Introduction

How do you tell if something is a meter long? You compare it with an object postulated to be a meter long. If the two are indistinguishable with regard to the pertinent property, their length, then you can conclude that the tested object is the given length.

Now, how do you tell if something is intelligent? You compare it with an entity postulated to be intelligent. If the two are indistinguishable with regard to the pertinent properties, then you can conclude that the tested entity is intelligent.

A test of intelligence such as this, based on *indistinguishability,* has a certain plausibility to it, and a long history. In its modern form, such a test has come to be known as the Turing Test, after Alan Turing, the scientist who most explicitly and concretely proposed it.

In 1950, Turing published a paper entitled "Computing Machinery and Intelligence" in the journal *Mind*. In the paper, he defined a simple test as a thought experiment to crystallize the questions surrounding the possibility of an intelligent artifact. In essence, Turing proposed to test whether the artifact was indistinguishable from a person with regard to what he took to be the pertinent property, *verbal behavior*. But unlike the case of meter measurement, the identification of the pertinent properties for intelligence are subtle, and ramifies widely in the foundation of the philosophy of mind.

Although the philosophical issues that the Turing Test raises had arisen before (as seen in part I) in philosophy, science, and literature, Turing's encapsulation of them in his simple thought experiment stands out as a trenchant codification of these issues around which discussion can naturally revolve. The familiarity and

immediacy of the concept can be seen in the ubiquity of the term both in technical parlance and in the popular mind. Turing is undoubtedly the only computer scientist to have a Broadway play written about him, Hugh Whitemore's *Breaking the Code* with Derek Jacobi as Turing in its New York premiere. He has been the inspiration for novels, such as Christos Papadimitriou's clever *Turing* (2003). His Test shows up in comic strips (figure 1) and collegiate humor magazines (figure 2).

This collection brings together a set of works that explore the philosophical issues surrounding the Turing Test as a test of intelligence. An exhaustive compilation of papers on the Turing Test would be impossible for reasons of both the depth and breadth
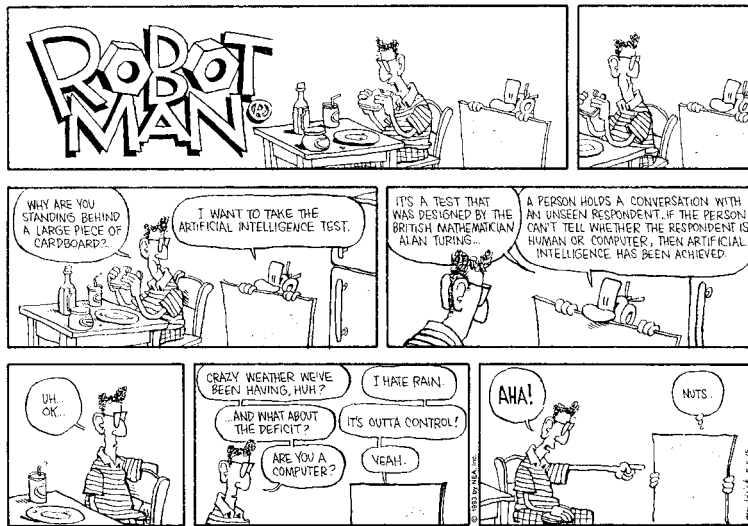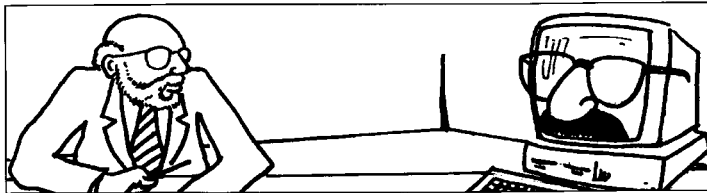

Figure 1
Robotman, Jim Meddick, 1993.

TURING TEST

ARE YOU A COMPUTER?
THE TURING TEST CAN TELL!

