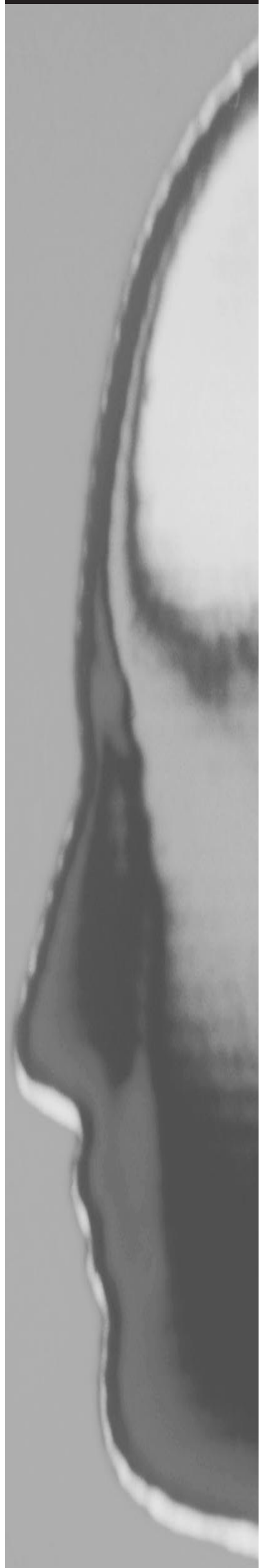
mode. *Proc. Natl. Acad. Sci. U. S. A.* (in press)
20 Rugg, M.D. *et al.* (1998) Neural correlates of memory retrieval during recognition memory and cued recall. *NeuroImage* 8, 262–273
21 Rugg, M.D *et al.* (1996) Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval. *Brain* 119, 2073–2083
22 Fletcher, P.C. *et al.* (1997) The functional neuroanatomy of episodic memory. *Trends Neurosci.* 20, 213–218
23 Rugg, M.D. *et al.* (1999) The role of the prefrontal cortex in recognition memory and memory for source: an fMRI study. *NeuroImage* 10, 520–529
24 Wagner, A.D. *et al.* (1998) Prefrontal cortex and recognition memory: functional MRI evidence for context-dependent retrieval processes. *Brain* 121, 1985–2002
25 Kutas, M. and Dale, A. (1997) Electrical and magnetic readings of mental functions. In *Cognitive Neuroscience* (Rugg, M.D., ed.), pp. 169–242, Psychology Press
26 Wilding, E.L. and Rugg, M.D. (1996) An event-related potential study of recognition memory with and without retrieval of source. *Brain* 119, 889–905
27 Donaldson, D.I. and Rugg, M.D. (1998) Recognition memory for new associations: electrophysiological evidence for the role of recollection. *Neuropsychologia* 36, 377–395
28 Düzel, E. *et al.* (1997) Event-related brain potential correlates of two states of conscious awareness in memory. *Proc. Natl. Acad. Sci. U. S. A.* 94, 5973–5978
29 Rugg, M.D. *et al.* (1998) Dissociation of the neural correlates of implicit and explicit memory. *Nature* 392, 595–598
30 Paller, K.A. and Kutas, M. (1992) Brain potentials during retrieval provide neurophysiological support for the distinction between conscious recollection and priming. *J. Cogn. Neurosci.* 4, 375–391
31 Smith, M.E. (1993) Neurophysiological manifestations of recollective experience during recognition memory judgments. *J. Cogn. Neurosci.* 5, 1–13
32 Rugg, M.D. *et al.* (1995) Event-related potentials and the recollection of low and high frequency words. *Neuropsychologia* 33, 471–484
33 Schacter, D.L. *et al.* (1996) Conscious recollection and the human hippocampal formation: evidence from positron emission tomography. *Proc. Natl. Acad. Sci. U. S. A.* 93, 321–325
34 Buckner, R.L. *et al.* (1998) Functional-anatomic study of episodic retrieval using fMRI: 1. Retrieval effort versus retrieval success. *NeuroImage* 7, 151–162
35 Buckner, R.L. *et al.* (1998) Functional-anatomic study of episodic retrieval using fMRI: 2. Selective averaging of event-related fMRI trials to test the success hypothesis. *NeuroImage* 7, 163–175
36 Nolde, S.F. *et al.* (1998) Left prefrontal activation during episodic remembering: an event-related fMRI study. *NeuroReport* 9, 3509–3514
37 Henson, R.N.A. *et al.* (1999) Right prefrontal cortex and episodic memory retrieval: a functional MRI test of the monitoring hypothesis. *Brain* 122, 1367–1381
38 Nolde, S.F. *et al.* (1998) The role of prefrontal cortex during tests of episodic memory. *Trends Cognit. Sci.* 2, 399–406
39 Johnson, M.K. *et al.* (1996) Electrophysiological brain activity and source memory. *NeuroReport* 7, 2929–2932
40 Senkfor, A. and Van Petten, C. (1998) Who said what? An event-related potential investigation of source and item memory. *J. Exp. Psychol: Learn. Mem. Cognit.* 24, 1005–1025
41 Ranganath, C. and Paller, K.A. (1999) Frontal brain potentials during recognition are modulated by requirements to retrieve perceptual detail. *Neuron* 22, 605–613
42 Ranganath, C. and Paller, K.A. Neural correlates of memory retrieval and evaluation. *Cognit. Brain Res.* (in press)
43 Burgess, P.W. and Shallice, T. (1996) Confabulation and the control of recollection. *Memory* 4, 359–411
44 Koriat, A. and Goldsmith, M. (1996) Monitoring and control processes in the strategic regulation of memory accuracy. *Psychol. Rev.* 103, 490–517

The Turing Test: the first 50 years

Robert M. French



The Turing Test, originally proposed as a simple operational definition of intelligence, has now been with us for exactly half a century. It is safe to say that no other single article in computer science, and few other articles in science in general, have generated so much discussion. The present article chronicles the comments and controversy surrounding Turing's classic article from its publication to the present. The changing perception of the Turing Test over the last 50 years has paralleled the changing attitudes in the scientific community towards artificial intelligence: from the unbridled optimism of 1960s to the current realization of the immense difficulties that still lie ahead. I conclude with the prediction that the Turing Test will remain important, not only as a landmark in the history of the development of intelligent machines, but also with real relevance to future generations of people living in a world in which the cognitive capacities of machines will be vastly greater than they are now.

