Turing Test Considered Harmful

Patrick Hayes
Beckman Institute
University of Illinois
Urbana, Illinois 61801
U. S. A.

Kenneth Ford
Institute for Human & Machine Cognition
University of West Florida
Pensacola, Florida 32514
U. S. A.

Abstract

Passing the Turing Test is not a sensible goal for Artificial Intelligence. Adherence to Turing's vision from 1950 is now actively harmful to our field. We review problems with Turing's idea, and suggest that, ironically, the very cognitive science that he tried to create must reject his research goal.

1 Introduction

Alan Turing was one of the greatest scientists of this century. His paper [1950], "Computing Machinery and Intelligence" inspired the creation of our field, giving it a vision, a philosophical charter and its first great challenge, the Turing Test.

The Turing Test has been with AI since its inception, and has always partly defined the field. Some AI pioneers seriously adopted it as a long-range goal, and some long-standing research programs are still guided by it; for others it has come to provide more of a vision to define and motivate the whole field. For example, in his recent text, Ginsberg [1993] defines AI as "the enterprise of constructing a physical symbol system that can reliably pass the Turing Test."

Passing the Turing Test is now often understood to mean something like "making an artificial intelligence," without paying too much attention to the details. In this article, however, we will take Turing seriously. We do not think he was being merely metaphorical or speaking in some loose, inspirational way. He seems to have been suggesting the imitation game as a definite goal for a program of research. It was supposed to be a concrete and relatively well-defined goal and hence to avoid the philosophical quagmire that Turing (correctly) predicted would result from debates about whether a computer could properly be described as "intelligent."

But taken this seriously, we will argue, it is no longer a useful idea. The Turing Test had a historical role in getting AI started, but it is now a burden to the field, damaging its public reputation and its own intellectual coherence. We must explicitly reject the Turing Test in order to find a more mature description of our goals; it is time to move it from the textbooks to the history books.

2 Head Games

As the reader probably knows, the Turing Test comprises an imitation game which involves a man, a woman, and a judge, all communicating (but unable to see one another) in a three-way conversation. (The sex of the judge is not specified, and we will use "he" for reasons purely of grammatical simplicity.) The immediate task of the judge is to decide which of the other two is the woman, and the task of each of the players is to persuade the judge that he or she is the woman and that the other is the man. Thus, the game is a test of the ability of a man to pretend to be a woman, and of a woman to resist being judged a man.[1] To make the game more exact, Turing proposes to use an average score over many conversations and to limit the length of each conversation to, say, 10 minutes. Turing then simply says that we should try to make a machine which could successfully "take the place of a man" in this game 70% of the time.

Turing is usually understood to mean that the game should be played with the question of gender (e.g., being female) replaced by the question of species (e.g., being human), so that the judge is faced with the task of differentiating a human participant from a machine pretending to be human. We will call this the species test.

However, Turing does not mention any change to the rules of the imitation game, and there is no need to interpret him as meaning to do so. If we take him at his word, the test is rather clever. It has a woman and a machine each trying to convince the judge that they are a woman, and the judge's task is still to decide which is the woman and which, therefore, is not. But this judge is not thinking about the differences between women and machines, but between women and men. 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. We will call this the gender test.[2]

There are some standard objections to Turing Tests in either version. For example, they seem likely to be extremely difficult: so difficult, indeed, that someone who declared one as his immediate research goal would now probably not be taken seriously. They ignore or sidestep many aspects of

[1] Ruling out vision avoids such complexities as skill in cross-dressing; but in any case Turing thought that machine vision was likely to be very difficult and was irrelevant to the goal of his project.

[2] Judith Genova [1994] argues similarly. We use her 'species/gender' terminology.

current AI research, such as vision and robotics, and seem too closely bound up with natural language understanding to now be a beacon for the entire field. These are familiar objections, but there are deeper ones. The imitation game itself has some basic design flaws, which are inherited by any version of the Turing Test. Later, we will argue that AI should not be defined as an imitation of human abilities in any case.

