argument from Continuity of the Nervous System

The human brain and nervous system is not much like a digital computer. In particular, there are reasons for being skeptical of the claim that the brain is a discrete-state machine. Turing observes that a small error in the information about the size of a nervous impulse impinging on a neuron may make a large difference to the size of the outgoing impulse. From this, Turing infers that the brain is likely

to be a continuous-state machine; and he then notes that, since discrete-state machines are not continuous-state machines, there might be reason here for thinking that no discrete-state machine can be intelligent.

Turing's response to this kind of argument seems to be that a continuous-state machine can be imitated by discrete-state machines with very small levels of error. Just as differential analyzers can be imitated by digital computers to within quite small margins of error, so too, the conversation of human beings can be imitated by digital computers to margins of error that would not be detected by ordinary interrogators playing the imitation game. It is not clear that this is the right kind of response for Turing to make. If someone thinks that real thought (or intelligence, or mind, or whatever) can only be located in a continuous-state machine, then the fact—if, indeed, it is a fact—that it is possible for discrete-state machines to pass the Turing Test shows only that the Turing Test is no good. A better reply is to ask why one should be so confident that real thought, etc. can only be located in continuous-state machines (if, indeed, it is right to suppose that we are not discrete-state machines). And, before we ask this question, we would do well to consider whether we really do have such good reason to suppose that, from the standpoint of our ability to think, we are not essentially discrete-state machines. (As Block (1981) points out, it seems that there is nothing in our concept of intelligence that rules out intelligent beings with quantised sensory devices; and nor is there anything in our concept of intelligence that rules out intelligent beings with digital working parts.)

2.8 Argument from Informality of Behavior

This argument relies on the assumption that there is no set of rules that describes what a person ought to do in every possible set of circumstances, and on the further assumption that there is a set of rules that describes what a machine will do in every possible set of circumstances. From these two assumptions, it is supposed to follow—somehow!—that people are not machines. As Turing notes, there is some slippage between "ought" and "will" in this formulation of the argument. However, once we make the appropriate adjustments, it is not clear that an obvious difference between people and digital computers emerges.

Suppose, first, that we focus on the question of whether there are sets of rules that describe what a person and a machine "will" do in every possible set of circumstances. If the world is deterministic, then there are such rules for both persons and machines (though perhaps it is not possible to write down the rules). If the world is not deterministic, then there are no such rules for either persons or machines (since both persons and machines can be subject to non-deterministic processes in the production of their behavior). Either way, it is hard to see any reason for supposing that there is a relevant difference between people and machines that bears on the description of what they will do in all possible sets of circumstances. (Perhaps it might be said that what the objection invites us to suppose is that, even though the world is not deterministic, humans differ from digital machines precisely because the operations of the latter are indeed deterministic. But, if the world is non-deterministic, then there is no reason why digital machines cannot be programmed to behave non-deterministically, by allowing them to access input from non-deterministic features of the world.)

Suppose, instead, that we focus on the question of whether there are sets of rules that describe what a person and a machine "ought" to do in every possible set of circumstances. Whether or not we suppose that norms can be codified—and quite apart from the question of which kinds of norms are in question—it is hard to see what grounds there could be for this judgment, other than the question-begging claim that machines are not the kinds of things whose behavior could be subject to norms. (And, in that case, the initial argument is badly mis-stated: the claim ought to be that, whereas there are sets of rules that describe what a person ought to do in every possible set of circumstances, there are no sets of rules that describe what machines *ought* to do in all possible sets of circumstances!)

2.9 Argument from Extra-Sensory Perception

The strangest part of Turing's paper is the few paragraphs on ESP. Perhaps it is intended to be tongue-in-cheek, though, if it is, this fact is poorly signposted by Turing. Perhaps, instead, Turing was influenced by the apparently scientifically respectable results of J. B. Rhine. At any rate, taking the text at face value, Turing seems to have thought that there was overwhelming empirical evidence for telepathy (and he was also prepared to take clairvoyance, precognition and psychokinesis seriously). Moreover, he also seems to have thought that if the human participant in the game was telepathic, then the interrogator could exploit this fact in order to determine the identity of the machine—and, in order to circumvent this difficulty, Turing proposes that the competitors should be housed in a "telepathy-proof room." Leaving aside the point that, as a matter of fact, there is no current statistical support for telepathy—or clairvoyance, or precognition, or telekinesis—it is worth asking what kind of theory of the nature of telepathy would have appealed to Turing. After all, if humans can be telepathic, why shouldn't digital computers be so as well? If the capacity for telepathy were a standard feature of any sufficiently advanced system that is able to carry out human conversation, then there is no in-principle reason why digital computers could not be the equals of human beings in this respect as well. (Perhaps this response assumes that a successful machine participant in the imitation game will need to be equipped with sensors, etc. However, as we noted above, this assumption is not terribly controversial. A plausible conversationalist has to keep up to date with goings-on in the world.)

After discussing the nine objections mentioned above, Turing goes on to say that he has "no very convincing arguments of a positive nature to support my views. If I had I should not have taken such pains to point out the fallacies in contrary views." (454) Perhaps Turing sells himself a little short in this self-assessment. First of all—as his brief discussion of solipsism makes clear—it is worth asking what grounds we have for attributing intelligence (thought, mind) to other people. If it is plausible to suppose that we base our attributions on behavioral tests or behavioral criteria, then his claim about the appropriate test to apply in the case of machines seems apt, and his conjecture that digital computing machines might pass the test seems like a reasonable—though controversial—empirical conjecture. Second, subsequent developments in the philosophy of mind—and, in particular, the fashioning of functionalist theories of the mind—have provided a more secure theoretical environment in which to place speculations about the possibility of thinking machines. If mental states are functional states—and if mental states are capable of realisation in vastly different kinds of materials—then there is some reason to think that it is an empirical question whether minds can be realised in digital computing machines. Of course, this kind of suggestion is open to challenge; we shall consider some important philosophical objections in the later parts of this review.

3. Some Minor Issues Arising

There are a number of much-debated issues that arise in connection with the interpretation of various parts of Turing (1950), and that we have hitherto neglected to discuss. What has been said in the first two sections of this document amounts to our interpretation of what Turing has to say (perhaps bolstered with what we take to be further relevant considerations in those cases where Turing's remarks can be fairly readily improved upon). But since some of this interpretation has been contested, it is probably worth noting where the major points of controversy have been.

3.1 Interpreting the Imitation Game

Turing (1950) introduces the imitation game by describing a game in which the participants are a man, a woman, and a human interrogator. The interrogator is in a room apart from the other two, and is set the task of determining which of the other two is a man and which is a woman. Both the man and the woman are set the task of trying to convince the interrogator that they are the woman. Turing recommends that the best strategy for the woman is to answer all questions truthfully; of course, the best strategy for the man will require some lying. The participants in this game also use teletypewriter to communicate with one another—to avoid clues that might be offered by tone of voice, etc. Turing

then says: "We now ask the question, 'What will happen when a machine takes the part of A in this 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?" (434).

