Chapter 3
Computing Machinery and Intelligence

Alan M. Turing

Editors' Note: The following is the article that started it all – the article by Alan Turing which appeared in 1950 in the British journal, *Mind*. Accompanying the article are three running commentaries by Kenneth Ford, Clark Glymour, and Pat Hayes of the University of West Florida; Stevan Harnad of the University of Southampton; and Ayse Pinar Saygin of the University of California, San Diego, designated respectively by the symbols:♠,♣, and♥. A fourth commentary by John Lucas of Merton College, Oxford, is found in Chapter 4.

3.1 The Imitation Game

I propose to consider the question, "Can machines think?"♣ This should begin with definitions of the meaning of the terms "machine" and "think". The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by

Manchester University

♣Harnad: Turing starts on an equivocation. We know now that what he will go on to consider is not whether or not machines can think, but whether or not machines can do what thinkers like us can do – and if so, how. Doing is performance capacity, empirically observable. Thinking is an internal state. It correlates empirically observable as neural activity (if we only knew which neural activity corresponds to thinking!) and its associated quality introspectively observable as our own mental state when we are thinking. Turing's proposal will turn out to have nothing to do with either observing neural states or introspecting mental states, but only with generating performance capacity indistinguishable from that of thinkers like us.

R. Epstein et al. (eds.), *Parsing the Turing Test.*

another, which is closely related to it and is expressed in relatively unambiguous words.♠♣♥

The new form of the problem can be described in terms of a game which we call the "imitation game".♠♣ It is played with three people, a man (A), a woman

♠ FORD, GLYMOUR, AND HAYES: Turing derides deciding the question by an empirical survey of which sorts of objects the word "think" or its synonyms are positively applied to. Presumably, in 1950 people rarely if ever said of machines that they think, and few people in 1950 would have said that any machine, then or in the future, could *possibly* be said to think. The procedure is absurd because what people say, even what almost everyone agrees in saying, is often wildly wrong: a century before almost everyone would have agreed to the proposition that angels think.

♣ HARNAD: "Machine" will never be adequately defined in Turing's paper, although what will eventually be known as the "Turing Machine," the abstract description of a computer, will be. This will introduce a systematic ambiguity between a real physical system, doing something in the world, and another physical system, a computer, simulating the first system formally, but not actually doing what it does: an example would be the difference between a real airplane – a machine, flying in the real world – and a computer simulation of an airplane, not really flying, but doing something formally equivalent to it, in a (likewise simulated) "virtual world."

A reasonable definition of machine, rather than Turing Machine, might be any dynamical, causal system. That makes the universe a machine, a molecule a machine, and also waterfalls, toasters, oysters, and human beings. Whether or not a machine is man-made is obviously irrelevant. The only relevant property is that it is "mechanical" – i.e., behaves in accordance with the cause-effect laws of physics.

"Think" will never be defined by Turing at all; it will be replaced by an operational definition to the effect that "thinking is as thinking does." This is fine, for thinking cannot be defined in advance of knowing how thinking systems do it, and we do not yet know how. But we do know that we thinkers do it, whatever it is, when we think and we know when we are doing it (by introspection). So thinking, a form of consciousness, is already ostensively defined by just pointing to that experience we all have and know.

Taking a statistical survey like a Gallup poll instead, to find out people's opinions of what thinking is would indeed be a waste of time, as Turing points out – but then later in the paper he needlessly introduces the equivalent of a statistical survey as his criterion for having passed his Turing Test!

♥ SAYGIN: Operational definition: a definition of a theoretical construct that is stated in terms of concrete, observable procedures (Pelham, 1999). While some readers believe the imitation game is only a thought experiment, I think it is pretty clear that Turing is proposing an operational definition for machine thought. One could argue whether this is the best way to test for machine intelligence, but that would be a discussion of construct validity, i.e., the quality of someone's operational definitions, not the existence or lack thereof.

♠ FORD, GLYMOUR, AND HAYES: Turing's use of the singular here may be misleading, as we will see. There are many versions of "the" imitation game, and Turing himself seems to slide between them without giving the reader adequate notice. It might be best to take this paragraph as a description of a family of "games" that share a common theme: a real exemplar and an imitator, respectively trying to help and to fool a judge.

♣ HARNAD: Another unfortunate terminological choice: "Game" implies caprice or trickery, whereas Turing in fact means serious empirical business. The game is science, the future science of cognition – actually a branch of reverse bioengineering. "Imitation" has connotations of fakery or deception too, whereas what Turing will be proposing is a rigorous empirical methodology for testing theories of human cognitive performance capacity (and thereby also theories of the thinking that presumably engenders it). Calling this an "imitation game" (instead of a methodology for reverse-engineering human cognitive performance capacity) has invited generations of needless misunderstandings (Harnad, 1992).

(B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman.♣ He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A". The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

Now suppose X is actually A, then A must answer. It is A's object in the game to try and cause C to make the wrong identification. His answer might therefore be

"My hair is shingled, and the longest strands are about nine inches long."

In order that tones of voice may not help the interrogator the answers should be written, or better still, typewritten.♣ The ideal arrangement is to have a teleprinter communicating between the two rooms. Alternatively the question and answers can be repeated by an intermediary. The object of the game for the third player (B) is to help the interrogator. The best strategy for her is probably to give truthful answers. She can add such things as "I am the woman, don't listen to him!" to her answers, but it will avail nothing as the man can make similar remarks.

We now ask the question, "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is

---

♣ HARNAD: The man/woman test is an intuitive "preparation" for the gist of what will eventually be the Turing Test, namely, an empirical test of performance capacity. For this, it is first necessary that all non-performance data be excluded (hence the candidates are out of sight). This sets the stage for what will be Turing's real object of comparison, which is a thinking human being versus a (nonthinking) machine, a comparison that is to be unbiased by appearance.

Turing's criteria, as we know, will turn out to be two (though they are often confused or conflated): (1) Do the two candidates have identical performance capacity? (2) Is there any way we can distinguish them, based only on their performance capacity, so as to be able to detect that one is a thinking human being and the other is just a machine? The first is an empirical criterion: Can they both do the same things? The second is an intuitive criterion, drawing on what decades later came to be called our human "mind-reading" capacities (Frith and Frith, 1999): Is there anything about the way the candidates go about doing what they can both do that cues me to the fact that one of them is just a machine?

Turing introduces all of this in the form of a party game, rather like 20-Questions. He never explicitly debriefs the reader to the effect that what is really at issue is no less than the game of life itself, and that the "interrogator" is actually the scientist for question (1), and, for question (2), any of the rest of us, in every one of our daily interactions with one another. The unfortunate party-game metaphor again gave rise to needless cycles of misunderstandings in later writing and thinking about the Turing Test.

♣ HARNAD: This restriction of the test exclusively to what we would today call email interactions is, as noted, a reasonable way of preparing us for its eventual focus on performance capacity alone, rather than appearance, but it does have the unintended further effect of ruling out all direct testing of performance capacities other than verbal ones; and that is potentially a much more serious equivocation, to which we will return. For now, we should bear in mind only that if the criterion is to be Turing-indistinguishable performance-capacity, we can all do a lot more than just email!

played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?"♠♣♥

♠ FORD, GLYMOUR, AND HAYES: This question can be understood in several ways, depending on what one takes Turing to mean by "this game." It is usually understood to mean a version of the game where, in the terminology of 1951 (published in 1950, exposition in 1951), a "man" – and a machine each try to persuade the judge that they are the human being. However, taken literally, and we see no reason not to, "the game" refers to the game just described, and then Turing seems to be proposing a comparative test of the ability of a man to pretend to be a woman, as against the ability of a computer to pretend to be a woman (the contestant in each case being a real woman, of course). This reading – the "gender test," in contradistinction to the "species test" usually assumed – may seem strange, but it has a number of subtle advantages, including the fact that the judge is not given a predisposition to be particularly alert for signs of non-human behaviour, and the fact that the players, man and machine, both have imitation tasks.

The critical question is whether, in typed conversation, a computer can pass as a woman as convincingly – and as *unconvincingly* – as can a man. In the gender test version, as posed, the test is not only of conversational competence, but also of a special kind of knowledge: the computer must "know" what it is like for a man to try to converse like a woman (Hayes and Ford, 1995).

Psychological speculations aside, one might reasonably object that different men and women and judges would yield widely varying accuracies of judgement, or that a sufficiently skilled judge, given sufficient time, would be able to distinguish most men from most women, so that to qualify as thoughtful, the computer would have a very low bar.

Many writers assume the game should be played with the question of gender (being female) replaced by the question of species (being human), so that the judge is faced with the task of differentiating a human participant from a machine pretending to be human. Notice that under this most common interpretation, the Turing Test slides from a test for intelligence, to a test of the ability of humans to distinguish members of their own species from mechanical impostors. This version is often called the "species version," and is the most popular understanding of the Turing Test, but it was not Turing's. In the gender test, the judge is still thinking about the differences between women and men, not humans and machines. The hypothesis that one of his subjects is not human is not even in his natural space of initial possibilities. This judge has exactly the same problem to solve as a judge in the original imitation game and could be expected to bring the same attitudes and skills to the problem. For a discussion of the two versions and the advantages of Turing's own version see (Genova, 1994; Sterrett, 2000).

♣ HARNAD: Here, with a little imagination, we can already scale up to the full Turing Test, but again we are faced with a needless and potentially misleading distraction: Surely the goal is not merely to design a machine that people mistake for a human being statistically more often than not! That would reduce the Turing Test to the Gallup poll that Turing rightly rejected in raising the question of what "thinking" is in the first place! No, if Turing's indistinguishability criterion is to have any empirical substance, the performance of the machine must be equal to that of a human being – to anyone and everyone, for a lifetime.

♥ SAYGIN: Note that here the machine takes the part of A, the man. The man was trying to convince the interrogator that he actually was the woman. Now that the machine takes the place of the man in the game, will it be trying to convince the interrogator that it is a woman? The answer could be yes or no, depending on interpretation (Piccinini, 2000; Saygin et al., 2000; Traiger, 2000). As it is now generally understood, the Turing Test tries to assess a machine's ability to imitate a human being, rather than its ability to simulate a woman in an imitation game. Most subsequent remarks on Turing's work in the following 50 years, as reviewed in Saygin et al. (2000), ignore the gender issue, and if they discuss the imitation game at all, consider a game that is played between a machine (A), a human (B), and an interrogator (C) whose aim is to determine which one of the two entities he/she

3.2 Critique of the New Problem

As well as asking, "What is the answer to this new form of the question", one may ask, "Is this new question a worthy one to investigate?" This latter question we investigate without further ado, thereby cutting short an infinite regress.♠

The new problem has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a man.♣ No engineer or chemist claims to be able to produce a material which is indistinguishable from the human skin. It is possible that at some time this might be done, but even supposing this invention available we should feel there was little point in trying to make a "thinking

---

is conversing with is the human. In many cases the imitation game is considered irrelevant and the discussion revolves around the vague idea of a digital computer "passing for" a human and the relation this possibility bears to Artificial Intelligence (AI). The imitation game is not an analogy Turing introduced but failed to properly execute, nor is it a joke or a commentary on gender roles in society (unlike that of [Genova, 1994; Lassègue, 1996]). But I will suggest below that the seemingly quirky gender-based imitation game is in fact an ideal and fair test for machine intelligence.

♠ FORD, GLYMOUR, AND HAYES: Turing's intellectual strategy is to replace questions of a traditional philosophical form, e.g., provide necessary and sufficient conditions for something to be intelligent, with related questions for which there is a hope of providing an answer empirically or mathematically.

♣ HARNAD: It would have had that advantage, if the line had only been drawn between appearance and performance or between structure and function. But if the line is instead between verbal and non-verbal performance capacities, then it is a very arbitrary line indeed and a very hard one to defend. As there is no explicit or even inferable defence of this arbitrary line in any of Turing's paper (nor of an equally arbitrary line between those of our "physical capacities" that do and do not depend on our "intellectual capacities"), I will take it that Turing simply did not think this through. Had he done so, the line would have been drawn between the candidate's physical appearance and structure on the one hand, and its performance capacities, both verbal and non-verbal, on the other. Just as (in the game) the difference, if any, between the man and the woman must be detected from what they do, and not what they look like, so the difference, if any, between human and machine must be detected from what they do, and not what they look like. This would leave the door open for the robotic version of the Turing Test that we will discuss later, and not just for the email version.

But before a reader calls my own dividing line between structure and function just as arbitrary, let me quickly agree that Turing has in fact introduced a hierarchy of Turing Tests here, but not an infinity of them. The relevant levels of this hierarchy will turn out to be only the following 5:

T1: The local indistinguishable capacity to perform some arbitrary task, such as chess. T1 is not really a Turing Test at all, because it is so obviously subtotal; hence the machine candidate is easily distinguished from a human being by seeing whether it can do anything else, other than play chess. If it cannot, it fails the test.

T2: The indistinguishable performance capacity in email (verbal) exchanges. This seems like a self-contained performance module, for one can talk about anything and everything, and language has the same kind of universality that computers (Turing Machines) turned out to have. T2 even subsumes chess-playing. But does it subsume star-gazing, or even food-foraging? Can the machine go and see and then tell me whether the moon is visible tonight and can it go and unearth truffles and then let me know how it went about it? These are things that a machine with email capacity alone cannot do, yet every human being can.

machine" more human by dressing it up in such artificial flesh.♣ The form in which we have set the problem reflects this fact in the condition which prevents the interrogator from seeing or touching the other competitors, or hearing their voices.♣ Some other advantages of the proposed criterion may be shown up by specimen questions and answers. Thus:

Q: Please write me a sonnet on the subject of the Forth Bridge.
A: Count me out on this one. I never could write poetry.
Q: Add 34957 to 70764.
A: (Pause about 30 s and then give as answer) 105621.
Q: Do you play chess?
A: Yes.
Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?
A: (After a pause of 15 s) R-R8 mate.

The question and answer method seems to be suitable for introducing almost any one of the fields of human endeavour that we wish to include.♣ We do not wish to penalize the machine for its inability to shine in beauty competitions,♣ nor to penalize a man for losing in a race against an aeroplane.♣ The conditions of our

---

T3: The indistinguishable performance capacity in robots (sensorimotor). This subsumes T2, and is (I will argue) the level of test that Turing really intended (or should have!).

