The Turing Test

William J. Rapaport

Department of Computer Science and Engineering,

Department of Philosophy, and Center for Cognitive Science

State University of New York at Buffalo, Buffalo, NY 14260-2000

rapaport@cse.buffalo.edu

http://www.cse.buffalo.edu/∼rapaport/

January 18, 2005

Abstract

This document is a draft of an article for the *Encyclopedia of Language and Linguistics, 2nd Edition* (Elsevier, forthcoming).

This article describes the Turing Test for determining whether a computer can think. It begins with a description of an "imitation game" for discriminating between a man and a woman, discusses variations of the Test, standards for passing the Test, and experiments with real Turing-like tests (including Eliza and the Loebner competition). It then considers what a

computer must be able to do in order to pass a Turing Test, including whether written linguistic behavior is a reasonable replacement for “cognition”, what counts as understanding natural language, the role of world knowledge in understanding natural language, and the philosophical implications of passing a Turing Test, including whether passing is a sufficient demonstration of cognition, briefly discussing two counterexamples: a table-lookup program and the Chinese Room Argument.

Keywords and Cross-References: AI-complete, Alan Turing, artificial intelligence, B.F. Skinner, chatterbots, Chinese-Room Argument, cognition, commonsense knowledge, computers, cultural literacy, Daniel Dennett, Descartes, Eliza, functionalism, Hilary Putnam, intelligence, intentional stance, John Searle, Loebner prize competition, metaphor, natural-language processing, natural-language understanding, Noam Chomsky, Parry, thinking, Turing machine, Turing Test, verbal behavior, world knowledge

1 Introduction

“The Turing Test” is the name given to any of several variations of an experiment proposed by the mathematician and computer scientist Alan M. Turing in his essay, “Computing Machinery and Intelligence”, which appeared in the philosophy journal *Mind* in 1950. The Test, very briefly put, is for an interrogator to decide

whether a participant in a natural-language conversation is a human or a computer. The participant passes the test to the extent that it convinces the interrogator that it is human (even if it is really a computer). (For simplicity, I will usually call the one (or two) participant(s) in a Turing Test other than the interrogator "the participant". In different versions of the Test, the participant(s) might be a male and/or female human, a computer, or perhaps some other candidate for the ascription of "intelligence".)

The experiment, a version of a parlor game that Turing called "the imitation game", was designed to help answer the question "Can machines think?" by replacing that informal question (containing the vague terms 'machines' and 'think') by a more precise question: Can a computer "pass" a Turing Test? Thus, the vague terms 'machines' and 'think' are replaced by, respectively, the more precise notions of a digital computer (or a Turing machine) and the more precise—or at least behaviorally determinable—notion of passing the Test. (This strategy—replacing a vague, informal question that is difficult to answer by a more precise question that is easier to answer—is similar to the strategy that Turing (1936) adopted in giving his extremely influential answer to the question of what 'computable' means.)

1.1 The imitation game

The original "imitation game" consists of three players: a man, a woman, and an interrogator, each in a separate room, communicating only via teletype (or what today would be considered an "instant messaging" interface), such that none of the players can see or hear the others, thus limiting their interactions to typed (i.e., written) linguistic communication. (In §3.2, we will consider the extent to which this is a limitation.) The woman's task is to convince the interrogator that she is the woman; the man's task is to **imitate** a woman so as to convince the interrogator that he **is** the woman; and the interrogator's task is to determine who is which. The man wins the imitation game (i.e., "passes" the imitation-game "test") if he succeeds in convincing the interrogator (e.g., perhaps by deception) that he is the woman; thus, the woman succeeds (as does the interrogator) to the extent that the man fails. This version of the imitation game seems to be computationally solvable: A recent computational study has suggested that it is possible to identify stylistically whether the author of a given text is male or female (Argamon et al. 2003).

1.2 The Turing Test and its variations

Several versions of "The Turing Test" may be found both in Turing's essay and in subsequent literature. Turing's first version merely replaces the man in the

imitation game by a computer. If the interrogator fails as often as he or she did on the imitation game, then the computer is said to have "passed the Turing Test". But does this mean that the computer passes the Test if it is able to convince the interrogator that it is a man who, in turn, is trying to convince the interrogator that he is the woman?

It is reasonably clear from Turing's essay that he had in mind something less convoluted. Thus, the second version of the Test replaces the man by a computer and the woman by another **man**, and so the computer's task is to convince the interrogator that it is a man (see Colby et al. 1972: 202 and Shieber 2004a: 100–104 for further discussion of the gender issue).