Now, of course, it is *possible* to interpret Turing as here intending to say what he seems literally to say, namely, that the new game is one in which the computer must pretend to be a woman, and the other participant in the game is a woman. (See, for example, Genova (1994), and Traiger (2000).) And it is also *possible* to interpret Turing as intending to say that the new game is one in which the computer must pretend to be a woman, and the other participant in the game is a man who must also pretend to be a woman. However, as Copeland (2000), Piccinini (2000), and Moor (2001) convincingly argue, the rest of Turing's article, and material in other articles that Turing wrote at around the same time, very strongly support the claim that Turing actually intended the standard interpretation that we gave above, viz. that the computer is to pretend to be a human being, and the other participant in the game is a human being of unspecified gender. Moreover, as Moor (2001) argues, there is no reason to think that one would get a better test if the computer must pretend to be a woman and the other participant in the game is a man pretending to be a woman (and, indeed, there is some reason to think that one would get a worse test). Perhaps it would make no difference to the effectiveness of the test if the computer must pretend to be a woman, and the other participant is a woman (any more than it would make a difference if the computer must pretend to be an accountant and the other participant is an accountant); however, this consideration is simply insufficient to outweigh the strong textual evidence that supports the standard interpretation of the imitation game that we gave at the beginning of our discussion of Turing (1950).

3.2 Turing's Predictions

As we noted earlier, Turing (1950) makes the claim that:

> 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.

Most commentators contend that this claim has been shown to be mistaken: in the year 2000, *no-one* was able to program computers to make them play the imitation game so well that an average interrogator had no more than a 70% chance of making the correct identification after five minutes of questioning. Copeland (2000) argues that this contention is seriously mistaken: "about fifty years" is by no means "exactly fifty years," and it remains open that we may soon be able to do the required programming. Against this, it should be noted that Turing (1950) goes on immediately to refer to how things will be "at the end of the century," which suggests that not too much can be read into the qualifying "about." However, as Copeland (2000) points out, there are other more cautious predictions that Turing makes elsewhere (e.g., that it would be "at least 100 years" before a machine was able to pass an unrestricted version of his test); and there are other predictions that are made in Turing (1950) that seem to have been vindicated. In particular, it is plausible to claim that, in the year 2000, educated opinion had altered to the extent that, in many quarters, one could speak of the possibility of machines' thinking—and of machines' learning—without expecting to be contradicted. As Moor (2001) points out, "machine intelligence" is not the oxymoron that it might have been taken to be when Turing first started thinking about these matters.

3.3 A Useful Distinction

There are two different theoretical claims that are run together in many discussions of The Turing Test

that can profitably be separated. One claim holds that the general scheme that is described in Turing's Imitation Game provides a good test for the presence of intelligence. (If something can pass itself off as a person under sufficiently demanding test conditions, then we have very good reason to suppose that that thing is intelligent.) Another claim holds that an appropriately programmed computer could pass the kind of test that is described in the first claim. We might call the first claim "The Turing Test Claim" and the second claim "The Thinking Machine Claim". Some objections to the claims made in Turing (1950) are objections to the Thinking Machine Claim, but not objections to the Turing Test Claim. (Consider, for example, the argument of Searle (1982), which we discuss further in Section 6.) However, other objections are objections to the Turing Test Claim. Until we get to Section 6, we shall be confining our attention to discussions of the Turing Test Claim.

3.4 A Further Note

In this article, we follow the standard philosophical convention according to which "a mind" means "at least one mind". If "passing the Turing Test" implies intelligence, then "passing the Turing Test" implies the presence of at least one mind. We cannot here explore recent discussions of "swarm intelligence", "collective intelligence", and the like. However, it is surely clear that two people taking turns could "pass the Turing Test" in circumstances in which we should be very reluctant to say that there is a "collective mind" that has the minds of the two as components.

4. Assessment of the Current Standing of The Turing Test

Given the initial distinction that we made between different ways in which the expression The Turing Test gets interpreted in the literature, it is probably best to approach the question of the assessment of the current standing of The Turing Test by dividing cases. True enough, we think that there is a correct interpretation of exactly what test it is that is proposed by Turing (1950); but a complete discussion of the current standing of The Turing Test should pay at least some attention to the current standing of other tests that have been mistakenly supposed to be proposed by Turing (1950).

There are a number of main ideas to be investigated. First, there is the suggestion that The Turing Test provides logically necessary and sufficient conditions for the attribution of intelligence. Second, there is the suggestion that The Turing Test provides logically sufficient—but not logically necessary— conditions for the attribution of intelligence. Third, there is the suggestion that The Turing Test provides "criteria"—defeasible sufficient conditions—for the attribution of intelligence. Fourth—and perhaps not importantly distinct from the previous claim—there is the suggestion that The Turing Test provides (more or less strong) probabilistic support for the attribution of intelligence. We shall consider each of these suggestions in turn.

4.1 (Logically) Necessary and Sufficient Conditions

It is doubtful whether there are very many examples of people who have explicitly claimed that The Turing Test is meant to provide conditions that are both logically necessary and logically sufficient for the attribution of intelligence. (Perhaps Block (1981) is one such case.) However, some of the objections that have been proposed against The Turing Test only make sense under the assumption that The Turing Test does indeed provide logically necessary and logically sufficient conditions for the attribution of intelligence; and many more of the objections that have been proposed against The Turing Test only make sense under the assumption that The Turing Test provides necessary and sufficient conditions for the attribution of intelligence, where the modality in question is weaker than the strictly logical, e.g., nomic or causal.

Consider, for example, those people who have claimed that The Turing Test is chauvinistic; and, in particular, those people who have claimed that it is surely logically possible for there to be something that possesses considerable intelligence, and yet that is not able to pass The Turing Test. (Examples:

Intelligent creatures might fail to pass The Turing Test because they do not share our way of life; intelligent creatures might fail to pass The Turing Test because they refuse to engage in games of pretence; intelligent creatures might fail to pass The Turing Test because the pragmatic conventions that govern the languages that they speak are so very different from the pragmatic conventions that govern human languages. Etc.) None of this can constitute objections to The Turing Test unless The Turing Test delivers *necessary* conditions for the attribution of intelligence.

French (1990) offers ingenious arguments that are intended to show that "the Turing Test provides a guarantee not of intelligence, but of culturally-oriented intelligence." But, of course, anything that has culturally-oriented intelligence *has* intelligence; so French's objections cannot be taken to be directed towards the idea that The Turing Test provides sufficient conditions for the attribution of intelligence. Rather—as we shall see later—French supposes that The Turing Test establishes sufficient conditions that no machine will ever satisfy. That is, in French's view, what is wrong with The Turing Test is that it establishes utterly uninteresting sufficient conditions for the attribution of intelligence.

4.2 Logically Sufficient Conditions

There are many philosophers who have supposed that The Turing Test is intended to provide logically sufficient conditions for the attribution of intelligence. That is, there are many philosophers who have supposed that The Turing Test claims that it is logically impossible for something that lacks intelligence to pass The Turing Test. (Often, this supposition goes with an interpretation according to which passing The Turing Test requires rather a lot, e.g., producing behavior that is indistinguishable from human behavior over an entire lifetime.)