The invention and development of the computer will undoubtedly rank as one of the twentieth century's most far-reaching achievements that will ultimately rival or even surpass that of the printing press. At the very heart of that development were three seminal contributions by Alan Mathison Turing. The first was theoretical in nature: in order to solve a major outstanding problem in mathematics, he developed a simple mathematical model for a universal computing machine (today referred to as a Turing Machine). The second was practical: he was actively involved in building

R.M. French is at Quantitative Psychology and Cognitive Science, Department of Psychology, University of Liege, Belgium.

tel: +32 4 221 05 42
fax: +32 4 366 28 59
e-mail: rfrench@ulg.ac.be

 PII: S1364-6613(00)01453-4

one of the very first electronic, programmable, digital computers. Finally, his third contribution was philosophical: he provided an elegant operational definition of thinking that, in many ways, set the entire field of artificial intelligence (AI) in motion. In this article, I will focus only on this final contribution, the Imitation Game, proposed in his classic article in *Mind* in 1950 (Ref. 1).

The Imitation Game

Before reviewing the various comments on Turing's article, I will briefly describe what Turing called the Imitation Game (called the Turing Test today). He began by describing a parlour game. Imagine, he says, that a man and a woman are in two separate rooms and communicate with an interrogator only by means of a teletype – the 1950s equivalent of today's electronic 'chat'. The interrogator must correctly identify the man and the woman and, in order to do so, he may ask *any question* capable of being transmitted by teletype. The man tries to convince the interrogator that he is the woman, while the woman tries to communicate her real identity. At some point during the game the man is replaced by a machine. If the interrogator remains incapable of distinguishing the machine from the woman, the machine will be said to have passed the Test and we will say that the machine is intelligent. (We see here why Turing chose communication by teletype – namely, so that the lack of physical features which Turing felt were not essential for cognition, would not count against the machine.)

The Turing Test, as it rapidly came to be described in the literature and as it is generally described today, replaces the woman with a person of either gender. It is also frequently described in terms of a single room containing either a person or a machine and the interrogator must determine whether he is communicating with a real person or a machine. These variations do, indeed, differ somewhat from Turing's original formulation of his imitation game. In the original test the man playing against the woman, as well as the computer that replaces him, are both 'playing out of character' (i.e. they are both relying on a theory of what women are like). The modern description of the Test simply pits a machine in one room against a person in another. It is generally agreed that this variation does not change the essence of Turing's operational definition of intelligence, although it almost certainly makes the Test more difficult for the machine to pass[2]. One significant point about the Turing Test that is often misunderstood is that *failing it proves nothing*. Many people would undoubtedly fail it if they were put in the role of the computer, but this certainly does not prove that they are not intelligent! The Turing Test was intended only to provide a sufficient condition for intelligence.

To reiterate, Turing's central claim is that there would be no reason to deny intelligence to a machine that could flawlessly imitate a human's unrestricted conversation. Turing's article has unquestionably generated more commentary and controversy than any other article in the field of artificial intelligence, and few papers in *any* field have created such an enduring reaction. Only 13 years after Turing's article appeared, Anderson had already counted over 1000 published papers on whether machines could think[3]. For half a century, references to the Turing Test have appeared regularly in artificial intelligence journals, philosophy journals, technical treatises, novels and the popular press. Type 'Turing Test' into any Web browser and you will have thousands of hits. Perhaps the reason for this high profile is partly our drive to build mechanical devices that imitate what humans do. However, there seems to be a particular fascination with mechanizing our ability to think. The idea of mechanized thinking goes back at least to the 17th century with the *Characteristica Universalis* of Leibnitz and extends through the work of La Mettrie to the writings of Hobbes, Pascal, Boole, Babbage and others. The advent of the computer meant that, for the first time, there was a realistic chance of actually achieving the goal of mechanized thought. It is this on-going fascination with mechanized thought that has kept the Turing Test in the forefront of discussions about AI for the past half century.

The value and the validity of the Turing Test

Opinions on the validity and, especially, the value of the Turing Test as a real guide for research vary widely. Some authors have maintained that it was precisely the operational definition of intelligence that was needed to sidestep the philosophical quagmire of attempting to define rigorously what was meant by 'thinking' and 'intelligence' (see Refs 4–7). At the other extreme, there are authors who believe that the Turing Test is, at best, *passé*[8] and, at worst, a real impediment to progress in the field of artificial intelligence[9,10]. Hayes and Ford[9] claim that abandoning the Turing Test as an ultimate goal is 'almost a requirement for any rational research program which declares itself interested in any particular part of cognition or mental activity'. Their not unreasonable view is that research time is better spent developing what they call 'a general science of cognition' that would focus on more restricted areas of cognition, such as analogy-making, vision, generalization and categorization abilities. They add, 'From a practical perspective, why would anyone want to build machines that could pass the Turing Test? Human cognition, even high-quality human cognition, is not in short supply. What extra functionality would such a machine provide?'