A more gender-neutral version replaces the man by a computer and the woman by a **human** (of either sex). The computer passes this version of the Test if it convinces the interrogator that it is the human.

In modern versions of the Test, this is simplified even more, by having an interrogator and only one participant: either a computer or a human. Again, the computer "passes" this Test if it convinces the interrogator that he or she is interacting with a human. Turing 1950 alludes to this version as a "**viva voce**", i.e., an oral examination; a subject passes an oral exam by convincing the examiner that he or she "understands" the material that he or she is being tested on.

Another variation (Stuart C. Shapiro, personal communication) depends on

whether the interrogator **knows** that there is a possibility that the participant is a computer or whether the interrogator is merely being asked to judge whether the participant is intelligent (or as intelligent as a normal person).

Turing (1950: 441) also mentions a version of the Test in which the interrogator must distinguish between a digital computer and a discrete-state machine. Other variations are described in §2.

1.3 Passing the Test

As a standard for deciding whether a participant has passed the Test, Turing (1950: 442) proposed that "an average interrogator will not have more than 70 per cent. [**sic**] chance of making the right identification after five minutes of questioning". Some have complained that this sets the standard too low, but the actual statistical standard to be applied could easily be varied (e.g., 70% increased to 95%, 5 minutes increased to several months, etc.). Turing's prediction was "that at the end of the century [i.e., by 2000] the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted" (Turing 1950: 442). (We will examine these two standards in §§3.1, 4.1. At the end of the century, in fact, several anthologies considered the status of Turing's prediction: Akman & Blackburn 2000; Moor 2000–2001, Moor 2003; Shieber 2004a.)

The Turing Test has been likened to two, more familiar situations:

1. The two-entity version (interrogator and participant) is like the situation in which a human corresponds with a pen-pal (Harnad 1989). In real life, the human would be unlikely to doubt that the pen-pal was a human. One point of the Turing test is that if the pen-pal were, in fact, a computer, still the human would not question its humanity (i.e., its human level of intelligence or its human ability to think).

2. As Stuart Shieber (2004a) points out, the interrogator's task is like that of someone determining whether something is 1 meter long by comparing it with a standard meter; both items share the same length. Here, the interrogator is determining whether something has a certain feature (namely, a human level of intelligence) by seeing if it shares that feature with (i.e., if it can be measured against) another entity, which, by hypothesis, is a "standard" for having that feature. And, in the case of the Turing Test, the standard is written linguistic communication (Shieber's term is "verbal behavior").

2 Experiments with Real Turing-Like Tests

Two early AI (artificial intelligence) programs in the areas of natural-language processing and cognitive modeling are often mentioned in connection with the Turing Test. Joseph Weizenbaum's (1966, 1967) computer program Eliza appears to converse in more or less fluent English with its user, apparently simulating a Rogerian psycho-therapist. Many anecdotes of people being fooled into thinking that they were conversing with a real human suggest that Eliza has passed a Turing-like Test. Yet Eliza's natural-language-processing techniques are mere pattern matching using little or no grammatical knowledge (e.g., swapping pronouns such as 'you' and 'me' in certain patterns)—i.e., the techniques are not cognitive (but cf. §5). Hence, Weizenbaum has claimed that Eliza is a counterexample to the Turing Test. But this has not been subject to controlled experiments.

On the other hand, Kenneth Mark Colby's computer program Parry is an Eliza-inspired program that, using more sophisticated natural-language-processing techniques than Eliza, simulates a paranoid patient. Parry has been the subject of controlled Turing-Test-like experiments (Colby et al. 1972; Colby 1981), and Colby, at least, claims that it has passed the Turing Test, having convinced human psycho-therapist judges that they were conversing with a human paranoid patient. On the other hand, that is all that Parry can discuss, and so it is not clear that the sort of test that it passes is a full-fledged Turing Test.

The Loebner Prize Competition, held annually since 1991, is a Turing-like test that has offered a monetary award "for the first computer whose responses were indistinguishable from a human's. Each year an annual prize of $2000 and a bronze medal is awarded to the most human computer. The winner of the annual contest is the best entry relative to other entries that year, irrespective of how good it is in an absolute sense" (Loebner 2003). Most of the winners have been elaborate Eliza-like programs (sometimes called "chatterbots"). However, one notable exception was David Levy's CONVERSE program (Batacharia et al. 1999), which won the Loebner Prize in 1997, and which was developed by a team including at least one well-known researcher in computational linguistics (Yorick Wilks). The Loebner Competition has been critically analyzed by the computational linguist Stuart Shieber, who has argued that the Competition, unlike other competitions for professional computational linguists (e.g., the Message Understanding Conferences and the Text Retrieval Conferences), has not fostered research and development, but merely encouraged unintelligent, Eliza-like "chatterbot" programs whose sole goal is to fool the judges (Shieber 1994ab; for Loebner's reply, see Loebner 1994).