T4: The indistinguishable external performance capacity, as well as internal structure and function. This subsumes T3 and adds all data that a neuroscientist might study. This is no longer strictly a Turing Test, because it goes beyond performance data, but it correctly embeds the Turing Hierarchy in a larger empirical hierarchy. Moreover, the boundary between T3 and T4 really is fuzzy: Is T3 or T4 blushing?

T5: The indistinguishable physical structure and function. This subsumes T4 and rules out any functionally equivalent but synthetic nervous systems: The T5 candidate must be indistinguishable from other human beings right down to the last molecule.

♣ HARNAD: Here Turing correctly rejects T5 and T4 – but certainly not T3.

♣ HARNAD: Yes, but using T2 as the example has inadvertently given the impression that T3 is excluded too.

♣ HARNAD: This correctly reflects the universal power of natural language (to say and describe anything in words). But "almost" does not fit the Turing criterion of identical performance capacity.

♣ HARNAD: This is the valid exclusion of appearance (moreover, most of us could not shine in beauty competitions either).

♣ HARNAD: Most of us could not beat Deep Blue at chess, nor even attain ordinary grandmaster level. It is only generic human capacities that are at issue, not those of any specific individual. On the other hand, just about all of us can walk and run. And even if we are handicapped (an anomalous case, and hardly the one on which to build one's attempts to generate positive performance capacity), we all have some sensorimotor capacity. (Neither Helen Keller nor Stephen Hawking are disembodied email-only modules.)

game make these disabilities irrelevant.♣ The "witnesses" can brag, if they consider it advisable, as much as they please about their charms, strength, or heroism, but the interrogator cannot demand practical demonstrations.♣

The game may perhaps be criticized on the ground that the odds are weighted too heavily against the machine. If the man were to try and pretend to be the machine he would clearly make a very poor showing. He would be given away at once by slowness and inaccuracy in arithmetic. May not machines carry out something which ought to be described as thinking but which is very different from what a man does? This objection is a very strong one, but at least we can say that if, nevertheless, a machine can be constructed to play the imitation game satisfactorily, we need not be troubled by this objection.♠♥

It might be urged that when playing the "imitation game" the best strategy for the machine may possibly be something other than imitation of the behaviour of a man. This may be, but I think it is unlikely that there is any great effect of this kind.

♣ HARNAD: Disabilities and appearance are indeed irrelevant. But non-verbal performance capacities are certainly not. Indeed, our verbal abilities may well be grounded in our non-verbal abilities (Cangelosi, 2001; Harnad, 1990; Steels and Kaplan, 1999). Actually, by "disability," Turing means non-ability, i.e., absence of an ability; he does not really mean being disabled in the sense of being physically handicapped, although he does mention Helen Keller later.

♣ HARNAD: This would definitely be a fatal flaw in the Turing Test, if Turing had meant it to exclude T3 – but I doubt that is what he meant. He was just arguing that it is performance capacity that is decisive (for the empirical problem that future cognitive science would eventually address), not something else that might depend on irrelevant features of structure or appearance. He merely used verbal performance as his intuition-priming example, without meaning to imply that all "thinking" is verbal and only verbal performance capacity is relevant.

♠ FORD, GLYMOUR, AND HAYES: It is clear that Turing intended passing the Turing Test to be an uncontroversial criterion *sufficient* for thought, not a necessary one. He allows machines to be inhumanly capable, for example. Indeed, the few electronic computers which existed at the time he was writing were already inhumanly capable in doing arithmetic, which is of course why large funds were expended in designing and constructing them.

The Turing Test is, however, a poorly designed experiment, depending entirely on the competence of the judge. As Turing noted above, a human would be instantly revealed by his comparative inadequacies in arithmetic unless, of course, the computer were programmed to be arithmetically incompetent. Likewise, according to media reports, some judges at the first Loebner competition (1991), a kind of Turing test contest held at the Computer Museum in Boston, rated a human as a machine on the grounds that she produced extended, well-written paragraphs of informative text at dictation speed without typing errors. (Apparently, this is now considered an inhuman ability in parts of our culture.)

♥ SAYGIN: Turing did not live long enough to reply to most critiques of his work, but in this paper he foresees many criticisms he thinks may be made by others and formulates some advance arguments against them (§6). Nevertheless, even those issues Turing diligently addresses have been raised in subsequent discussions. For example, he has subsequently been criticized both for his test being too anthropomorphic and limited (Millar, 1973), and on the basis that playing the imitation game is just one thing an intelligent machine can do and is not general enough for purposes of intelligence granting (Gunderson, 1964).

In any case there is no intention to investigate here the theory of the game, and it will be assumed that the best strategy is to try to provide answers that would naturally be given by a man.

3.3 The Machines Concerned in the Game

The question which we put in §1 will not be quite definite until we have specified what we mean by the word "machine". It is natural that we should wish to permit every kind of engineering technique to be used in our machines.♣ We also wish to allow the possibility than an engineer or team of engineers may construct a machine which works, but whose manner of operation cannot be satisfactorily described by its constructors because they have applied a method which is largely experimental.♣♥ Finally, we wish to exclude from the machines men born in the usual manner.♣ It is difficult to frame the definitions so as to satisfy these three conditions. One might for instance, insist that the team of engineers should be all of one sex, but this would not really be satisfactory, for it is probably possible to rear a complete individual from a single cell of the skin (say) of a man.♠ To do so would be a feat of biological technique deserving of the very highest praise, but we would not be inclined to regard it as a case of "constructing a thinking machine".♣ This prompts

---

♣ HARNAD: This passage (soon to be withdrawn!) implies that Turing did not mean only computers: any dynamical system we build is eligible (as long as it delivers the performance capacity). But we do have to build it, or at least have a full causal understanding of how it works. A cloned human being cannot be entered as the machine candidate (because we did not build it and do not know how it works), even though we are all "machines" in the sense of being causal systems (2000, 2003).

♣ HARNAD: Here is the beginning of the difference between the field of AI, whose goal is merely to generate a useful performance tool, and cognitive modelling (CM), whose goal is to explain how human cognition is generated. A device we built without knowing how it works would suffice for AI but not for CM.

♥ SAYGIN: Turing would clearly allow many kinds of machines to pass the test, and more importantly, through various means. Several researchers opposed this idea, especially the latter point, holding that restrictions should be placed on internal information processing if a machine is to be granted thought (Block, 1981; Gunderson, 1964). Is Turing happy to grant intelligence to any old hack that may be programmed to play the imitation game? Or is he so confident that the problem is too hard that he is willing to take the risk of having a quick and dirty solution?

♣ HARNAD: This does not, of course, imply that we are not machines, but only that the Turing Test is about finding out what kind of machine we are, by designing a machine that can generate our performance capacity, but by a functional means that we understand because we designed them.

♠ FORD GLYMOUR, AND HAYES: Turing's anticipation of cloning was not out of the blue. In the period in which this paper was written, he had a strong interest in mathematical biology; especially in morphogenesis. He published one paper on the topic and wrote a number of others. They are available in his Collected Papers.

♣ HARNAD: This is because we want to explain thinking capacity, not merely duplicate it.

us to abandon the requirement that every kind of technique should be permitted. We are the more ready to do so in view of the fact that the present interest in "thinking machines" has been aroused by a particular kind of machine, usually called an "electronic computer" or "digital computer". Following this suggestion we only permit digital computers to take part in our game.♣

♣ HARNAD: This is where Turing withdraws the eligibility of all engineering systems but one, introducing another arbitrary restriction – one that would again rule out T3. Turing said earlier (correctly) that any engineering device ought to be eligible. Now he says it can only be a computer. His motivation is partly, of course, the fact that the computer (Turing Machine) has turned out to be universal, in that it can simulate any other kind of machine. But here we are squarely in the T2/T3 equivocation, for a simulated robot in a virtual world is neither a real robot, nor can it be given a real robotic Turing Test, in the real world. Both T2 and T3 are tests conducted in the real world. But an email interaction with a virtual robot in a virtual world would be T2, not T3.

To put it another way, with the Turing Test we have accepted, with Turing, that thinking is as thinking does. But we know that thinkers can and do more than just talk. And it remains what thinkers can do that our candidate must likewise be able to do, not just what they can do verbally. Hence, just as flying is something that only a real plane can do, and not a computer-simulated virtual plane, be it ever so Turing-equivalent to the real plane – so passing T3 is something only a real robot can do, not a simulated robot tested by T2, be it ever so Turing-equivalent to the real robot. (I also assume it is clear that Turing Testing is testing in the real world: a virtual-reality simulation [VR] would be no kind of a Turing Test; it would merely be fooling our senses in the VR chamber, rather than testing the candidate's real performance capacity in the real world.)

So the restriction to computer simulation, though perhaps useful for planning, designing and even pretesting the T3 robot, is merely a practical methodological strategy. In principle, any engineered device should be eligible, and it must be able to deliver T3 performance, not just T2.

It is of interest that contemporary cognitive robotics has not gotten as much mileage out of computer-simulation and virtual-worlds as might have been expected, despite the universality of computation. "Embodiment" and "situatedness" (in the real world) have turned out to be important ingredients in empirical robotics (Brooks, 2002; Steels and Kaplan, 1999), with the watchword being that the real world is better used as its own model (rather than virtual robots having to simulate, hence second-guess in advance, not only the robot, but the world too).

The impossibility of second-guessing the robot's every potential "move" in advance, in response to every possible real-world contingency, also points to a latent (and I think fatal) flaw in T2 itself: Would it not be a dead giveaway if one's email T2 pen pal proved incapable of commenting on the analogue family photos we kept inserting with our text? (If he can process the images, he is not just a computer, but at least a computer plus A/D peripheral devices, already violating Turing's arbitrary restriction to computers alone.) Or if one's pen pal was totally ignorant of contemporaneous real-world events, apart from those we describe in our letters? Would not even its verbal performance break down if we questioned it too closely about the qualitative and practical details of sensorimotor experience? Could all of that really be second-guessed purely verbally in advance?

This restriction appears at first sight to be a very drastic one. I shall attempt to show that it is not so in reality. To do this necessitates a short account of the nature and properties of these computers.♠♣

It may also be said that this identification of machines with digital computers, like our criterion for "thinking", will only be unsatisfactory if (contrary to my belief), it turns out that digital computers are unable to give a good showing in the game.♣

There are already a number of digital computers in working order, and it may be asked, "Why not try the experiment straight away? It would be easy to satisfy the conditions of the game. A number of interrogators could be used, and statistics compiled to show how often the right identification was given." The short answer is that we are neither asking whether all digital computers would do well in the game nor whether the computers at present available would do well, but whether there are imaginable computers which would do well.♠ But this is only the short answer. We shall see this question in a different light later.

♠ FORD, GLYMOUR, AND HAYES: The "universal machine" idea that a computer can be designed that can in principle simulate *all* other computers, is now widely understood, but it was not at all obvious when Turing was writing, and indeed the idea was widely derided or rejected as ludicrous. The multitude of purposes that computers could serve was little appreciated. A senior British government scientific advisor asserted that the entire country would only need four or five computers, on the grounds that they could be used only for generating elevation tables for naval gunnery. Even von Neumann thought the most important application of computers in mathematics would be to compute examples that would then give mathematicians intuitions about proofs. It seems safe to say that nobody, probably not even Turing, could have foreseen the many uses to which computers have been put in modern society.

The next few pages are a tour de force of exposition for the time Turing was writing, but will seem obvious to many people in this and future generations.

♣ HARNAD: The account of computers that follows is useful and of course correct, but it does not do anything at all to justify restricting the Turing Test to candidates that are computers. Hence this arbitrary restriction is best ignored.

♣ HARNAD: This is the "game" equivocation again. It is not doubted that computers will give a good showing, in the Gallup poll sense. But empirical science is not just about a good showing: An experiment must not just fool most of the experimentalists most of the time! If the performance-capacity of the machine must be indistinguishable from that of the human being, it must be totally indistinguishable, not just indistinguishable more often than not. Moreover, some of the problems that I have raised for T2 – the kinds of verbal exchanges that draw heavily on sensorimotor experience – are not even likely to give a good showing if the candidate is only a digital computer, regardless of how rich a database it is given in advance.

♠ FORD, GLYMOUR, AND HAYES: Again, a simple point that has often been misunderstood since.

3.4 Digital Computers♥

The idea behind digital computers may be explained by saying that these machines are intended to carry out any operations which could be done by a human computer.♠ The human computer is supposed to be following fixed rules♣; he has no authority to deviate from them in any detail. We may suppose that these rules are supplied in a book, which is altered whenever he is put on to a new job. He has also an unlimited supply of paper on which he does his calculations. He may also do his multiplications and additions on a "desk machine", but this is not important.

If we use the above explanation as a definition we shall be in danger of circularity of argument. We avoid this by giving an outline of the means by which the desired effect is achieved. A digital computer can usually be regarded as consisting of three parts:

(i) Store
(ii) Executive unit
(iii) Control

The store is a store of information, and corresponds to the human computer's paper, whether this is the paper on which he does his calculations or that on which his book of rules is printed. In so far as the human computer does calculations in his head a part of the store will correspond to his memory.

The executive unit is the part which carries out the various individual operations involved in a calculation. What these individual operations are will vary from machine to machine. Usually fairly lengthy operations can be done such as

---

♥ SAYGIN: Turing's treatment here and in the next section is one of the most concise, but clear explanations of basic theory of computing that exists – I think it could be useful for teaching purposes. Although Turing is one of the fathers of computer science, being a pioneer in a field does not in itself mean that one is able to speak coherently about that field at an introductory level. I think his success is partly due to the fact that he himself is able to see, in a way characteristic of an interdisciplinary scientist, the relations between the abstract notions of computation, different levels of application, behavioural manifestation, and philosophical analysis.

♠ FORD, GLYMOUR, AND HAYES: It is often said that computers were invented around 1940, but this claim would have sounded strange at that date. The bare term "computer" then meant a human being who (often aided by an electromechanical calculator) performed computations for a living, or in support of some other intellectual activity such as theoretical physics, astronomy, or code-breaking. Computational skill was highly prized, and to be called a "computer" in 1940 was a professional compliment, as it had been since at least the 1850s.