There are well-known arguments against the claim that passing The Turing Test—or any other purely behavioral test—provides logically sufficient conditions for the attribution of intelligence. *The* standard objection to this kind of analysis of intelligence (mind, thought) is that a being whose behavior was produced by "brute force" methods ought not to count as intelligent (as possessing a mind, as having thoughts).

Consider, for example, Ned Block's *Blockhead*. Blockhead is a creature that looks just like a human being, but that is controlled by a "game-of-life look-up tree," i.e. by a tree that contains a programmed response for every discriminable input at each stage in the creature's life. If we agree that Blockhead is logically possible, and if we agree that Blockhead is not intelligent (does not have a mind, does not think), then Blockhead is a counterexample to the claim that the Turing Test provides a logically sufficient condition for the ascription of intelligence. After all, Blockhead could be programmed with a look-up tree that produces responses identical with the ones that *you* would give over the entire course of *your* life (given the same inputs).

There are perhaps only two ways in which someone who claims that The Turing Test offers logically sufficient conditions for the attribution of intelligence can respond to Block's argument. First, it could be denied that Blockhead is a logical possibility; second, it could be claimed that Blockhead would be intelligent (have a mind, think).

In order to deny that Blockhead is a logical possibility, it seems that what needs to be denied is the commonly accepted link between conceivability and logical possibility: it certainly seems that Blockhead is *conceivable*, and so, if (properly circumscribed) conceivability is sufficient for logical possibility, then it seems that we have good reason to accept that Blockhead is a logical possibility. Since it would take us too far away from our present concerns to explore this issue properly, we merely note that it remains a controversial question whether (properly circumscribed) conceivability is sufficient for logical possibility. (For further discussion of this issue, see Crooke (2002).)

The question of whether Blockhead is intelligent (has a mind, thinks) may seem straightforward, but —despite Block's confident assertion that Blockhead "has all of the intelligence of a toaster"—it is not

completely obvious that we should deny that Blockhead is intelligent. True enough, Blockhead is a particularly inefficient processor of information; but it is at least a processor of information, and that —in combination with the behavior that is produced as a result of the processing of information— might well be taken to be sufficient grounds for the attribution of *some* level of intelligence to Blockhead.

4.3 Criteria

In his *Philosophical Investigations*, Wittgenstein famously writes: "An 'inner process' stands in need of outward criteria" (580). Exactly what Wittgenstein meant by this remark is unclear, but one way in which it might be interpreted is as follows: in order to be justified in ascribing a "mental state" to some entity, there must be some true claims about the observable behavior of that entity that, (perhaps) together with other true claims about that entity (not themselves couched in "mentalistic" vocabulary), entail that the entity has the mental state in question. If no true claims about the observable behavior of the entity can play any role in the justification of the ascription of the mental state in question to the entity, then there are no grounds for attributing that kind of mental state to the entity.

The claim that, in order to be justified in ascribing a mental state to an entity, there must be some true claims about the observable behavior of that entity that alone—i.e. without the addition of any other true claims about that entity—entail that the entity has the mental state in question, is a piece of philosophical behaviorism. It may be—for all that we are able to argue—that Wittgenstein was a philosophical behaviorist; it may be—for all that we are able to argue—that Turing was one, too. However, if we go by the letter of the account given in the previous paragraph, then all that need follow from the claim that the Turing Test is criterial for the ascription of intelligence (thought, mind) is that, when other true claims (not themselves couched in terms of mentalistic vocabulary) are conjoined with the claim that an entity has passed the Turing Test, it then follows that the entity in question has intelligence (thought, mind).

(Note that the parenthetical qualification that the additional true claims not be couched in terms of mentalistic vocabulary is only one way in which one might try to avoid the threat of trivialization. The difficulty is that the addition of the true claim that an entity has a mind will always produce a set of claims that entails that that entity has a mind, no matter what other claims belong to the set!)

To see how the claim that the Turing Test is merely criterial for the ascription of intelligence differs from the logical behaviorist claim that the Turing Test provides logically sufficient conditions for the ascription of intelligence, it suffices to consider the question of whether it is *nomically* possible for there to be a "hand simulation" of a Turing Test program. Many people have supposed that there is good reason to deny that Blockhead is a nomic (or physical) possibility. For example, in *The Physics of Immortality*, Frank Tipler provides the following argument in defence of the claim that it is physically impossible to "hand simulate" a Turing-Test-passing program:

> If my earlier estimate that the human brain can code as much as 10^{15} bits is correct, then since an average book codes about 10^{6} bits … it would require more than 100 million books to code the human brain. It would take at least thirty five-story main university libraries to hold this many books. We know from experience that we can access any memory in our brain in about 100 seconds, so a hand simulation of a Turing Test-passing program would require a human being to be able to take off the shelf, glance through, and return to the shelf all of these 100 million books in 100 seconds. If each book weighs about a pound (0.5 kilograms), and on the average the book moves one yard (one meter) in the process of taking it off the shelf and returning it, then in 100 seconds the energy consumed in just moving the books is 3 x 10^{19} joules; the rate of energy consumption is 3 x 10^{11} megawatts. Since a human uses energy at a normal rate of 100 watts, the power

required is the bodily power of 3×10^{15} human beings, about a million times the current population of the entire earth. A typical large nuclear power plant has a power output of 1,000 megawatts, so a hand simulation of the human program requires a power output equal to that of 300 million large nuclear power plants. As I said, a man can no more hand-simulate a Turing Test-passing program than he can jump to the Moon. In fact, it is far more difficult. (40)

While there might be ways in which the details of Tipler's argument could be improved, the general point seems clearly right: the kind of combinatorial explosion that is required for a look-up tree for a human being is ruled out by the laws and boundary conditions that govern the operations of the physical world. But, if this is right, then, while it may be true that Blockhead is a *logical* possibility, it follows that Blockhead is not a *nomic* or *physical* possibility. And then it seems natural to hold that The Turing Test does indeed provide *nomically* sufficient conditions for the attribution of intelligence: given everything else that we already know—or, at any rate, take ourselves to know—about the universe in which we live, we would be fully justified in concluding that anything that succeeds in passing The Turing Test is, indeed, intelligent (possessed of a mind, and so forth).

There are ways in which the argument in the previous paragraph might be resisted. At the very least, it is worth noting that there is a serious gap in the argument that we have just rehearsed. Even if we can rule out "hand simulation" of intelligence, it does not follow that we have ruled out all other kinds of mere simulation of intelligence. Perhaps—for all that has been argued so far—there are nomically possible ways of producing mere simulations of intelligence. But, if that's right, then passing The Turing Test need not be so much as criterial for the possession of intelligence: it need not be that given everything else that we already know—or, at any rate, take ourselves to know—about the universe in which we live, we would be fully justified in concluding that anything that succeeds in passing The Turing Test is, indeed, intelligent (possessed of a mind, and so forth).