Turning the tables, some recent projects have used computers as "interrogators" in a Turing-like test to unmask computer programs that attempt to convince other computer programs that they are human and hence entitled, say, to e-mail accounts.

However, these tests are not based on natural-language understanding or even verbal behavior; they usually involve visual images containing distorted words that only humans can read (Ahn et al. 2004; [http://www.captcha.net/]). One formal investigation that replaces the interrogator with a Turing machine claims to show that no computer can be a perfect interrogator in the standard Turing Test (Sato & Ikegami 2004). And there have been other investigations of the formal, logical, and computational structure of the Turing Test (e.g., Bradford & Wollowski 1994, Shieber 2004b).

3 What Must a Computer Be Able to Do in Order to Pass a Turing Test?

A significant feature of the Test is the central role of language. In order to pass the Test, a participant must be able to communicate in the written natural language of the interrogator to such an extent that the interrogator cannot distinguish its linguistic behavior from that of a human using that language. (To simplify matters, we will assume that all participants speak that language equally well.)

This feature echoes a historical antecedent to the Turing Test: In his **Discourse on Method** (1637, Part V), the philosopher René Descartes argued that there was a test that could show that a machine that looked and behaved like a human was

not a real human, namely, that even if it could use language in certain appropriate, responsive ways (e.g., to “ask what we wish to say to it” if “touched in a particular part”, or to “exclaim that it is being hurt” if touched somewhere else), it could not “arrange its speech … in order to reply appropriately to everything that may be said in its presence” (Descartes 1637/1970: 116).

There are two issues to consider: (1) Is linguistic ability a reasonable replacement for the informal notions of “thinking” or “intelligence”? (2) Must the test be limited to linguistic ability?

3.1 Is written linguistic behavior a reasonable replacement for “cognition”?

‘Thinking’ is a vague term, and ‘intelligence’ is not necessarily, or even probably, intended in the sense of the IQ-test notion of intelligence, but merely as a convenient synonym for ‘cognition’. Therefore, we will use the latter term, which is less controversial.

To approach an answer to the question whether written linguistic behavior can replace “cognition”, we can look at what Turing thought the computer might have to be able to do. In his examples of the computer’s behavior and in his replies to a series of potential objections, Turing suggests that the computer might have to do the following (kinds of) things:

- Answer questions about its physical appearance and abilities.
- Answer questions about its personality and experiences.
- Write and discuss a poem on a given topic.
- Do arithmetic.
- Solve chess problems.
- Have access to, and be able to use, "commonsense" information about the seasons, Christmas, familiar fictional characters, etc. (This is sometimes called "world knowledge".)
- Learn.

For Turing, the last of these was the most important. Presumably, this learning would have to come from reading, not from doing, although Turing also suggests that the machine might have other "experience[s], not to be described as education". In any case, Turing seems to have been willing to restrict the Test-passing capabilities of the machine to "all purely intellectual fields", which, presumably, manifest themselves through language.

Insofar as ordinary cognition is surely involved in these sorts of activities, being able to do them does seem to be a reasonable test of human-level cognition. Surely, if a machine could do such things, an interrogator would be justified in being unable

to distinguish between it and a human who could do them equally well. Hence (perhaps) the interrogator would be justified in concluding that the participant had cognition. Turing, however, does not say that the interrogator needs to draw the conclusion that the participant really has cognition (i.e., really can do these things); perhaps this is why he wrote that "the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted" (see §§1.3, 4.1).

Thus, a computer must have at least the following two capacities in order to pass a Turing Test: (1) The ability to understand and generate natural language. (2) Lots of world knowledge. Let us consider each of these.

3.1.1 Understanding natural language

What counts as "understanding" (including generating) natural language? Turing apparently thought that it simply meant behaving in such a way that the interrogator **believes** that the participant understands the language, i.e., that the interrogator **attributes** natural-language understanding to the participant (cf. Dennett's "intentional stance"). Cognitivists hold that it means something more than mere behavior. Perhaps any entity that exhibits such behavior must have the further capacity to be able to parse linguistic input, construct a (mental) representation of it, integrate or assimilate it into "background knowledge",

determine an appropriate linguistic response, and formulate and express that response in a grammatical utterance. Or **understanding** natural language might mean something even more than behavior together with this natural-language-**processing** capacity: Perhaps the entity must have "intrinsic intentionality" or consciousness (Searle 1980).