In fact, the famous astronomer Simon Newcomb wrote a recommendation letter for one of his calculators in which he said, "His mind is more like a mathematical machine than any I have ever encountered," which was high praise indeed. To explain the operation of an electronic computer (the adjective, now seeming redundant, is commonly dropped) in terms of rule-books used by human beings, was therefore a perfectly natural expository device. However, this device can be misleading when read with hindsight, since it can suggest that the "thinking" part of the computer is the part of it which corresponds in this expositional metaphor to the human computer, i.e., the "executive unit" or CPU, which is nowadays simply a piece of etched silicon. A similar misunderstanding is the layman's objection – which Turing mentions later – that computers "can only obey instructions."

♣ HARNAD: This goes on to describe what has since become the standard definition of computers as rule-based symbol-manipulating devices (Turing machines).

"Multiply 3540675445 by 7076345687" but in some machines only very simple ones such as "Write down 0" are possible.

We have mentioned that the "book of rules" supplied to the computer is replaced in the machine by a part of the store. It is then called the "table of instructions". It is the duty of the control to see that these instructions are obeyed correctly and in the right order. The control is so constructed that this necessarily happens.

The information in the store is usually broken up into packets of moderately small size. In one machine, for instance, a packet might consist of ten decimal digits. Numbers are assigned to the parts of the store in which the various packets of information are stored, in some systematic manner. A typical instruction might say:

"Add the number stored in position 6809 to that in 4302 and put the result back into the latter storage position."

Needless to say it would not occur in the machine expressed in English. It would more likely be coded in a form such as 6809430217. Here, 17 says which of various possible operations is to be performed on the two numbers. In this case the operation is that described above, viz. "Add the number…" It will be noticed that the instruction takes up ten digits and so forms one packet of information, very conveniently. The control will normally take the instructions to be obeyed in the order of the positions in which they are stored, but occasionally an instruction such as

"Now obey the instruction stored in position 5606, and continue from there" may be encountered, or again

"If position 4505 contains 0 obey next the instruction stored in 6707, otherwise continue straight on."

Instructions of these latter types are very important because they make it possible for a sequence of operations to be replaced over and over again until some condition is fulfilled, but in doing so to obey, not fresh instructions on each repetition, but the same ones over and over again. To take a domestic analogy. Suppose Mother wants Tommy to call at the cobbler's every morning on his way to school to see if her shoes are done, she can ask him afresh every morning. Alternatively she can stick up a notice once and for all in the hall which he will see when he leaves for school and which tells him to call for the shoes, and also to destroy the notice when he comes back if he has the shoes with him.

The reader must accept it as a fact that digital computers can be constructed, and indeed have been constructed, according to the principles we have described, and that they can in fact mimic the actions of a human computer very closely.♠♥

♠ FORD, GLYMOUR, AND HAYES: One can almost sense the frustration that Turing may have felt when trying to find a convincing way to persuade a sceptical audience that mechanical computation was indeed possible. Again, all these metaphors about Mother and Tommy seem curiously antiquated to a contemporary sensibility.

♥ SAYGIN: The analogies here are overly simplified for purposes of explanation. However, I think Turing does believe at some level that most human behaviour is guided by "programs" of the sort that one prepares to make machines perform actions. It is easy to criticize this view, claiming AI has not produced much in terms of intelligent behaviour based on programs of this sort, and that

The book of rules which we have described our human computer as using is of course a convenient fiction. Actual human computers really remember what they have got to do. If one wants to make a machine mimic the behaviour of the human computer in some complex operation one has to ask him how it is done, and then translate the answer into the form of an instruction table.♠ Constructing instruction tables is usually described as "programming". To "programme a machine to carry out the operation A" means to put the appropriate instruction table into the machine so that it will do A.♠

An interesting variant on the idea of a digital computer is a "digital computer with a random element". These have instructions involving the throwing of a die or some equivalent electronic process; one such instruction might for instance be, "Throw the die and put the resulting number into store 1000". Sometimes such a machine is described as having free will (though I would not use this phrase

---

there is much more to being human than just following rules. These latter views are not inconsistent with Turing's thought, and a careful reading of his work will reveal he is also aware that human behaviour is guided by a program much more complex than these analogies suggest, even when random elements are thrown into the picture. It is also likely to be a rather opaque, if not cryptic, program since it will be based on a lifetime of perception, sensation, action, learning and buildup on little innate substrate in a rather experience-driven manner over a long period of time. But it does not follow from the complexity and opacity of the "human behavior program" that runs on the brain that is qualitatively different from a computer program of the sort discussed here. The point here is not to defend what is sometimes called the "computational view of the mind," which, in the light of recent research in cognitive neuroscience, is too symbolic and restricted to account for the level of complexity needed to model human minds – I am pretty sure Turing would not subscribe to that view either. But creating such a program based on ideas from modern cognitive science research and theory (e.g., based on connectionism, dynamical systems theory, embodied cognition and theoretical neuroscience) could be consistent with Turing's views.

♠ Ford, Glymour, and Hayes: This sentence is prescient. Turing was probably thinking of iterative numerical computations of the kind that human computers did indeed perform, but in fact (with a generous interpretation of "instruction table") this is exactly how "knowledge-based systems" are constructed, which have proven capable of performing many tasks which were not previously considered to lie within the province of human computation, but instead to require other human abilities such as "intuition" or "judgement."

♠ Ford, Glymour, and Hayes: Again, describing programming as the construction of look-up tables now seems very archaic. We are now much more familiar with programming as centrally concerned with *language*: programs typically manipulate expressions which themselves may be further interpreted as code, and the actual physical machine may be many levels below all this programming, almost invisible to the human user and even to the programmer. What Turing is describing, and what was at the time the only method of programming available, is what we would now call "assembly-code" programming, an activity that only a few specialists ever practice. In most modern computers, virtually every instruction executed by the CPU was generated by some other piece of code rather than written by a human programmer. Writing assembly code requires an intimate knowledge of the inner workings of the computer's hardware. Turing was what would now be called a wizard or a hacker. Given his views on programming technique and hardware design, he would probably be horrified by the wastefulness of modern programming, in which billions of machine cycles are wasted waiting for human typists' fingers to hit the next key.

myself).♣ It is not normally possible to determine from observing a machine whether it has a random element, for a similar effect can be produced by such devices as making the choices depend on the digits of the decimal for π.

Most actual digital computers have only a finite store. There is no theoretical difficulty in the idea of a computer with an unlimited store. Of course, only a finite part can have been used at any one time. Likewise only a finite amount can have been constructed, but we can imagine more and more being added as required. Such computers have special theoretical interest and will be called infinitive capacity computers.

The idea of a digital computer is an old one. Charles Babbage, Lucasian Professor of Mathematics at Cambridge from 1828 to 1839, planned such a machine, called the Analytical Engine, but it was never completed. Although Babbage had all the essential ideas, his machine was not at that time such a very attractive prospect. The speed which would have been available would be definitely faster than a human computer but something like 100 times slower than the Manchester machine, itself one of the slower of the modern machines. The storage was to be purely mechanical, using wheels and cards.

The fact that Babbage's Analytical Engine was to be entirely mechanical will help us to rid ourselves of a superstition. Importance is often attached to the fact that modern digital computers are electrical, and that the nervous system also is electrical.♠ Since Babbage's machine was not electrical, and since all digital computers are in a sense equivalent, we see that this use of electricity cannot be of theoretical importance. Of course, electricity usually comes in where fast signalling is concerned, so that it is not surprising that we find it in both these connections. In the nervous system chemical phenomena are at least as important as electrical. In certain computers the storage system is mainly acoustic. The feature of using electricity is thus seen to be only a very superficial similarity. If we wish to find such similarities we should look rather for mathematical analogies of function.

3.5 Universality of Digital Computers

The digital computers considered in the last section may be classified amongst the "discrete state machines". These are the machines which move by sudden jumps or clicks from one quite definite state to another. These states are sufficiently different

---

♣ HARNAD: Nor would I. But surely an even more important feature for a Turing Test candidate than a random element or statistical functions would be autonomy in the world – which is something a T3 robot has a good deal more of than a T2 pen pal. The ontic side of free will – namely, whether we ourselves, real human beings, actually have free will – rather exceeds the scope of Turing's paper (Harnad, 1982b). So too does the question of whether a Turing test-passing machine would have any feelings at all (whether free or otherwise; Harnad, 1995). What is clear, though, is that computational rules are not the only ways to "bind" and determine performance: ordinary physical causality can do so too.

♠ FORD, GLYMOUR, AND HAYES: A similar superstition is the view that brains cannot be thought of as computers because they are made of organic material.

for the possibility of confusion between them to be ignored. Strictly speaking there are no such machines. Everything really moves continuously. But there are many kinds of machine which can profitably be *thought* of as being discrete-state machines. For instance, in considering the switches for a lighting system it is a convenient fiction that each switch must be definitely on or definitely off. There must be intermediate positions, but for most purposes we can forget about them. As an example of a discrete-state machine we might consider a wheel which clicks round through 120° once a second, but may be stopped by a lever which can be operated from outside; in addition a lamp is to light in one of the positions of the wheel. This machine could be described abstractly as follows. The internal state of the machine (which is described by the position of the wheel) may be q_1, q_2, or q_3. There is an input signal i_0 or i_1 (position of lever). The internal state at any moment is determined by the last state and input signal according to the table

		Last State	
		q_1	q_2	q_3
Input	i_0	q_2	q_3	q_1
	i_1	q_2	q_3	q_1

The output signals, the only externally visible indication of the internal state (the light) are described by the table

State	q_1	q_2	q_3
Output	o_0	o_0	o_1

This example is typical of discrete-state machines. They can be described by such tables provided they have only a finite number of possible states.

It will seem that given the initial state of the machine and the input signals it is always possible to predict all future states.[♣] This is reminiscent of Laplace's view that from the complete state of the universe at one moment of time, as described by the positions and velocities of all particles, it should be possible to predict all future states. The prediction which we are considering is, however, rather nearer to practicability than that considered by Laplace. The system of the "universe as a whole" is such that quite small errors in the initial conditions can have an overwhelming effect at a later time.[♠] The displacement of a single electron by a

[♣] HARNAD: The points about determinism are probably red herrings. The only relevant property is performance capacity. Whether either the human or the machine is completely predictable is irrelevant. (Both many-body physics and complexity theory suggest that neither causal determinacy nor following rules guarantee predictability in practise – and this is without even invoking the arcana of quantum theory.)

[♠] FORD, GLYMOUR, AND HAYES: In more modern terminology, the universe is in some sense *chaotic*. Chaos theory had not been developed when Turing was writing, of course.

billionth of a centimetre at one moment might make the difference between a man being killed by an avalanche a year later, or escaping. It is an essential property of the mechanical systems which we have called "discrete state machines" that this phenomenon does not occur. Even when we consider the actual physical machines instead of the idealized machines, reasonably accurate knowledge of the state at one moment yields reasonably accurate knowledge any number of steps later.

As we have mentioned, digital computers fall within the class of discrete-state machines. But the number of states of which such a machine is capable is usually enormously large. For instance, the number for the machine now working at Manchester is about $2^{165,000}$, i.e., about $10^{50,000}$. Compare this with our example of the clicking wheel described above, which had three states. It is not difficult to see why the number of states should be so immense. The computer includes a store corresponding to the paper used by a human computer. It must be possible to write into the store any one of the combinations of symbols which might have been written on the paper. For simplicity suppose that only digits from 0 to 9 are used as symbols. Variations in handwriting are ignored. Suppose the computer is allowed 100 sheets of paper each containing 50 lines each with room for 30 digits. Then the number of states is $10^{100 \times 50 \times 30}$, i.e., $10^{150,000}$. This is about the number of states of three Manchester machines put together. The logarithm to the base two of the number of states is usually called the "storage capacity" of the machine. Thus the Manchester machine has a storage capacity of about 165,000 and the wheel machine of our example about 1.6. If two machines are put together their capacities must be added to obtain the capacity of the resultant machine. This leads to the possibility of statements such as "The Manchester machine contains 64 magnetic tracks each with a capacity of 2,560, eight electronic tubes with a capacity of 1,280. Miscellaneous storage amounts to about 300 making a total of 174,380."♠♥

Given the table corresponding to a discrete-state machine it is possible to predict what it will do. There is no reason why this calculation should not be carried out by means of a digital computer. Provided it could be carried out sufficiently quickly the digital computer could mimic the behaviour of any discrete-state machine. The imitation game could then be played with the machine in question (as B) and the mimicking digital computer (as A) and the interrogator would be unable to distinguish them.♠ Of course, the digital computer must have an adequate storage

---

♠ Ford, Glymour, and Hayes: It is hard to compare this accurately with modern machines, but a typical laptop computer may have an active memory capacity of approximately 10|9, and a hard disc capacity of perhaps a hundred times more. Of course, not all of this huge capacity may be being used in a way that Turing would have thought sensible.

♥ Saygin: Revisiting the question of whether Turing was proposing the game as a real operational definition or test. It seems highly unlikely to me that a man proposing a thought experiment would spend such time, space and energy to explain not only what he means by "thinking" but also exactly what kind of machine a digital computer is.

♠ Ford, Glymour, and Hayes: Here, Turing seems to be using the term "imitation game" in a very generic sense.

capacity as well as working sufficiently fast. Moreover, it must be programmed afresh for each new machine which it is desired to mimic.♥

This special property of digital computers, that they can mimic any discrete-state machine, is described by saying that they are *universal* machines. The existence of machines with this property has the important consequence that, considerations of speed apart, it is unnecessary to design various new machines to do various computing processes. They can all be done with one digital computer, suitably programmed for each case. It will be seen that as a consequence of this all digital computers are in a sense equivalent.♣

We may now consider again the point raised at the end of §3. It was suggested tentatively that the question, "Can machines think?" should be replaced by "Are there imaginable digital computers which would do well in the imitation game?" If we wish we can make this superficially more general and ask "Are there discrete-state machines which would do well?" But in view of the universality property we see that either of these questions is equivalent to this, "Let us fix our attention on one particular digital computer C. Is it true that by modifying this computer to have an adequate storage, suitably increasing its speed of action, and providing it with an appropriate program, C can

♥ SAYGIN: Turing reminds us of the imitation game here. He is trying to emphasize the universality aspect of the discrete-state machine, but he does so using the original indistinguishability test we started with. There is an interesting twist: What is the interrogator looking for in this instantiation of the game? Which one is the machine? Which one is the mimicking digital computer? What would you even ask in order to differentiate between the two? It does not make much sense, unless Turing means they will play the gender-based imitation game. He never says we change the game into anything other than the gender-based game anyway. It may sound silly or pointless to compare how well two entities imitate a woman in a teletype conversation, but as I will elaborate below, tests and experiments often construct situations that do not have direct equivalents in real life (i.e., they do not always have high ecological validity).