Taking a historical view, Whitby[8] describes four phases in the evolving interest in the Turing Test:

1950–1966: a source of inspiration for all concerned with AI
1966–1973: a distraction from some more promising avenues of AI research
1973–1990: by now a source of distraction mainly to philosophers, rather than AI workers
1990 onwards: consigned to history

I am not sure exactly what Whitby means by 'consigned to history', but if he means 'forgotten', I personally doubt that this will be the case. I believe that in 300 years' time people will still be discussing the arguments raised by Turing in his paper. It could even be argued that the Turing Test will take on an even greater significance several centuries in the future when it might serve as a moral yardstick in a world where machines will move around much as we do, will use natural language, and will interact with humans in ways that are almost inconceivable today. In short, one of the questions facing future generations may well be, 'To what extent do machines have to act like humans before it becomes immoral to damage or destroy them?' And the very essence of the Turing Test is our judgment of how well machines act like humans.

Shift in perception of the Turing Test

It is easy to forget just how high the optimism once ran for the rapid achievement of artificial intelligence. In 1958, a mere eight years after the appearance of Turing's article, when computers were still in their infancy and even high-level programming languages had only just been invented, Simon and Newell[11], two of the founders of the field of artificial intelligence, wrote, '…there are now in the world machines that think, that learn and that create. Moreover, their ability to do these things is going to increase rapidly until – in a visible future – the range of problems they can handle will be co-extensive with the range to which the human mind has been applied'. Minsky, head of the MIT AI Laboratory, wrote in 1967, 'Within a generation the problem of creating "artificial intelligence" will be substantially solved'[12].

During this period of initial optimism, most of the authors writing about the Turing Test shared with the founders of AI the belief that a machine could actually be built that would be able to pass the Test in the foreseeable future. The debate, therefore, centered almost exclusively around Turing's operational definition of disembodied intelligence – namely, did passing the Turing Test constitute a sufficient condition for intelligence or not? As it gradually dawned on AI researchers just how difficult it was going to be to produce artificial intelligence, the focus of the debate on the Turing Test shifted. By 1982, Minsky's position regarding artificial intelligence had undergone a radical shift from one of unbounded optimism 15 years earlier to a far more sober assessment of the situation: 'The AI problem is one of the hardest ever undertaken by science'[13]. The perception of the Turing Test underwent a parallel shift. At least in part because of the great difficulties being experienced by AI, there was a growing realization of just how hard it would be for a machine to pass the Turing Test. Thus, instead of discussing whether or not a machine that had passed the Turing Test was really intelligent, the discussion shifted to whether it would even be possible for any machine to pass such a test.

Turing's comments of the Imitation Game

The first set of comments on the Imitation Game were voiced by Turing himself. I will briefly consider three of the most important. The first is the 'mathematical objection' based on Gödel's Theorem[14], which proves that there are truths that can be expressed in any sufficiently powerful formal system, that we humans can recognize as truths, but that cannot be proved within that system (i.e. a computer could not recognize them as truths, because it would have to prove them in order to recognize them as such). This then would provide a limitation for the computer, but not for humans. This argument was taken up and developed in detail a decade later in a well-known paper by Lucas[15]. Turing replied that humans are not perfect formal systems and, indeed, may also have a limit to the truths they can recognize.

The second objection is the 'argument from consciousness' or the 'problem of other minds'. The only way to know if anything is thinking is to *be* that thing, so we cannot know if anything else really thinks. Turing's reply was that if we adopt this solipsistic position for a machine, we must also adopt it for other people, and few people would be willing to do that.

Finally, the most important objection that Turing raised was what he calls 'Lady Lovelace's objection'. The name of this objection comes from a remark by Lady Lovelace concerning Charles Babbage's 'Analytical Engine', and was paraphrased by Turing as 'the machine can only do what we know how to order it to do'[1]. In other words, machines, unlike humans, are incapable of creative acts because they are only following the programmer's instructions. His answer is, in essence, that although we may program the basics, a computer, especially a computer capable of autonomous learning (see section 7 of Turing's article[1], 'Learning Machines'), may well do things that could not have been anticipated by its programmer.

A brief chronicle of early comments on the Turing Test

Mays wrote one of the earliest replies to Turing, questioning the fact that a machine designed to perform logical operations could actually capture 'our intuitive, often vague and imprecise, thought processes'[16]. Importantly, this paper contained a first reference to a problem that would take center stage in the artificial intelligence community three decades later: 'Defenders of the computing machine analogy seem implicitly to assume that the whole of intelligence and thought can be built up summatively from the warp and woof of atomic propositions'[16]. This objection, in modified form, would re-appear in the 1980s as one of the fundamental criticisms of traditional artificial intelligence.

In Scriven's first article[17], he arrived at the conclusion that merely imitating human behaviour was certainly not enough for consciousness. Then, a decade later, apparently seduced by the claims of the new AI movement, he changed his mind completely, saying, 'I now believe that it is possible so to construct a supercomputer as to make it wholly unreasonable to deny that it had feelings'[3].

Gunderson clearly believed that passing the Turing Test would *not* necessarily be a proof of real machine intelligence[18,19]. Gunderson's objection was that the Test is based on a behaviouristic construal of thinking, which he felt must be rejected. He suggested that thinking is a very broad concept and that a machine passing the Imitation Game is merely exhibiting a single skill (which we might dub 'imitation-game playing'), rather than the all-purpose abilities defined by thinking. Further, he claimed that playing the Imitation Game successfully could well be achieved in ways other than by thinking, without saying precisely what these other ways might be. Stevenson, writing a decade later when the difficulties with AI research had become clearer, criticized Gunderson's single-skill objection, insisting that to play the game would require 'a very large range of other properties'[20].

In articles written in the early 1970s we see the first shift away from the acceptance that it might be possible for a machine to pass the Turing Test. Even though Purtill's basic objection[21] to the Turing Test was essentially the Lady Lovelace objection (i.e. that any output is determined by what the programmer explicitly put into the machine, and therefore can be explained in this manner), he concluded his paper in a particularly profound manner, thus: '…if a

Box 1. The Human Subcognitive Profile

Let us designate as 'subcognitive' any question capable of providing a window on low-level (i.e. unconscious) cognitive or physical structure. By 'low-level cognitive structure', we mean the subconscious associative network in human minds that consists of highly overlapping, activatable representations of experience (Refs a–c).