3 On Not Detecting Anything

One of the first lessons learned by a graduate student in psychology is never to design an experiment to detect nothing. This is such a fundamental error that it has been given a title: confirming the null hypothesis. It is impossible either to completely define the experimental conditions (how hard should one look for the thing that might not be there?) or to come to a firm conclusion (what if one had looked harder, or differently?). The imitation game is precisely such a design, in which a difference between two behaviors is what isn't being detected. Assume for a moment that one accepts the Turing Test as valid: if an artificial intelligence could reliably pass a given instantiation of the test, it would have demonstrated either that its intelligence was genuine or that the judge was not clever enough to ask sufficiently telling questions. But this raises the problem of what exactly *are* the telling questions? Ironically then, the issue that the Turing Test was supposed to avoid remains in force: would it be an adequate criterion for intelligence?

The imitation game conditions say nothing about the judge, but the success of the game depends crucially on how clever, knowledgeable, and insightful the judge is. A clever judge will be looking out for subtle signs of femininity. For example, sociolinguistic studies by Robin Lakoff have shown that adult American women tend to use a wider range of color words than men do. A woman will typically distinguish crimson and scarlet, where a man will usually describe them simply as red, a word which many adult women regard as ambiguous. A good imitation-game judge would know this and be alert for this sign of womanhood; and therefore, a successful player must also be expected to know it, and use it. And of course this applies to any other detectable sexual differences in word usage. But how many such differences are there? The question will always be a matter for research. The imitation game does not have a stable endpoint.

The zero-sum competitive design of Turing's game has more odd consequences. It would not be enough simply to exhibit female use of color words, for example. If a female player notices her opponent (whom she knows to be male) using words like "puce" or "magenta," she might challenge him to engage in an explicit debate about color to test his knowledge and explicitly draw the judge's attention to such attempts to mislead, and a male player would need to able to deal with this. To be a successful player it would not be sufficient simply to have, and therefore exhibit the linguistic symptoms of, a feminine attitude to color; one would have to consciously know of those symptoms and use this knowledge in tactical planning. To be successful at the imitation game, one would have to be thinking all the time about techniques of female impersonation.

Note that such conscious strategic use of sociolinguistics is quite different from exhibiting a symptom of some underlying cognitive difference. Some writers have objected to such things as an implemented model of female use of color vocabulary on the grounds that any such model involving "knowledge representation" can be only understood as modeling conscious thought. The difference between someone who, quite unwittingly, uses a rich color vocabulary and someone who consciously uses knowledge of sociolinguistics to improve his or her imitation game performance provides a vivid illustration of the necessary distinction.

Another problem with null-effect experiments is that they cannot measure anything. The imitation game can test only for complete success. A man who failed to seem feminine in, say, 10% of what he said would almost always fail the imitation game: to pass, one has to be totally convincing almost all the time. This is a criticism of these tests not only as a guide to research—they provide no way to measure partial progress toward the goal—but also of the goal itself. Even in humans we recognize the possibility that an intellectual talent need not correlate with conversational skill or debating ability; but using any kind of imitation game as our research goal denies this simple insight and declares that we must strive to create a fully human-like collection of abilities, organized to succeed in winning an argument.

4 Turing Test Problems

All of these criticisms apply directly to the Turing Tests. This is what we would have to make our program able to do: not talk like a human because it thinks like a human, or even talk like a woman because it thinks like a woman, but rather to talk like a woman as a result of thinking about how best to talk like a woman. The gender test is not a test of making an artificial human, but of making a mechanical transvestite.

Our point here, to emphasize, is not a moral one; rather it is concerned with what the program would have to be thinking about in order to be successful in these artificial games. Human players would also be forced into these artificial frames of mind, which arise simply from the tactical pressures of the games themselves. For example, to succeed at the species test, a machine must not just pass as human, it must succeed in persuading the judge that its human opponent is a machine. To do this would require more than ordinary conversational abilities. The competitive nature of this test makes it essential that the machine give a human-like impression in every possible way and be alert for any way in which its opponent might seem mechanical. To pass this test, a machine would have to not just give a human-like impression, but also be an expert on making a good impression, be always aware of the impression it was giving, and be ready to defend itself against accusations of giving the wrong impression. It would have to take care not to exhibit any inhuman talents which it might have; it would have to always cleverly lie, cheat, and dissemble. The winner of the Loebner competition, for example, sometimes deliberately "mistyped" a word, then backspaced to correct it at human typing speed. This strategy is clever, but surely such tricks should not be central to our subject. To pass the species test we must make not an artificial intelligence, but an artificial con artist.

