Artificial Intelligence 175 (2011) 950–957

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Anthropomorphism and AI: Turing's much misunderstood imitation game

Diane Proudfoot

Department of Philosophy and the Turing Archive for the History of Computing, University of Canterbury, New Zealand

ARTICLE INFO

Article history:
Received 5 April 2010
Accepted 31 October 2010
Available online 21 January 2011

Keywords:
Turing test
Anthropomorphism
Human-level AI
Mindless intelligence

ABSTRACT

The widespread tendency, even within AI, to anthropomorphize machines makes it easier to convince us of their intelligence. How can any putative demonstration of intelligence in machines be trusted if the AI researcher readily succumbs to make-believe? This is (what I shall call) *the forensic problem of anthropomorphism*. I argue that the Turing test provides a solution. This paper illustrates the phenomenon of misplaced anthropomorphism and presents a new perspective on Turing's imitation game. It also examines the role of the Turing test in relation to the current dispute between human-level AI and 'mindless intelligence'.

1. Introduction

A fundamental dispute is alive again in AI. Several computer scientists once more regard human-level artificial intelligence—or 'artificial general intelligence' (AGI)—as a central and achievable goal. Conferences and workshops have again focused on human-level AI and AGI.[1] In a recent article in *Artificial Intelligence*, John McCarthy said, 'Human-level AI will be achieved … I'd be inclined to bet on this 21st century'.[2] Some in the field, however, see 'human-level' AI as an ill-defined goal[3] and others expressly oppose it—Jordan Pollack recently claimed that 'AI's great mistake is its assumption that *human-level intelligence is the greatest intelligence that exists*', and called for AI to concentrate instead on 'mindless intelligence'.[4]

Critics of AI's traditional focus on human-level (or human-like) intelligence typically blame the Turing test. For example:

> [T]he Turing Test has served to focus much AI research on those facets of human behavior which are least susceptible to useful generalization precisely because they are not shared by other species … The influence of the Turing Test vision … is a tragedy for AI.[5]

> "Computing Machinery and Intelligence" has led AI into a blind alley. … If we focus future work in AI on the imitation of human abilities, such as might be required to succeed in the imitation game, we are in effect building "intellectual statues" when what we need are "intellectual tools".[6]

Turing's imitation game has been described as 'virtually useless', 'obsolete', and 'impotent', and machines that (supposedly) can do well in the game as 'dead ends in artificial intelligence research'.[7] Opponents say that '[a]dherence to Turing's vision … is … actively harmful' and that 'Turing's legacy alienates maturing subfields'.[8] For many, the moral is clear:

> [T]he Turing test should be relegated to the history of science … [W]ith its focus on imitating human performance, [it] is … directly at odds with the proper objectives of AI.[9]

E-mail address: diane.proudfoot@canterbury.ac.nz.

0004-3702/$ – see front matter
doi:10.1016/j.artint.2011.01.006

Some theorists even argue that Turing did not intend to propose a *test* at all. Aaron Sloman, for example, claims that Turing was 'far too intelligent to do any such thing' and that this widespread misinterpretation of 'Computing Machinery and Intelligence' has led to 'huge amounts of wasted effort' discussing the purely 'mythical' Turing test.[10] According to Sloman, Turing introduced the imitation game, 'not in order to give a criterion for thinking or for intelligence', but to provide 'a basis for attacking arguments' against the possibility that a computer could succeed in the game.[11]

However, Turing did make it clear that he was proposing a test of intelligence in machines. In his famous 1950 paper he referred to the imitation game as a 'test' and as providing a 'criterion for "thinking"' (he even said that his opponent Geoffrey Jefferson probably 'would be quite willing to accept the imitation game as a test').[12] In his 1952 radio broadcast, 'Can Automatic Calculating Machines Be Said To Think?', Turing said 'I would like to suggest a particular kind of *test* that one might apply to a machine'. He described the imitation game and continued:

> Well, that's my test. Of course I am not saying at present either that machines really could pass the test, or that they couldn't. My suggestion is just that this is the question we should discuss. It's not the same as 'Do machines think?', but it seems near enough for our present purpose, and raises much the same difficulties.[13]

Turing also made it clear that he introduced the imitation game in order to illustrate his approach to intelligence—rather than, as Sloman claims, to make a technological prediction or rebut objections to machine intelligence. Turing first described the computer-imitates-human game in 'Intelligent Machinery' (his 1948 report for the National Physical Laboratory) as a 'little experiment' (which, he said, he had actually conducted) based on his thesis that '[t]he extent to which we regard something as behaving in an intelligent manner is determined as much by our own state of mind and training as by the properties of the object under consideration'.[14]

I shall argue that Turing's test is of considerable importance for AI today—and not solely as an intriguing thought experiment or as part of a far-off scientific goal. Advocates of both human-level AI and of mindless intelligence face the question: how do we test for intelligence in machines? Both also face a difficult problem that results from the human tendency to *anthropomorphize* artificial systems. This tendency makes it too easy to convince us of the intelligence, human-level or otherwise, of a machine. Turing provided a test of intelligence in machines that solves this problem.

2. Anthropomorphism in AI

Computer scientists often complain that the public expects intelligent androids to appear any day. Yet science fiction and make-believe can also be found *within* AI—even at its very beginnings.[15] Turing compared what he called his 'child-machine' to Helen Keller, said that the machine could not 'be sent to school without the other children making excessive fun of it' but it would have 'homework', and he suggested that 'the education of the machine should be entrusted to some highly competent schoolmaster'.[16] Turing was discussing his 'P-type' unorganized machines (modified Turing machines, with no tape and with 'pain' and 'pleasure' inputs).[17] These existed only as 'paper machines'—simulations of machine behaviour by a human being using paper and pencil[18]—but Turing anthropomorphized them nonetheless.