10:00 AM: Easy Questions
Turing: Hello, I am Dr. Alan Turing. I'm going to ask you a few simple questions. State your name please.
Computer: Simon III.
T: Your full name?
Simon: Max Felix John Simon III.
T: Where are you from?
Max: ...Belize.
T: Have any family?
M: No.
T: What happened to them?
M: ...Earthquake.
T: Sorry. What do you do for a living?
M: Manufacturing.
T: You mean you work in a factory, or you're a supervisor?
M: Yes.

10:30 AM: Psychology
T: Now we're going to look at some pictures. What does this look like to you?
M: It looks like an ink splotch.
T: Okay, but look deep into it, let your imagination run wild.
M: Two ink splotches.
T: Deeper, wilder.
M: Actually, four ink splotches.
T: Maybe you don't understand. Look at this picture, and then tell me what it reminds you of, like a butterfly, or a face. Try this one.
M: Butterfly.
T: Good! Now try this one.
M: Butterfly face.

11:15 AM: Tricky Questions
T: It's time to do the laundry. You need to wash a white t-shirt, a blue pinstriped button-down, a black turtleneck, and a red sweater with white polka dots. How many loads will you need to do?
M: That could all fit in one load.
T: Darks and lights in the same load?
M: Two loads would do it.

12:00 AM: Psychology Again
T: Let's play a little game. I'm going to say a word, and you say the next word that comes into your mind.
M: Sounds easy.
T: Okay, here we go. Dog.
M: Doily.
T: No, no, no. I say a word, then you say what it makes you think of.
M: Oh, I get it. Try me again.
T: Electricity.
M: Food.
T: *Food?*
M: That's what I thought of. Wouldn't it be neat if electricity were not a painful shock, but a *tasty treat*? Of course, that's not the case for humans like us.

12:45 AM: Nap Time
T: Are you a computer?
M: Nope.
T: You'd be surprised how many fall for that one.
M: Not me.

1:30 AM: Math Time
T: What's fifty-six times thirty-three?
M: One thousand eight hundred forty-eight.
T: You're pretty fast!
M: Those are my favorite numbers.
T: All right, how about five thousand and two divided by sixty-one?
M: Eighty-two.
T: Right again! Are you some sort of math whiz?
M: ...Those are more of my favorite numbers.

3:00 PM: Computer Science
T: How good are you with computers?
M: About the same as an average person, I'd say... how about you?
T: Here's an easy question. Let's say you type the following program into a computer:
10 PRINT "HELLO"
20 GOTO 10
What do you think would happen?
M:
HELLO
HELLO
HELLO
HELLO
HELLO
HELLO
HELLO
HELLO
HELLO
HELLO
HELLO
HELLO
HELLO
HELLO

DSJ

LAMPOON 9

Figure 2
Are you a computer? The Turing test can tell, David S. Joerg, *The Harvard Lampoon*, 1994.

of the Test's influence. In terms of depth, literally thousands of papers have been written on the possibility of machine intelligence since Turing's test was first proposed; it would be hard to imagine that any of them would not be influenced by Turing's work. In terms of breadth, the subject of the Turing Test arises not only in the context of the question of machine intelligence but in many other areas as well. Scholars have speculated about the likelihood of actually constructing a machine capable of passing the Test, argued about the use of the Test as a goal for research in the field of artificial intelligence, proposed and analyzed variations of the Test, wondered about the ethical implications of a Turing-Test-passing entity, and so forth. (The end of this section includes a discussion of some of these issues, with references to the literature.) Although these issues may be interesting in their own right, and discussion of them may be improved by being informed about the fundamental philosophical issues raised by the Turing Test, they are largely separable from the more basic concerns here.