The species test further reveals the poor experimental design of the imitation game in the difficulty of obtaining an unbiased judge. The general perception of what are essentially human talents keeps shifting. As AI progresses and more and more tasks previously considered to involve human abilities are performed by machines, a judge in the naive Turing Test will gain more and more subtle ways of detecting the behavior of nonhuman machines, just as a skilled doctor will become more adept at recognizing subtle symptoms. Three hundred years ago, when Pascal described his "calculator," (a machine roughly similar to an automobile odometer), European academics were astonished that a machine could perform arithmetic, an intellectual ability that only few humans possessed. Even as late as the second world war, when Turing was working, the ability to perform complex mental calculations rapidly was considered evidence of intellectual talent, found useful throughout science and engineering, and given academic recognition. The ability to perform simultaneous translation may soon be reduced to the "merely mechanical."

When Eliza first appeared, some people found its conversational abilities quite human-like. No machine until then could have reacted even in such a simple way to what had been said to it. But during the Loebner competition, many programs were instantly revealed as nonhuman precisely by the first hint of resemblance of their behavior to that of Eliza. Amusingly, the boundary shifts in both directions. Some judges in the Loebner competition rated a human as a machine on the grounds that she produced extended, well-written paragraphs of informative text, which is now apparently considered to be an inhuman ability in parts of our culture. The Loebner competition illustrates very clearly how the imitation game inevitably slides from a concern with cognitive status to being a test of the ability of the human species to discriminate its members from mechanical imposters.

Turing Tests suffer from another flaw: it is not clear what exactly they can be failing to detect. While Turing was careful to not suggest the test as a definition of humanity or intelligence, the fact that it is often described that way is revealing. Let us say that a Turing Test is a test of "human conversational competence." But what is that, exactly? The only answer is the ability to pass the corresponding Turing Test. The tests are circular: they define the qualities they are claiming to be evidence for. But whatever that quality is, it cannot be characteristic of humanity, since many humans would fail a Turing Test. Since one of the players must be judged to be a machine, half the human population would fail the species test.

5 Inhuman Intelligence

These have all been criticisms of the design of Turing's test considered as some kind of experiment. One might argue, however, that it should be regarded more as a spur to technological progress, much as the goal of getting a man on the moon was undertaken to spur the development of the US space program. However, we will argue that using this test to define our field, even loosely, now leads the field to disown and reject its own successes.

Notice first how parochial, one might even say arrogant, a perspective is assumed by the imitation game. Why should we take it as our goal to build something which is just like us? A dog would never win any imitation game; but there seems to be no doubt that dogs exhibit cognition, and a machine with the cognitive and communicative abilities of a dog would be an interesting challenge for AI (and might usefully be incorporated, for example, into automobiles.)

Likewise, the species detection aspect of the Turing Test has served to focus much AI research on those facets of human behavior which are least susceptible to useful generalization precisely because they are not shared by other species (silicon-based or otherwise). As we develop a general science of cognition, it is the aspects of human thought which are *not* distinctively human that seem the most fundamental. As others have emphasized, cognitive science is a science of cognition, not particularly of human cognition; but we cannot expect to be able to understand human cognition without first having a firm grasp of the basic principles of cognition.

From a practical perspective, why would anyone want to build machines that could pass the Turing Test? As many have observed, there is no shortage of humans, and we already have well-proven ways of making more of them. Human cognition, even high-quality human cognition, is not in short supply. What extra functionality would such a machine provide, even if we could build it?

One answer is that if we could make a human intelligence then we could make a superhuman intelligence just by getting a better processor and extra memory. This vision—the HAL vision of AI—is cited by several AI pioneers (e.g., McCarthy, Feigenbaum, Minsky) and by many people who are worried about what these superintelligences might do.[3] But if we abandon the Turing Test vision, the goal naturally shifts from making artificial superhumans which can replace us, to making superhumanly intelligent artifacts which we can use to amplify and support our own cognitive abilities, just as people use hydraulic power to amplify their muscular abilities. This is in fact occurring, of course, and has been clearly forseen and articulated by others; our point here is only to emphasize how different this goal is from the one that Turing left us with. AI should play a central role in this exciting new technology, but to do so it must turn its back on Turing's dream.