Illustrative cases of anthropomorphism also include Valentino Braitenberg's descriptions of his robot vehicles. These machines are said to dream, sleepwalk, have free will, and 'ponder over their decisions'; they are 'inquisitive', 'optimistic', and 'friendly'. They even have the 'a priori concept of 2-dimensional space'.[19] David Hogg, Fred Martin, and Mitchel Resnick gave their Braitenberg-like 'creatures' (constructed from LEGO bricks containing electronic circuits) names such as 'Timid', 'Indecisive', 'Paranoid', 'Insecure', and 'Inhumane'. Frantic and Observant are, they said, 'philosophical'; Frantic 'does nothing but think'.[20] Masaki Yamamoto too described his robot vacuum cleaner SOZZY as 'friendly'—and as having 'four emotions ... joy, desperation, fatigue, sadness'.[21] Daniel Dennett said that Cog is to have an 'infancy and childhood' and will be designed 'to want to keep its mother's face in view'. Cog

> must somehow delight in learning, abhor error, strive for novelty, recognize progress. It must be vigilant in some regards, curious in others, and deeply unwilling to engage in self-destructive activity. While we are at it, we might as well try to make it crave human praise and company, and even exhibit a sense of humour.[22]

Often the researchers building social robots such as Cog (and especially 'face' robots) attempt to avoid anthropomorphism—for example, by using scare-quotes or other notational devices when describing a face robot's 'emotion' system, and by claiming that the robot has mere analogues of human emotions.[23] In these cases they typically deny that the robot has—'real' or human—emotions.[24] (This denial need not presuppose a *definition* of 'emotion'. To use a famous example, we do not need a definition of a 'game' to be justified in saying that football is a game and dermatology is not. Nor can we provide one[25]: instead we use prototypes and exemplars. Likewise, we do not need a definition of 'emotion' to be justified in saying that you have emotions and SOZZY does not.)

Nevertheless, the same researchers who deny that their robots have emotions attribute *expressive behaviours* to the machines literally and without qualification; in this way they unwittingly anthropomorphize the machines. Kismet, for example, is said (without scare-quotes) to have 'a smile on [its] face', to 'frown', and to have a 'happy and interested expression' and a 'sorrowful expression'.[26] Kismet is also said to 'glance towards' a toy, 'look in another direction', and 'settl[e] its gaze on the person's face'.[27] This is not to claim merely that the robot has certain physical configurations or behaviours. Smiling, for

example, is a complex conventional gesture. A facial display is a *smile* only if it has a certain meaning—the meaning that distinguishes a smile from a human grimace or facial tic, and from a chimpanzee's bared-teeth display. Similarly, saying that Kismet has a *happy expression* (rather than a 'happy expression') is claiming that the meaning of the robot's 'facial' display is *happiness* (and not mere 'happiness'). In saying that Kismet smiles and frowns (as opposed to saying merely that its facial display is, like an emoticon or photograph, a *representation* of a smile or frown), Kismet's creators are claiming that the robot has a certain communicative intent—the intent possessed by creatures that smile, namely human beings.

In the 1970s Drew McDermott ridiculed the use of 'wishful mnemonics' to refer to programs and data structures. He said that, if the AI programmer

> calls the main loop of his program "UNDERSTAND", he is (until proven innocent) merely begging the question. He may mislead a lot of people, most prominently himself, and enrage a lot of others. What he should do instead is refer to this main loop as "G0034", and see if he can *convince* himself or anyone else that G0034 implements some part of understanding.[28]

McDermott's aim was to eliminate the use of wishful mnemonics: he said, 'If we are to retain any credibility, this should stop'. His recommendation to the 'disciplined' AI researcher is to use 'colourless' or 'sanitized' terms—those with a 'humble, technical meaning'.[29]

Anthropomorphism in AI goes far beyond wishful mnemonics, though. Some of the examples given above could be seen as mnemonics; for example, Turing's 'pain' or 'punishment' signal could instead be referred to as a 'cancellation' instruction (it cancels tentative entries in the P-type's machine table). In contrast, the claim that a Braitenberg machine has the 'a priori concept of 2-dimensional space' cannot be cashed out in this way. Likewise, although some instances of anthropomorphism in AI are question-begging, as McDermott points out, other examples give rise to a different problem. Describing SOZZY as friendly is to engage in make-believe and make a plainly *false*, rather than a question-begging, claim.[30]

McDermott requires that AI researchers, before using psychological terms, demonstrate that their programs do implement psychological properties. But how *can* a researcher's effort to 'convince himself or anyone else' of intelligence in machines be trusted if the researcher readily succumbs to anthropomorphism and make-believe—ascribing joy to a robot vacuum cleaner, for example? This is (what I shall call) *the forensic problem of anthropomorphism*. This is not the traditional metaphysical problem of the relation between the physical and the mental, nor the ethical question of how we should treat artificial agents; nor does it assume that humans are exceptional. And it poses a different problem from that identified by McDermott. The forensic problem of anthropomorphism is this: anthropomorphizing risks introducing bias (in favour of the machine) into judgements of intelligence in machines—unless the risk is mitigated, these judgements are suspect.

McDermott's prescription for the 'disease' of wishful mnemonics is abstinence.[31] Yet fiction can be useful (the *centre of gravity* and *light travels in a straight line* are useful fictions in physics), and researchers in social robotics point to advantages in anthropomorphizing machines—for example, it facilitates human-machine interaction and machine learning. Moreover, anthropomorphism may be, not eliminable, but natural and inevitable (even McDermott, at the very same time as ridiculing wishful mnemonics, calls GPS a 'particularly stupid' program).[32] The forensic problem of anthropomorphism requires that anthropomorphism be *managed*, rather than purged. Turing's imitation game, I shall argue, achieves this.