3.1.2 World knowledge

Understanding natural language is probably not something that can be done merely on the basis of linguistic knowledge (e.g., knowledge of a grammar and lexicon). It probably requires much world knowledge—not only semantic and pragmatic knowledge, but also something like what the educational reformer E.D. Hirsch (1987) calls "cultural literacy" or what the AI researcher Douglas B. Lenat (1995) calls "commonsense knowledge", i.e., lots of facts about the world, and perhaps also some specialized domain knowledge (e.g., knowledge of how to play chess; cf. §3.1).

The computational linguist Terry Winograd argued for the claim that natural-language understanding requires such knowledge: Consider these two sentences, which differ in only one word: (a) "The city councilmen refused the demonstrators a permit because they advocated violence." and (b) "The city councilmen refused the demonstrators a permit because they feared violence." Arguably, it is "world

knowledge" or "commonsense knowledge" about the roles of city councilmen and demonstrators, as well as semantic knowledge about the meanings of 'advocated' and 'feared', that aids in determining that 'they' refers to the demonstrators in (a) but the city councilmen in (b).

Moreover, someone reading *The Diary of Anne Frank* who does not understand the word 'Gestapo' might be told that it was Hitler's secret police, but if the reader does not know who Hitler was, this will not be of much help. The reader needs to know that Hitler was the Nazi dictator of Germany during World War II, who the Nazis were, where Germany is, what World War II was, etc.

3.2 Is written linguistic communication a limitation?

Must the test be limited to linguistic behavior? After all, cognition in general is usually taken to consist of more than just linguistic ability; e.g., it includes learning, perception, reasoning, memory, planning (and acting on those plans), etc. In the original imitation game, the restriction to linguistic interrogation eliminated the need to deal with the physical make-up of the participants. Limiting the interrogation to linguistic communication does not seem debilitating: Surely the sorts of conversations that Turing suggested cover a wide range of intellectual activities—enough to convince an interrogator of the participant's cognition (or, more precisely, the participant's cognitive-behavioral indistinguishability from a

human).

Linguistic ability by itself is surely a mark of cognition. Moreover, it may be that linguistic ability is the essential core of the kind of cognition that is in question, in the sense that any entity that had such linguistic ability is highly likely to have—or require—most other kinds of cognition (Rapaport 1988, Harnad 1989, Rey 2002). Not only does natural-language understanding require world knowledge, but it is arguably the case that natural-language understanding is "AI-complete"; i.e., "solving" the natural-language-understanding problem may require solving all other problems in AI (artificial intelligence). If so, then, insofar as passing the Turing Test requires natural-language understanding, it would be a sufficient condition of cognition. In any case, the psychologist Stevan Harnad (1989) has proposed a "Total Turing Test" variation that extends the Test to allow for demonstrations of perception, action, and other sorts of non-linguistic cognition (or cognitive indistinguishability from humans).

If passing the Turing Test requires, not merely the ability to fool or (statistically) convince an interrogator that he or she is conversing with a human (or at least with an entity with human-level cognition), but actual, human-level linguistic abilities, then what kinds of abilities are required? Research in AI and computational linguistics since the mid-1950s suggests that any natural-language understander—including a Turing-Test-passing program—must be able (at least)

to:

1. take coherent discourse (as opposed to isolated sentences) as input,
2. understand ungrammatical input,
3. make inferences and revise beliefs,
4. make plans (for speech acts, to ask and answer questions, and to initiate conversation),
5. understand plans (including the speech acts of all interlocutors in the conversation),
6. construct user models (i.e., “theories of mind”, or psychological theories, of all interlocutors),
7. learn about the world and about language, and do this in part via a natural-language interaction,
8. have background (or “world”, or “commonsense”) knowledge,
9. and remember what it heard, learned, inferred, and revised.

Arguably, any entity having these abilities could be said to be able to understand language or, more generally, to be cognitive, or to think. The open research question is whether **computational** theories of these cognitive abilities can be developed.

4 What Does Passing a Turing Test Mean?

4.1 Does passing the Turing Test demonstrate real cognition or only the appearance of cognition?

Is a machine that **appears** to understand natural language or to have cognition **really** understanding natural language; does it really have cognition? Perhaps the early-to-mid-20th-century behavioral bias made it seem reasonable to Turing that his test answers this question in the only way possible. But the modern late-20th-to-early-21st-century cognitive bias has kept the question open.