(Perhaps it is worth noting that we cannot see how Tipler arrived at his figure of 3×10^{19} joules in the calculation that he provides. Making what seem to us to be plausible estimates, we get a figure of around 3×10^{14} joules. Even on this revised figure, the argument that Tipler is running still goes through—and it may very well be that the larger figure that he quotes can be justified. It is a shame that the further details for his calculation are not provided in his text.)

4.4 Probabilistic Support

When we look at the initial formulation that Turing provides of his test, it is clear that he thought that the passing of the test would provide probabilistic support for the hypothesis of intelligence. There are at least two different points to make here. First, the *prediction* that Turing makes is itself probabilistic: Turing predicts that, in about fifty years from the time of his writing, it will be possible to programme digital computers to make them play the imitation game so well that an average interrogator will have no more than a seventy per cent chance of making the right identification after five minutes of questioning. Second, the probabilistic nature of Turing's prediction provides good reason to think that the *test* that Turing proposes is itself of a probabilistic nature: a given level of success in the imitation game produces—or, at any rate, should produce—a specifiable level of increase in confidence that the participant in question is intelligent (has thoughts, is possessed of a mind). Since Turing doesn't tell us how he supposes that levels of success in the imitation game correlate with increases in confidence that the participant in question is intelligent, there is a sense in which The Turing Test is greatly underspecified. Relevant variables clearly include: the length of the period of time over which the questioning in the game takes place (or, at any rate, the "amount" of questioning that takes place); the skills and expertise of the interrogator (this bears, for example, on the "depth" and "difficulty" of the questioning that takes place); the skills and expertise of the third player in the game; and the number of independent sessions of the game that are run (particularly when the other participants in the game differ from one run to the next). Clearly, a machine that is very successful in many different runs of

the game that last for quite extended periods of time and that involve highly skilled participants in the other roles has a much stronger claim to intelligence than a machine that has been successful in a single, short run of the game with highly inexpert participants. That a machine has succeeded in one short run of the game against inexpert opponents might provide some reason for increase in confidence that the machine in question is intelligent: but it is clear that results on subsequent runs of the game could quickly overturn this initial increase in confidence. That a machine has done much better than chance over many long runs of the imitation game against a variety of skilled participants surely provides much stronger evidence that the machine is intelligent. (Given enough evidence of this kind, it seems that one could be quite confident indeed that the machine is intelligent, while still —of course—recognizing that one's judgment could be overturned by further evidence, such as a series of short runs in which it does much worse than chance against participants who use the same strategy over and over to expose the machine as a machine.)

The probabilistic nature of The Turing Test is often overlooked. True enough, Moor (1976, 2001)— along with various other commentators—has noted that The Turing Test is "inductive," i.e. that "The Turing Test" provides no more than defeasible evidence of intelligence. However, it is one thing to say that success in "a rigorous Turing test" provides no more than defeasible evidence of intelligence; it is quite another to note the probabilistic features to which we have drawn attention in the preceding paragraph. Consider, for example, Moor's observation (Moor 2001:83) that "… inductive evidence gathered in a Turing test can be outweighed by new evidence. … If new evidence shows that a machine passed the Turing Test by remote control run by a human behind the scenes, then reassessment is called for." This—and other similar passages—seems to us to suggest that Moor supposes that a "rigorous Turing test" is a one-off event in which the machine either succeeds or fails. But this interpretation of The Turing Test is vulnerable to the kind of objection lodged by Bringsjord (1994): even on a moderately long single run with relatively expert participants, it may not be all that unlikely that an unintelligent machine serendipitously succeeds in the imitation game. In our view, given enough sufficiently long runs with different sufficiently expert participants, the likelihood of serendipitous success can be made as small as one wishes. Thus, while Bringsjord's "argument from serendipity" has force against some versions of The Turing Test, it has no force against the most plausible interpretation of the test that Turing actually proposed.

It is worth noting that it is quite easy to construct more sophisticated versions of "The Imitation Game" that yield more fine-grained statistical data. For example, rather than getting the judges to issue Yes/No verdicts about both of the participants in the game, one could get the judges to provide probabilistic answers. ("I give a 75% probability to the claim that A is the machine, and only 25% probability to the claim that B is the machine.") This point is important when one comes to consider criticisms of the "methodology" implicit in "The Turing Test". (For further discussion of the probabilistic nature of "The Turing Test", see Shieber (2007).)

5. Alternative Tests

Some of the literature about The Turing Test is concerned with questions about the framing of a test that can provide a suitable guide to future research in the area of Artificial Intelligence. The idea here is very simple. Suppose that we have the ambition to produce an artificially intelligent entity. What tests should we take as setting the goals that putatively intelligent artificial systems should achieve? Should we suppose that The Turing Test provides an appropriate goal for research in this field? In assessing these proposals, there are two different questions that need to be borne in mind. First, there is the question whether it is a useful goal for AI research to aim to make a machine that can pass the given test (administered over the specified length of time, at the specified degree of success). Second, there is the question of the appropriate conclusion to draw about the mental capacities of a machine that does manage to pass the test (administered over the specified length of time, at the specified degree of success).

Opinion on these questions is deeply divided. Some people suppose that The Turing Test does not

provide a useful goal for research in AI because it is far too difficult to produce a system that can pass the test. Other people suppose that The Turing Test does not provide a useful goal for research in AI because it sets a very narrow target (and thus sets unnecessary restrictions on the kind of research that gets done). Some people think that The Turing Test provides an entirely appropriate goal for research in AI; while other people think that there is a sense in which The Turing Test is not really demanding enough, and who suppose that The Turing Test needs to be extended in various ways in order to provide an appropriate goal for AI. We shall consider some representatives of each of these positions in turn.

5.1 The Turing Test is Too Hard

Some people have claimed that The Turing Test doesn't set an appropriate goal for current research in AI because we are plainly so far away from attaining this goal. Amongst these people there are some who have gone on to offer reasons for thinking that it is doubtful that we shall ever be able to create a machine that can pass The Turing Test—or, at any rate, that it is doubtful that we shall be able to do this at any time in the foreseeable future. Perhaps the most interesting arguments of this kind are due to French (1990); at any rate, these are the arguments that we shall go on to consider. (Cullen (2009) sets out similar considerations.)

According to French, The Turing Test is "virtually useless" as a real test of intelligence, because nothing without a "human subcognitive substrate" could pass the test, and yet the development of an artificial "human cognitive substrate" is almost impossibly difficult. At the very least, there are straightforward sets of questions that reveal "low-level cognitive structure" and that—in French's view—are almost certain to be successful in separating human beings from machines.