The area generally called "expert systems" has been hugely successful as a technology but is widely perceived, both in the academic community and in the commercial marketplace, as having somehow failed to achieve its goal. The systems produced are often described as "brittle," for example, which is a way of saying that they perform only in their intended domain. That it only performs its intended task would hardly be considered a criticism of most machines, however; and we suggest that it seems a valid criticism here solely because of the lingering influence of the Turing Test measure of AI success. Specialized AI systems are sometimes criticized as being "idiot savants"; but if we abandon the goal of making artificial people, we can rejoice in making useful idiot savants.

[3] The worries seem to arise from the idea that the superhuman intelligence would not just be smart, but would also have superhuman political ambition and be vulnerable to human moral temptation. This might indeed be unwise, but not even the imitation game requires this.

The authors look forward happily to having several such idiots for lawn mowing, tax preparation, etc... We are not here simply reiterating the widespread "hype" complaint that AI has promised too much and found itself unable to deliver. In this area, in fact, AI systems have delivered the goods very well, sometimes spectacularly well. But even this success is often somehow sicklied over with the pale cast of Turing Test insufficiency.

Low fidelity simulations of human behavior are quite a different goal from systems which complement, surpass, and extend our cognitive attributes. The Turing Test does not admit of weaker, different, or even stronger forms of intelligence than those deemed human. This puts AI engineering in a rather ridiculous position. Our most useful computer applications (including AI programs) are often valuable exactly by virtue of their lack of humanity. A truly human-like program would be nearly useless.

One can detect a trend in the marketplace in which instead of selling "intelligence," even a limited version of it called "expertise," engineers are incorporating what might be called cognitive functionality into products whose overall behavior would often not be thought of as particularly intelligent. As AI progresses, we become able to make computers do more and more things, and some of these would be regarded as requiring intelligence—or at any rate cognitive ability of some kind—if a human did them. But this functionality is not made into a special category called "AI ability," or taken to the market as anything having to do with human beings. In fact, the AI is often quite invisible in the final product. There are cameras, copiers, televisions, automobiles, battery rechargers and laptop operating systems all with algorithms incorporated into them which use AI ideas and techniques, but they are not usually advertised as "intelligent" or "expert." They certainly could not pass a Turing Test and there is no particular reason to suppose that they represent an application of a part or component of something that might one day pass a Turing Test. The designers of these systems construe AI as an enabling technology, and reject the Turing Test as a criterion for success. The influence of the Turing Test vision is so pervasive, however, that such work is often not called artificial intelligence just for this reason. This is a tragedy for AI. Our subject is fuelling technical revolutions and changing the world, but Turing's ghost orders us to disinherit these successes.

One is not going to get something which can pass the Turing Test by eventually assembling a collection of these techniques. It would be both far too good and far too bad. It would be lightning-fast and superhumanly accomplished in some ways, curiously inept in others. We could find ways of disguising its inhuman talents, of course; Turing considers this kind of problem explicitly, observing that a machine can always pretend to be worse at arithmetic than it really is. But if one's aim is to provide better machines for people to use, what a silly business to get involved in! It is like gluing a beak and feathers onto an airplane to make it look more like a bird.

6 On Computational Wings

This flight metaphor is quite precise and worth pursuing in more detail. Early attempts to make flying machines often did things like attaching a beak onto the front, or trying to make a wing which would flap like a bird's wing (This extraordinarily persistent idea is found in Leonardo's notebooks and in a textbook on airplane design published in 1911). It is easy for us to smile at such naivete, but one should realize that it made good sense at the time. What birds did was incredible, and nobody really knew how they did it. It always seemed to involve feathers and flapping. Maybe the beak was critical for stability. When one's ignorance was almost total, it made good sense to copy as much of the natural thing as one could, if only to find out what aspects were essential and which were not. A few hundred years ago the idea of artificial flight could have been defined—indeed, often was so characterized—as the idea of making a machine that could fly like a bird. Birds were the only available exemplars for flight then, just as humans were the only exemplars for cognition when Turing was writing. The Turing Test version of artificial flight is just that: make a machine which would be indistinguishable from a bird, if all you could see of it was how it flew. This bird making was the goal of artificial flight for centuries. Most early attempts to make gliders copied aspects of bird structure. As late as 1880, Lielenthal's pioneering experiments with man-carrying gliders used wings and tails clearly based on bird anatomy, and a US patent was issued at the turn of the century for a "flying suit" with wing-linkages covered with feathers.