♣ HARNAD: All true, but all irrelevant to the question of whether a digital computer alone could pass T2, let alone T3. The fact that eyes and legs can be simulated by a computer does not mean a computer can see or walk (even when it is simulating seeing and walking). So much for T3. But even just for T2, the question is whether simulations alone can give the T2 candidate the capacity to verbalize and converse about the real world indistinguishably from a T3 candidate with autonomous sensorimotor experience in the real world.

(I think yet another piece of unnoticed equivocation by Turing – and many others – arises from the fact that thinking is not directly observable, which helps us imagine that computers think. But even without having to invoke the other-minds problem (Harnad, 1991), one needs to remind oneself that a universal computer is only formally universal: it can describe just about any physical system, and simulate it in symbolic code, but in doing so, it does not capture all of its properties: Exactly as a computer-simulated airplane cannot really do what a plane does (i.e., fly in the real world), a computer-simulated robot cannot really do what a real robot does (act in the real world) – hence there is no reason to believe it is really thinking. A real robot may not really be thinking either, but that does require invoking the other-minds problem, whereas the virtual robot is already disqualified for exactly the same reason as the virtual plane: both fail to meet the Turing Test criterion itself, which is real performance capacity, not merely something formally equivalent to it!).

be made to play satisfactorily the part of A in the imitation game, the part of B being taken by a man?"♠♥

♠ FORD, GLYMOUR, AND HAYES: Turing implicitly uses what has become known as the Church-Turing thesis: The computable functions are all and only those computable by a Universal Turing Machine. The idea that one can make a given computer act like any other just by reprogramming it, given enough processor speed and memory capacity, is now a familiar idea in our culture; but persuading his audience of its reasonableness was probably one of Turing's most difficult tasks of exposition in 1951.

♥ SAYGIN: Notice that the woman has disappeared from the game altogether. But the objectives of A, B, and C remain unaltered; at least Turing does not explicitly state any change. To be precise, what we have is a digital computer and a man both trying to convince an interrogator that they are the real woman.

Why the fuss about the woman, the man, and the replacement? Turing does not seem the type of person who would beat around the bush for no reason. What is going on here?

One could say the gender-based imitation game is merely an analogy, serving the purpose of making the paper easier to understand. But in a paper that starts with the sentence, "Can machines think?" would something like "Let us take a machine and a man and see if the machine can convince interrogators that it is a human being via teletype conversations" be really much harder to process or understand? Or maybe Turing was simply careless and forgot to clarify that we are no longer talking about the gender-based imitation game. Given the level of detail and precision in Turing's writing (see Sections 4 and 5 of this paper), this is unlikely to be the explanation. Also bear in mind Turing is a successful mathematician, a discipline characterized by precision of definition and rigor in argument and generalization, which would make it unlikely that he is being sloppy.

Here is my explanation for the quirks of the game – I cannot guarantee that this is what Turing intended, but I think it is consistent with the way he thinks and writes. Neither the man in the gender-based imitation game nor any kind of machine is a woman. Furthermore what Turing proposes is essentially to compare the machine's success against that of the man – not to look at whether it actually "beats" the woman. The man and the machine are measured in terms of their respective performances and their performances are comparable because they are both simulating something which they are not. Even though it is regarded as obscure by many, the imitation game could be a carefully planned experimental design. It provides a fair basis for comparison: the woman (either as a participant in the game or as a concept) acts as a neutral point so that the two imposters can be assessed in how well they perform the imitation. In other words, Turing gives us a question that is to be examined via a carefully defined task, an experimental group (digital computers) and a control group (men). This setup looks more like an operational definition given in terms of an experimental design than anything else.

It might seem that we are a long way from such relatively minor methodological points being relevant. But at least two studies have shown that people's judgments of computers' conversational performance are substantially influenced by whether or not they know in advance that their conversational partners may be machines. In the 1970s, a group of scientists devised an electronic interviewing environment where experienced psychiatrists corresponded with both real-life paranoid patients and computer programs simulating paranoid behaviour through teletype. The judges were not told that some of the interviewees could be computer programs. Details can be found in Colby et al. (1972), but to summarize, the finding was that the psychiatric judges did not do better than chance guessing at identifying the computers from the human patients.

In a more recent study, we carried out an experiment to examine possible relationships between pragmatic violations and imitation game performance, using real excerpts from human–computer conversations (Saygin and Cicekli, 2002). Due to the design of the experiment, some subjects made pragmatic judgments on a set of conversations without being told there were computers involved

3.6 Contrary Views on the Main Question

We may now consider the ground to have been cleared and we are ready to proceed to the debate on our question, "Can machines think?" and the variant of it quoted at the end of the last section. We cannot altogether abandon the original form of the problem, for opinions will differ as to the appropriateness of the substitution and we must at least listen to what has to be said in this connexion.

It will simplify matters for the reader if I explain first my own beliefs in the matter. Consider first the more accurate form of the question. I believe that in about 50 years' time it will be possible, to program computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70% chance of making the right identification after 5 min of questioning.♠♣♥ The original

---

before they were told about the imitation game and asked to evaluate the computers' performance in the same conversations, while other subjects knew computers' involvement from the outset. We noted that even something seemingly trivial like having read the conversations only once without any bias prior to being asked to make decisions regarding the computers' behaviour had a differential effect on people's judgments. In particular, the analysis revealed that when people were faced with anomalies in the conversations, those who knew about computers' involvement tended to automatically think these were indicative of the conversational partner's identity (i.e., by the fact that it is a machine). On the other hand, unbiased subjects tried to work through the problematic exchanges in the same way they would in a pragmatically confusing conversation between humans. Now note that the gender-based imitation game is immune to the bias that knowledge of computer participation may bring. It allows the interrogators to work out pragmatic violations (and in general, exercise their naive psychology) the way they normally do; therefore, this design allows us to give the digital computers a fairer shot at performing well.

In sum, the gender-based imitation game is a good experimental design. It provides an operational definition (i.e., a larger question is replaced by a task we can evaluate). It is controlled; the task is simulating something both the experimental and control subjects are not. Furthermore, any bias the interrogator (which may be thought of as a measurement device) brings in will be based on gender expectations, which will tend not to affect the two groups differentially. Viewed in this light, the quirky imitation game may well be one of the few ways to fairly and experimentally assess machine thought.

♠ Ford, Glymour, and Hayes: Turing was right about the memory capacity of modern computers, but it is widely claimed that he was wrong in his Turing Test prediction: here we are, 50 years later, and where are the passers of his imitation game? However, notice that Turing says that it will be *possible*. That question is still moot: maybe it is possible. Certainly, computers have already performed many tasks that were previously thought of as requiring human sagacity of some kind. But in any case, very few contemporary AI researchers are seriously trying to build a machine to play Turing's imitation game. Instead they are concerned with exploring the computational machinery of intelligence itself, whether in humans, dogs, computers, or aliens. The scientific aim of AI research is to understand intelligence as computation, and its engineering aim is to build machines that surpass or extend human mental abilities in some useful way. Trying to imitate a human conversation (however "intellectual" it may be) contributes little to either ambition. Progress in AI is not measured by checking fidelity to a human conversationalist. And yet many critics of AI are complaining of a lack of progress toward this old ambition. But perhaps we should forgive the critics, as even many AI textbooks still offer the Turing Test as AI's ultimate goal, which seems akin to starting a textbook on aeronautical engineering with an explanation that the goal of the field is to make machines that fly so exactly like pigeons that they can even fool other pigeons (Ford and Hayes, 1998).

question, "Can machines think?" I believe to be too meaningless to deserve discussion.♣ Nevertheless, I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of

---

This is of course a huge area of controversy, far larger than we have space here to survey, but one point may be worth making. In making this 50-year prediction, Turing may have meant that a determined 50-year effort devoted to this single aim could succeed, a kind of imitation-game Manhattan project, or a fivefold expansion of the decade of national effort it took to get a human being to the moon and back. Certainly, given his wartime experiences of large-scale government-sponsored projects, this interpretation is not implausible; and later in the paper he suggests explicitly that it would be a 3,000-man-year project.

♣ HARNAD (p. 41): No doubt this party-game/Gallup poll criterion can be met by today's computer programs – but that remains as meaningless a demographic fact today as it was when predicted 50 years ago. Like any other science, cognitive science is not the art of fooling most of the people for some or most of the time! The candidate must really have the generic performance capacity of a real human being – capacity that is totally indistinguishable from that of a real human being to any real human being (for a lifetime, if need be!). No tricks: real performance capacity.

♥ SAYGIN (p. 41): More than 50 years have gone by since Turing wrote these often-quoted words, yet we are nowhere near "the goal." How could Turing, a man with such great vision and intellect, so grossly underestimate the time it would take to tackle the problems he left behind? I grant it that Turing underestimated either how hard the task at hand is, or how long it takes to carry out such a task. But I wonder sometimes if that is the whole story. Could he, in addition, have overestimated how hard future researchers would work at the problems? I think the latter has played more of a role than is commonly considered in the fact that we have a gaping hole between Turing's expectations and the current state of AI. Think about it: Can we really say we followed Turing's advice, gave it our all and it did not work? Or did we try shortcuts and little hacks and cheats and gave up in pursuit of "useful" AI when they did not work? The point here is not to criticize AI researchers for working on this or that topic. I only want to note that we do not know how much closer we would have been at developing AI systems that can communicate using natural language had we actually pursued it as a serious, full-time goal. Turing's tone in this paper leads me to think that the future he envisioned is based on scientists, philosophers, and programmers working hard and wholeheartedly towards the goal, patiently overcoming obstacles and making steady progress. What really happened in the AI arena was a buzz, a wave of optimism with many researchers believing that successful AI was right around the corner, finding the whole endeavor challenging but "cool," and wanting to make it work and make it work fast. However, when the problem proved too difficult to yield fruit soon, there was an ensuing burnout, which soon led to a lack of serious interest in endeavors such as the Turing Test. Some AI researchers even went as far as outwardly refusing to work on the Turing Test, defending that it belongs in history books rather than current research agendas, indeed calling it "harmful for AI" (Hayes and Ford, 1995).

♣ HARNAD: It is not meaningless, it is merely indecisive: What we mean by "think" is, on the one hand, what thinking creatures can do and how they can do it, and, on the other hand, what it feels-like to think. What thinkers can do is captured by the Turing Test. A theory of how they do it is provided by how our man-made machine does it. (If there are several different successful machines, it is a matter of normal inference-to-the-best-theory.) So far, nothing is meaningless. Now we ask: Do the successful candidates really feel, as we do when we think? This question is not meaningless; it is merely unanswerable – in any other way than by being the candidate. It is the familiar old other-minds problem (Harnad, 1991).

machines thinking without expecting to be contradicted.♠♣ I believe further that no useful purpose is served by concealing these beliefs. The popular view that scientists proceed inexorably from well-established fact to well-established fact, never being influenced by any improved conjecture, is quite mistaken. Provided it is made clear which are proved facts and which are conjectures, no harm can result. Conjectures are of great importance since they suggest useful lines of research.♠♣

I now proceed to consider opinions opposed to my own.

3.6.1 The Theological Objection

Thinking is a function of man's immortal soul.♣ God has given an immortal soul to every man and woman, but not to any other animal or to machines. Hence no animal or machine can think.

♠ FORD, GLYMOUR, AND HAYES: That prediction has in a sense been vindicated: the idea of thinking machines has indeed entered popular culture, and is widely discussed as either a present reality or an imminent one. What is more interesting, however, and what Turing apparently did not foresee, is the emergence of a kind of linguistic creepage, where the boundaries of "real thinking" are redrawn so as to exclude whatever it is that machines become able to do. When electronic computers were new, the ability to perform mental arithmetic rapidly and accurately was widely admired as a human mental ability; now it is "merely" mechanical. Now that a computer has beaten the world chess champion, skill at chess is becoming perceived as "merely" mechanical. This gradual but irresistible cultural shift in meaning also bears on the utility of the imitation game: One has to ask, to which generation does the judge belong? Behaviour that someone of Turing's generation would have found convincing may completely fail to impress someone who grew up with talking teddy bears.

♣ HARNAD: Yes, but only at a cost of demoting "thinking" to meaning only "information processing" rather than what you or I do when we think, and what that feels-like.

♠ FORD, GLYMOUR, AND HAYES: One can make out a reasonable case that this paper, and its bold conjectures, played a central role in the emergence of AI and cognitive science in the 1960s and 1970s.

♣ HARNAD: This is mistaken. Yes, science proceeds by a series of better approximations, from empirical theory to theory. But the theory here would be the actual design of a successful Turing Test candidate, not the conjecture that computation (or anything else) will eventually do the trick. Turing is confusing formal conjectures (such as that the Turing machine and its equivalents capture all future notions and instances of what we mean by "computation" – the "Church/Turing Thesis") and empirical hypotheses, such as that thinking is just computation. Surely the Turing Test is not a license for saying that we are explaining thinking better and better as our candidates fool more and more people longer and longer. On the other hand, something else that sounds superficially similar to this could be said about scaling up the Turing Test empirically by designing a candidate that can do more and more of what we can do. And Turing testing certainly provides a methodology for such cumulative theory-building and theory-testing in cognitive science.