First, if interrogators are allowed to draw on the results of research into, say, *associative priming*, then there is data that will very plausibly separate human beings from machines. For example, there is research that shows that, if humans are presented with series of strings of letters, they require less time to recognize that a string is a word (in a language that they speak) if it is preceded by a related word (in the language that they speak), rather than by an unrelated word (in the language that they speak) or a string of letters that is not a word (in the language that they speak). Provided that the interrogator has accurate data about average recognition times for subjects who speak the language in question, the interrogator can distinguish between the machine and the human simply by looking at recognition times for appropriate series of strings of letters. Or so says French. It isn't clear to us that this is right. After all, the design of The Turing Test makes it hard to see how the interrogator will get reliable information about response times to series of strings of symbols. The point of putting the computer in a separate room and requiring communication by teletype was precisely to rule out certain irrelevant ways of identifying the computer. If these requirements don't already rule out identification of the computer by the application of tests of associative priming, then the requirements can surely be altered to bring it about that this is the case. (Perhaps it is also worth noting that administration of the kind of test that French imagines is not ordinary conversation; nor is it something that one would expect that any but a few expert interrogators would happen upon. So, even if the circumstances of The Turing Test do not rule out the kind of procedure that French here envisages, it is not clear that The Turing Test will be impossibly hard for machines to pass.)

Second, at a slightly higher cognitive level, there are certain kinds of "ratings games" that French supposes will be very reliable discriminators between humans and machines. For instance, the "Neologism Ratings Game"—which asks participants to rank made-up words on their appropriateness as names for given kinds of entities—and the "Category Rating Game"—which asks participants to rate things of one category as things of another category—are both, according to French, likely to prove highly reliable in discriminating between humans and machines. For, in the first case, the ratings that humans make depend upon large numbers of culturally acquired associations (which it would be well-nigh impossible to identify and describe, and hence which it would (arguably) be well-nigh impossible to program into a computer). And, in the second case, the ratings that people actually

make are highly dependent upon particular social and cultural settings (and upon the particular ways in which human life is experienced). To take French's examples, there would be widespread agreement amongst competent English speakers in the technologically developed Western world that "Flugblogs" is not an appropriate name for a breakfast cereal, while "Flugly" is an appropriate name for a child's teddy bear. And there would also be widespread agreement amongst competent speakers of English in the developed world that pens rate higher as weapons than grand pianos rate as wheelbarrows. Again, there are questions that can be raised about French's argument here. It is not clear to us that the data upon which the ratings games rely is as reliable as French would have us suppose. (At least one of us thinks that "Flugly" would be an entirely inappropriate name for a child's teddy bear, a response that is due to the similarity between the made-up word "Flugly" and the word "Fugly," that had some currency in the primarily undergraduate University college that we both attended. At least one of us also thinks that young children would very likely be delighted to eat a cereal called "Flugblogs," and that a good answer to the question about ratings pens and grand pianos is that it all depends upon the pens and grand pianos in question. What if the grand piano has wheels? What if the opponent has a sword or a sub-machine gun? It isn't obvious that a refusal to play this kind of ratings game would necessarily be a give-away that one is a machine.) Moreover, even if the data is reliable, it is not obvious that any but a select group of interrogators will hit upon this kind of strategy for trying to unmask the machine; nor is it obvious that it is impossibly hard to build a machine that is able to perform in the way in which typical humans do on these kinds of tests. In particular, if—as Turing assumes—it is possible to make learning machines that can be "trained up" to learn how to do various kinds of tasks, then it is quite unclear why these machines couldn't acquire just the same kinds of "subcognitive competencies" that human children acquire when they are "trained up" in the use of language.

There are other reasons that have been given for thinking that The Turing Test is too hard (and, for this reason, inappropriate in setting goals for current research into artificial intelligence). In general, the idea is that there may well be features of human cognition that are particularly hard to simulate, but that are not in any sense essential for intelligence (or thought, or possession of a mind). The problem here is not merely that The Turing Test really does test for *human* intelligence; rather, the problem here is the fact—if indeed it is a fact—that there are quite inessential features of human intelligence that are extraordinarily difficult to replicate in a machine. If this complaint is justified—if, indeed, there are features of human intelligence that are extraordinarily difficult to replicate in machines, *and* that could and would be reliably used to unmask machines in runs of The Turing Test —then there is reason to worry about the idea that The Turing Test sets an appropriate direction for research in artificial intelligence. However, as our discussion of French shows, there may be reason for caution in supposing that the kinds of considerations discussed in the present section show that we are already in a position to say that The Turing Test does indeed set inappropriate goals for research in artificial intelligence.

5.2 The Turing Test is Too Narrow

There are authors who have suggested that The Turing Test does not set a sufficiently broad goal for research in the area of artificial intelligence. Amongst these authors, there are many who suppose that The Turing Test is too easy. (We go on to consider some of these authors in the next sub-section.) But there are also some authors who have supposed that, even if the goal that is set by The Turing Test is very demanding indeed, it is nonetheless too restrictive.

Objection to the notion that the Turing Test provides a logically sufficient condition for intelligence can be adapted to the goal of showing that the Turing Test is too restrictive. Consider, for example, Gunderson (1964). Gunderson has two major complaints to make against The Turing Test. First, he thinks that success in Turing's Imitation Game might come for reasons other than the possession of intelligence. But, second, he thinks that success in the Imitation Game would be but one example of the kinds of things that intelligent beings can do and—hence—in itself could not be taken as a reliable indicator of intelligence. By way of analogy, Gunderson offers the case of a vacuum cleaner salesman

who claims that his product is "all-purpose" when, in fact, all it does is to suck up dust. According to Gunderson, Turing is in the same position as the vacuum cleaner salesman *if* he is prepared to say that a machine is intelligent merely on the basis of its success in the Imitation Game. Just as "all purpose" entails the ability to do a range of things, so, too, "thinking" entails the possession of a range of abilities (beyond the mere ability to succeed in the Imitation Game).

There is an obvious reply to the argument that we have here attributed to Gunderson, viz. that a machine that is capable of success in the Imitation Game is capable of doing a large range of different kinds of things. In order to carry out a conversation, one needs to have many different kinds of cognitive skills, each of which is capable of application in other areas. Apart from the obvious general cognitive competencies—memory, perception, etc.—there are many particular competencies—rudimentary arithmetic abilities, understanding of the rules of games, rudimentary understanding of national politics, etc.—which are tested in the course of repeated runs of the Imitation Game. It is inconceivable that that there be a machine that is startlingly good at playing the Imitation Game, and yet unable to do well at *any* other tasks that might be assigned to it; and it is equally inconceivable that there is a machine that is startlingly good at the Imitation Game and yet that does not have a wide range of competencies that can be displayed in a range of quite disparate areas. To the extent that Gunderson considers this line of reply, all that he says is that there is no reason to think that a machine that can succeed in the Imitation Game *must* have more than a narrow range of abilities; we think that there is no reason to believe that this reply should be taken seriously.