The Turing Test interrogator prepares a long list of subcognitive questions (the Subcognitive Question List) and produces a profile of answers to these questions from a representative sample of the general population; for example:

'On a scale of 0 (completely implausible) to 10 (completely plausible):

- Rate *Flugblogs* as the name of a start-up computer company
- Rate *Flugblogs* as the name of air-filled bags that you tie on your feet and use to cross swamps
- Rate *banana splits* as *medicine*
- Rate *purses* as *weapons*.

Other questions might include:

- Someone calls you a *trubhead*. Is this a compliment or an insult?
- Which word do you find prettier: *blutch* or *farfaletta?*
- Does holding a gulp of Coca-Cola in your mouth feel more like having pins and needles in your foot or having cold water poured on your head?

We can imagine many more questions that would be designed to test not only for subcognitive associations, but for internal physical structure. These would include questions whose answers would be, for example, a product of the spacing of the candidate's eyes, would involve visual aftereffects, would be the results of little self-experiments involving tactile sensations on their bodies or sensations after running in place, and so on.

The interrogator would then come to the Turing Test and asks both candidates the questions on her Subcognitive Question List. The candidate most closely matching the average answer profile from the human population will be the human.

The essential idea here is that the 'symbols-in/symbols-out' level specified in Turing's original article (Harnad's level T2; see Ref. d and Box 2) can indirectly, but reliably, probe much deeper subcognitive and even physical levels of the two candidates. The clear boundary between the symbolic level and the physical level that Turing had hoped to achieve with his teletype link to the candidates all but disappears (Refs b,e). People's answers to subcognitive questions are produced by our lifetime of experiencing the world with our human bodies, our human behaviors (whether culturally or genetically engendered), our human desires and needs, etc. (See Harnad for a discussion of the closely related 'symbol grounding problem', Ref. f.) It does not matter if we are confronted with made-up words or conceptual juxtapositions that never normally occur (e.g. *banana splits* and *medicine*), we can still respond and, moreover, these responses will show statistical regularities over the population. Thus, by surveying the population at large with an extensive set of these questions, we draw up a Human Subcognitive Profile for the population. It is precisely this profile that could not be reproduced by a machine that had not experienced the world as the members of the sampled human population had. The Subcognitive Question List that was used to produce the Human Subcognitive Profile gives the interrogator a tool for eliminating machines from a Turing test in which humans are also participating.

References

a French, R.M. (1988) Subcognitive probing: hard questions for the Turing Test. In *Proc. Tenth Annu. Cognit. Sci. Soc. Conf.*, pp. 361–367, Erlbaum

b French, R.M. (1990) Subcognition and the limits of the Turing Test. *Mind* 99, 53–65

c French, R.M. (1996) The Inverted Turing Test: how a simple (mindless) program could pass it. *Psycoloquy* 7 (39), turing-test.6.french

d Harnad, S. (1994) Levels of functional equivalence in reverse bioengineering: the Darwinian Turing Test for artificial life. *Artif. Life* 1, 293–301

e Davidson, D. (1990) Turing's test. In *Modelling the Mind* (Said, K.A. *et al.*, eds), pp. 1–11, Oxford University Press

f Harnad, S. (1990) The symbol grounding problem. *Physica D* 42, 335–346

computer could play the complete, "any question" imitation game it might indeed cause us to consider that perhaps that computer was capable of thought. But that any computer might be able to play such a game in the foreseeable future is so immensely improbable as to make the whole question academic'. Sampson replied that low-level determinism (i.e. the program and its inputs) does not imply predictable high-level behaviour[22]. Two years later, Millar presented the first explicit discussion of the Turing Test's anthropocentrism: 'Turing's test forces us to ascribe typical human objectives and human cultural background to the machine, but if we are to be serious in contemplating the use of such a term [intelligence] we should be open-minded enough to allow computing machinery or Martians to display their intelligence by means of behaviour which is well-adapted for achieving their own specific aims'[23].

Moor agreed that passing the test would constitute a sufficient proof of intelligence[24]. He viewed the Test as 'a potential source of good inductive evidence for the hypothesis that machines can think', rather than as a purely operational definition of intelligence. However, he suggested that it is of little value in guiding real research on artificial intelligence. Stalker replied that an explanation of how a computer passes the Turing Test would require an appeal to mental, not purely mechanistic notions[25]. Moor then countered that these two explanations are not necessarily competitors[26].

Comments from the 1980s

Numerous papers on the Turing Test appeared at the beginning of the 1980s, among them one by Hofstadter[27]. This paper covers a wide range of issues and includes a particularly interesting discussion of the ways in which a computer simulation of a hurricane differs or does not differ from a real hurricane. (For a further discussion of this point, see Ref. 28.) The two most often cited papers from this period were by Block[29] and Searle[30]. Instead of following up the lines of inquiry opened by Purtill[21] and Millar[23], these authors continued the standard line of attack on the Turing Test, arguing that even if a machine passed the Turing Test, it still might not be intelligent. The explicit assumption was, in both cases, that it was, in principle, possible for machines to pass the Test.

Block claimed that the Test is testing merely for behaviour, not the underlying mechanisms of intelligence[29]. He suggested that a mindless machine could pass the Turing Test in the following way: the Test will be defined to last an hour; the machine will then memorize *all possible conversational exchanges that could occur during an hour*. Thus, wherever the questions of the interrogator lead, the machine will be ready with a perfect conversation. But for a mere hour's worth of conversation such a machine would have to store at least 10^{1500} 20-word strings, which is far, far greater than the number of particles in the universe. Block drops all pretence that he is talking about real computers in his response to this objection: 'My argument requires only that the machine be *logically* possible, not that it be feasible or even nomologically possible'. Unfortunately, Block is no longer talking about the Turing Test because, clearly, Turing was talking about real computers (cf. sections 3 and 4 of Turing's article). In addition, a real interrogator might throw in questions with invented words in them like, 'Does the word *splugpud* sound very pretty to you?' A perfectly legitimate question, but impossible for the Block machine to answer. Combinatorial explosion brings the walls down around Block's argument.