♣ HARNAD: The real theological objection is not so much that the soul is immortal but that it is immaterial. This view also has non-theological support from the mind/body problem: no one – theologian, philosopher, or scientist – has even the faintest hint of an idea of how mental states

I am unable to accept any part of this, but will attempt to reply in theological terms. I should find the argument more convincing if animals were classed with men, for there is a greater difference, to my mind, between the typical animate and the inanimate than there is between man and the other animals.♣ The arbitrary character of the orthodox view becomes clearer if we consider how it might appear to a member of some other religious community. How do Christians regard the Moslem view that women have no souls?♠ But let us leave this point aside and return to the main argument. It appears to me that the argument quoted above implies a serious restriction of the omnipotence of the Almighty. It is admitted that there are certain things that He cannot do such as making one equal to two, but should we not believe that He has freedom to confer a soul on an elephant if He sees fit? We might expect that He would only exercise this power in conjunction with a mutation which provided the elephant with an appropriately improved brain to minister to the needs of this soul. An argument of exactly similar form may be made for the case of machines. It may seem different because it is more difficult to "swallow". But this really only means that we think it would be less likely that He would consider the circumstances suitable for conferring a soul. The circumstances in question are discussed in the rest of this paper. In attempting to construct such machines we should not be irreverently usurping His power of creating souls, any more than we are in the procreation of children: rather we are, in either case, instruments of His will be providing mansions for the souls that He creates.

However, this is mere speculation. I am not very impressed with theological arguments whatever they may be used to support. Such arguments have often been found unsatisfactory in the past. In the time of Galileo it was argued that the texts, "And the sun stood still… and hasted not to go down about a whole day" (Joshua x. 13) and "He laid the foundations of the earth, that it should not move at any time" (Psalm cv. 5) were an adequate refutation of the Copernican theory. With our present knowledge such an argument appears futile. When that knowledge was not available it made a quite different impression.♠

---

can be material states (or, as I prefer to put it, how functional states can be felt states). This problem has been dubbed "hard" (Chalmers in Shear, 1997). It may be even worse: it may be insoluble (Harnad, 2001). But this is no objection to Turing Testing which, even if it will not explain how thinkers can feel, does explain how they can do what they can do.

♣ HARNAD: Yes, and this is why the other-minds problem comes into its own in doing Turing testing of machines rather than in doing mind reading of our own species and other animals. ("Animate" is a weasel word, though, for vitalists are probably also animists; Harnad, 1994a.)

♠ FORD, GLYMOUR, AND HAYES: Turing's source for this view is unknown. The contrary opinion is given in the Qu'ran.

♠ FORD, GLYMOUR, AND HAYES: The last three sentences are a bit odd. We only acquired our present knowledge because many people (Galileo himself, Bruno before him, Kepler, Newton, etc.) already found the argument futile.

3.6.2 The "Heads in the Sand" Objection

"The consequences of machines thinking would be too dreadful. Let us hope and believe that they cannot do so."

This argument is seldom expressed quite so openly as in the form above. But it affects most of us who think about it at all. We like to believe that Man is in some subtle way superior to the rest of creation. It is best if he can be shown to be *necessarily* superior, for then there is no danger of him losing his commanding position. The popularity of the theological argument is clearly connected with this feeling. It is likely to be quite strong in intellectual people, since they value the power of thinking more highly than others, and are more inclined to base their belief in the superiority of Man on this power.

I do not think that this argument is sufficiently substantial to require refutation. Consolation would be more appropriate: perhaps this should be sought in the transmigration of souls.♠

3.6.3 The Mathematical Objection

There are a number of results of mathematical logic which can be used to show that there are limitations to the powers of discrete-state machines. The best known of these results is known as Gödel's theorem (1931), and shows that in any sufficiently powerful logical system statements can be formulated which can neither be proved nor disproved within the system, unless possibly the system itself is inconsistent. There are other, in some respects similar, results due to Church (1936), Kleene (1935), Rosser, and Turing (1937). The latter result is the most convenient to

♠ Ford, Glymour, and Hayes: Turing raises an issue still central in our own time: the limitation of scientific inquiry by religious dogma, and in particular by the doctrine of souls. The fundamental religious objection to embryonic stem cell research is that when a human sperm cell and an ovum form an embryonic cell, the cell is "ensouled," it supernaturally acquires a soul. In this subsection, irritation or exasperation seems to have overwhelmed Turing's usual ingenuity in argument. While many current advocates of "Heads in the Sand" may be utterly thoughtless, there is a history of arguments for the position, all of which Turing ignores. William James, in *The Will to Believe*, argued roughly as follows: we should not believe that human intelligence has a purely biological, chemical, and physical explanation, for if we did so believe, we would conclude there is no basis for moral assessment; the world would be a worse place if we believed there is no basis for moral assessment, and it is rational not to act to bring about the worse case. The argument is in the spirit of Pascal's Wager. Pascal, the Turing of the 17th century, argued that one should act so as to cause oneself to believe in God, because the expected payoff of believing is infinitely greater than the expected payoff of not believing. We think neither James' argument nor Pascal's is sound, but the arguments deserve at least as much consideration as others Turing does respond to.

consider, since it refers directly to machines, whereas the others can only be used in a comparatively indirect argument: for instance, if Gödel's theorem is to be used we need in addition to have some means of describing logical systems in terms of machines, and machines in terms of logical systems. The result in question refers to a type of machine which is essentially a digital computer with an infinite capacity. It states that there are certain things that such a machine cannot do. If it is rigged up to give answers to questions as in the imitation game, there will be some questions to which it will either give a wrong answer, or fail to give an answer at all however much time is allowed for a reply. There may, of course, be many such questions, and questions which cannot be answered by one machine may be satisfactorily answered by another. We are of course, supposing for the present that the questions are of the kind to which an answer "Yes" or "No" is appropriate, rather than questions such as "What do you think of Picasso?" The questions that we know the machines must fail on are of this type, "Consider the machine specified as follows.… Will this machine ever answer "Yes" to any question?" The dots are to be replaced by a description of some machine in a standard form, which could be something like that used in §5. When the machine described bears a certain comparatively simple relation to the machine which is under interrogation, it can be shown that the answer is either wrong or not forthcoming. This is the mathematical result: it is argued that it proves a disability of machines to which the human intellect is not subject.

The short answer to this argument is that although it is established that there are limitations to the powers of any particular machine, it has only been stated, without any sort of proof, that no such limitations apply to the human intellect.♣♥ But I do not think this view can be dismissed quite so lightly. Whenever one of these machines is asked the appropriate critical question, and gives a definite answer, we know that this answer must be wrong, and this gives us a certain feeling of superiority. Is this feeling illusory? It is no doubt quite genuine, but I do not think too much importance should be attached to it. We too often give wrong answers to questions ourselves to be justified in being very pleased at such evidence of fallibility on the part of the machines. Further, our superiority can only be felt on such an occasion in relation to the one machine over which we have scored our petty triumph. There would be no question of triumphing simultaneously over *all* machines. In short, then, there might be men cleverer than any

---

♣ HARNAD: Gödel's theorem shows that there are statements in arithmetic that are true, and we know are true, but their truth cannot be computed. Some have interpreted this as implying that "knowing" (which is just a species of "thinking") cannot be just computation. Turing replies that maybe the human mind has similar limits, but it seems to me it would have been enough to point out that "knowing" is not the same as "proving". Gödel shows the truth is unprovable, not that it is unknowable. There are far better reasons for believing that thinking is not computation.

♥ SAYGIN: I do not see this issue discussed much anywhere, but I think it is a profound idea. How do we know that the human brain-computer would "halt" given the description of another human brain-computer and asked what it would reply to a given input?

given machine, but then again there might be other machines cleverer again, and so on.♠

Those who hold to the mathematical argument would, I think, mostly be willing to accept the imitation game as a basis for discussion. Those who believe in the two previous objections would probably not be interested in any criteria.♠

3.6.4 The Argument from Consciousness

This argument is very well expressed in *Professor Jefferson's* Lister Oration for 1949, from which I quote. "Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain♣ – that is, not only write it but know that it had written it. No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be

♠FORD, GLYMOUR, AND HAYES: It is interesting that the two men who might be said to have provided the intellectual machinery of this objection, Gödel and Turing, came to opposite philosophical conclusions. Turing's conclusion was that since human thinkers are equivalent in inferential power to machines, there must be truths that they – we – are incapable of proving. Gödel started with the assumption that there were no mathematical truths that human beings could not, in principle, grasp, and concluded that human beings could not be machines. Penrose, who has a similarly grand view of the power of human thought, cites Gödel with approval.

The argument can be further illustrated by considering the fact that many people believe themselves to be consistent: It follows from the Gödel result mentioned that no consistent system of a reasonable deductive power can conclusively establish its own consistency. Turing concluded that human beings have inconsistent beliefs, a view that Gödel and Penrose apparently reject.

It is hard to see how this disagreement could be resolved conclusively, since the central issue could not be determined empirically; there is no finite amount of information that could conclusively establish that human beings, or indeed anything else in the universe, are capable of such all-encompassing deductive powers. Further, if the inferential abilities of humans are in fact bounded by Turing computability, a human could not even reliably converge to the truth about whether a system (including humans) is or not a computer from its responses to inputs.

♠FORD, GLYMOUR, AND HAYES: Turing accurately predicted and then succinctly rebutted the many subsequent resurrections of the mathematical argument, including much of what Roger Penrose has written against the computational conception of mind. This debate has now taken place a number of times; the "mathematical objection" seems to get resuscitated every 15 years or so. For a more thorough analysis of the technical issues involved in this debate see (Laforte, Hayes, and Ford, 1998).

♣HARNAD: This standard argument against the Turing Test (repeated countless times in almost exactly the same way until the present day) is merely a restatement of the other-minds problem: THERE IS NO WAY TO KNOW WHETHER EITHER HUMANS OR MACHINES DO WHAT THEY DO BECAUSE THEY FEEL like it – or whether they feel anything at all, for that matter. But there is a lot to be known from identifying what can and cannot generate the capacity to do what humans can do. (The limits of symbol-manipulation [computation] are another matter, and one that can be settled empirically, based on what sorts of machine can and cannot pass the Turing Test; Harnad, 2003.)

warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants."♠

This argument appears to be a denial of the validity of our test. According to the most extreme form of this view the only way by which one could be sure that machine thinks is to *be* the machine and to feel oneself thinking. One could then describe these feelings to the world, but of course, no one would be justified in taking any notice. Likewise according to this view the only way to know that a *man* thinks is to be that particular man. It is in fact the solipsist point of view.♠ It may be the most logical view to hold but it makes communication of ideas difficult.♣ A is liable to believe "A thinks but B does not" whilst B believes "B thinks but A does not". Instead of arguing continually over this point it is usual to have the polite convention that everyone thinks.

I am sure that Professor Jefferson does not wish to adopt the extreme and solipsist point of view. Probably he would be quite willing to accept the imitation game as a test.♠

♠ FORD, GLYMOUR, AND HAYES: Professor Jefferson's view is a special case of the view argued subsequently by Keith Gunderson, that to be intelligent, a thing must not only do some of what we do intelligently (e.g., converse or imitate), but must do *all* of what we do, and must do it *as* we do. Essentially, this position demands that any intelligent machine actually *be* human or, more precisely, it requires that any intelligent machine have a humanoid phenomenology: it must have the rather elusive quality that being it *feels like* being human. Of course, any such criterion rejects the "behavioural" quality of any test such as the one Turing proposes. Unfortunately, at least until we develop a theory of phenomenology linked adequately to the rest of science, it rejects *any* objective test of any kind.

♠ FORD, GLYMOUR, AND HAYES: This is a slight oversimplification. It is possible to hold this view rationally without adopting solipsism. In fact, the current mainstream view of consciousness, aptly described by David Chalmers in "The Conscious Mind" as "the hard problem," is that the presence of such an "inner view" is indeed characteristic of consciousness, and that any kind of behaviour might, in principle, be produced by a "zombie" which has no such inner life; and that therefore, no such behavioural criterion can be taken to be definitive evidence for consciousness. Of course, this is not in itself an objection to the presence of thought itself, as Chalmers himself is at pains to point out, since a zombie may indeed be thinking without being conscious of thinking. (The distinctions being applied in this area have become much more delicate than they were when Turing was writing.)

♣ HARNAD: Turing is dead wrong here. This is not solipsism (i.e., not the belief that only I exist and all else is my dream). It is merely the other-minds problem (Harnad, 1991); and it is correct, but irrelevant – or rather put into perspective by the Turing Test: there is no one else we can know has a mind but our own private selves, yet we are not worried about the minds of our fellow-human beings, because they behave just like us and we know how to mind read their behaviour. By the same token, we have no more or less reason to worry about the minds of anything else that behaves just like us – so much so that we cannot tell them apart from other human beings. Nor is it relevant what stuff they are made out of, since our successful mind reading of other human beings has nothing to do with what stuff they are made out of. It is based only on what they do.

♠ FORD, GLYMOUR, AND HAYES: This is the second time that Turing assumes confidently that an intellectual opponent would "probably" be willing to accept his imitation game as a valid test. This optimism seems misplaced if it is supposed to indicate a willingness to accede to Turing's philosophical position; what he may have meant, however, is that if faced with an actual machine which passed the test, Jefferson would probably agree that it was, in fact, thinking intelligently.

The game (with the player B omitted) is frequently used in practice under the name of viva voce to discover whether some one really understands something or has "learnt it parrot fashion". Let us listen in to a part of such a *viva voce*♥:

Interrogator:	In the first line of your sonnet which reads "Shall I compare thee to a summer's day", would not "a spring day" do as well or better?
Witness:	It wouldn't scan.
Interrogator:	How about "a winter's day" That would scan all right.
Witness:	Yes, but nobody wants to be compared to a winter's day.
Interrogator:	Would you say Mr. Pickwick reminded you of Christmas?
Witness:	In a way.
Interrogator:	Yet Christmas is a winter's day, and I do not think Mr. Pickwick would mind the comparison.
Witness:	I don't think you're serious. By a winter's day one means a typical winter's day, rather than a special one like Christmas.

And so on. What would Professor Jefferson say if the sonnet-writing machine was able to answer like this in the viva voce? I do not know whether he would regard the machine as "merely artificially signalling" these answers, but if the answers were as satisfactory and sustained as in the above passage I do not think he would describe it as "an easy contrivance". This phrase is, I think, intended to cover such devices as the inclusion in the machine of a record of someone reading a sonnet, with appropriate switching to turn it on from time to time.