More recently, Erion (2001) has defended a position that has some affinity to that of Gunderson. According to Erion, machines might be "capable of outperforming human beings in limited tasks in specific environments, [and yet] still be unable to act skillfully in the diverse range of situations that a person with common sense can" (36). On one way of understanding the claim that Erion makes, he too believes that The Turing Test only identifies one amongst a range of independent competencies that are possessed by intelligent human beings, and it is for this reason that he proposes a more comprehensive "Cartesian Test" that "involves a more careful examination of a creature's language, [and] also tests the creature's ability to solve problems in a wide variety of everyday circumstances" (37). In our view, at least when The Turing Test is properly understood, it is clear that anything that passes The Turing Test must have the ability to solve problems in a wide variety of everyday circumstances (because the interrogators will use their questions to probe these—and other—kinds of abilities in those who play the Imitation Game).

5.3 The Turing Test is Too Easy

There are authors who have suggested that The Turing Test should be replaced with a more demanding test of one kind or another. It is not at all clear that any of these tests actually proposes a better goal for research in AI than is set by The Turing Test. However, in this section, we shall not attempt to defend that claim; rather, we shall simply describe some of the further tests that have been proposed, and make occasional comments upon them. (One preliminary point upon which we wish to insist is that Turing's Imitation Game was devised against the background of the limitations imposed by then current technology. It is, of course, not essential to the game that tele-text devices be used to prevent direct access to information about the sex or genus of participants in the game. We shall not advert to these relatively mundane kinds of considerations in what follows.)

5.3.1 The Total Turing Test

Harnad (1989, 1991) claims that a better test than The Turing Test will be one that requires responses to all of our inputs, and not merely to text-formatted linguistic inputs. That is, according to Harnad, the appropriate goal for research in AI has to be to construct a robot with something like human sensorimotor capabilities. Harnad also considers the suggestion that it might be an appropriate goal for AI to aim for "neuromolecular indistinguishability," but rejects this suggestion on the grounds that once we know how to make a robot that can pass his Total Turing Test, there will be no problems

about mind-modeling that remain unsolved. It is an interesting question whether the test that Harnad proposes sets a more appropriate goal for AI research. In particular, it seems worth noting that it is not clear that there could be a system that was able to pass The Turing Test and yet that was not able to pass The Total Turing Test. Since Harnad himself seems to think that it is quite likely that "full robotic capacities [are] … necessary to generate … successful linguistic performance," it is unclear why there is reason to replace The Turing Test with his extended test. (This point against Harnad can be found in Hauser (1993:227), and elsewhere.)

5.3.2 The Lovelace Test

Bringsjord et al. (2001) propose that a more satisfactory aim for AI is provided by a certain kind of meta-test that they call the Lovelace Test. They say that an artificial agent *A*, designed by human H, passes the Lovelace Test just in case three conditions are jointly satisfied: (1) the artificial agent *A* produces output *O*; (2) *A*'s outputting *O* is not the result of a fluke hardware error, but rather the result of processes that *A* can repeat; and (3) *H*—or someone who knows what *H* knows and who has *H*'s resources—cannot explain how *A* produced *O* by appeal to *A*'s architecture, knowledge-base and core functions. Against this proposal, it seems worth noting that there are questions to be raised about the interpretation of the third condition. If a computer program is long and complex, then no human agent can explain in *complete* detail how the output was produced. (Why did the computer output 3.16 rather than 3.17?) But if we are allowed to give a highly schematic explanation—the computer took the input, did some internal processing and then produced an answer—then it seems that it will turn out to be very hard to support the claim that human agents ever do anything genuinely creative. (After all, we too take external input, perform internal processing, and produce outputs.) What is missing from the account that we are considering is any suggestion about the appropriate *level* of explanation that is to be provided. It is quite unclear why we should suppose that there is a relevant difference between people and machines at any level of explanation; but, if that's right, then the test in question is trivial. (One might also worry that the proposed test rules out *by fiat* the possibility that creativity can be best achieved by using genuine *randomising* devices.)

5.3.3 The Truly Total Turing Test

Schweizer (1998) claims that a better test than The Turing Test will advert to the evolutionary history of the subjects of the test. When we attribute intelligence to human beings, we rely on an extensive historical record of the intellectual achievements of human beings. On the basis of this historical record, we are able to claim that human beings are intelligent; and we can rely upon this claim when we attribute intelligence to individual human beings on the basis of their behavior. According to Schweizer, if we are to attribute intelligence to machines, we need to be able to advert to a comparable historical record of cognitive achievements. So, it will only be when machines have developed languages, written scientific treatises, composed symphonies, invented games, and the like, that we shall be in a position to attribute intelligence to individual machines on the basis of their behavior. Of course, we can still use The Turing Test to determine whether an individual machine is intelligent: but our answer to the question won't depend merely upon whether or not the machine is successful in The Turing Test; there is the further "evolutionary" condition that also must be satisfied. Against Schweizer, it seems worth noting that it is not at all clear that our reason for granting intelligence to other humans on the basis of their behavior is that we have prior knowledge of the collective cognitive achievements of human beings.

5.4 Should the Turing Test be Considered Harmful?

Perhaps the best known attack on the suggestion that The Turing Test provides an appropriate research goal for AI is due to Hayes and Ford (1995). Among the controversial claims that Hayes and Ford make, there are at least the following:

a. Turing suggested the imitation game as a definite goal for program of research.

b. Turing intended The Turing Test to be a gender test rather than a species test.

c. The task of trying to make a machine that is successful in The Turing Test is so extremely difficult that no one could seriously adopt the creation of such a machine as a research goal.

d. The Turing Test suffers from the basic design flaw that it sets out to confirm a "null hypothesis", viz. that there is no difference in behavior between certain machines and humans.

e. No null effect experiment can provide an adequate criterion for intelligence, since the question can always arise that the judges did not look hard enough (and did not raise the right kinds of questions). But, if this question is left open, then there is no stable endpoint of enquiry.

f. Null effect experiments cannot measure anything: The Turing Test can only test for complete success. ("A man who failed to seem feminine in 10% of what he said would almost always fail the Imitation game.")

g. The Turing Test is really a test of the ability of the human species to discriminate its members from human imposters. ("The gender test … is a test of making a mechanical transvestite.")

h. The Turing Test is circular: what it fails to detect cannot be "intelligence" or"humanity", since many humans would fail The Turing Test. Indeed, "since one of the players must be judged to be a machine, half the human population would fail the species test".

i. The perspective of The Turing Test is arrogant and parochial: it mistakenly assumes that we can understand human cognition without first obtaining a firm grasp of the basic principles of cognition.

j. The Turing Test does not admit of weaker, different, or even stronger forms of intelligence than those deemed human.

Some of these claims seem straightforwardly incorrect. Consider (h), for example. In what sense can it be claimed that 50% of the human population would fail "the species test"? If "the species test" requires the interrogator to decide which of two people is a machine, why should it be thought that the verdict of the interrogator has any consequences for the assessment of the intelligence of the person who is judged to be a machine? (Remember, too, that one of the conditions for "the species test"—as it is originally described by Hayes and Ford—is that one of the contestants *is* a machine. While the machine can "demonstrate" its intelligence by winning the imitation game, a person cannot "demonstrate" their lack of intelligence by failing to win.)