Searle replaced the Turing Test with his now-famous 'Chinese Room' thought experiment[30]. Instead of the Imitation Game we are asked to imagine a closed room in which there is an English-speaker who knows not a word of Chinese. A native Chinese person writes a question in Chinese on a piece of paper and sends it into the room. The room is full of symbolic rules specifying inputs and outputs. The English-speaker then matches the symbols in the question with symbols in the rule-base. This does not have to be a direct table matching of the string of symbols in the question with symbols in the rule base, but can include any type of look-up program, regardless of its structural complexity. The English-speaker is blindly led through the maze of rules to a string of symbols that constitutes an answer to the question. He copies this answer on a piece of paper and sends it out of the room. The Chinese person on the outside of the room would see a perfect response, even though the English-speaker understood no Chinese whatsoever. The Chinese person would therefore believe that the person inside the room understands Chinese. Many replies have been made to this argument[31] and I will not include them here. One simple refutation would be to ask how the room could possibly contain answers to questions that contained caricaturally distorted characters. So, for example, assume the last character in a question had been distorted in a very phallic manner (but the character is still clearly recognizable to a native Chinese person). The question sent into the room is: 'Would the last character in this question be likely to embarrass a very shy young woman?' Now, to answer this question, all possible inputs, including all possible *distortions* of those inputs, would have to be contained in the rules in the room. Combinatorial explosion, once again, brings down this line of argument.

Could any machine ever pass the Turing Test?

In the mid-1980s, Dennett emphasized the sheer difficulty of a machine's passing the Turing Test[32]. He accepted it as a sufficient condition for intelligence, but wrote that, 'A failure to think imaginatively about the test actually proposed by Turing has led many to underestimate its severity…' He suggests that the Turing Test, when we think of just how hard it would be to pass, also shows why AI has turned out to be so hard.

As the 1980s ended, a new type of discussion about the Turing Test appeared, one that reflected not only the difficulties of traditional, symbolic AI but also the surge of interest in sub-symbolic AI fuelled by the ideas of connectionism[33,34]. These new ideas were the basis of work by French[35,36] that sought to show, by means of a technique based on 'subcognitive' questions (see Box 1), that 'only a computer that had acquired adult human intelligence by experiencing the world as we have could pass the Turing Test'[36]. Further, he argued that any attempt to fix the Turing Test 'so that it could test for intelligence in general and not just human intelligence is doomed to failure because of the completely interwoven and interdependent nature of the human physical, subcognitive, and cognitive levels'[36]. French also emphasized the fact that the Turing Test, when rigorously administered, probes deep levels of the associative concept networks of the candidates and that these 'networks are the product of a lifetime of interaction with the world which *necessarily involves* human sense organs, their location on the body, their sensitivity to various stimuli, etc'[36]. A similar conclusion was reached by Davidson, who wrote, 'Turing wanted his Test to draw "a fairly sharp line between the physical and the intellectual capacities of man." There is no such line'[37].

In the past decade, Harnad has been one of the most prolific writers on the Turing Test[38–42]. Most importantly he has proposed a 'Total Turing Test' (TTT) in which the screen provided by the teletype link between the candidates and the interrogator is removed[38]. This is an explicit recognition of the importance of bodies in an entity's interaction with the environment. The heart of Harnad's argument is that mental semantics must be 'grounded', in other words, the meanings of internal symbols must derive, at least partly, from interactions with the external environment[43]. Shanon also recognized the necessity of an interaction with the environment[44]. However, Hauser argued that the switch from the normal Turing Test to the TTT is unwarranted[45]. In later papers, Harnad extended this notion by defining a hierarchy of Turing Tests (see Box 2) of which the second (T2: the symbols-in/symbols-out Turing Test) corresponds to the standard Turing Test. T3 (the Total Turing Test) is the Robotic Turing Test in which the interrogator directly, visually, tactically, addresses the two candidates – the teletype 'screening' mechanism is eliminated. But we might still be able to detect some internal differences, even if the machine passed T3. Therefore, Harnad proposes T4: Internal Microfunctional Indistinguishability. And finally, T5: Grand Unified Theories of Everything, where the two candidates would be microfunctionally equivalent by every test relevant to a neurologist, neurophysiologist, and neurobiophysicist (for example, both fully obey the Hodgkin–Huxley equations governing neuronal firing) but would nonetheless be distinguishable to a physical chemist.

Harnad clearly recognizes the extreme difficulty of achieving even T2 and stresses the impossibility of implementing disembodied cognition. Schweizer wishes to improve the Robotic Turing Test (T3) by proposing a Truly Total Turing Test in which a long-term temporal dimension is added to

Box 2. The Turing Test hierarchy

Stevan Harnad has proposed a five-level Turing Test (TT) hierarchy (Refs a–c). This hierarchy attempts to encompass various levels of difficulty in playing an Imitation Game. The levels are t1, T2, T3, T4, and T5. The Harnad hierarchy works as follows:

Level t1

The 'toy-model' level. These are models ('toys', hence the lower case 't') that only handle a fragment of our cognitive capacity. So, for example, Colby's program designed to imitate a paranoid schizophrenic would fall into this category, because 'the TT is predicated on total functional indistinguishability, and toys are most decidedly distinguishable from the real thing.'

Harnad designates this level as 't1', essentially the level of current AI research, and adds that 'research has not even entered the TT hierarchy yet'.