In short then, I think that most of those who support the argument from consciousness could be persuaded to abandon it rather than be forced into the solipsist position. They will then probably be willing to accept our test.♠

I do not wish to give the impression that I think there is no mystery about consciousness. There is, for instance, something of a paradox connected with any attempt to localize it.♠ But I do not think these mysteries necessarily need to be solved before we can answer the question with which we are concerned in this paper.

♥ SAYGIN: One final comment about the gender-based game: The game actually ended up being ecologically valid after all! Rest assured that at this very moment a variant is being played in many Internet chat rooms across the world, with participants trying to guess each others' gender (often with the hope of forming romantic relations) based on "teletype" connections.

♠ FORD, GLYMOUR, AND HAYES: Perhaps Professor Jefferson need not be driven to solipsism at all. He might reply to Turing with an argument from similarity: "I am composed of the same kind of tissue and cells as any other human. When a human feels pain, from a burning finger, say, we know there is a course of nerve signals from the digit to the brain, and as neuroscience advances we will be able to follow the trace in more detail. It is alike in every human as far as we can tell, and so also in me. But I, Jefferson, know I *feel* pain. Since other humans are composed and function as I do, I can reasonably infer that they feel pain in like circumstances. But I have no such assurance for the digital computer."

♠ FORD, GLYMOUR, AND HAYES: Indeed there is. For examples, see Daniel Dennett's delightful essay "Where Am I?"

3.6.5 Arguments from Various Disabilities

These arguments take the form, "I grant you that you can make machines do all the things you have mentioned but you will never be able to make one to do X". Numerous features X are suggested in this connexion. I offer a selection:

Be kind, resourceful, beautiful, friendly, have initiative, have a sense of humour, tell right from wrong, make mistakes, fall in love, enjoy strawberries and cream, make some one fall in love with it, learn from experience, use words properly, be the subject of its own thought, have as much diversity of behaviour as a man, do something really new. (Some of these disabilities are given special consideration as indicated [on the next few pages]).♣

No support is usually offered for these statements. I believe they are mostly founded on the principle of scientific induction. A man has seen thousands of machines in his lifetime. From what he sees of them he draws a number of general conclusions. They are ugly, each is designed for a very limited purpose, when required for a minutely different purpose they are useless, the variety of behaviour of any one of them is very small, etc. Naturally he concludes that these are necessary properties of machines in general. Many of these limitations are associated with the very small storage capacity of most machines. (I am assuming that the idea of storage capacity is extended in some way to cover machines other than discrete-state machines. The exact definition does not matter as no mathematical accuracy is claimed in the present discussion.) A few years ago, when very little had been heard of digital computers, it was possible to elicit much incredulity concerning them, if one mentioned their properties without describing their construction. That was presumably due to a similar application of the principle of scientific induction. These applications of the principle are of course, largely unconscious. When a burnt child fears the fire and shows that he fears it by avoiding it, I should say that he was applying scientific induction. (I could, of course, also describe his behaviour in many other ways.) The works and customs of mankind do not seem to be very suitable material to which to apply scientific induction. A very large part of space-time must be investigated, if reliable results are to be obtained. Otherwise we may (as most English children do) decide that everybody speaks English, and that it is silly to learn French.♠

There are, however, special remarks to be made about many of the disabilities that have been mentioned. The inability to enjoy strawberries and cream may have struck the reader as frivolous. Possibly a machine might be made to enjoy this

---

♣ HARNAD: Turing rightly dismisses this sort of scepticism (which I have dubbed "Granny Objections" by pointing out that these are empirical questions about what computers (and other kinds of machines) will eventually be shown to be able to do. The performance items on the list, that is. The mental states (feelings), on the other hand, are moot because of the other-minds problem.

♠ FORD, GLYMOUR, AND HAYES: Could it be that the French actually converse by secretly passing notes in English back and forth?

delicious dish, but any attempt to make one do so would be idiotic. What is important about this disability is that it contributes to some of the other disabilities, e.g., to the difficulty of the same kind of friendliness occurring between man and machine as between white man and white man, or between black man and black man.♠

The claim that "machines cannot make mistakes" seems a curious one. One is tempted to retort, "Are they any the worse for that?" But let us adopt a more sympathetic attitude, and try to see what is really meant. I think this criticism can be explained in terms of the imitation game. It is claimed that the interrogator could distinguish the machine from the man simply by setting them a number of problems in arithmetic. The machine would be unmasked because of its deadly accuracy. The reply to this is simple. The machine (programmed for playing the game) would not attempt to give the *right* answers to the arithmetic problems. It would deliberately introduce mistakes in a manner calculated to confuse the interrogator.♠ A mechanical fault would probably show itself through an unsuitable decision as to what sort of a mistake to make in the arithmetic. Even this interpretation of the criticism is not sufficiently sympathetic. But we cannot afford the space to go into it much further. It seems to me that this criticism depends on a confusion between two kinds of mistake. We may call them "errors of functioning" and "errors of conclusion". Errors of functioning are due to some mechanical or electrical fault which causes the machine to behave otherwise than it was designed to do. In philosophical discussions one likes to ignore the possibility of such errors; one is therefore discussing "abstract machines". These abstract machines are mathematical fictions rather than physical objects. By definition they are incapable of errors of functioning. In this sense we can truly say that "machines can never make mistakes".♥ Errors of conclusion can only arise when some meaning is attached to the output signals from the machine. The machine might, for instance, type out mathematical equations, or sentences in English. When a false proposition is typed we say that the machine has committed an error of conclusion. There is clearly no reason at all for saying that a machine cannot make this kind of mistake. It might do nothing but type out repeatedly "0 = 1". To take a less perverse example, it might have some method for drawing

---

♠ FORD, GLYMOUR, AND HAYES: It is hard to know whether it is Turing's own racism in evidence in the preceding sentence, or merely a realistic acknowledgement of the state of race relations in England and her colonies in 1950 (and alas, often since). Given the post-WW2 history of racial relations in the USA, this racial reference may convey an unfortunate impression, but one needs to view it both from a 50-year perspective and a transatlantic shift. Turing was probably making a point about cross-*cultural* difficulties of communication.

♠ FORD, GLYMOUR, AND HAYES: This point is obvious, but seems to undermine the value of imitation game to research in computer science. As Turing seems to realize, playing the imitation game – in any version – is really an exercise in mendacity. When one bears this in mind, it is hardly surprising that relatively little effort has in fact been expended in seriously trying to succeed at such a very silly and pointless goal, when there are so many more interesting and useful applications available for computer technology and AI.

♥ SAYGIN: In a similar vein, can we say neurons can or cannot make mistakes?

conclusions by scientific induction. We must expect such a method to lead occasionally to erroneous results.♠♥

The claim that a machine cannot be the subject of its own thought can of course only be answered if it can be shown that the machine has *some* thought with *some* subject matter. Nevertheless, "the subject matter of a machine's operations" does seem to mean something, at least to the people who deal with it. If, for instance, the machine was trying to find a solution of the equation $x^2 - 40x - 11 = 0$ one would be tempted to describe this equation as part of the machine's subject matter at that moment. In this sort of sense a machine undoubtedly can be its own subject matter. It may be used to help in making up its own programs, or to predict the effect of alterations in its own structure. By observing the results of its own behaviour it can modify its own programs so as to achieve some purpose more effectively. These are possibilities of the near future, rather than Utopian dreams.♠♥

The criticism that a machine cannot have much diversity of behaviour is just a way of saying that it cannot have much storage capacity. Until fairly recently a storage capacity of even a thousand digits was very rare. ♠

♠ FORD, GLYMOUR, AND HAYES: Such programs have been written. The most famous examples are probably Lenat's AM, which learned a number of mathematical concepts, and Simon's BACON program. Inductive learning is now a generally useful technique in many application areas. The same basic techniques are used in televisions to guess likely programs to record, to rapidly detect likely credit card fraud, and in industrial control applications.

♥ SAYGIN: Computers are now routinely used in several real-world applications that are best addressed by inductive learning from large amounts of data. A lot of solicitation we receive (e.g., credit card and loan offers) is based on computations carried out by computer programs that try to predict our interests and spending habits. They make mistakes – otherwise I would not be getting all this junk information about how I should finance my "second home."

♠ FORD, GLYMOUR, AND HAYES: This was not merely a conjecture by Turing. In some sense it was obvious that a machine's interactions with the environment could alter the machine's program, and, equally, that a machine could have a representation of its own program and use that representation as an object of computation – i.e., make inferences about it. Optimizing compilers have been doing this kind of thing for decades. It is amusing to recall that one of the papers presented at one of the first AI meetings ever held, in 1956, was about the design of a Fortran complier.

♥ SAYGIN: Turing is correct about this prediction. The field of machine learning is one of the most fruitful lines of research in AI. Furthermore, computer programs that modify themselves are also used (e.g., many types of genetic algorithm and neural network systems).

♠ FORD, GLYMOUR, AND HAYES: Now that machines have gigantic capacities, it is sometimes objected that they still are rather unenterprising in their behaviour: they do not strike out for newground, form new concepts, or behave unpredictably. This criticism misses the fact that most computer programs are *designed* to be unenterprising because they are more useful that way. It is not hard to make a laptop computer extremely unpredictable in its behaviour, but – like a servant with attention-deficit disorder – that also makes it much less useful. What Turing seems to have in mind is the objection that computers do not seem to adaptively violate their own behavioural regularities in appropriate circumstances, a capacity that Sterrett has suggested is a mark of genuine intelligence. There is, however, no argument that computers cannot be designed to do as much.

The criticisms that we are considering here are often disguised forms of the argument from consciousness. Usually if one maintains that a machine *can* do one of these things, and describes the kind of method that the machine could use, one will not make much of an impression. It is thought that the method (whatever it may be, for it must be mechanical) is really rather base. Compare the parentheses in Jefferson's statement quoted on page 47.♠

3.6.6 Lady Lovelace's Objection

Our most detailed information of Babbage's Analytical Engine comes from a memoir by Lady Lovelace (1842). In it she states, "The Analytical Engine has no pretensions to *originate* anything. It can do *whatever we know how to order it* to perform" (her italics). This statement is quoted by *Hartree* (1949) who adds: "This does not imply that it may not be possible to construct electronic equipment which will "think for itself", or in which, in biological terms, one could set up a conditioned reflex, which would serve as a basis for "learning". Whether this is possible in principle or not is a stimulating and exciting question, suggested by some of these recent developments. But it did not seem that the machines constructed or projected at the time had this property".

I am in thorough agreement with Hartree over this. It will be noticed that he does not assert that the machines in question had not got the property, but rather that the evidence available to Lady Lovelace did not encourage her to believe that they had it. It is quite possible that the machines in question had in a sense got this property. For suppose that some discrete-state machine has the property. The Analytical Engine was a universal digital computer, so that, if its storage capacity and speed were adequate, it could by suitable programming be made to mimic the machine in question. Probably this argument did not occur to the Countess or to Babbage.♠ In any case there was no obligation on them to claim all that could be claimed.

This whole question will be considered again under the heading of learning machines.

A variant of Lady Lovelace's objection states that a machine can "never do anything really new".♣ This may be parried for a moment with the saw, "There is

---

♠ FORD, GLYMOUR, AND HAYES: Exactly. As many have noted, if we know how it works, we are reluctant to call it intelligent.

♠ FORD, GLYMOUR, AND HAYES: To be fair, it is not a very good argument in any case, as the analytical engine had intrinsic speed limitations due to its mechanical construction. Turing here slips a little too quickly between theoretical computability and practical implementability.

♣ HARNAD: This is one of the many Granny objections. The correct reply is that (1) all causal systems are describable by formal rules (this is the equivalent of the Church/Turing Thesis), including ourselves; (2) we know from complexity theory as well as statistical mechanics that the fact that a system's performance is governed by rules does not mean we can predict everything it does; (3) it is not clear that anyone or anything has "originated" anything new since the Big Bang.

nothing new under the sun". Who can be certain that "original work" that he has done was not simply the growth of the seed planted in him by teaching, or the effect of following well-known general principles.♠ A better variant of the objection says that a machine can never "take us by surprise". This statement is a more direct challenge and can be met directly. Machines take me by surprise with great frequency. This is largely because I do not do sufficient calculation to decide what to expect them to do, or rather because, although I do a calculation, I do it in a hurried, slipshod fashion, taking risks. Perhaps I say to myself, "I suppose the voltage here ought to be the same as there: anyway let's assume it is". Naturally I am often wrong, and the result is a surprise for me for by the time the experiment is done, these assumptions have been forgotten. These admissions lay me open to lectures on the subject of my vicious ways, but do not throw any doubt on my credibility when I testify to the surprises I experience.

I do not expect this reply to silence my critic. He will probably say that such surprises are due to some creative mental act on my part, and reflect no credit on the machine. This leads us back to the argument from consciousness, and far from the idea of surprise. It is a line of argument we must consider closed, but it is perhaps worth remarking that the appreciation of something as surprising requires as much of a "creative mental act" whether the surprising event originates from a man, a book, a machine, or anything else.