3. A new perspective on the Turing test

Turing was well aware of how easy it is to anthropomorphize an artificial system; he said that playing chess against even a paper machine gives 'a definite feeling that one is pitting one's wits against something alive'.[33] (The 'little experiment' that, Turing indicated, he had actually performed two or so years before publishing 'Computing Machinery and Intelligence' was a chess-playing imitation game, involving a judge and two contestants—a human player and a paper machine.[34])

It is impossible now to determine exactly why Turing designed the imitation game as he did. However, the design of the game does avoid the forensic problem of anthropomorphism. This is because the game includes both a *disincentive* to anthropomorphize and a *control* to screen for any anthropomorphic bias in favour of the machine.

The disincentive to anthropomorphize results from the fact that imitation-game interrogators risk making a real and readily-exposed error—misidentifying a computer as a human being. The likelihood of this mistake can only be increased by the tendency to anthropomorphize; hence there is a disincentive to anthropomorphism built into Turing's game. This disincentive can be seen at work in the Loebner Prize Contest in Artificial Intelligence. Judges in this competition are suspicious, even hostile—rather than behaving like the humans playing with Kismet or chatting with Eliza. For example, they use nonsense utterances (e.g. 'Wljso lwjejdo wleje elwjeo wjeol, isn't it?'), deliberate misspellings ('what do you think of the whether?'), slang ('Yo, whassup?'), and questions probing common-sense knowledge ('I drove here this morning. Which side of the road should I have driven on?') in order to identify the machine.[35] Prior to 2004, the Loebner contest differed from the test in Turing's 1950 paper, in that it did not have the form of a 3-player game. Instead each judge interviewed each contestant individually (there were several judges and several contestants); the judges were unaware of the ratio of computer to human contestants. Judges in this 2-player form of the game appeared to behave strategically, to avoid error: in the 2000 contest, for example, no judge classed a computer as a human—in fact, in 10 interviews a *human* was classed as a computer.[36] In the 2003 competition, no judge classed a machine as 'definitely a human', and in 4 interviews a human was classed as 'definitely a machine'.[37] The judges refused to anthropomorphize, even when interviewing human beings.

Turing's 3-player imitation game (as formulated in 'Computing Machinery and Intelligence') contains an additional means of managing bias resulting from anthropomorphism. This game is a blind controlled trial of a machine's ability to answer questions, in a human-like manner, on a wide range of topics. Including a human contestant in simultaneous interviews turns the tendency to anthropomorphize into a controlled variable. We may assume that any tendency the interrogator has to anthropomorphize an unseen contestant is independent of the actual identity of the contestant.[38] The consequence is that the interrogator's tendency to anthropomorphize does *not* advantage the machine.

On this new understanding of Turing's test, a popular objection to the test is wrong-headed. Kenneth Ford and Patrick Hayes, for example, claim that the test is 'a poorly designed experiment ... depending too much on the subjectivity of the judge'.[39] Likewise, the philosopher Ned Block asserts that the test makes whether or not a machine thinks 'depend on how gullible human interrogators tend to be'.[40] This criticism of Turing's test is influential but gets things back-to-front. The imitation game *mitigates* anthropomorphic bias introduced by 'gullible' judges; it is a *well*-designed experiment.

4. Mindless intelligence and Turing's test

Mindless intelligence, in Pollack's sense, seems to be intelligence *without symbols*. It is plain that his main target is the physical symbol system hypothesis; he describes Lisp interpreters, symbols, grammars, and logic or inference engines as 'the accoutrements of cognition'. A 'mindless process', we may infer, will have none of these.

In this sense of 'mindless', it is in fact an error to *oppose* human-level intelligence and mindless intelligence; it is an entirely open question whether a mindless machine can achieve human-level intelligence. In consequence, there is a role for the Turing test as a criterion of human-level mindless intelligence in machines (the imitation game does not test for any particular cognitive architecture).

Pollack, of course, calls for computer science to stop chasing human-level intelligence—the 'same old AI goals' are, he says, 'red herrings that promise the practically impossible'. But why is human-level AI *impossible*? Pollack's reason is that '[s]ymbolic conscious reasoning is ... a myth'. However, this is to identify the goal of building a human-level intelligence with that of building a symbolic mind—and this is a mistake, since human-level AI can exploit different architectures.[41] (Turing himself recommended trying both symbolic and behaviour-based approaches.[42]) In fact, Pollack's call for mindless intelligence *presupposes* a criterion of human-level intelligence in machines, since he claims that dynamical physical processes can become 'more intelligent than' a smart adult human.

Turing shared several of the goals of mindless intelligence.[43] What is now called 'Turing Test AI' was not his only aim; his research was wide-ranging.[44] Crucially, the common assertion that the Turing test 'assumes that human thought is the final, highest pinnacle of thinking against which all others must be judged' is mistaken.[45] Turing recognized that his test measures human-like intelligence and that an intelligent non-human system might behave very differently.[46] His hope was that the imitation game would provide an *existence proof* of the hypothesis that machines can think. If a machine were to do well in the imitation game (played numerous times, with different interrogators, and on diverse topics), wouldn't that be a convincing demonstration of artificial intelligence—even to the die-hard naysayer?

5. Blockheads, zombies, and other objections

Any argument for the importance of Turing's test must address standard objections to the test. The most influential philosophical objection is: if a machine does well in the imitation game, this demonstrates merely that the machine *behaves as if* it is intelligent (or thinks)—not that it is *really* intelligent.[47] Why say this? One reason proposed is that manifestly *un*intelligent programs can pass the test. Some critics point to the success of primitive programs in Loebner's contests, especially the early competitions.[48] Yet these fail as counterexamples to Turing's test. Loebner's test is much easier than Turing's (for example, in the protocol for scoring the test); the early competitions also restricted the topics of the questions and forbade the judges to use tricks—unlike Turing's test.