Level T2

This is the level described in Turing's original article. Harnad refers to it as the 'pen-pal version' of the Turing Test, because all exchanges are guaranteed by the teletype link to occur in a symbols-in/symbols-out manner. Thus, T2 calls for a system that is indistinguishable from us in its symbolic (i.e. linguistic) capacities. This is also the level for which Searle's Chinese Room experiment is written. One central question is to what extent questions at this level can be used successfully, but indirectly, to probe the deep levels of cognitive, or even physical structure of the candidates.

Level T3: The 'Total Turing Test' (or the robotic Turing Test)

At this level the teletype 'screen' is removed. T3 calls for a system that is not only indistinguishable from us in its symbolic capacities, but it further requires indistinguishability in all of our robotic capacities: in other words, total indistinguishability in external (i.e. behavioral) function. At this level, physical appearance and directly observable behaviour matter.

Level T4: 'Microfunctional Indistinguishability'

This level would call for internal indistinguishability, right down to the last neuron and neurotransmitter. These could be synthetic neurons, of course, but they would have to be functionally indistinguishable from real ones.

Level T5: 'Grand Unified Theories of Everything (GUTE)'

At this level the candidates are 'empirically identical in kind, right down to the last electron', but there remain unobservable-in-principle differences at the level of their designers' GUTEs.

Harnad feels that T3 is the right level for true cognitive modeling. He writes, 'My own guess is that if ungrounded T2 systems are underdetermined and hence open to overinterpretation, T4 systems are overdetermined and hence include physical and functional properties that may be irrelevant to cognition. I think T3 is just the right empirical filter for mind-modeling.'

References

a Harnad, S. (1991) Other bodies, other minds: a machine incarnation of an old philosophical problem. *Minds and Machines* 1, 43–54
b Harnad, S. (1994) Levels of functional equivalence in reverse bioengineering: the Darwinian Turing Test for Artificial Life. *Artif. Life* 1, 293–301
c Harnad, S. Turing on reverse-engineering the mind. *J. Logic Lang. Inf.* (in press)

the Test[46]. He wants the historical record of our achievements (in inventing chess, in developing languages, etc.) also to match those of the machine.

One important question is: to what extent is the level specified by Turing in 1950 (i.e. Harnad's T2, symbols-in/symbols-out) sufficient to probe adequately the deeper subcognitive and even physical levels of the candidates? If we ask enough carefully worded questions (Box 1) even low-level physical differences in the human and machine candidates can be revealed. Questions such as, 'Rate on a scale of 1 to 10 how much keeping a gulp of Coca-Cola in your mouth feels like having pins-and-needles in your feet', indirectly test for physical attributes and past experiences; in this case, the presence of a mouth and limbs that fall asleep from time to time and the experience of having held a soft drink in one's mouth[47]. And while it might be possible for the computer to guess correctly on one or two questions of this sort, it would have no way of achieving the same overall profile of answers that humans will effortlessly produce. The machine can guess (or lie), to be sure, but it must guess (or lie) *convincingly* and, not just once or twice, but over and over again. In this case, guessing convincingly and systematically would mean that the machine's answer profile for these questions would be very similar overall to the human answer profile in the possession of the interrogator. But how could the machine be able to achieve this for a broad range of questions of this type if it had not experienced the world as we had?

Many of these objections concerning the difficulty of making an actual machine that could pass the Turing Test are also voiced by Crockett in his discussion of the relationship of the Turing Test to the famous frame problem in AI (i.e. the problem of determining exactly what information must remain unchanged at a representational level within a system after the system has performed some action that affects its environment)[48]. In essence, Crockett claims that passing the Turing Test is essentially equivalent to solving the frame problem (see also Ref. 49). Crockett arrives at essentially the same conclusion as French: 'I think it is unlikely that a computer will pass the test…because I am particularly impressed with the test's difficulty [which is] more difficult and anthropocentric than even Turing fully appreciated'[48].

Mitchie introduced the notion of 'superarticulacy' into the debate[50]. He claims that for certain types of phenomena that we view as purely intuitive, there are, in fact, rules that can explain our behaviour, even if we are not consciously aware of them. We could unmask the computer in a Turing Test because, if we gave the machine rules to answer certain types of subcognitive questions – for example, 'how do you pronounce the plurals of the imaginary English words 'platch', 'snorp' and 'brell'?' (Answer: 'platchez', 'snorps' and 'brellz') – the machine would be able to explain how it gave these answers, but we humans could not, or at least our explanation would not be the one given by the computer. In this way we could catch the computer out and it would fail the Turing Test. The notion of superarticulacy is particularly relevant to current cognitive science research. Our human ability to know something without being able to articulate that knowledge, or to learn something (as demonstrated by an ability to perform a particular task) without being aware that we have learned it, is at present a very active line of research in cognitive science.

In a recent and significant comment on the Turing Test, Watt proposed the Inverted Turing Test (ITT) based on considerations from 'naive psychology'[51] – our human tendency and ability to ascribe mental states to others and to themselves. In the ITT, the machine must show that its tendency to ascribe mental states is indistinguishable from that of a real human. A machine will be said to pass the ITT if it is 'unable to distinguish between two humans, or between a human and a machine that can pass the normal TT, but which can discriminate between a human and a machine that can be told apart by a normal TT with a human observer'[51]. There are numerous replies to this proposal[52–55]. It can be shown, however, that the ITT can be simulated by the standard Turing Test[52,55]. French used the technique of a 'Human Subcognitive Profile' (i.e. a list of subcognitive questions whose answers have been gathered from people in the larger population, see Box 1) to show that a mindless program using the Profile could pass this variant of the Turing Test[55]. Ford and Hayes[54] renewed their appeal to reject this type of test as any kind of meaningful yardstick for AI. Collins suggested his own type of test, the Editing Test[53], based on 'the skillful way in which humans "repair" deficiencies in speech, written texts, handwriting, etc., and the failure of computers to achieve the same interpretative competence'[53].