But progress was actually made when this aim of imitating nature was abandoned. The technology of flight advanced rapidly once workers gave themselves clearly-defined functional goals, separated from any notion of imitating biology, and strove to achieve these goals by any means available. The Wright brothers clearly separated the problems of power-to-weight ratio, lift, lateral stability, pitch and yaw control and solved them one at a time, using such unnatural devices as box kites, launch catapults and vertical fin surfaces. The first successful flyers were very unlike birds, and did not fly like birds. Likewise, the new science of aerodynamics made rapid progress only once it had artifacts with which to perform controlled experiments. The idea of the airfoil (which is crucial to the performance of all birds except the hummingbirds) was not discovered in nature. The shape of real bird wings is far too complex and flexible to suggest the idea of the airfoil; but once it had been discovered by experiments with artificial flyers, and its basic role understood from theory, a gull's wing can easily be recognized as one. Birds are incredibly efficient and clever flyers, and aeronautical engineers still look to them for inspiration; but this productive interaction between technology and biology did not come about by the engineers taking as their goal the task of imitating nature. Indeed, it happened as a direct result of abandoning that naive notion and seeking instead to identify general principles of stable flight and create machines based on them. Similar things are happening now in cognitive science where computational ideas originating in AI are being successfully applied in cognitive psychology, linguistics, and neuroscience.

Artificial flight both transcends and lags far behind natural flight, just as AI machines both surpass and lag behind human intelligence. Airplanes fly at Mach 3, miles above the clouds; but we doubt if an airplane will ever be able to land on the

branch of a tree, or scoop a swimming fish from the ocean. Machines can't lay eggs, of course, but even if we restrict ourselves to matters of flying, birds have talents that will probably always escape technology. No aircraft will pass the Turing Test for flight.

Perhaps human conversation will always be beyond computer abilities in its complexity and subtlety. If so, we should not think that AI has failed, even if the aim of our science is to understand intelligence and of our technology to amplify and extend it. Neither of them should be trying to reproduce it. That is unnecessary for the science and insufficient for the technology.

Even if one's primary goal is essentially psychological, to understand human intelligence, attempting to build a replica of a human is not a sensible approach. But there is no reason why AI, or more generally cognitive science, should define itself in terms of human intelligence or cognition. While this was a natural way to begin, just as flight pioneers began by trying to imitate bird wings, the science itself provides explanations for cognition which deny the uniqueness of any biologically defined categories. Its general insights and ideas apply equally well to electronic computers as to nervous systems, much as aerodynamics applies equally well to airflow over a metal wing as to one covered with feathers. This is not a new observation, but we have only recently begun to understand the extent to which it implies a rejection of imitation-game criteria for success in AI, and how pervasive the consequences of these criteria are.

7 Turing's Ghost

Two venerable intellectual threads weave through history and converge on Alan Turing. To his great credit he grasped them and started knitting what has become the rich tapestry of motivations and ideas that comprise AI. One thread is the idea that machines might somehow process meanings, which runs through Hobbes, Pascal, Leibniz, Boole, Babbage, and many others. The other is the ancient ambition, which is probably older than civilization, to steal divine power by making something come alive. It is reflected, for example, in the Greek myth of Pygmalion, the Golem legend, and the Frankenstein story. Turing was perhaps the first person to be in a position to see how these ancient themes might be brought together. Whether or not he intended it, his insight that technology might, at long last, be able to reach a kind of divine power almost certainly played a key role in motivating early AI projects. Viewed in this historical context, Turing's suggestion of an imitation game seems more understandable; but the same historical view suggests strongly that we must now distinguish legend from science. AI is the proud heir of Boole, Babbage, and Turing, but not of Mary Shelley.