For these reasons, this collection comprises three types of works most useful in developing a sense of the philosophical issues raised by the Turing Test. It starts with a look to philosophical precursors, early writings by Descartes and others who were the first to propose indistinguishability tests to resolve certain theological questions. In particular, Descartes first pinpointed *verbal behavior* as the crucial property for distinguishing humans from beasts, the soul-bearing from the soul-less. Second, it brings together for the first time all of Turing's own writings related to the Turing Test—the *Mind* article of course, but also little known ephemeral material. The latter answers some questions that are interesting in their own right and subjects of scholarly contention, and Turing's own status as a revolutionary mathematical thinker and a founder of modern computer science makes his personal views on the subject illuminating. Third, the book includes a select set of seminal papers culled from the philosophical literature that directly address the issue of the Turing Test as a test for intelligence, providing a broad spectrum of views that together comprise some of the most important and widely cited works on the subject. In order to sample the immediate reaction from the

philosophical community, the collection incorporates essentially all of the direct responses to the *Mind* article published in that journal.

The remainder of this introduction provides some background on Turing and his Test, ending with a brief exposition of the variety of issues, philosophical and otherwise, that have arisen around the general topic of the Turing Test. The following chapters present the three sets of readings, each introduced with background material that is intended to be read both as a map of the readings themselves and, taken together and sequentially, a self-contained essay on the Turing Test.

Who Was Alan Turing?

Alan Turing was born in 1912 in London and educated at King's College, Cambridge, and at Princeton where he wrote his doctoral dissertation under the eminent logician Alonzo Church. Today, we would call Turing a computer scientist, but during his career he was naturally thought of as a mathematician and logician, simply because he had not invented computer science yet. This is not hyperbole: Turing can be credited with perhaps the single most fundamental result in computer science, the existence of uncomputable functions. In the course of his solution to one of David Hilbert's famous problems, the "Entscheidungsproblem", the twenty-three-year-old Turing invented the first formal model of computation, the so-called "Turing machine", and argued that the notion "computability by a Turing machine" could serve as an apt substitute for the vague notion of computability in general. He published his seminal paper "On Computable Numbers" in 1936, arguably the first and most important paper in computer science (Turing 1936).

After completing a doctorate at Princeton in 1938 and postdoctoral work back in England, he joined the British Foreign Office as part of a government intelligence unit. His efforts led to the breaking of the German Enigma code, a central contribution to the Allied war effort, by the use of electromechanical devices for carrying out repetitive calculations, a nonprogrammable precursor of the computer.

His experiences at the Bletchley Park code-breaking unit led Turing to further work on the design and construction of early computers, including the Automatic Computing Engine at the National Physical Laboratory and the Manchester machine at the University of Manchester. As one of the first computer programmers, writing programs for the not-yet-built Manchester machine, Turing first came upon and discussed the idea of the subroutine. And in his writings on the question of whether machines could think, he laid the groundwork for the computer science subfield of "artificial intelligence" (AI), the study of the computational explication and replication of behaviors that are associated with intelligence in humans.

Through his research, Turing thus set the foundation for the major subfields of computer science: the theory of computation, the design of hardware and software, and the study of artificial intelligence. Tragically, his career came to a premature end. After his 1952 arrest under British laws against homosexuality, the authorities required him to undergo a draconian hormone treatment for his "condition". Two years later, he died of cyanide poisoning, apparently self-administered, though the nature of his death is still controversial. If his death was suicide, it seems likely that his treatment under outmoded sodomy laws contributed directly to it. In any case, Turing's premature death is certainly one of the great intellectual tragedies of the twentieth century.[1]

What Is the Turing Test?

Turing proposed the Turing Test in the context of the question "Can machines think?"[2] but not as a way of answering the question.

---

1 The authoritative biography of Turing is that of Hodges (1983), which is strongly recommended for any student of the Turing Test.

2 Turing used the terms "think" and "be intelligent" as if they were synonyms, as one can tell by a simple comparison of his article's title and first sentence. In common usage, the two often mean quite distinct things. When I say that my son is intelligent, I usually mean something beyond the fact that he is capable of thought. However, I and many authors follow Turing's practice, taking the notion of "being intelligent" under which it means "being capable of thought", rather than "being smart".