Loebner Prize

An overview of the Turing Test would not be complete without briefly mentioning the Loebner Prize[56,57], which originated in 1991. The competition stipulates that the first program to pass an unrestricted Turing Test will receive $100,000. For the Loebner Prize, both humans and machines answer questions by the judges. The competition, however, is among the various machines, each of which attempts to fool the judges into believing that it is a human. The machine that best plays the role of a human wins the competition. Initially, restrictions were placed on the form and content of the questions that could be asked. For example, questions were restricted to specific topics, judges who were computer scientists were disallowed, and 'trick questions' were not permitted.

There have been numerous attempts at 'restricted' simulations of human behaviour over the years, the best known probably being Colby's PARRY[58,59], a program that simulates a paranoid schizophrenic by means of a large number of canned routines, and Weizenbaum's ELIZA[60], which simulates a psychiatrist's discussion with patients.

Aside from the fact that restricting the domain of allowable questions violates the spirit of Turing's original 'anything-goes' Imitation Game, there are at least two major problems with domain restrictions in a Turing Test. First, there is the virtual impossibility of clearly defining what does and does not count as being part of a particular real-world domain. For example, if the domain were International Politics, a question like, 'Did Ronald Reagan wear a shirt when he met with Mikhail Gorbachev?' would seem to qualify as a 'trick question', being pretty obviously outside of the specified domain. But now change the question to, 'Did Mahatma Ghandi wear a shirt when he met with Winston Churchill?' Unlike the first, the latter question is squarely within the domain of international politics because it was Ghandi's practice, in order to make a political/cultural statement, to be shirtless when meeting with British statesmen. But how can we differentiate these two questions a priori, accepting one as within the domain of international politics, while rejecting the other as outside of it? Further, even if it were somehow possible to clearly delineate domains of allowable questions, what would determine whether a domain were too restricted? In a tongue-in-cheek response to Colby's claims that PARRY had passed something that could rightfully be called a legitimate Turing Test, Weizenbaum claimed to have written a program for another restricted domain: infant autism[61]. His program, moreover, did not even require a computer to run on; it could be implemented on an electric typewriter. Regardless of the question typed into it, the typewriter would just sit there and hum. In terms of the domain-restricted Turing Test, the program was indistinguishable from a real autistic infant. The deep point of this example is the problem with domain restrictions in a Turing Test.

To date, nothing has come remotely close to passing an unrestricted Turing Test and, as Dennett, who agreed to chair the Loebner Prize event for its first few years, said, '…passing the Turing Test is not a sensible research and development goal for serious AI'[62]. Few serious scholars of the Turing Test, myself included, take this competition seriously and Minsky has even publicly offered $100 for anyone who can convince Loebner to put an end to the competition![63] (For those who wish to know more about the Loebner Competition, refer to Ref. 57.)

There are numerous other commentaries on the Turing Test. Two particularly interesting comments on actually building truly intelligent machines can be found in Dennett[64] and Waltz[65].

Conclusions

For 50 years the Turing Test has been the object of debate and controversy. From its inception, the Test has come under fire as being either too strong, too weak, too anthropocentric, too broad, too narrow, or too coarse. One thing, however, is certain: gradually, ineluctably, we are moving into a world where machines will participate in all of the activities that have heretofore been the sole province of humans. While it is unlikely that robots will ever perfectly simulate human beings, one day in the far future they might indeed have sufficient cognitive capacities to pose certain ethical dilemmas for us, especially regarding their destruction or exploitation. To resolve these issues, we will be called upon to consider the question: 'how much are these machines really like us?' and I predict that the yardstick that will be used measure this similarity will look very much like the test that Alan Turing invented at the dawn of the computer age.

Acknowledgements
The present paper was supported in part by research grant IUAP P4/19 from the Belgian government. I am grateful to Dan Dennett and Stevan Harnad for their particularly helpful comments on an earlier draft of this review.