To produce an effective counterexample, critics must invent hypothetical entities. For example, a *blockhead* (after Ned Block) is a hypothetical program that incorporates a vast look-up table; according to Block, the program he calls 'Aunt Bubbles' can do well in the imitation game, but has the intelligence only of a 'jukebox'.[49] However, this famous counterexample does not work.[50] *In the real world* no look-up table device would fool a Turing-test interrogator; given the constraints of processing speed and storage capacity, the candidate would take too long to answer the questions. Block acknowledges that a device like Aunt Bubbles is 'only logically possible, not physically possible'—the machine is, he admits, 'too vast to exist'. In his view this does not matter: 'because we are considering a proposed definition of intelligence that is supposed to capture the *concept* of intelligence, conceptual possibility will do the job'.[51] However, Turing did *not* use his test as a definition of (i.e., a logically necessary and sufficient condition for) intelligence.[52] It is perverse to insist on this interpretation of the test, in the face of Turing's saying (in his 1952 radio broadcast) 'I don't want to give a definition of thinking',[53] his acknowledgement (in his 1950 paper) that success in the test is not a necessary condition of intelligence, and his remarks (in his 1952 broadcast) demonstrating his concern with real-world machines—machines that can solve problems, he said, 'within a reasonable time'.[54] This last excludes look-up table devices of the sort described by Block and others.

Several critics of the Turing test argue as follows: 'real' intelligence (or thinking) essentially involves *consciousness*, and the imitation game cannot test for this.[55] John Searle, for example, appears to take this view; he claims that Turing's test aims to detect 'the presence of mind' but fails, since it tests only 'external behaviour'.[56] This criticism can be made out

using a *zombie* thought experiment. A zombie is a hypothetical entity that lacks all conscious awareness (i.e., qualitative awareness or 'feeling'), but is otherwise indistinguishable from a human being. The objection runs: a zombie could pass Turing's test, but isn't it obvious that such a creature does not really think? However, this objection begs two crucial questions. First, it simply assumes that thinking essentially involves consciousness—ignoring the fact that humans have non-conscious thoughts. Second, it assumes that a zombie *would* do well in the imitation game. But it may be that only an entity capable of conscious thought will pass Turing's test. This is an entirely open question, which is not settled by the fact that we can *imagine* a zombie passing the test (if indeed we can imagine this).

Turning from philosophical to scientific and engineering objections to Turing's test, many of these do not attack the test itself (for example, those objections quoted at the beginning of this paper). Some criticize the use of the test to stipulate a *core goal* for AI (namely, the imitation of human performance). Others criticize attempts by researchers working in *Good Old-Fashioned AI* to build machines that will pass the test. And the true target of some other objections is *Loebner's* annual contest, on the ground that it encourages programmers to use tricks. All these criticisms could be conceded without any damage to Turing's test.

Those scientific and engineering objections that attack the test directly make, in effect, two related criticisms. The first is: the test fails to assist researchers to *build* a thinking machine (let alone a mindless 'insect' or an 'intelligence amplifier' for humans).[57] For example, some critics argue that the test is too difficult and fails to provide a practical way forward for AI.[58] Turing recognized the difficulty in constructing a machine that would succeed in the game; in his 1952 broadcast he predicted that it would be 'at least 100 years' before a machine 'stand[s] any chance [of passing the test] with no questions barred'.[59] Nevertheless, if AI researchers are to build a thinking machine, they must have some criterion of *when* they have reached this goal. Turing's test provides such a criterion.

The second criticism is: the Turing test fails to assist researchers to *understand*—that is, construct a computational theory of—intelligence. For example, some critics argue that the focus on human-like systems will not help us to understand intelligence in a generic sense.[60] Turing might have agreed; he made it clear that his test is intended to provide a criterion of specifically human-like intelligence. Some critics, though, judge Turing's test too narrow even as a test of human-like intelligence.[61] Here Turing would not have agreed; he chose 'something like a viva-voce examination', he said, because it 'has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a man' and 'seems to be suitable for introducing almost any one of the fields of human endeavour that we wish to include'.[62] (The imitation game can also be used to test for different human intellectual capacities or levels.) Sloman objects that 'the very idea of a Turing test or any other test of intelligence is muddled [because] there is no binary divide between things that are and things that are not intelligent'.[63] However, even if the concept of intelligence is fuzzy, the imitation game can be used as a test of (human-like) intelligence; this is because, as Turing said, we cannot conclude from a contestant's failing to do well in the game that the contestant is *not* intelligent.

The complaint that the imitation game does not help researchers to theorize about intelligence overlooks Turing's own very different approach to intelligence. Sloman, for example, claims that a 'deep' understanding of human-level intelligence requires that we identify the mechanisms underlying intelligent human behaviour as solutions to problems posed in human evolutionary development (clearly the imitation game does not provide this).[64] In contrast, Turing said in his 1952 broadcast, 'As soon as one can see the cause and effect working themselves out in the brain, one regards it as not being thinking, but a sort of unimaginative donkey-work'.[65] In his view, the ordinary concept of intelligence is linked, not to the notion of underlying mechanisms, but to the notion of people's responses in specified conditions (see Section 1). Intelligence in this sense cannot be understood by the sort of investigation Sloman proposes—but it can be tested in the imitation game, which explores the human's response to the machine in carefully circumscribed circumstances.

6. 'Intelligence is in the eye of the observer'

If AI does abandon the Turing test, what will take its place? One popular approach (with some similarity to Turing's own view) is simply to say that 'intelligence is in the eye of the observer'. Rodney Brooks, for example, uses these very words (and Pollack's definition of 'mindless intelligence' specifies that it is intelligent behaviour ascribed by an observer).[66] This approach raises in a graphic way the forensic problem of anthropomorphism.