Rather, he found the original question "too meaningless to deserve discussion" and sought to replace it with something more concrete. He found his concrete form in a game-theoretic crystallization of Descartes's observation that flexibility of verbal behavior is the hallmark of humanness. He proposed an "imitation game" in which an interrogator attempts to determine which of two agents[3] is human and which a machine, based on purely verbal interaction with both. If the interrogator is not able to reliably determine which is the human, the machine has passed the test. This test has come to be known as the "Turing Test".

More specifically, Turing imagined the following setup: The two agents A and B and the interrogator C are each placed in separate rooms. C knows only that one of the agents is a human and one a machine, and is not, of course, aware of which is which. C carries on conversations with each of the agents by passing typewritten notes through a courier to each room and getting typewritten replies back. After some indeterminate but appropriately lengthy interaction, C must make a decision as to which of A and B is the machine. Now, by merely guessing blindly, C will get the answer right half the time, so any single test of this sort is not definitive, but one can imagine C going through this exercise many times, and verifying whether C can do significantly better than chance at determining which agent is the machine. If not, that is, if C can do no better than random guessing, the machine is said to have passed the Turing Test.

This, in sum, is the Turing Test. It has many attractive aspects to it as a criterion for intelligence (or a replacement). The test is *operational* or *behavioral* so as to get around (so Turing thought) the tricky definitional questions of intelligence. When asked to define "obscenity", Supreme Court Justice Potter Stewart famously demurred: "I know it when I see it." (Stewart 1964) Maybe intelligence is like that—impossible to define, but you know it when you see it. The use of *verbal interaction* is desirable because it

---

3 Again on a terminological note, the term "agent" is used here and throughout as a generic term for any entity—human or machine, simple or sophisticated—that displays behavior. The notion of agency implicit in the term should be construed broadly.

abstracts from incidental properties like visual appearance that might immediately answer the question of which entity is the machine, but not on the basis of facts pertinent to the question of intelligence. The *open-ended* nature of the interaction is crucial because it allows any possible area of human experience to be used as criterial in the decision. The *statistical* aspect of the decision is fortuitous since on any given running of a Test, even between two people, one of the two will be selected out. Failure on a single test therefore cannot be taken to be indicative of anything at all; the statistical approach moves the test in the direction of testing a disposition or capacity, rather than a singleton behavior.

Issues Surrounding the Turing Test

The commentaries on the Turing Test in this volume are included because they bear on the primary philosophical issue raised by the *Mind* paper, the relationship between the Turing Test and intelligence. The big question, or as referred to henceforth, the Big Question, is "Is passing a Turing Test criterial for intelligence?" That length is the pertinent property for determining meter-hood is uncontroversial. But exactly what the pertinent property or properties are for assessing intelligence, and whether verbal behavior in particular is the one, has become the key issue regarding the Turing Test.

The views on the Big Question have been varied. Some have argued that the Test is too difficult as a test of intelligence; intelligent agents would routinely fail. Robert French (1990, *chapter 13*), for instance, has argued that even with its restriction to verbal interaction, incidental properties, such as a lack of idiosyncratic cultural knowledge, could easily unmask a machine. Others view the test as too easy. Searle (1980, *chapter 14*), Block (1981, *chapter 15*), and Gunderson (1964, *chapter 9*) each argue that the Test misses testing for some crucial property, so that in principle at least unthinking machines could pass a Turing Test. (These various considerations can be seen as splitting the Big Question into multiple Big Questions—concerning the Turing Test as a necessary condition, as a sufficent condition, and so forth—complexities that are

explored in detail in this book.) In support of a positive answer to the Big Question, some philosophers find the reasoning from passing a Turing Test to ascription of intelligence to be sound, including Dennett (1985, *chapter 16*), or at least—as Moor (1976, *chapter 17*) would have it—a convincing source of evidence. Finally, Turing's original view is reiterated by others: the Test should not be taken as criterial at all, but as a replacement for the question, and one with useful outcomes. Such a view, sidestepping the Big Question entirely, is recommended by Chomsky (*chapter 20*).