One position is that if the output of the machine is not produced by "the same" methods that humans use, then although its external behavior might be similar to a human's, its internal mechanism is not and, hence, it is not really doing them (see §5). Another position is that anything that is capable of these complex sorts of linguistic behavior is indeed communicating in language (at least, no less than a human is) and not merely appearing to, no matter what internal mechanism produces the output. That is, it is not merely simulating the use of language; rather, at least in the case of language use (and other cognitive phenomena), this sort of simulation **is** the real thing (Rapaport 1988).

Another line of argument, however, suggests that even in the case in which it is felt that the appearance is not reality, it is near enough so as not to matter. Some

support for this is given by Turing's statement "that at the end of the century **the use of words** and **general educated opinion** will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted." One way that this might come about is if "the use of words" were to change. E.g., the meaning of 'think' might be extended metaphorically to cover both machines and humans. Enlarging the scope of meaning of a word is something that has happened many times before: "To fly" was once something that only birds and bats were said to do; now, airplanes, helicopters, space shuttles, and their human riders are all said to fly. A "computer" was once a **human** whose job was to carry out complex or lengthy mathematical calculations; now, it is a **machine** that does what those humans once did. Such semantic enlargements can come about by metaphorical extension: Perhaps planes don't literally fly (in the bird-sense), just as philosophers don't literally shoot down an opponent's argument, but we eventually come to apply the metaphor so often that it becomes an unquestioningly acceptable—even unavoidable—predication.

In addition to "the use of words" changing, it is also possible for "general educated opinion" about the nature of cognition to change, perhaps as a result of progress in AI, philosophical analysis, or psychological research. E.g., the meaning of the word 'think' might be abstracted to an essence that allows it to be predicated of both AI-programmed computers and humans (not to mention

other animals). Our current theoretical understanding of heavier-than-air flight encompasses both bird flight and airplane flight. And our current theoretical understanding of computation—due, in fact, to Turing (1936)—has its foundation in an analysis of what human computers do. Extending this idea, progress in AI and cognitive science could cause us to have a better understanding of cognition in such a way that the new understanding applies to both humans and machines.

4.2 Is passing a Turing Test a sufficient demonstration of cognition?

In what sense does passing the Turing Test **replace** the vague notion of "thinking"? Logically, there are two possibilities: It is a necessary condition, or it is a sufficient condition (or both).

To say that passing the Turing Test is a **necessary** condition of "thinking" is to mean that if something thinks, then it can pass the Turing Test. It might, of course, be the case that something can think but fails the Test, perhaps because the interrogator is not up to par or because the thinking agent simply is having an "off day". But, presumably, under normal circumstances, a real cognitive agent that takes the Test will pass it.

More controversial is the proposition that passing the Turing Test is a **sufficient** condition of "thinking". To say this is to mean that if something passes the Test, then it "thinks". Several objections to the Turing Test take the form of showing that

it is not a sufficient condition for thinking.

4.2.1 The "Aunt Bertha" table-lookup objection

One, due to the philosopher Ned Block (1981), considers a computer program that stores a finite, but extremely large, number of possible conversations—all the one-hour conversations that the programmer's "Aunt Bertha" might possibly have during her entire life, on any subject whatever. Such a program would, presumably, pass the Test, yet would not normally be considered to "think". The point is that the methods whereby the responses are produced are behaviorally, but not psychologically, plausible. Thus, we have a case of something that passes the Test but does not think, which is a counterexample to the test's sufficiency. Thinking, according to this objection, is not merely a matter of behavior, but also a matter of how that behavior is produced, and certain ways of producing it are not suitably cognitive (such as memorizing a large lookup table), whereas certain others might be (such as requiring reasoning; cf. Shannon & McCarthy 1956).

One possible response is to deny the possibility of such a "pre-compiled" lookup table (e.g., such a system could not learn new words; cf. §§3.1, 4.1). To this, it can be replied that there is only a finite (albeit astronomically large) number of possible words, all of which could be pre-compiled. On behalf of the Test, another response to Block's objection is to admit this, but to point out that the Test

still provides considerable evidence supportive of a claim to cognition and, in the absence of any ability to know how the cognitive behavior was produced (as in the pen-pal case), that might be all the evidence one could possibly have.

4.2.2 The Chinese-Room objection

Perhaps the most well-known objection to the Test's sufficiency is the philosopher John Searle's (1980) Chinese-Room Argument. Here, a human who knows no Chinese is placed in a room with an instruction book that tells him how to manipulate certain symbols that are input to him in the room, and then how to output other symbols on the basis of these manipulations. The interrogator is a native speaker of Chinese who inputs questions written in Chinese characters and reads the output from the room. By hypothesis, the interrogator judges that the entity with which he is fluently conversing understands Chinese, whereas, again by hypothesis, the human in the room in fact knows no Chinese and is merely manipulating marks that are "meaningless" to him. Thus, we seem to have a case in which something does not have a certain sort of cognition (in this case, understanding natural language) yet passes the Turing Test (and, moreover, via "reasoning" as opposed to "memorization").