Brooks claims that 'it is only an external observer that has anything to do with cognition, by way of attributing cognitive abilities to a system that works well in the world'.[67] This recalls Dennett's classic stipulation that 'a particular thing is an intentional system only in relation to the strategies of someone who is trying to explain and predict its behavior'.[68] These claims raise the possibility that *any* machine is an intentional system, just because the observer anthropomorphizes the machine. How is this possibility to be avoided? Dennett has proposed that an observer is justified in treating a machine as an intentional system where it is *useful* ('convenient, explanatory, pragmatically necessary for prediction') to do so.[69] Brooks has also suggested that, if the illusion of genuine communication with a machine is shattered less and less, the machine should be counted as a thinking thing.[70]

Yet humans may *always* find it 'convenient' to adopt the intentional stance when giving common-sense accounts of a machine's behaviour. The extravagance with which even AI researchers anthropomorphize machines suggest that Dennett's condition (that the observer find the intentional stance useful) is too easily satisfied. The same problem arises for Brooks's suggestion. Many researchers argue that human beings have an evolved tendency to personalize the world (we even see faces in the clouds)[71]; if so, the illusion of communication with a machine may be too readily generated.

In an ideal world the forensic problem of anthropomorphism would not arise. Such a world would include an *ideal observer*—a spectator without any inclination to misplaced anthropomorphism, or who is able (despite this inclination) to make unbiased decisions. However, in the real world ideal observers are hard to find, if not impossible. Turing's imitation game solves the problem without invoking an ideal observer.

7. Conclusion

Irrespective of the theoretical merits of Turing's test, there are the familiar economic arguments against employing it. David Waltz, for example, recently claimed that in the future 'few agencies or industries are likely to fund research whose primary goal is to pass some variant of the Turing Test'.[72] Yet even if Waltz is correct, those researchers who do want to develop machines with human-level intelligence will require some standard for determining success. Without a criterion of human-level intelligence in machines, it is impossible even to theorize about building such devices. Moreover, building human-level AI requires a test of intelligence in machines that is not undermined by our tendency to anthropomorphize.

The Turing test is anthropomorphism-proofed.

Acknowledgements

I would like to thank Jack Copeland, Aaron Sloman, Randy Goebel, and an anonymous AIJ reviewer for valuable detailed comments and discussion.

Notes

[1] See, for example, N. Cassimatis, E.T. Mueller, and P.H. Winston, "Achieving human-level intelligence through integrated systems and research: introduction to this Special Issue", *AI Magazine,* vol. 27, no. 2, 2006; also E.A. Feigenbaum, "Some Challenges and Grand Challenges for Computational Intelligence", *Journal of the ACM*, vol. 50, no. 1, 2003; D.B. Lenat, "The voice of the turtle: whatever happened to AI?", *AI Magazine*, vol. 29, no. 2, 2008; B. Goertzel, "Human-level artificial general intelligence and the possibility of a technological singularity: A reaction to Ray Kurzweil's *The Singularity Is Near*, and McDermott's critique of Kurzweil", *Artificial Intelligence*, vol. 171, 2007; J. McCarthy, "From Here to Human-Level AI", *Artificial Intelligence*, vol. 171, 2007, and "Formalizing common sense knowledge in mathematical logic", *The Rutherford Journal: The New Zealand Journal for the History and Philosophy of Science and Technology,* vol. 3, www.rutherfordjournal.org; M. Minsky, P. Singh, and A. Sloman, "The St. Thomas common sense symposium: designing architectures for human-level intelligence", *AI Magazine,* vol. 25, no. 2, 2004; L.A. Zadeh, "Toward Human Level Machine Intelligence—Is It Achievable? The Need for a Paradigm Shift", *IEEE Computational Intelligence Magazine*, vol. 3, no. 3, 2008.

[2] J. McCarthy, "From Here to Human-Level AI", p. 1174. In the same issue Goertzel makes an analogous claim: 'AGI should be the focus of a significant percentage of contemporary AI research . . . [and] dramatic progress on AGI in the near future is something that's reasonably likely' ("Human-level artificial general intelligence and the possibility of a technological singularity: A reaction to Ray Kurzweil's *The Singularity Is Near*, and McDermott's critique of Kurzweil", p. 1163).

[3] See, for example, D. McDermott, "Level-headed", *Artificial Intelligence,* vol. 171, 2007; A. Sloman, "The well-designed young mathematician", *Artificial Intelligence*, vol. 172, 2008. Goertzel criticises the goal of human-level AI but is happier with AGI ("Human-level artificial general intelligence and the possibility of a technological singularity: A reaction to Ray Kurzweil's *The Singularity Is Near*, and McDermott's critique of Kurzweil"). Several of McDermott's and Sloman's objections to the notion of human-level AI seem to apply also to AGI.

[4] J.B. Pollack, "Mindless Intelligence", *IEEE Intelligent Systems*, vol. 21, no. 3, 2006, pp. 50–56. All references to Pollack are to this paper.

[5] P.J. Hayes and K.M Ford, "Turing Test Considered Harmful", IJCAI-95 *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, Montreal, Quebec, August 20–25, vol. 1, Morgan Kaufman, 1995, p. 974.

[6] B. Whitby, "The Turing Test: AI's Biggest Blind Alley?", in P. Millican and A. Clark eds, *The Legacy of Alan Turing, Vol. I, Machines and Thought*, Oxford Univ. Press, 1996, p. 58.

[7] R.M. French, "Subcognition and the Limits of the Turing Test", in P. Millican and A. Clark eds, 1996, p. 12; D. Michie, "Turing's Test and Conscious Thought", in P. Millican and A. Clark eds, 1996, p. 36; P.R. Cohen, "If not Turing's test, then what?", *AI Magazine*, vol. 26, no. 4 (April), 2005, p. 62; N. Block, "The Mind as the Software of the Brain", in E.E. Smith and D.N. Osherson eds, *An Invitation to Cognitive Science, Vol. 3, Thinking*, 2nd edition, MIT Press, 1995, p. 378.

[8] P.J. Hayes and K.M. Ford, 1995, pp. 972, 976.