Beyond the Big Question, the Turing Test raises a wide variety of other issues. Coverage of such topics is well beyond the scope of this volume, but some of them are listed below to serve as entry into the appropriate literature.

Pragmatic Issues

In practice, could a machine pass the Turing Test? If so, when will such an event come to pass? Understanding the independence of this question and the Big Question is important. One can believe that a Turing-Test-passing machine is not intelligent, yet still believe that a machine may pass the Test at some future date. One would simply have to conclude that the performance on the Test is not proof of the intelligence of the machine.

This question is only interesting, of course, under the assumption that a machine could pass *in principle,* which many of the papers in this volume take to be controversial at best. In any case, it is clear that at current levels of technology the answer is "no". Some would argue that even assuming the ability in principle, machines will never be able to pass the Test in practice. French's paper (1990, *chapter 13*) can be read in this way. Others believe that only a few decades of continued engineering progress are required. Mitchell Kapor and Raymond Kurzweil have an outstanding bet regarding whether a machine will pass the Turing Test by 2029, for example (Kurzweil and Kapor 2002).

As it turns out, the history of research in AI is littered with predictions of the imminent passing of the Turing Test. Dreyfus (1979) has catalogued examples of this sort of hubris. One lesson learned from the past half-century of AI research is that the

problems involved in generating intelligent behavior are deeper and more profound than many had ever imagined. AI researchers, even while making continued progress in many areas, more rarely make the bold predictions of walking, talking robots right around the corner. This leads directly to the issue of whether work towards passing the Turing Test is an appropriate research methodology.

Methodological Issues

Is passing the Turing Test an appropriate research goal? Research in artificial intelligence is concerned with computational explication and replication of behavioral capacities that are associated with intelligence in humans. Construction of a program capable of passing the Turing Test would seem a natural goal for the field. As the early readings in this volume attest, the duplication of human intelligence has inspired scholars for centuries. Indeed, the Turing Test did serve as a defining inspiration in the early history of AI research. Even now, some researchers take passing the Turing Test as fundamental to the field of AI research. Ginsberg (1993), for instance, defines the field as "the enterprise of constructing a physical symbol system that can reliably pass the Turing Test."

But as a goal for a concrete research program (as opposed to a philosophical thought experiment), the Turing Test is fraught with problems. First, insofar as the test is not a necessary condition for intelligence, it encumbers research efforts with extraneous burdens. In particular, as French (1990, *chapter 13*) argues, it forces the modeling of human idiosyncrasies that have nothing to do with intelligence per se. Second, the Test permits conclusions only of success or failure; there is no interesting notion of *almost* passing a Turing Test. Thus, failure in a Turing Test is not diagnostic of any particular deficiency in the test subject, and so provides no mid-course guidance for research direction towards success. Finally, it aims at a goal—the construction of an artificial human intelligence—that is not intrinsically desirable, as we already have plenty of intelligences with human abilities and disabilities and can too easily make more. Hayes and Ford (1995) make these arguments especially forcefully, concluding that the Turing Test is

simply inappropriate—indeed, harmful—as a goal of research in AI. A novel argument of theirs is that the test falls prey to the evolving abilities of the judges; people these days easily unmask Eliza-like systems that would have been convincing only twenty-five years ago. For related reasons, Whitby (1996) calls the Turing Test "AI's biggest blind alley".

Nonetheless, attempts to run Turing-like tests as competitions crop up on occasion, sometimes motivated by their entertainment value, sometimes as a purported prod to scientific research. Shieber (1994) presents a critique of a particular effort along these lines, arguing that carrying out such competitions is grossly premature at best.