We suspect that several subfields of AI have tended to reject their association with their parent precisely because they found it necessary to develop methodologies which are inconsistent with any kind of Turing Test. Vision, for example, is a perfectly well defined area of scientific investigation or technological ambition within which one can work without feeling obliged to also thereby accept a larger goal of creating a complete intelligent machine. Just as the Turing Tests allow for no degree of partial success, the research programs they define cannot be sensibly taken apart into subfields without an implicit agreement to conform to some kind of grand intellectual architecture, which is not a reasonable constraint to put on either a science or a technology. Turing's legacy alienates maturing subfields with methodological inconsistencies. Abandoning the Turing test as an ultimate goal is almost a requirement for any rational research program which declares itself interested in any particular part of cognition or mental ability. Let us emphasize again, we do not deny the need for this abandonment. The harm is done when this is perceived as abandoning AI.

Allowing our field to be defined by a Turing Test also harms its reputation, in direct and subtle ways. Perhaps not surprisingly, many lay critics of AI assume the field to be defined by Turing's goal, and by this light it does not seem to be doing very well. For just one example, Frederick Allen, writing in 'The Atlantic' [1994]:

> Today traditional artificial intelligence, or AI, is a backwater at best, and the confidence with which it was once pursued seems unimaginable. Nobody has ever designed a program that can converse at all convincingly on a single subject, and the field has splintered into disparate parts. .... The grand vision has nearly vanished.

Such a pessimistic summary of a flourishing research area like ours may seem merely to reflect ignorance. But if one identifies AI with the goal of passing a Turing Test—the "grand vision" to which Allen refers—then he is perfectly correct. The mistake lies in allowing that identification.

Another much-cited attack on AI may arise in part from an insight into the Turing Test. As we have emphasized, in order to succeed at the imitation game even a human player would be obliged to think consciously, to an unnatural extent, about what effects his utterances might be having on a listener. It is quite natural to go from this insight to Searle's idea that AI programs can be only a simulation of cognition, leading to his notorious distinction between strong and weak AI. The Turing Test indeed challenges a computer to simulate a woman, rather than be one.

Finally, perhaps the most subtle kind of damage which the Turing Test has done to AI is by limiting its sights. Ironically, Turing's daring vision may in fact be too restrictive. All versions of the Turing Test are based on a massively anthropocentric view of the nature of intelligence. Turing correctly insisted that his test was not meant to define intelligence. Nevertheless, in giving us this touchstone of success, he chose human intelligence—in fact, even more peculiarly, the arguing skill of the educated English middle class in playing a kind of party game—as our goal. But the very science which Turing directed us towards provides a perspective from which a much broader and more satisfying account of intelligence is emerging. The Turing Test focuses our attention on the most human details of behavior, rather than general computational principles of cognition.

8 What Is AI?

AI has always wondered how to define itself, and has engaged in a long-running territorial battle with other parts of computer science. Techniques are often developed in AI and later absorbed into mainstream computer science. Unlike most subdisciplines of computer science, AI seems to be defined

not by its methods but by the source of their inspiration.[4]

So, which parts of computer science are part of AI? We suggest a rather radical answer to this question: all of them. AI is not a part of computer science in the way that compiler design, object-oriented programming or genetic algorithms are. AI is the business of using computation to make machines act more intelligently, or to somehow amplify human intelligence. It is not a particular collection of methods, or a programming style. Any technique can be used by a program to do something intelligent, or to display a cognitive ability.

Until perhaps a decade ago many computational methods were pioneered in AI or in close association with it. One of the first compilers was reported at a meeting on intelligent machines. Larry Tesler has suggested that AI be defined to be the part of computer science where things don't work properly yet; the edge of the ice, as it were. But this may have been simply a historical consequence of the fact that many of the creative pioneers of computer science had accepted Turing's dream, and were struggling to make computers act "intelligently." It isn't true any longer, and many of the most exciting new ideas in computation are now being developed in other parts of computer science which have quite different aims. But it would be foolish to regard these methods as somehow excluded from AI.