[9] K.M. Ford and P.J. Hayes, "On Computational Wings: Rethinking the Goals of Artificial Intelligence", *Scientific American Presents*, vol. 9, no. 4, 1998, p. 79.

[10] A. Sloman, "The Mythical Turing Test", 26 May 2010, http://www.cs.bham.ac.uk/research/projects/cogaff/misc/turing-test.html. Whitby makes the same claim, saying 'Turing himself was always careful to refer to "the game". The suggestion that it might be some sort of test involves an important extension of Turing's claims' ("The Turing Test: AI's Biggest Blind Alley?", p. 54). Likewise Narayanan states that 'Turing did not originally intend his imitation game to be a test' ("The Intentional Stance and the Imitation Game", in P. Millican and A. Clark eds, 1996, p. 66). For criticism of these claims (and other objections to Turing in this and the companion volume), see B.J. Copeland and D. Proudfoot, 'The Legacy of Alan Turing', *Mind*, vol. 108, 1998.

[11] A. Sloman, "The Mythical Turing Test". Sloman says, '[F]ar from proposing a test to answer the question whether machines can think or whether machines are intelligent, [Turing] actually decides (rightly) that the question is absurd' (ibid.). However, what Turing states is 'absurd' is any attempt to answer this question by means of 'a statistical survey such as a Gallup poll' ("Computing Machinery and Intelligence", *Mind*, vol. 59, 1950, p. 433). Despite his notorious remark that the question 'Can machines think?' is 'too meaningless to deserve discussion' (ibid., p. 442), Turing discussed this question at length (Copeland points this out in B.J. Copeland ed., *The Essential Turing*, Oxford Univ. Press, 2004, pp. 476–477). He discussed it in his 1950 paper (where, after all, he said that the question 'Are there imaginable digital computers which would do well in the imitation game?' is a 'variant' of the question 'Can machines think?' (p. 442)), and in his radio broadcasts 'Can Digital Computers Think?' (A.M. Turing, 1951, "Can Digital Computers Think?", in B.J. Copeland ed., 2004) and "Can Automatic Calculating Machines Be Said To Think?" (A.M. Turing, R. Braithwaite, G. Jefferson, and M. Newman, 1952, in B.J. Copeland ed., 2004). In the first of these broadcasts, for example, Turing speaks of 'programming a machine to think' and 'the attempt to make a thinking machine', and he makes it clear that he is in favour of 'the theory that machines could be made to think' (pp. 485–486).

[12] A.M. Turing, 1950, pp. 452, 436, 446.

[13] A.M. Turing, et al., 1952, p. 495.

[14] A.M. Turing, 1948, "Intelligent Machinery", in B.J. Copeland ed., 2004, p. 431. For my analysis of Turing's approach to intelligence, see D. Proudfoot, "Rethinking Turing's test", forthcoming; D. Proudfoot and B.J. Copeland, "Artificial Intelligence", in E. Margolis, R. Samuels, and S. Stich eds, *Oxford Hand-*

book of Philosophy and Cognitive Science, Oxford Univ. Press, forthcoming; B.J. Copeland and D. Proudfoot, "Artificial Intelligence: History, Foundations, and Philosophical Issues", in P. Thagard ed., *Handbook of the Philosophy of Psychology and Cognitive Science*, Elsevier, 2006.

[15]On anthropomorphism in AI, see e.g. D. Proudfoot, "How Human Can They Get?", *Science*, vol. 248, 1999, p. 745 and "The Implications of an Externalist Theory of Rule-Following Behaviour for Robot Cognition", *Minds and Machines*, vol. 14, 2004, pp. 295, 302; B.J. Copeland and D. Proudfoot, 2006, pp. 445–446.

[16]A.M. Turing, 1950, pp. 456–457; "Intelligent Machinery: a Heretical Theory", c. 1951, in B.J. Copeland ed., 2004, p. 473.

[17]A.M. Turing, 1948, pp. 424–429.

[18]A.M. Turing, 1948, p. 412.

[19]V. Braitenberg, *Vehicles: Experiments in Synthetic Psychology*, MIT Press, 1984, pp. 83, 68, 19, 46, 83, 46, 41.

[20]D.W. Hogg, F. Martin, and M. Resnick, "Braitenberg Creatures", Epistemology and Learning Group Memo No. 13, MIT Media Laboratory, 1991, pp. 1–8.

[21]M. Yamamoto, "'SOZZY': A Hormone-Driven Autonomous Vacuum Cleaner", *Proceedings of the International Society for Optical Engineering*, vol. 2058, 1993, pp. 212–213.

[22]D.C. Dennett, "When HAL Kills, Who's to Blame? Computer Ethics", in D.G. Stork ed., *Hal's Legacy: 2001's Computer as Dream and Reality*, MIT Press, 1997, p. 358; "The Practical Requirements for Making a Conscious Robot", *Philosophical Transactions of the Royal Society of London*, vol. 349, 1994, pp. 140, 141.

[23]See, e.g., C. Breazeal, "Regulating Human–Robot Interaction using 'emotions', 'drives' and facial expressions", 1998; http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.56.6651.

[24]See, e.g., C. Breazeal and J. Velásquez, "Toward Teaching a Robot 'Infant' using Emotive Communication Acts", in *Proceedings of 1998 Simulation of Adaptive Behavior, workshop on Socially Situated Intelligence*; http://www.ai.mit.edu/projects/sociable/publications.html.

[25]L. Wittgenstein, *Philosophical Investigations*, 2nd edition, eds. G.E.M. Anscombe and R. Rhees, Blackwell, 1972.