It seems wrong to say that The Turing Test is defective because it is a "null effect experiment". True enough, there is a sense in which The Turing Test does look for a "null result": if ordinary judges in the specified circumstances fail to identify the machine (at a given level of success), then there is a given likelihood that the machine is intelligent. But the point of insisting on "ordinary judges" in the specified circumstances is precisely to rule out irrelevant ways of identifying the machine (i.e. ways of identifying the machine that are not relevant to the question whether it is intelligent). There might be all kinds of irrelevant differences between a given kind of machine and a human being—not all of them rendered undetectable by the experimental set-up that Turing describes—but The Turing Test will remain a good test provided that it is able to ignore these irrelevant differences.

It also seems doubtful that it is a serious failing of The Turing Test that it can only test for "complete success". On the one hand, if a man has a one in ten chance of producing a claim that is plainly not feminine, then we can compute the chance that he will be discovered in a game in which he answers N questions—and, if N is sufficiently small, then it won't turn out that "he would almost always fail to win". On the other hand, as we noted at the end of Section 4.4 above, if one were worried about the "YES/NO" nature of "The Turing Test", then one could always get the judges to produce probabilistic verdicts instead. This change preserves the character of The Turing Test, but gives it scope for greater statistical sophistication.

While there are (many) other criticisms that can be made of the claims defended by Hayes and Ford (1995), it should be acknowledged that they are right to worry about the suggestion that The Turing Test provides the defining goal for research in AI. There are various reasons why one should be loathe

to accept the proposition that the one central ambition of AI research is to produce artificial people. However it is worth pointing out that there is no reason to think that Turing supposed that The Turing Test defined the field of AI research (and there is not much evidence that any other serious thinkers have thought so either). Turing himself was well aware that there might be non-human forms of intelligence—cf. (j) above. However, all of this remains consistent with the suggestion that it is quite appropriate to suppose that The Turing Test sets *one* long term goal for AI research: one thing that we might well aim to do *eventually* is to produce artificial people. If—as Hayes and Ford claim—that task is almost impossibly difficult, then there is no harm in supposing that the goal is merely an *ambit* goal to which few resources should be committed; but we might still have good reason to allow that it is *a* goal.

6. The Chinese Room

There are many different objections to The Turing Test which have surfaced in the literature during the past fifty years, but which we have not yet discussed. We cannot hope to canvass all of these objections here. However, there is one argument—Searle's "Chinese Room" argument—that is mentioned so often in connection with the Turing Test that we feel obliged to end with some discussion of it.

In *Minds, Brains and Programs* and elsewhere, John Searle argues against the claim that "appropriately programmed computers literally have cognitive states" (64). Clearly enough, Searle is here disagreeing with Turing's claim that an appropriately programmed computer could think. There is much that is controversial about Searle's argument; we shall just consider *one* way of understanding what it is that he is arguing for.

The basic structure of Searle's argument is very well known. We can imagine a "hand simulation" of an intelligent agent—in the case described, a speaker of a Chinese language—in circumstances in which we might well be very reluctant to allow that there is any appropriate intelligence lying behind the simulated behavior. (Thus, what we are invited to suppose is a logical possibility is not so very different from what Block invites us to suppose is a logical possibility. However, the argument that Searle goes on to develop is rather different from the argument that Block defends.) *Moreover*—and this is really the key point for Searle's argument—the "hand simulation" in question is, in all relevant respects, simply a special kind of digital computation. So, there is a possible world—doubtless one quite remote from the actual world—in which a digital computer simulates intelligence but in which the digital computer does not itself possess intelligence. But, if we consider any digital computer in the actual world, it will not differ from the computer in that remote possible world in any way which could make it the case that the computer in the actual world is more intelligent than the computer in that remote possible world. Given that we agree that the "hand simulating" computer in the Chinese Room is not intelligent, we have no option but to conclude that digital computers are simply not the kinds of things that *can* be intelligent.

So far, the argument that we have described arrives at the conclusion that no appropriately programmed computer can think. While this conclusion is not one that Turing accepted, it is important to note that it is compatible with the claim that The Turing Test is a good test for intelligence. This is because, for all that has been argued, it may be that it is not *nomically* possible to provide any "hand simulation" of intelligence (and, in particular, that it is not possible to simulate intelligence using any kind of computer). In order to turn Searle's argument—at least in the way in which we have developed it—into an objection to The Turing Test, we need to have some reason for thinking that it is at least *nomically* possible to simulate intelligence using computers. (If it is nomically impossible to simulate intelligence using computers, then the alleged fact that digital computers cannot genuinely possess intelligence casts no doubt at all on the usefulness of the Turing Test, since digital computers are nomically disqualified from the range of cases in which there is mere simulation of intelligence.) In the absence of reason to believe this, the most that Searle's argument yields is an objection to Turing's confidently held belief that digital computing machines will one day pass The Turing Test. (Here, as

elsewhere, we are supposing that, for any kind of creature C, there is a version of The Turing Test in which C takes the role of the machine in the specific test that Turing describes. This general format for testing for the presence of intelligence would not necessarily be undermined by the success of Searle's Chinese Room argument.)

There are various responses that might be made to the argument that we have attributed to Searle. One kind of response is to dispute the claim that there is no intelligence present in the case of the Chinese Room. (Suppose that the "hand simulation" is embedded in a robot that is equipped with appropriate sensors, etc. Suppose, further, that the "hand simulation" involves updating the process of "hand simulation," etc. If enough details of this kind are added, then it becomes quite unclear whether we do want to say that we still haven't described an intelligent system.) Another kind of response is to dispute the claim that digital computers in the actual world could not be relevantly different from the system that operates in the Chinese Room in that remote possible world. (If we suppose that the core of the Chinese Room is a kind of giant look-up table, then it may well be important to note that digital computers in the actual world do not work with look-up tables in that kind of way.) Doubtless there are other possible lines of response as well. However, it would take us out of our way to try to take this discussion further. (One good place to look for further discussion of these matters is Braddon-Mitchell and Jackson (1996).)

7. A Brief Note on Measurement of Intelligence

There are radically different views about the measurement of intelligence that have not been canvassed in this article. Our concern has been to discuss Turing (1950) and its legacy. But, of course, a more wide-ranging discussion would also consider, for example, research on the measurement of intelligence using the mathematical and computational resources of Algorithmic Information Theory, Kolmogorov Complexity Theory, Minimum Message Length (MML) Theory, and so forth. (For an introduction to this literature, see Hernandez-Orallo and Dowe (2010), and the list of references contained therein.)

Bibliography

Abramson, D. (2008), "Turing's Responses to Two Objections," *Minds and Machines*, 18: 147–67.

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

Boolos, G. and Jeffrey, R. (1980), *Computability and Logic*, Second Edition, Cambridge: Cambridge University Press.

Bowie, L. (1982), "Lucas's Number is Finally Up," *Journal of Philosophical Logic*, 11: 279–85.

Braddon-Mitchell, D. and Jackson, F. (1996), *The Philosophy of Mind and Cognition*, Oxford: Blackwell.