One response to Searle's proposed counterexample is to object that the solo human is not the Test-taker; rather, the Test is being taken—and passed—by the

system consisting of the human combined with the instruction book, and it is possible that that **system** does understand Chinese even though its **parts** don't. In any case, denying that the **system** understands Chinese is not the same as denying that the **human** in the room understands Chinese.

This "systems reply" is often supplemented by a "robot reply", a version of the Total Turing Test in which the participant memorizes the instruction book and is endowed with other kinds of cognition, such as those involving sensors (e.g., perception) and effectors (e.g., acting). Finally, as in the AI-complete situation (§3.2), perhaps a suitably rich sort of symbol manipulation (combining linguistically-derived concepts with perceptually-derived concepts) could provide the apparently missing semantic component (Rapaport 2000).

4.2.3 Is cognition subjective?

Both Block's and Searle's objections also raises the issue of whether cognition (or natural-language understanding) is a "subjective" phenomenon: A participant passes a Turing Test **with respect to a given interrogator**. Thus, a given lookup table might not work for a given interrogator, and different interrogators might disagree on whether a participant in a Chinese Room demonstrates understanding.

5 Is Cognitive Behavior Really Cognition?

A participant will either pass the Turing Test or not. If **no** computer ever passes it—that is, if only humans (or other entities independently judged to be capable of cognition, such as extraterrestrials) pass it—then the Test succeeds in the sense that it offers a (behavioral) mark of (human) cognition. If, on the other hand, a computer **does** pass it, that could be for two reasons: Either it is **really** cognitive (and able to **convince** the interrogator of this), or else it isn't but only **appears** to be.

One way of phrasing the fundamental problem posed by the Turing Test is whether mere cognitive behavior **is** real cognition. There are three possibilities: (1) Real cognition might be "mere" input-output behavior (statistical correlation). (2) Or real cognition might be input-output behavior **plus** any intervening process (including the "Aunt Bertha" kinds of table-lookup processes or Eliza-like natural-language-processing tricks). (3) Or real cognition might be input-output behavior with an intervening process **of a "cognitive" kind**, such as a reasoning process, a symbolic computational process, a connectionist computational process, or a "dynamic systems" process, implementing some psychological theory.

The Turing Test doesn't seem to distinguish among these three possibilities. Should it? I.e., is it only external behavior that counts as cognition, or is the internal process (if any) important, too (as Shannon & McCarthy 1956 suggested;

cf. §4.2.1)? Consider a "magical" natural-language-processing machine: a pure input-output machine that takes natural language input (from the interrogator) and "magically" yields (appropriate) natural-language output (i.e., a "function machine" with no internal processes). Suppose this magical machine passes a Turing Test. Is it cognitive? Turing might have argued that it was, even though we couldn't know why or how it managed to understand natural language (after all, it's magic!). On the other hand, although we know that our brain is a non-magical natural-language processor, we are not yet at the stage where we can say with certainty that we know why or how **it** manages to understand natural language.

Thus, the fundamental problem of the Turing Test is the classic problem of intension vs. extension: Is the **way** in which a result is obtained or an act is performed (i.e., the result or act viewed intensionally) more important than the **result** itself (i.e., the result or act viewed extensionally)? From the perspective of linguistics, is the ability to respond to stimuli with "fluent" verbal behavior that is indistinguishable from "normal" human verbal behavior different from a "deeper", "cognitive" ability of "understanding language"? (This is very nearly the debate between B.F. Skinner's theory of verbal behavior and Noam Chomsky's theories of grammar.)

Thus, there are two possibilities of why a participant might pass a Turing Test: Its cognitive behavior is (merely) extensionally equivalent to (real, human)

cognition, or its cognitive behavior is (also) intensionally equivalent to real (human) cognition. Should the first possibility vitiate the Test? Suppose that the interrogator (an external observer) takes the "intentional stance" and ascribes cognition to the participant. It doesn't follow that the participant is **not** really (intensionally) cognitive. After all, one reason that ascribing an internal, intensional cognitive state or process might be the best explanation for an entity's behavior is that the entity really produced its behavior via that process.