The view that machines cannot give rise to surprises is due, I believe, to a fallacy to which philosophers and mathematicians are particularly subject. This is the assumption that as soon as a fact is presented to a mind all consequences of that fact spring into the mind simultaneously with it. It is a very useful assumption under many circumstances, but one too easily forgets that it is false. A natural

♠ FORD, GLYMOUR, AND HAYES: Turing's point applies to many philosophical cavils about machine learning. John Norton, at the University of Pittsburgh, once described a machine that learned Newton's laws from empirical regularities of the solar system such as Kepler's laws. The procedure used simple heuristics that could have been applied in other domains, but perhaps not always successfully. The same could be said of the discovery procedures proposed by Pat Langley, Herb Simon, and their collaborators in their *Scientific Discovery*, 1985, and elsewhere. A common objection is that these programs do not really discover anything, because the context, data, and framework are all kludged – built in by the programmer. Newton and a few others (e.g., Einstein) showed enormous flexibility and inventiveness for constructing new mathematical representations framing families of possible theories, and for inventing heuristics for inference to theories so represented. But even they started with an enormous background of tools they did not invent (as even Newton acknowledged: "I have stood on the shoulders of giants."). Einstein, for example, learned electrodynamics as an undergraduate, and his textbook emphasized the key puzzle behind the special theory of relativity, the induction of current by relative motion of conductor and magnetic field. When he turned to attempts to extend relativity to gravitation he learned differential geometry and field theory and used these tools in a heuristic search, over 8 years and many failed attempts, for a satisfactory theory. The mystification of Einstein by some notable historians notwithstanding, it is not implausible to think a program could simulate Einstein's search for the general theory of relativity.

consequence of doing so is that one then assumes that there is no virtue in the mere working out of consequences from data and general principles.♠♣

3.6.7 Argument from Continuity in the Nervous System

The nervous system is certainly not a discrete-state machine. A small error in the information about the size of a nervous impulse impinging on a neuron, may make a large difference to the size of the outgoing impulse. It may be argued that, this being so, one cannot expect to be able to mimic the behaviour of the nervous system with a discrete-state system.♣

It is true that a discrete-state machine must be different from a continuous machine. But if we adhere to the conditions of the imitation game, the interrogator will not be able to take any advantage of this difference. The situation can be made clearer if we consider some other simpler continuous machine. A differential analyser will do very well. (A differential analyser is a certain kind of machine not of the discrete-state type used for some kinds of calculation.) Some of these provide their answers in a typed form, and so are suitable for taking part in the game. It would not be possible for a digital computer to predict exactly what answers the differential analyser would give to a problem, but it would be quite capable of giving the right sort of answer. For instance, if asked to give the value of π (actually about 3.1416) it would be reasonable to choose at random between the values 3.12, 3.13, 3.14, 3.15, 3.16 with the probabilities of 0.05, 0.15, 0.55, 0.19, 0.06 (say). Under these circumstances it would be very difficult for the interrogator to distinguish the differential analyser from the digital computer.

♠ FORD, GLYMOUR, AND HAYES: This tendency is evident in several methodological critics of AI. For example, both Jerry Fodor and Hilary Putnam, interviewed in "Speaking Minds" (Baumgartner and Payr, 1996), seem to feel that the "engineering details" of the actual mechanisms of mechanical thinking are of no real interest (Hayes and Ford, 1997).

♣ HARNAD: Turing is quite right to point out that knowing something is true does not mean knowing everything it entails; this is especially true of mathematical conjectures, theorems, and axioms.

But I think Lady Lovelace's preoccupation with freedom from rules and novelty is even more superficial than this. It takes our introspective ignorance about the causal basis of our performance capacities at face value, as if that ignorance demonstrated that our capacities are actually sui generis acts of our psychokinetic will – rather than being merely the empirical evidence of our functional ignorance, for future reverse-engineering (cognitive science) to remedy.

♣ HARNAD: According to the Church/Turing Thesis, there is almost nothing that a computer cannot simulate, to as close an approximation as desired, including the brain. But, as noted, there is no reason computers should be the only machines eligible for Turing testing. Robots can have analogue components as well. Any dynamical causal system is eligible, as long as it delivers the performance capacity.

3.6.8 The Argument from Informality of Behaviour

It is not possible to produce a set of rules purporting to describe what a man should do in every conceivable set of circumstances.♣ One might for instance, have a rule that one is to stop when one sees a red traffic light, and to go if one sees a green one, but what if by some fault both appear together? One may perhaps decide that it is safest to stop. But some further difficulty may well arise from this decision later. To attempt to provide rules of conduct to cover every eventuality, even those arising from traffic lights, appears to be impossible. With all this I agree.

From this it is argued that we cannot be machines. I shall try to reproduce the argument, but I fear I shall hardly do it justice. It seems to run something like this. "If each man had a definite set of rules of conduct by which he regulated his life he would be no better than a machine. But there are no such rules, so men cannot be machines." The undistributed middle is glaring. I do not think the argument is ever put quite like this, but I believe this is the argument used nevertheless. There may however be a certain confusion between "rules of conduct" and "laws of behaviour" to cloud the issue. By "rules of conduct" I mean precepts such as Stop if you see red lights", on which one can act, and of which one can be conscious. By "laws of behaviour" I mean laws of nature as applied to a man's body such as "if you pinch him he will squeak".♠ If we substitute "laws of behaviour which regulate his life" for "laws of conduct by which he regulates his life" in the argument quoted the undistributed middle is no longer insuperable. For we believe that it is not only true that being regulated by laws of behaviour implies being some sort of machine (though not necessarily a discrete-state machine), but that conversely being such a machine implies being regulated by such laws. However, we cannot so easily convince ourselves of the absence of complete laws of behaviour as of complete rules of conduct. The only way we know of for finding such laws is scientific observation, and we certainly know of no circumstances under which we could say, "We have searched enough. There are no such laws."

♣ HARNAD: First, the successful Turing Test candidate need not be just computational (rule based); all the arguments for T3 robots and their need of real-world sensorimotor capacities, mechanisms, and experience suggest that more is required in a successful candidate than just computation. The impossibility of second-guessing a set of rules that predicts every contingency in advance is probably also behind the "Frame Problem" in (AI) (Harnad, 1993). But it will still be true, because of the Church/Turing Thesis, that the successful hybrid computational/dynamic T3 robot will still be computer-simulable in principle – a virtual robot in a virtual world. So the rule-based system can describe what a T3 robot would do under all contingencies; that simulation would simply not be a T3 robot, any more than its virtual world would be the real world.

♠ FORD, GLYMOUR, AND HAYES: This dichotomy seems, with hindsight, to omit a number of intermediate possibilities, in particular "laws" or "rules" which are purely psychological, but of which we are unconscious.

We can demonstrate more forcibly that any such statement would be unjustified; for suppose, we could be sure of finding such laws if they existed. Then given a discrete-state machine it should certainly be possible to discover by observation sufficient about it to predict its future behaviour, and this within a reasonable time, say a thousand years. But this does not seem to be the case. I have set up on the Manchester computer a small program using only 1,000 units of storage, whereby the machine supplied with one 16-figure number replies with another within 2 s. I would defy anyone to learn from these replies sufficient about the program to be able to predict any replies to untried values.♠

3.6.9 The Argument from Extra-sensory Perception

I assume that the reader is familiar with the idea of extrasensory perception, and the meaning of the four items of it, viz. telepathy, clairvoyance, precognition, and psychokinesis. These disturbing phenomena seem to deny all our usual scientific ideas. How we should like to discredit them! Unfortunately the statistical evidence, at least for telepathy, is overwhelming.♣ It is very difficult to rearrange one's ideas so as to fit these new facts in. Once one has accepted them it does not seem a very big step to believe in ghosts and bogies. The idea that our bodies move simply according to the known laws of physics, together with some others not yet discovered but somewhat similar, would be one of the first to go.

This argument is to my mind quite a strong one. One can say in reply that many scientific theories seem to remain workable in practice, in spite of clashing with E.S.P.; that in fact one can get along very nicely if one forgets about it. This is rather cold comfort, and one fears that thinking is just the kind of phenomenon where E.S.P. may be especially relevant.

A more specific argument based on E.S.P. might run as follows: "Let us play the imitation game, using as witnesses a man who is good as a telepathic receiver, and a digital computer. The interrogator can ask such questions as 'What suit does the

♠ FORD, GLYMOUR, AND HAYES: This seems to be a similar point to that made earlier about prediction and complexity. Turing has testified that machines often surprise him: Here, he seems to be saying, he has a small machine that will surprise *you*. More seriously, however, Turing's general point here seems to be well supported by the last 50 years of cognitive science. Indeed, there are many "laws of behavior" which seem to apply to the workings of our minds, and of which we are quite unconscious. The many fringe areas of consciousness revealed by studies such as those made popular by the writings of such authors as Vilayanur Ramachandran and Oliver Sacks are also eloquent testimonials to the imperfect nature of our own introspections.

♣ HARNAD: It is a pity that at the end Turing reveals his credulousness about these dubious phenomena, for if psychokinesis (mind over matter) were genuinely possible, then ordinary matter/energy engineering would not be enough to generate a thinking mind; and if telepathy (true mind-reading) were genuinely possible, then that would definitely trump the Turing Test.

card in my right hand belong to?' The man by telepathy or clairvoyance gives the right answer 130 times out of 400 cards. The machine can only guess at random, and perhaps gets 104 right, so the interrogator makes the right identification." There is an interesting possibility which opens here. Suppose the digital computer contains a random number generator. Then it will be natural to use this to decide what answer to give. But then the random number generator will be subject to the psychokinetic powers of the interrogator. Perhaps this psychokinesis might cause the machine to guess right more often than would be expected on a probability calculation, so that the interrogator might still be unable to make the right identification. On the other hand, he might be able to guess right without any questioning, by clairvoyance. With E.S.P. anything may happen.

If telepathy is admitted it will be necessary to tighten our test up. The situation could be regarded as analogous to that which would occur if the interrogator were talking to himself and one of the competitors was listening with his ear to the wall. To put the competitors into a "telepathy-proof room" would satisfy all requirements.♠

3.7 Learning Machines♣

The reader will have anticipated that I have no very convincing arguments of a positive nature to support my views. If I had I should not have taken such pains to point out the fallacies in contrary views. Such evidence as I have I shall now give.

Let us return for a moment to Lady Lovelace's objection, which stated that the machine can only do what we tell it to do. One could say that a man can "inject" an idea into the machine, and that it will respond to a certain extent and then drop into quiescence, like a piano string struck by a hammer. Another simile would be an atomic pile of less than critical size: an injected idea is to correspond to a neutron entering the pile from without. Each such neutron will cause a certain disturbance which eventually dies away. If, however, the size of the pile is sufficiently increased,

♠ FORD, GLYMOUR, AND HAYES: This section reads strangely now, but the conviction that the statistical evidence for telepathy was "overwhelming" was not uncommon at the time. (Compare, e.g., Michael Scriven's similar conclusion only a few years later.) But that conviction was seriously shaken when J. B. Rhine's successor, who tried unsuccessfully to guarantee the reality of extrasensory perception by using a computer to randomly generate targets and score subjects' hits and misses, was found to have jimmied the computer to produce faked positive results. There is a less anecdotal argument: even the tiniest quantum effects, when real, can be promulgated through multipliers to reliably produce big effects (Turing hinted at an example earlier). But no ESP phenomenon, by anyone, has ever been so multiplied (Glymour, 1987).

♣ HARNAD: Turing successfully anticipates machine learning, developmental modelling and evolutionary modelling in this prescient section.

the disturbance caused by such an incoming neutron will very likely go on and on increasing until the whole pile is destroyed. Is there a corresponding phenomenon for minds, and is there one for machines? There does seem to be one for the human mind. The majority of them seem to be "sub-critical", i.e., to correspond in this analogy to piles of subcritical size. An idea presented to such a mind will on average give rise to less than one idea in reply. A smallish proportion is supercritical. An idea presented to such a mind may give rise to a whole "theory" consisting of secondary, tertiary, and more remote ideas. Animals minds seem to be very definitely subcritical. Adhering to this analogy we ask, "Can a machine be made to be super-critical?"

The "skin of an onion" analogy is also helpful. In considering the functions of the mind or the brain we find certain operations which we can explain in purely mechanical terms. This we say does not correspond to the real mind: it is a sort of skin which we must strip off if we are to find the real mind. But then in what remains we find a further skin to be stripped off, and so on. Proceeding in this way do we ever come to the "real" mind, or do we eventually come to the skin which has nothing in it? In the latter case the whole mind is mechanical. (It would not be a discrete-state machine however. We have discussed this.)

These last two paragraphs do not claim to be convincing arguments. They should rather be described as "recitations tending to produce belief".♠

The only really satisfactory support that can be given for the view expressed at the beginning of §6, will be that provided by waiting for the end of the century and then doing the experiment described. But what can we say in the meantime? What steps should be taken now if the experiment is to be successful?

As I have explained, the problem is mainly one of programming. Advances in engineering will have to be made too, but it seems unlikely that these will not be adequate for the requirements. Estimates of the storage capacity of the brain vary from 10^{10} to 10^{15} binary digits. I incline to the lower values and believe that only a very small fraction is used for the higher types of thinking.♠ Most of it is probably used for the retention of visual impressions. I should be surprised if more than 10^9 were required for satisfactory playing of the imitation game, at any rate against a blind man. (Note: The capacity of the *Encyclopaedia Britannica*, 11th edition, is 2×10^9.) A storage capacity of 10^7 would be a very practicable possibility even by present techniques. It is probably not necessary to increase the speed of operations of the machines at all.♠ Parts of modern machines which can be regarded as

♠ FORD, GLYMOUR, AND HAYES: What Daniel Dennett later referred to as "intuition pumps." It seems that both sides in these debates are often reduced to pumping.

♠ FORD, GLYMOUR, AND HAYES: These estimates were based on very little neurological data. The current view is that it may be hard to express the storage capacity of the human brain in such simple terms.

♠ FORD, GLYMOUR, AND HAYES: That view now seems unrealistic. It may be that Turing underestimated the computational costs involved in running realistically reliable software simulations. At the time of writing, no one had even attempted to write a program of the complexity of a modern operating system.

analogues of nerve cells work about a thousand times faster than the latter.♠ This should provide a "margin of safety" which could cover losses of speed arising in many ways. Our problem then is to find out how to program these machines to play the game. At my present rate of working I produce about a thousand digits of program a day, so that about 60 workers, working steadily through the 50 years might accomplish the job, if nothing went into the waste-paper basket.♠ Some more expeditious method seems desirable.

In the process of trying to imitate an adult human mind we are bound to think a good deal about the process which has brought it to the state that it is in. We may notice three components:

(a) The initial state of the mind, say at birth
(b) The education to which it has been subjected
(c) Other experience, not to be described as education, to which it has been subjected

Instead of trying to produce a program to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain. Presumably the child-brain is something like a notebook as one buys it from the stationer's. Rather little mechanism, and lots of blank sheets.♠ (Mechanism and writing are from our point of view almost synonymous.) Our hope is that there is so little mechanism in the child-brain that something like it can be easily programmed. The amount of

---

♠ FORD, GLYMOUR, AND HAYES: Again, this comment seems naïve in hindsight, although it was one commonly made at that time. Neurons, we now realize, are far more complex than the simple switch-like components that they were once thought to be. Moreover, as the neurologist Valentino Braitenberg has observed, the average connectivity of the mammalian cortex is so high that it would be impossible to physically assemble a brain even if one were given all the neurons and a connection diagram. This does not argue against Turing's basic point, but it does suggest that the complexity estimates might need to be revised.