Bringsjord, S., Bello, P. and Ferrucci, D. (2001), "Creativity, the Turing Test, and the (Better) Lovelace Test," *Minds and Machines*, 11: 3–27.

Chalmers, D. (1995), "On Implementing a Computation," *Minds and Machines*, 4: 391–402.

Churchland, P. M. and Churchland, P. S. (1990), "Could a Machine Think?" *Scientific American*, 262 (1): 32–37.

Clark, A. (1997), *Being There: Putting Brain, Body and World Together Again*, Cambridge: MIT Press.

Copeland, J. (ed.) (1999), "A Lecture and Two Radio Broadcasts on Machine Intelligence by Alan Turing," in K. Furukawa, D. Michie, and S. Muggleton (eds.), *Machine Intelligence* 15, Oxford: Oxford University Press.

Copeland, J. (2000), "The Turing Test," *Minds and Machines*, 10: 519–39.

Copeland, J. and Sylvan, R. (1999), "Beyond the Universal Turing Machine," *Australasian Journal of Philosophy*, 77: (1): 46–66.

Crooke, A. (2002), *Confabulating Consciousness*, Ph.D. Dissertation, Philosophy Department, Monash University.

Cullen, J. (2009), "Imitation Versus Communication: Testing for Human-Like Intelligence," *Minds and Machines*, 19: 237–54.

Dennett, D. (1985), "Can Machines Think?" in M. Shafto (ed.), *How We Know*, Cambridge, MA: Harper and Row.

Dietrich, E. (ed.) (1994), *Thinking Computers and Virtual Persons: Essays on the Intentionality of Machines*, San Diego: Academic Press.

Dreyfus, H & Dreyfus, S. (1986), *Mind Over Machine*, New York: Free Press.

Erion, G. (2001), "The Cartesian Test for Automatism," *Minds and Machines*, 11: 29–39.

Feferman, S. (1996), "Penrose's Gödelian Argument,", *Psyche*, 2: 21–32.

French, R. (1990), "Subcognition and the Limits of the Turing Test," *Mind*, 99: 53–65.

Genova, J. (1994), "Turing's Sexual Guessing Game," *Social Epistemology*, 8: 313–26.

Gunderson, K. (1964), "Descartes, La Mettrie, Language and Machines," *Philosophy*, 39: 193–222.

Gunderson, K. (1985), *Mentality and Machines*, 2nd edition, Minneapolis: University of Minnesota Press.

Harnad, S. (1991), "Other Bodies, Other Minds: A Machine Incarnation of an Old Philosophical Problem," *Minds and Machines*, 1: 43–54.

Haugeland, J. (1981), "Semantic Engines: An Introduction to Mind Design," in J. Haugeland (ed.), *Mind Design: Philosophy, Psychology, Artificial Intelligence*, Cambridge: MIT Press, 1–34.

Hauser, L. (1993), "Reaping the Worldwind: Reply to Harnad's Other Bodies, Other Minds," *Minds and Machines*, 3: 219–38.

Hauser, L. (2001), "Look Who's Moving the Goalposts Now," *Minds and Machines*, 11: 41–51.

Hayes, P., and Ford, K. (1995), "Turing Test Considered Harmful," *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, Montreal: Morgan Kaufmann, 972–977.

Hernandez-Orallo, J., and Dowe, D. L. (2010), "Measuring Univeral Intelligence: Towards an Anytime Intelligence Test," *Artificial Intelligence*, 174: 1508–39.

Hodges, A. (1983), *Alan Turing: The Enigma*, London: Burnett with Hutchinson.

Hofstadter, D. (1982), "The Turing Test: A Coffee-House Conversation," in D. Hofstadter and D. Dennett (eds.), *The Mind's I: Fantasies and Reflections on Self and Soul*, London: Penguin, 69–95.

Lewis, D. (1969), "Lucas against mechanism," *Philosophy*, 44: 231–233.

Lewis, D. (1979), "Lucas against mechanism II," *Canadian Journal of Philosophy*, 9: 373–376.

Lucas, J. (1961), "Minds, Machines and Gödel," *Philosophy*, 36: 120–4.

Moor, J. (1976), "An Analysis of Turing's Test," *Philosophical Studies*, 30: 249–57.

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

Penrose, R. (1989), *The Emperor's New Mind*, Oxford: Oxford University Press.

Piccinini, G. (2000), "Turing's Rules for the Imitation Game," *Minds and Machines*, 10: 573–85.

Saygin, A., Cicekli, I., and Akman, V. (2000), "Turing Test: 50 Years Later," *Minds and Machines*, 10: 463–518.

Schweizer, P. (1998), "The Truly Total Turing Test," *Minds and Machines*, 8: 263–72.

Searle, J. (1981), "Minds, Brains, and Programs," *Behavioral and Brain Sciences*, 3: 417–57.

Shieber, S. (1994), "Lessons from a restricted Turing Test," *Communications of the Association for Computing Machinery*, 37: 70–8. [Preprint available online]

Shieber, S. (2007), "The Turing Test as Interactive Proof," *Nous*, 41: 686–713.

Sterrett, S. (2000), "Turing's Two Tests for Intelligence," *Minds and Machines*, 10: 541–59.

Traiger, S. (2000), "Making the Right Identification in the Turing Test," *Minds and Machines*, 10: 561–572.

Turing, A. (1950), "Computing Machinery and Intelligence," *Mind*, 59 (236): 433–60.

Weizenbaum, J. (1966), "ELIZA-A Computer Program for the Study of Natural Language Communication Between Men and Machines," *Communications of the ACM*, 9: 36–45.

Whitby, B. (1996), "The Turing Test: AI's Biggest Blind Alley?" in P. Millican and A. Clark (eds.), *Machines and Thought: The Legacy of Alan Turing*, Volume 1, Oxford: Clarendon.

Zdenek, S. (2001), "Passing Loebner's Turing Test: A Case of Conflicting Discourse Functions," *Minds and Machines*, 11: 53–76.

Academic Tools

    How to cite this entry.

    Preview the PDF version of this entry at the Friends of the SEP Society.

    Look up this entry topic at the Indiana Philosophy Ontology Project (InPhO).

    Enhanced bibliography for this entry at PhilPapers, with links to its database.

Other Internet Resources

- Alan Turing Home Page (Andrew Hodges, Wadham College, Oxford)
- "Computing machinery and intelligence" by Alan Turing (1950) (hosted by Hugh Loebner)
- The Loebner Prize
- A.L.I.C.E. AI Foundation
- Machine Intelligence Part 1: The Turing Test and Loebner Prize (Ashley Dunn)
- CAPTCHAS, (Dr. Manuel Blum, CS, Carnegie Mellon. CAPTCHA is an acronym: Completely Automated Public Turing Test to Tell Computers and Humans Apart.)

Related Entries

artificial intelligence | Chinese room argument | functionalism | logic: provability | Turing, Alan

Acknowledgments

Open access to the SEP is made possible by a world-wide funding initiative.
Please Read How You Can Help Keep the Encyclopedia Free

Stanford | Center for the Study of Language and Information