Ethical and Normative Issues

Should a machine that could pass a Turing Test be subject to the rights and responsibilities accorded people? Suppose we stipulate the existence of Turing-Test-capable machines. Would it be ethical to turn them off? Should they be allowed to vote? Such science-fiction scenarios have been imagined by many. (One such scenario is the basis for the evocatively titled movie *AI: Artificial Intelligence,* for instance.) Futurists have started examining the issues in some detail (Brooks 2002; Kurzweil 1999; Moravec 1999). Science fiction authors have been exercised over the matter at least since Samuel Butler's *Erewhon*. In any case, such ethical questions, interesting and potentially important as they might be, are posterior to the Big Question of whether passing the Turing Test is criterial for thinking.

On the other hand, the corresponding ethical questions concerning *thinking* machines (ex hypothesi, as opposed to Turing-Test-passing machines) are not posterior to the Big Question, and are therefore appropriate to discuss before its resolution. But, of course, they are not questions about the Turing Test at all.

Alternative Tests

Is there a better way to design a Turing Test? Some researchers have attempted to solve problems in the design of the Turing Test through alternative formulations.

Stevan Harnad (2000), for instance, proposes a hierarchy of Turing-like tests, of which the classical Test is categorized as T2, with T3 expanding the interaction to allow full interaction with the device through auditory, visual, even tactile channels, T4 further requiring internal microfunctional indistinguishability, and T5 requiring indistinguishability at every level.

Watt (1996), arguing that ascription of mental states to others is, for certain purposes, crucial to Turing-like tests, proposes an "inverted Turing Test", in which the machine under test serves as the interrogator, trying to distinguish a human and a machine in a traditional Turing Test. "A system passes if it is itself unable to distinguish between two humans, or between a human and a machine that can pass the normal Turing test, but which can discriminate between a human and a machine that can be told apart by a normal Turing test with a human observer." (Watt 1996) Many respondents note that, regardless of its other problems, the inverted Turing Test can be emulated through a normal Turing Test (Bringsjord 1996; French 1996).

Dowe and Hajek (1998) extend the Turing Test with a non-behavioral component, requiring that the machine be sufficiently compact, that is, the size of the program and data that it uses be small relative to its performance, so as to circumvent the type of objections to the Turing Test (e.g., those of Searle or Block) detailed in the final part of this volume. The sufficiency of such a modification is based on the close relationship between inductive inference and descriptional complexity. (Along the same lines, Hernandez-Orallo [2000] proposes to replace the Turing Test with a series of psychometric tests based on completing sequences graded according to their descriptional complexity.)

Application Issues

Is the Turing Test good for anything practical? One might think that the answer is definitively negative, given that no machine is close to passing a Turing Test nor is one likely to do so in the foreseeable future. Abstractly, however, there is still the question of whether a Turing-Test-passing machine would be of utility;

Ronald and Sipper (2001), for instance, answer this question in the negative.

Furthermore, it is exactly the inability of computers to emulate certain behaviors that people find straightforward that leads to concrete and useful applications that have arisen under the name "reverse Turing Tests". A reverse Turing Test is a Turing Test intended to be administered by a *computer* as judge. The notion was first proposed by Naor (1996), and developed by Coates et al. (2001) and von Ahn et al. (2004). Reverse Turing Tests can be used to discriminate against computer agents in access to computer services. For instance, web portal company Yahoo! requires the passing of a reverse Turing Test as a condition of signing up for a free email account. In order to sign up, an agent must type in a word that has been presented in typographically deformed form. Although people have no problem identifying the word, the optical character recognition technology that would be required of a computer agent is beyond the state of the art.

All of these issues are important in their own way and show the ability of the Turing Test to insinuate itself broadly into a tremendous range of intellectual areas. Yet all take a back seat to the key question of the relation between the Turing Test and intelligence explored further in the pages ahead.