[26]C. Breazeal, "Affective Interaction between Humans and Robots", in J. Kelemen and P. Sosik eds, *ECAL 2001, LNAI 2159*, Springer-Verlag, 2001, p. 585; C. Breazeal and B. Scassellati, "Challenges in Building Robots That Imitate People", in K. Dautenhahn and C. Nehaniv eds, *Imitation in Animals and Artifacts*, MIT Press, 2001, www.ai.mit.edu/projects/humanoid-robotics-group/publications.html; C. Breazeal, 2001, p. 584. Not all human smiles and frowns are genuine, but non-genuine ('Duchenne') smiles nevertheless have meaning, and the person smiling has a certain communicative intent.

[27]C. Breazeal and P. Fitzpatrick, "That Certain Look: social amplification of animate vision", *Proceedings of the AAAI Fall Symposium, Socially Intelligent Agents—The Human in the Loop*, 2000, http://people.csail.mit.edu/paulfitz/publications.html.

[28]D. McDermott, "Artificial Intelligence Meets Natural Stupidity", *SIGART Newsletter*, no. 57, April 1976, p. 4. Sloman makes the similar claim that referring to hypothesized parts of brains as 'perceiving' or 'deciding' is *circular* (A. Sloman, "The Design-Based Approach to the Study of Mind (in humans, other animals, and machines), Including the Study of Behaviour Involving Mental Processes", http://www.cs.bham.ac.uk/research/projects/cogaff/misc/design-based-approach.html, February 28, 2010). I am indebted to Aaron Sloman for emphasizing the relevance of McDermott's paper to the discussion of anthropomorphism in AI.

[29]D. McDermott, 1976, pp. 4, 5.

[30]Another example of plainly false ascriptions in AI is the use of personal pronouns ('he' and 'she') to refer to machines. A recent case is E.S. Brunette, R.C. Flemmer, and C.L. Flemmer, "A review of artificial intelligence", ICARA 2009 – *Proceedings of the Fourth International Conference on Autonomous Robots and Agents*, 2009, IEEE. S.P. Franklin goes so far as to use personal pronouns to refer, not only to robots and software agents, but also to processors and computer models (in *Artificial Minds*, Bradford Books, 1995).

[31]'Disease' is McDermott's term (1976, p. 5).

[32]D. McDermott, 1976, p. 4.

[33]A.M. Turing, 1948, p. 412.

[34]A.M. Turing, 1948, p. 431.

[35]The judges' inputs quoted here are from the transcripts of, respectively, the 2005, 2003, and 2008 competitions; http://www.loebner.net/Prizef/loebner-prize.html.

[36]B.J. Copeland, "The Turing Test", in J.H. Moor ed., *The Turing Test: The Elusive Standard of Artificial Intelligence*, Kluwer, 2003, p. 7; J.H. Moor, "The Status and Future of the Turing Test", in J.H. Moor ed., 2003, p. 204. The results of the 2000 contest are set out in "The Status and Future of the Turing Test", p. 205.

[37]http://www.loebner.net/Prizef/loebner-prize.html.

[38]Might an interrogator's tendency to anthropomorphize somehow be triggered only (or more strongly) in her interviews with the machine? This possibility points to the fact that the imitation game, however interpreted, cannot be used as a one-off test. Individual interrogators may have quirks and may make surprising decisions; to obtain a convincing result, the imitation game must be played several times (see B.J. Copeland ed., 2004, pp. 528–529). There is no reason to think that, over a series of games using different interrogators chosen randomly, the tendency to anthropomorphize will favour the machine against the human contestant.

[39]K.M. Ford and P.J. Hayes, 1998, p. 79.

[40]N. Block, "Psychologism and Behaviorism", *Philosophical Review*, vol. 90, no. 1, 1981, p. 10.

[41]See, e.g., M. Minsky, P. Singh, and A. Sloman, 2004.

[42]A.M. Turing, 1950, p. 460.

[43]For example, Turing—just as Pollack advocates—sought to understand how the human brain might emerge naturally; he said that '[t]he brain structure has to be one which can be achieved by the genetic embryological mechanism' and hoped that his work on neuron-like computation might clarify this process. (Letter from Turing to the biologist J.Z. Young. A copy is in the Modern Archive Centre, King's College, Cambridge (catalogue reference K1.78).)

[44]For example, in addition to his ground-breaking work on mechanized search in the Bombe, Turing anticipated computation by neural networks and was the first to use computer simulation to investigate the development of pattern in living things. On the Bombe, see B.J. Copeland ed., 2004, ch. 6 and pp. 353–355; B.J. Copeland and D. Proudfoot, 2006. On neural networks, see Turing, 1948; B.J. Copeland and D. Proudfoot, "On Alan Turing's Anticipation of Connectionism", *Synthese*, vol. 108, 1996 (reprinted in R. Chrisley ed., *Artificial Intelligence: Critical Concepts in Cognitive Science*, *Volume 2: Symbolic AI*, Routledge, 2000); B.J. Copeland and D. Proudfoot, "Alan Turing's Forgotten Ideas in Computer Science", *Scientific American*, vol. 280 (April), 1999. On simulation, see A.M. Turing, "The Chemical Basis of Morphogenesis", *Philosophical Transactions of the Royal Society of London*, *Series B*, vol. 237, 1952 (reprinted in B.J. Copeland ed., 2004; see also pp. 507–518); B.J. Copeland and D. Proudfoot, "Turing and the computer", in B.J. Copeland ed., *Alan Turing's Automatic Computing Engine*, Oxford Univ. Press, 2005.

[45]K.M. Ford and P.J. Hayes, 1998, p. 79.

[46]A.M. Turing, 1950, p. 435.