Consider the Aunt Bertha memorization machine. Suppose that the table lookups can be done in real time. We seem to be faced with two choices: assert that it is a counterexample to the Turing Test or accept that the Aunt Bertha machine really is cognitive (and not merely **behaving** cognitively). On the side of real cognition, we could argue that the Aunt Bertha machine wouldn't work—that only a real cognitive agent could pass the Turing Test. The Chinese-Room Argument would **not** be a counterexample in this case, because the Chinese-Room system consisting of the human together with the instruction book might only pass if the instruction book were AI-complete, containing a full computational cognitive theory. To assert that it couldn't contain such a theory is to go beyond the empirical evidence, especially since we don't have that evidence yet.

Or, also on the side of real cognition, we could bite the bullet and say that the Aunt Bertha machine **is** cognitive **despite** being a mere lookup table, i.e., that

appropriate lookup tables are just one way of being cognitive (cf. Putnam's (1960) Turing-machine lookup-table analogy to the human mind). In other words, the **appearance** of cognition **is** real cognition; i.e., the external ascription of cognition is all that matters (as in the pen-pal case, §1.3). Arguably, our ordinary notion of "understanding" applies to an omniscient deity that would never have to draw inferences or in any way "process" information. Thus, if table lookup is not cognition, then such omniscience would not be cognitive, which seems to violate our pre-theoretical intuitions.

Consider an analogy: a Turing Test for seeing. Under what conditions would it be fair to say that an external observer can determine whether an entity sees? Is mere "seeing behavior" enough? E.g., is it enough that the entity reacts appropriately to visual stimuli whether or not it has any internal representation of them? Plants not only respond tropistically to light, but there is even some evidence that they can sense the presence of distal stimuli by means of light reflection. So, do plants see? Probably not, in the sense that they probably have no internal representation of the external stimuli—but can that be determined behaviorally? Consider "blindsight": There is evidence that some blind humans can respond appropriately to visual stimuli. Do they see (despite their protestations that they cannot)? This case is especially puzzling, since it is conceivable that such blindsighted humans have internal representations that they use in processing the

external stimuli, yet they are not aware of them.

6 Conclusion

Turing hoped that his Test would **replace** the question, "Can machines think?". It hasn't—in large part because of disagreement over whether it is an **appropriate** replacement. Any formal idea that is proposed to replace an informal one is bound to suffer from this problem, for the formal idea will always include some feature that the informal one appears to exclude, and vice versa. In the case of the Turing Test, this has given rise to the debate over its sufficiency for thinking and for cognition in general. Nevertheless, over 50 years after its appearance, the Turing Test continues to help focus debate on issues concerning the relation of cognition to computation.

Acknowledgments

I am grateful to Paula Chesley, Frances L. Johnson, Jean-Pierre Koenig, Shakthi Poornima, Stuart C. Shapiro, and other members of the SNePS Research Group for comments on an earlier draft.

Bibliography

Ahn, L. von; Blum, M.; Langford, J. (2004). "Telling humans and computers apart automatically". *Communications of the ACM* 47(2) (February), 56–60.

Akman, V.; & Blackburn, P. (eds.) (2000). Special issue on Alan Turing and artificial intelligence, *Journal of Logic, Language and Information*, Vol. 9, No. 4.

Argamon, S.; Koppel, M.; Fine, J.; et al. (2003). "Gender, genre, and writing style in formal written texts". *Text* 23(3), 321–346.

Batacharia, B.; Levy, D.; Catizone, R.; et al. (1999). "CONVERSE: A Conversational Companion". In Wilks, Y. (ed.), *Machine conversations*. Boston: Kluwer Academic Publishers. 205–215.

Block, N. (1981). "Psychologism and Behaviorism". *Philosophical Review* 90(1), 5–43.

Bradford, P.G., & Wollowski, M. (1994). "A formalization of the Turing test". *Technical Report 399*. Bloomington, IN: Indiana University Department of Computer Science.

Colby, K.M. (1981). "Modeling a Paranoid Mind". *Behavioral and Brain Sciences* 4, 515–560.

Colby, K.M.; Hilf, F.D.; Weber, S.; et al. (1972). "Turing-Like Indistinguishability Tests for the Validation of a Computer Simulation of Paranoid Processes". *Artificial Intelligence* 3, 199–221.

Descartes, R. (1637). *Discourse on the Method of Rightly Conducting the Reason and Seeking for Truth in the Sciences*. In Haldane, E.S., & Ross, G.R.T. (trans.), *The Philosophical Works of Descartes*. Cambridge, UK: Cambridge University Press, 1970, Vol. I, pp. 79–130.