♠ FORD, GLYMOUR, AND HAYES: Turing's optimism about the management of large-scale software engineering projects now seems incredible.

♠ FORD, GLYMOUR, AND HAYES: This remark also seems now to be naive, although again it reflects a commonly held view at the time Turing was writing. The range of innate knowledge available to neonates, or occurring automatically in the course of maturation without the application of learning mechanisms, has been the subject of intense research and debate in developmental psychology in recent years. Nativists argue that knowledge of biological distinctions, native physics, and elements of "folk" psychology are innate; non-nativists argue that while some perceptual categorizations and dispositions to imitate may be innate, more sophisticated knowledge is produced by rapid learning mechanisms, so that the internal mental life of even very young children consists of private, and perhaps subconscious, theory construction and revision. Nativist arguments turn on experimental demonstrations of knowledge at ever earlier ages, with the tacit premise that children can not have learned specific ranges of knowledge with the requisite speed. But there has been very little research on learning mechanisms in early childhood. For a discussion of folk psychology, see (Gopnik and Meltzoff, 1996), and for a discussion of hypotheses about early learning mechanisms see (Glymour, 2001).

work in the education we can assume, as a first approximation, to be much the same as for the human child.♥

We have thus divided our problem into two parts. The child-program and the education process. These two remain very closely connected. We cannot expect to find a good child-machine at the first attempt. One must experiment with teaching one such machine and see how well it learns. One can then try another and see if it is better or worse. There is an obvious connection between this process and evolution, by the identifications

Structure of the child machine = Hereditary material
Changes of the child machine = Mutations
Natural selection = Judgment of the experimenter

One may hope, however, that this process will be more expeditious than evolution. The survival of the fittest is a slow method for measuring advantages. The experimenter, by the exercise of intelligence, should be able to speed it up. Equally important is the fact that he is not restricted to random mutations. If he can trace a cause for some weakness he can probably think of the kind of mutation which will improve it.

It will not be possible to apply exactly the same teaching process to the machine as to a normal child. It will not, for instance, be provided with legs, so that it could not be asked to go out and fill the coal scuttle. Possibly it might not have eyes.♥ But however, well these deficiencies might be overcome by clever engineering, one could not send the creature to school without the other children making excessive fun of it. It must be given some tuition. We need not be too concerned about the legs, eyes, etc. The example of Miss *Helen Keller* shows that education can take place provided that communication in both directions between teacher and pupil can take place by some means or other.

♥ SAYGIN: Most neurons can be thought of as having "very little mechanism." The problem is, there is just a lot of that little mechanism in the brain, running in parallel. The way digital computers prefer to process information is fast and serial. The way neurons do it is slower but massively parallel. Almost anything that is worth talking about in terms of "thought" arises out of some mysterious interaction of all these little mechanical events, the sum not just bigger, but somehow also different than the parts. Shall we model the mechanism literally and hope intelligence emerges, or shall we try to figure out meaningful "chunks" in their operation and encapsulate them to represent them the way digital computers prefer it? The former becomes computationally intractable for large numbers of neurons. The latter is prone to errors of interpretation and representation at several levels. Both approaches are being tried in current research, but it is unlikely that either approach alone will work for modeling a substantial component of human thought or behaviour.

♥ SAYGIN: To me, this is the only place in this paper where Turing is clearly wrong. Perceiving, sensing, and acting upon the environment is what knowledge is built upon (e.g., Barsalou, 2000). It might be the case that having some type of sensor and body with which to experience and act upon objects and events is necessary to be able to play the imitation game well (e.g., as argued by Harnad, 1990).

We normally associate punishments and rewards with the teaching process. Some simple child-machines can be constructed or programmed on this sort of principle. The machine has to be so constructed that events which shortly preceded the occurrence of a punishment-signal are unlikely to be repeated, whereas a reward-signal increased the probability of repetition of the events which led up to it. These definitions do not presuppose any feelings on the part of the machine.♥ I have done some experiments with one such child-machine, and succeeded in teaching it a few things, but the teaching method was too unorthodox for the experiment to be considered really successful.

The use of punishments and rewards can at best be a part of the teaching process. Roughly speaking, if the teacher has no other means of communicating to the pupil, the amount of information which can reach him does not exceed the total number of rewards and punishments applied. By the time a child has learnt to repeat "Casabianca" he would probably feel very sore indeed, if the text could only be discovered by a "Twenty Questions" technique, every "NO" taking the form of a blow. It is necessary therefore to have some other "unemotional" channels of communication. If these are available it is possible to teach a machine by punishments and rewards to obey orders given in some language, e.g., a symbolic language. These orders are to be transmitted through the "unemotional" channels. The use of this language will diminish greatly the number of punishments and rewards required.♠

Opinions may vary as to the complexity which is suitable in the child machine. One might try to make it as simple as possible consistently with the general principles.♠ Alternatively one might have a complete system of logical inference "built in". In the latter case the store would be largely occupied with definitions and propositions. The propositions would have various kinds of status, e.g., well-established facts, conjectures, mathematically proved theorems, statements given by an authority, expressions having the logical form of proposition but not belief-value. Certain propositions may be described as "imperatives". The machine

---

♥ SAYGIN: Interesting point… many philosophers of mind talk about beliefs, desires, and emotions as complex and "human" traits. However, emotional systems are one of the better understood systems in neuroscience and are not all that complex or opaque compared with many others. If our emotions, more or less, boil down to levels of a handful of chemicals, it may not be all that far-fetched to just say that a computer "feels" those emotions it is programmed to feel.

♠ FORD, GLYMOUR, AND HAYES: This passage seems oddly out of place. One wonders if Turing here accidentally strayed into reminiscence of his own schooldays for a while.

♠ FORD, GLYMOUR, AND HAYES: The basic idea of making a simple "child machine" which can learn about its own environment was a common trope in early AI. It is probably safe to say that the optimism expressed here has not been borne out in practice. The learning process gone through during a human childhood is extremely complex and still not well understood, but it is certainly not a simple matter of assembling conditioned reflexes under the influence of positive and negative feedbacks. Even language learning, for example, seems to involve very intricate built-in processes that are supplied by genetics.

should be so constructed that as soon as an imperative is classed as "well- established" the appropriate action automatically takes place. To illustrate this, suppose the teacher says to the machine, "Do your homework now". This may cause "Teacher says 'Do your homework now'" to be included amongst the well-established facts. Another such fact might be, "Everything that teacher says is true". Combining these may eventually lead to the imperative, "Do your homework now," being included amongst the well-established facts, and this, by the construction of the machine, will mean that the homework actually gets started, but the effect is very satisfactory. The processes of inference used by the machine need not be such as would satisfy the most exacting logicians. There might for instance be no hierarchy of types. But this need not mean that type fallacies will occur, any more than we are bound to fall over unfenced cliffs. Suitable imperatives (expressed *within* the systems, not forming part of the rules *of* the system) such as "Do not use a class unless it is a subclass of one which has been mentioned by teacher" can have a similar effect to "Do not go too near the edge".

The imperatives that can be obeyed by a machine that has no limbs are bound to be of a rather intellectual character, as in the example (doing homework) given above. Important amongst such imperatives will be ones which regulate the order in which the rules of the logical system concerned are to be applied. For at each stage when one is using a logical system, there is a very large number of alternative steps, any of which one is permitted to apply, so far as obedience to the rules of the logical system is concerned. These choices make the difference between a brilliant and a footling reasoner, not the difference between a sound and a fallacious one. Propositions leading to imperatives of this kind might be "When Socrates is mentioned, use the syllogism in Barbara" or "If one method has been proved to be quicker than another, do not use the slower method". Some of these may be "given by authority", but others may be produced by the machine itself, e.g., by scientific induction.♠

The idea of a learning machine may appear paradoxical to some readers. How can the rules of operation of the machine change? They should describe completely how the machine will react whatever its history might be, whatever changes it might undergo. The rules are thus quite time-invariant. This is quite true. The

♠ Ford, Glymour, and Hayes: This remarkable passage clearly outlines the research program that has been referred to as "good old-fashioned AI" (GOFAI) by John Haugeland; less derisively, it is often called "mainstream AI": the use of an explicit knowledge representation formalism which is manipulated by reasoning engines and linked to actions which are triggered by particular inference patterns. The details differ, but the same basic paradigm underlies McCarthy's logical reasoner-based research program, all of theorem-proving and computational logic, the production-system-based style of psychological modelling pioneered and developed by Simon and Newell, together with its successors, and most work in AI planning and natural language comprehension (Russell and Norvig, 2002). While the adequacy of this paradigm as a basic model for cognitive science is still controversial, the overall success of this research program is beyond dispute; most of these ideas are now part of mainstream computer science and engineering.

explanation of the paradox is that the rules which get changed in the learning process are of a rather less pretentious kind, claiming only an ephemeral validity. The reader may draw a parallel with the Constitution of the USA.

An important feature of a learning machine is that its teacher will often be very largely ignorant of quite what is going on inside, although he may still be able to some extent to predict his pupil's behaviour. This should apply most strongly to the later education of a machine arising from a child-machine of well-tried design (or program). This is in clear contrast with normal procedure when using a machine to do computations: one's object is then to have a clear mental picture of the state of the machine at each moment in the computation. This object can only be achieved with a struggle. The view that "the machine can only do what we know how to order it to do", appears strange in face of this. Most of the programs which we can put into the machine will result in its doing something that we cannot make sense of at all, or which we regard as completely random behaviour. Intelligent behaviour presumably consists in a departure from the completely disciplined behaviour involved in computation, but a rather slight one, which does not give rise to random behaviour, or to pointless repetitive loops. Another important result of preparing our machine for its part in the imitation game by a process of teaching and learning is that "human fallibility" is likely to be omitted in a rather natural way, i.e., without special "coaching". (The reader should reconcile this with the point of view [developed early in this essay].) Processes that are learnt do not produce a 100% certainty of result; if they did they could not be unlearnt.

It is probably wise to include a random element in a learning machine (see p. 438). A random element is rather useful when we are searching for a solution of some problem. Suppose, for instance, we wanted to find a number between 50 and 200 which was equal to the square of the sum of its digits, we might start at 51 then try 52 and go on until we got a number that worked. Alternatively we might choose numbers at random until we got a good one. This method has the advantage that it is unnecessary to keep track of the values that have been tried, but the disadvantage that one may try the same one twice, but this is not very important if there are several solutions. The systematic method has the disadvantage that there may be an enormous block without any solutions in the region which has to be investigated first. Now the learning process may be regarded as a search for a form of behaviour which will satisfy the teacher (or some other criterion). Since there is probably a very large number of satisfactory solutions the random method seems to be better than the systematic.♠ It should be noticed that it is used in the analogous process of evolution. But there the systematic method is not possible. How could one keep track of the different genetical combinations that had been tried, so as to avoid trying them again?

We may hope that machines will eventually compete with men in all purely intellectual fields. But which are the best ones to start with? Even this is a difficult

♠ Ford, Glymour, and Hayes: Indeed, so-called Markov methods which search large spaces starting from random points have proven extremely successful in many applications, and are now routinely used in AI and search processes more generally.

decision. Many people think that a very abstract activity, like the playing of chess, would be best.♠ It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English.♥ This process could follow the normal teaching of a child. Things would be pointed out and named, etc. Again I do not know what the right answer is, but I think both approaches should be tried.♠♥

We can only see a short distance ahead, but we can see plenty there that needs to be done.♠

♠ FORD, GLYMOUR, AND HAYES: As is well-known, early AI work followed Turing's suggestion. Checkers (draughts) was an early success in machine learning, when Samuels' program rapidly learned to outplay its creator. Since then, of course, Deep Blue has become the world's best chess player, and many personal computers can outplay most human amateurs. However, one can ask whether this effort has been as meaningful as its early enthusiasts felt it was. Certainly it seems to have been partly responsible for a rather unhealthy attitude towards AI in which it is assumed that the only purpose of making machines with intellectual capacity is to somehow beat human beings at some game or other. Turing's imitation game itself also has this character, of course. It would be a pity if this emphasis on needless competition were to obscure the far more useful, and we think ultimately far more interesting, goal of making machines which extend, rather than replace, human cognitive and intellectual powers. Most applied AI is in fact devoted to making systems that can be best described as servants, aids, or even cognitive prosthetics to humans, rather than artificial competitors. In beating Kasparov, Deep Blue was not attacking humanity. In fact, a better way to characterize the situation is to say that Deep Blue is a tool with which anyone, even a child, could be world chess champion. The winner, ultimately, is the person moving the chess pieces, no matter what kind of machine he or she is using.

♥ SAYGIN: I have always found it extremely interesting that Turing mentions "chess" and "buying the best sense organs money can buy and teach it to understand and speak English." The former is a completely disembodied task and computers have come to perform it rather well. The latter is done effortlessly by infants across the world but has proven to be very difficult to model on computers. But, we did not follow Turing's advice, we did not buy the best sense organs and let the machine acquire English – or any other language for that matter.

♠ FORD, GLYMOUR, AND HAYES: This is one of the most original sections of a very original paper, and perhaps the part that has drawn the least positive response. Turing proposes nothing less than a behavioural simulacrum of human cognitive development, motor activity aside. To the best of our knowledge, nothing like it has been attempted, and the computational theories that might be applied to the task are only beginning to emerge. Until very recently, developmental psychologists did not take up Turing's implicit challenge to describe algorithms by which children learn the myriad features and regularities of the everyday world. That is beginning to change, especially with the work of Alison Gopnik and her collaborators on young children's procedures for learning causal relationships (Gopnik and Meltzoff, 1997). But one might wonder: if we could program a computer to develop the knowledge of the world, the capacity for recognition, classification, categorization, prediction, learning, and control exhibited in the course of development by a normal human child; if we could do that, would we need the imitation game to convince ourselves that such a marvellous computer thinks?

♥ SAYGIN: Again, so is nature. For instance, the gist of the idea of "eye" has been around in evolution in even very simple early organisms.

♠ FORD, GLYMOUR, AND HAYES: Again, a call to arms! It seems clear that Turing hoped his bold conjecture would motivate research and interest. Certainly Turing could see further ahead than most.