[47]Discussions of philosophical objections to the Turing test include: J.R. Searle, "Minds, Brains, and Programs", *Behavioral and Brain Sciences*, vol. 3, 1980; R.M. French, "The Turing Test: The First 50 Years", *Trends in Cognitive Science*, vol. 4, no. 3, 2000; D. Lenat, "Building a Machine Smart Enough to Pass the Turing Test: Could We, Should We, Will We?"; P.M. Churchland, "On the Nature of Intelligence: Turing, Church, von Neumann, and the Brain", in R. Epstein, G. Roberts, and G. Beber eds, 2008; D. Proudfoot, "Wittgenstein's Anticipation of the Chinese Room", in J. Preston and M. Bishop eds, 2002 and "Review of James Moor (Ed.), *The Turing Test: The Elusive Standard of Artificial Intelligence*", *Philosophical Psychology*, vol. 19, no. 2, 2005; B.J. Copeland and D. Proudfoot, 2006 and "The Turing Test: A Philosophical and Historical Guide," in R. Epstein, G. Roberts, and G. Beber eds, 2008.

[48]See, e.g., N. Block, "The Mind as the Software of the Brain", p. 379. Block himself points out the discrepancy between the early Loebner contests and Turing's test, yet describes Loebner's test as a 'reasonable facsimile' of Turing's test.

[49] N. Block, "The Mind as the Software of the Brain", p. 383.

[50] The objection in this paragraph is based on B.J. Copeland, 2003, pp. 14–15.

[51] N. Block, "The Mind as the Software of the Brain", p. 381.

[52] On this point, see e.g. J.H. Moor, "Turing Test", in S.C. Shapiro ed., *Encyclopedia of Artificial Intelligence*, John Wiley, 1987, p. 1126.

[53] A.M. Turing, et al., 1952, p. 494; B.J. Copeland first cites this quotation, to underline the flaw in the orthodox interpretation of Turing's test, in his 2003, p. 6. Many critics misconstrue the test by treating it as an operational definition; a recent example is in Stevan Harnad's running commentary on "Computing Machinery and Intelligence", in R. Epstein, G. Roberts, and G. Beber eds, 2008, p. 378.

[54] A.M. Turing, et al., 1952, pp. 503–504. See B.J. Copeland, 2003, p. 15.

[55] The recent philosophical literature on consciousness, zombie thought experiments, and machine consciousness is extensive. See, for example, D. Chalmers, "Facing up to the problem of consciousness", *Journal of Consciousness Studies*, vol. 2, no. 3, 1995; N. Block, "On a confusion about the function of consciousness", *Behavioral and Brain Sciences*, vol. 18, no. 2, 1995; J. McCarthy, "Todd Moody's zombies", *Journal of Consciousness Studies*, vol. 2, no. 4, 1995; D. Dennett, 'The unimagined preposterousness of zombies', *Journal of Consciousness Studies,* vol. 2, no. 4, 1995; O. Flanagan and T. Polger, "Zombies and the function of consciousness', *Journal of Consciousness Studies,* vol. 2, no. 4, 1995; K. Balog, "Conceivability, Possibility, and the Mind-Body Problem", *Philosophical Review*, vol. 108, no. 4, 1999; R. Kirk, "The inconceivability of zombies", *Philosophical Studies*, vol. 139, 2008; A. Sloman, "Phenomenal and Access Consciousness and the 'Hard' Problem: A View from the Designer Stance', *International Journal of Machine Consciousness*, vol. 2, no. 1, 2010.

[56] J.R. Searle, "Twenty-One Years in the Chinese Room", in J. Preston and M. Bishop eds, *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*, Oxford Univ. Press, 2002, pp. 52, 64.

[57] 'Intelligence amplifier' is Lenat's term. D.B. Lenat, "Building a Machine Smart Enough to Pass the Turing Test: Could We, Should We, Will We?", in R. Epstein, G. Roberts, and G. Beber eds, *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, Springer, 2008, p. 281.

[58] See, for example, P.R. Cohen, "If not Turing's test, then what?".

[59] A.M. Turing, et al., 1952, p. 495. This prediction has generally been ignored in the literature on the Turing test, in favour of Turing's prediction in his 1950 paper, where he states 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 per cent. chance of making the right identification after five minutes of questioning' (A.M. Turing, 1950, p. 442). Emphasis on the much weaker 1950 prediction tends to distort the nature of Turing's claims concerning his test.

[60] See, for example, P. Hayes and K. Ford, 1995; J.B. Pollack, 2006; D.B. Lenat, "The voice of the turtle: whatever happened to AI?".

[61] For example, A. Sloman, "The Mythical Turing Test".

[62] A.M. Turing, c. 1951, p. 484; 1950, pp. 434, 435.

[63] A. Sloman, "The Mythical Turing Test".

[64] A. Sloman, "The well-designed young mathematician", *Artificial Intelligence*, vol. 172, 2008; "The Design-Based Approach to the Study of Mind (in humans, other animals, and machines), Including the Study of Behaviour Involving Mental Processes"; "An Alternative to Working on Machine Consciousness", http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#910, May 14, 2010; and "Requirements for Artificial Companions: It's harder than you think", http://www.cs.bham.ac.uk/research/projects/cogaff/07.html#711, April 25, 2010.

[65] A.M. Turing, et al., 1952, p. 500.

[66] R.A. Brooks, "Intelligence without Reason", in L. Steels and R.A. Brooks eds, *The Artificial Life Route to Artificial Intelligence*, Lawrence Erlbaum, 1995, p. 57.

[67] R.A. Brooks, *Cambrian Intelligence: The Early History of the New AI*, Bradford Books, 1999, p. x.

[68] D.C. Dennett, "Intentional Systems", in his *Brainstorms: Philosophical Essays on Mind and Psychology*, Harvester Press, 1979, pp. 3–4.

[69] D.C. Dennett, 1979, p. 8.

[70] Personal communication from Rodney Brooks.

[71] A classic early text is S.E. Guthrie, *Faces in the Clouds: A New Theory of Religion*, Oxford Univ. Press, 1993.

[72] D.L. Waltz, "Evolution, Sociobiology, and the Future of Artificial Intelligence", *IEEE Intelligent Systems*, vol. 21, no. 3, 2006, p. 68.