Harnad, S. (1989). "Minds, machines and Searle". *Journal of Experimental and Theoretical Artificial Intelligence* 1(1), 5–25.

Hirsch, Jr., E.D. (1987). *Cultural literacy: What every American needs to know*. Boston: Houghton Mifflin.

Lenat, D.B. (1995). "CYC: A large-scale investment in knowledge infrastructure". *Communications of the ACM* 38(11), 33–38.

Loebner, Hugh Gene (1994). "In Response [to Shieber 1994a]", *Communications of the ACM* 37(6): 79–82.

Loebner, H.G. (2003). "Home page of the Loebner prize—'the first Turing test' ", [http://www.loebner.net/Prizef/loebner-prize.html].

Moor, J.H. (ed.) (2000–2001). Special issues on the Turing test: Past, present and future, *Minds and Machines*, Vol. 10, No. 4, and Vol. 11, No. 1.

Moor, J.H. (ed.) (2003). *The Turing test: The elusive standard of artificial intelligence*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Putnam, H. (1960). "Minds and machines". In Hook, Sidney (ed.), *Dimensions of mind: A symposium*. New York: New York University Press, 148–179.

Rapaport, W.J. (1988). "Syntactic Semantics: Foundations of Computational Natural-Language Understanding". In Fetzer, J.H. (ed.), *Aspects of artificial intelligence*. Dordrecht, Holland: Kluwer Academic Publishers, 81–131.

Rapaport, W.J. (2000). "How to pass a Turing test: Syntactic semantics, natural-language understanding, and first-person cognition". In Akman & Blackburn 2000, 467–490.

Rey, G. (2002). "Searle's misunderstandings of functionalism and strong AI". In Preston, J., & Bishop, M. (eds.), *Views into the Chinese room: New essays on Searle and*

artificial intelligence. Oxford: Oxford University Press, 201–225.

Sato, Y., & Ikegami, T. (2004). "Undecidability in the Imitation Game". *Minds and Machines* 14(2), 133–143.

Searle, J.R. (1980). "Minds, brains, and programs". *Behavioral and Brain Sciences* 3, 417–457.

Shannon, C.E.; & McCarthy, J. (1956). "Preface". In Shannon, C., & McCarthy, J. (eds.), *Automata studies*. Princeton, NJ: Princeton University Press, v–viii.

Shieber, S.M. (1994a). "Lessons from a restricted Turing test". *Communications of the ACM* 37(6), 70–78.

Shieber, S.M. (1994b). "On Loebner's lessons". *Communications of the ACM* 37(6), 83–84.

Shieber, S.M. (ed.) (2004a). *The Turing test: verbal behavior as the hallmark of intelligence*. Cambridge, MA: MIT Press.

Shieber, S.M. (2004b). "The Turing test as interactive proof" (unpublished ms.). [http://www.eecs.harvard.edu/shieber/Biblio/Papers/turing-interactive-proof.pdf].

Turing, A.M. (1936). "On computable numbers, with an application to the *Entscheidungsproblem*". *Proceedings of the London Mathematical Society*, Ser. 2, Vol. 42, 230–265.

Turing, A.M. (1950). "Computing machinery and intelligence". *Mind* 59, 433–460.

Weizenbaum, J. (1966). "ELIZA—A computer program for the study of natural language communication between man and machine". *Communications of the ACM* 9, 36–45.

Weizenbaum, J. (1967). "Contextual understanding by computers". *Communications of the ACM* 10, 474–480.

Author's Biography

William J. Rapaport (B.A. (mathematics), University of Rochester (1968); M.A., Ph.D. (philosophy), Indiana University (1974, 1976); M.S. (computer science), SUNY Buffalo (1984)) is an Associate Professor in the Department of Computer Science and Engineering, Adjunct Professor in the Department of Philosophy, and a member of the Center for Cognitive Science, all at State University of New York at Buffalo. Previously, he was an Associate Professor in the Department of Philosophy at SUNY Fredonia. His research interests are in cognitive science, artificial intelligence, computational linguistics, knowledge representation and reasoning, contextual vocabulary acquisition, philosophy of mind, philosophy of language, critical thinking, and cognitive development. He has received grant support from the National Science Foundation, the National Endowment for the Humanities, the SUNY Research Foundation, and FIPSE. He is the co-author or co-editor of two books, has written over 100 articles in computer science, philosophy, cognitive science, and education, and has been Review Editor of the cognitive science journal *Minds and Machines* and on the editorial boards of journals in philosophy, computational linguistics, and cognitive science. He has supervised (or is supervising) 6 Ph.D. dissertations and over 25 master's theses or projects.