References

1 Turing, A. (1950) Computing machinery and intelligence. *Mind* 59, 433–460
2 Saygin, P. *et al.* (1999) *Turing Test: 50 Years Later*, Technical Report No. BU-CEIS-9905, Department of Computer Engineering, Bilkent University, Ankara, Turkey
3 Anderson, A. (1964) *Minds and Machines*, Prentice-Hall
4 Dreyfus, H. (1992) *What Computers Still Can't Do*, MIT Press
5 Haugeland, J. (1985) *Artificial Intelligence, the Very Idea*, MIT Press
6 Hofstadter, D. (1979) *Gödel, Escher, Bach*, Basic Books
7 Ginsberg, M. (1993) *Essentials of Artificial Intelligence*, Morgan Kaufmann
8 Whitby, B. (1996) The Turing Test: AI's biggest blind alley? In *Machines and Thought: The Legacy of Alan Turing* (Millican, P. and Clark, A. eds), pp. 53–63, Oxford University Press
9 Hayes, P. and Ford, K. (1995) Turing Test considered harmful. In *Proc. Fourteenth IJCAI-95, Montreal, Canada* (Vol. 1), pp. 972–977, Morgan Kaufmann
10 Johnson, W. (1992) Needed: a new test of intelligence. *Sigart Bull.* 3, 7–9
11 Simon, H. and Newell, A. (1958) Heuristic problem solving: the next advance in operations research. *Operations Res.* 6,
12 Minsky, M. (1967) *Computation: Finite and Infinite Machines*, p. 2, Prentice-Hall
13 Kolata, G. (1982) How can computers get common sense? *Science* 217, 1237
14 Gödel, K. (1931) Uber formal unentscheidbare Sätze der *Principia Mathematica* und Verwandter Systeme: I. *Monatshefte für Mathematik und Physik* 38, 173–198
15 Lucas, J. (1961) Minds, machines and Gödel. *Philosophy* 36, 112–127
16 Mays, W. (1952) Can machines think? *Philosophy* 27, 148–162
17 Scriven, M. (1953) The mechanical concept of mind. *Mind* 62, 230–240
18 Gunderson, K. (1964) The imitation game. *Mind* 73, 234–245
19 Gunderson, K. (1967) *Mentality and Machines*, Doubleday
20 Stevenson, J. (1976) On the imitation game. *Philosophia* 6, 131–133
21 Purtill, R. (1971) Beating the imitation game. *Mind* 80, 290–294
22 Sampson, G. (1973) In defence of Turing. *Mind* 82, 592–594
23 Millar, P. (1973) On the point of the Imitation Game. *Mind* 82, 595–597
24 Moor, J. (1976) An analysis of the Turing Test. *Philos. Stud.* 30, 249–257
25 Stalker, D. (1978) Why machines can't think: a reply to James Moor. *Philos. Stud.* 34, 317–320
26 Moor, J. (1978) Explaining computer behaviour. *Philos. Stud.* 34, 325–327
27 Hofstadter, D. (1981) The Turing Test: a coffee-house conversation. In *The Mind's I* (Hofstadter, D. and Dennett, D., eds), pp. 69–95, Basic Books
28 Anderson, D. (1987) Is the Chinese room the real thing? *Philosophy* 62, 389–393
29 Block, N. (1981) Psychologism and behaviourism. *Philos. Rev.* 90, 5–43
30 Searle, J. (1980) Minds, brains and programs. *Behav. Brain Sci.* 3, 417–424
31 Hofstadter, D. and Dennett, D. (1981) Reflections on 'Minds, Brains, and Programs'. In *The Mind's I* (Hofstadter, D. and Dennett, D., eds), pp. 373–382, Basic Books
32 Dennett, D. (1985) Can machines think? In *How We Know* (Shafto, M., ed.), pp. 121–145, Harper & Row
33 Rumelhart, D., McClelland, J. and the PDP Research Group, eds (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vols 1 and 2), MIT Press
34 Smolensky, P. (1988) On the proper treatment of connectionism. *Behav. Brain Sci.* 11, 1–74
35 French, R. (1988) Subcognitive probing: hard questions for the Turing Test. In *Proc. Tenth Annu. Cognit. Sci. Soc. Conf.*, pp. 361–367, Erlbaum
36 French, R. (1990) Subcognition and the limits of the Turing Test. *Mind* 99, 53–65
37 Davidson, D. (1990) Turing's test. In *Modelling the Mind* (Said, K.A. *et al.*, eds), pp. 1–11, Oxford University Press
38 Harnad, S. (1989) Minds, machines and Searle. *J. Exp. Theor. Artif. Intell.* 1, 5–25
39 Harnad, S. (1991) Other bodies, other minds: a machine incarnation of an old philosophical problem. *Minds Machines* 1, 43–54
40 Harnad, S. (1992) The Turing Test is not a trick: Turing indistinguishability is a scientific criterion. *Sigart Bull.* 3, 9–10
41 Harnad, S. (1994) Levels of functional equivalence in reverse bio-engineering: the Darwinian Turing Test for artificial life. *Artif. Life* 1, 293–301
42 Harnad, S. Turing on reverse-engineering the mind. *J. Logic Lang. Inf.* (in press)
43 Harnad, S. (1990) The symbol grounding problem. *Physica D* 42, 335–346
44 Shanon, B. (1989) A simple comment regarding the Turing Test. *J. Theory Soc. Behav.* 19, 249–256
45 Hauser, L. (1993) Reaping the whirlwind: reply to Harnad's 'Other bodies, other minds'. *Minds Machines* 3, 219–237
46 Schweizer, P. (1998) The Truly Total Turing Test. *Minds Machines* 8, 263–272
47 French, R. Peeking behind the screen: the unsuspected power of the standard Turing Test. *J. Exp. Theor. Artif. Intell.* (in press)
48 Crockett, L. (1994) *The Turing Test and the Frame Problem: AI's Mistaken Understanding of Intelligence*, Ablex
49 Harnad, S. (1993) Problems, problems: the frame problem as a symptom of the symbol grounding problem. *Psycoloquy* 4 (34)
50 Michie, D. (1993) Turing's test and conscious thought. *Artif. Intell.* 60, 1–22
51 Watt, S. (1996) Naive psychology and the Inverted Turing Test. *Psycoloquy* 7 (14)
52 Bringsjord, S. (1996) The Inverted Turing Test is provably redundant. *Psycoloquy* 7 (29)
53 Collins, H. (1997) The Editing Test for the deep problem of AI. *Psycoloquy* 8 (1)
54 Ford, K. and Hayes, P. (1996) The Turing Test is just as bad when inverted. *Psycoloquy* 7 (43)
55 French, R. (1996) The Inverted Turing Test: a simple (mindless) program that could pass it. *Psycoloquy* 7 (39)
56 Epstein, R. (1992) Can Machines Think? *AI Magazine* 13, 81–95
57 Shieber, S. (1994) Lessons from a restricted Turing Test. *Commun. ACM* 37, 70–78
58 Colby, K. (1981) Modeling a paranoid mind. *Behav. Brain Sci.* 4, 515–560
59 Colby, K. *et al.* (1971) Artificial paranoia. *Artif. Intell.* 2, 1–25
60 Weizenbaum, J. (1966) ELIZA: a computer program for the study of natural language communication between men and machines. *Commun. ACM* 9, 36–45
61 Weizenbaum, J. (1974) Reply to Arbib: more on computer models of psychopathic behaviour. *Commun. ACM* 17, 543
62 Dennett. D. (1998) *Brainchildren*, p. 28, MIT Press
63 Minsky, M. (1995) Article 24971 of comp.ai.philosophy, March 3, 1995
64 Dennett, D. (1994) The practical requirements for making a conscious robot. *Philos. Trans. R. Soc. London Ser.* A 349, 133–146 (Reprinted in Dennett, D., 1998, *Brainchildren.* MIT Press)
65 Waltz, D. (1988) The prospects for building truly intelligent machines. *Daedalus* 117, 191–212