For example, there has been a long-standing intellectual struggle in machine translation between methods based on explicit semantic representations and those which apply statistical techniques to large lexical corpora. This is often described as a battle between AI methods and other, non-AI, methods. While this may be an accurate account of the sociology of the two sides, it makes no scientific or technological sense. Our aim might be to model the skill of human translators, or to make an effective mechanical translator: either way, we should not have any *ad hoc* constraints on what computational methods to use. If it works, or seems plausible, try it. AI has difficult enough problems already, without also having its technical hands tied.

Consider again the analogy with flight. If cognitive psychology, psycholinguistics, etc. are like the study of natural flight in all its complexity, and AI is like aeronautical engineering, then computer science supplies the aerodynamic theory. The fundamental insight of cognitive science might be summarized by saying that computational science supplies, as it were, the dynamics of cognition. Just as Turing predicted almost half a century ago, the empirical sciences of natural cognition now share a computational vocabulary with the engineering discipline of AI.

This picture of our field defines it in a more useful and more mature way than Turing could give us. AI is the engineering of cognition based on the computational vision which runs through and informs all of cognitive science. We expect AI to produce cognitive artifacts; things that think, see, communicate, plan, play and argue in some way. Perhaps not in a human way, but somehow useful to humans. Exactly what counts as "cognitive" will shift and change, and be altered by the science itself, just as the meaning of words like "energy" has been changed by physics. But ultimately, this doesn't matter. Turing's ultimate aim, which we can happily share, was not to describe the difference between thinking people and unthinking machines, but to remove it.

[4] Attempts to define AI in terms of its computational methods never work properly. For example, if AI is the study of search then successful learning processes automatically remove themselves from the discipline.

9 Coda: The Human Condition

Colby [1975, 1981] has argued that we should consider variations of the gender test, where the judge is asked to make different kinds of distinction. For example, the judge might be asked to decide which of the interrogants was really a child, or really an Englishman; or, in a more familiar example, the judge might be a clinical psychologist trying to diagnose which of them is really paranoid. On this view, Turing's choice of sex as the topic of conversation had no particular significance. It may have been chosen simply because it seems impossible to give any *a priori* bounds on the subject matter of the resulting conversation.

However, Turing was a careful enough thinker that he would have suggested this diverse-topic interpretation of his game had this been what he had in mind, and we again propose to take him at face value. He seems to have chosen the topic of sexual identity deliberately. It is hard to avoid noticing that for Turing, the problem of how to convincingly display a sexual identity was more than just deliberately vague. It was a real problem at the very core of his emotional and social life. Turing was openly gay at a time when homosexuality was a crime in England and was widely regarded as unnatural and deviant. He was prosecuted for homosexuality, and avoided prison only by submitting to a six-month program of "rehabilitation" involving hormone injections which, among other things, caused his body to grow breasts. This bizarre and horrifying treatment is thought to have been part of the reason for his suicide in 1954.

We suspect that Turing chose this topic because he wanted the test to be about what it really means to be human. This is why he has set us up in this way. He tells us, quite clearly, to try to make a program which can do as well as a man at pretending to be a woman. If we really tried to do this, we might be forced into thinking very hard about what it really means to be not just a thinker, but a human being in a human society, with all its difficulties and complexities. If this was what Turing meant, then we need not reject it as our ultimate goal.

References

[Allen, 1994] Frederick Allen. Unreasonable facsimile: Do we really want computers to be more like us? *The Atlantic,* 274(2): 20(3), August 1994.

[Colby, 1981] Kenneth M. Colby. Modeling a Paranoid Mind. *Behavioral and Brain Sciences,* 4: 515-534, 1981.

[Colby, 1975] Kenneth M. Colby. *Artificial Paranoia.* Pergamon, New York, New York, 1975.

[Genova, 1994] Judith Genova. Turing's Sexual Guessing Game. *Social Epistemology,* 8(4): 313-326, 1994.

[Ginsberg, 1993] Matt Ginsberg. *Essentials of Artificial Intelligence.* Morgan Kaufmann, San Mateo, California, 1993.

[Turing, 1950] Alan M. Turing. Computing Machinery and Intelligence. *Mind,* 59: 433-460, 1950.