Turing Test: 50 Years Later

Ayse Pinar Saygin
Ilyas Cicekli
Varol Akman

Department of Computer Engineering and Information Science
Bilkent University
Bilkent, Ankara, 06533, Turkey
{psaygin,ilyas,akman}@cs.bilkent.edu.tr

SUBMITTED FOR PUBLICATION

Abstract

The Turing Test is one of the most disputed topics in Artificial Intelligence, Philosophy of Mind and Cognitive Science. This paper is a review of the past 50 years of the Turing Test. Philosophical debates, practical developments and repercussions in related disciplines are all covered. We discuss Turing's ideas in detail and present the important comments that have been made on them. Within this context, behaviorism, consciousness, the 'other minds' problem and similar topics in the philosophy of mind are discussed. We also cover the sociological and psychological aspects of the Turing Test. Finally, we take a look at the current situation and analyze the programs that have been developed with the aim of passing the Turing Test. We conclude that the Turing Test has been, and will continue to be, a very influential and controversial topic.

1 Introduction

This is the story of the Turing Test; a modest attempt to summarize its 50 years of existence.

Alan Turing[1], British mathematician, proposed the Turing Test (TT) as a replacement for the question "Can machines think?" in his 1950 *Mind* article 'Computing Machinery and Intelligence' [Tur50]. Since then, it has been a widely discussed topic. It has been attacked and defended over and over. At one extreme, Turing's paper has been considered to represent the "beginning" of AI and the TT was considered its ultimate goal. At the other, the TT has been called useless, even harmful. In between are arguments on consciousness, behaviorism, the 'other minds' problem, operational definitions of intelligence, necessary and sufficient conditions for intelligence-granting, and so on.

It will be the aim of this paper to present the 50 years of the TT. We have tried to make this survey as comprehensive and multi-disciplinary as possible. Important concepts are introduced, and discussed in an easy-to-understand manner. Familiarity with special terms and concepts is not assumed. The reader is directed to further references when they are available. While the

[1] For information on Turing refer to the excellent biography by Andrew Hodges [Hod83] or the Alan Turing page at http://www.turing.org.uk/turing, also maintained by Hodges.

review is not strictly chronological, we have tried to present related works in the order they appeared. Interdisciplinary readership is assumed and no particular aspect of the TT (e.g., philosophical or computational) is taken as a focal point.

In our attempt to make this survey complete, we have explored a large number of references. However, this does not mean that we have commented on each paper that mentions the TT. The reader will notice that we have devoted separate sections to certain papers, discussed some others briefly and merely cited the remaining. We made these decisions according to our opinions of what is to be expanded upon in a review of this sort. From this it should not be understood that the papers we spare less space are less important or interesting. In fact, we devoted more space to papers that are not discussed in detail elsewhere[2]. Some papers were explained in detail because they are *representative* of some important ideas.

The rest of the paper is organized as follows: Section 2 introduces the TT and analyzes 'Computing Machinery and Intelligence' [Tur50]. In this section, we also attempted to develop new ideas and probe side issues. Section 3 describes and explains some of the earlier (those from the 60's and the 70's) comments on the TT. In Section 4 we analyze the arguments that are more recent. We chose to study the repercussions of the TT in the social sciences separately in Section 5. Similarly, in Section 6, we give an overview of the concrete, computational studies directed towards passing the TT. Some natural language conversation systems and the annual Loebner Prize contests are overviewed in this section. Finally, Section 7 concludes our survey.

[2]For instance, the discussion of Searle's Chinese room is kept short (Section 4.2), not because it is irrelevant or unimportant, but because there is an abundance of excellent resources on the subject. Conversely, Ned Block's arguments are described in more detail (Section 4.1) because not many in-depth analyses of them were found in the literature.

2 Turing's 'Computing Machinery and Intelligence'

It makes sense to look at Turing's landmark paper 'Computing Machinery and Intelligence' [Tur50] before we begin to consider certain arguments defending, attacking or discussing the TT [3]. In analyzing the 50 years of the TT, it is important to distinguish what has been originally proposed by Turing himself and what has been added on afterwards. We aren't saying, by any means, that the TT is what (or should remain as) Turing proposed in 'Computing Machinery and Intelligence'. As any other concept, it has changed throughout the 50 years it has been around. In fact, one of the purposes of this paper is to trace the steps in this evolution. Thus, it is only natural that we are interested in the original version.

In Section 2.1, we will analyze Turing's original proposal. We summarize Turing's replies to certain objections to his ideas in Section 2.2. Turing's opinions on learning machines are briefly discussed in Section 2.3. Finally, we list some predictions of Turing in Section 2.4.

2.1 The Imitation Game

Turing's aim is to provide a method to assess whether a machine can think or not. He states at the beginning of his paper that the question "Can machines think?" is a highly ambiguous one. He attempts to transform this into a more concrete form by proposing what is called the Imitation Game (IG). The game is played with a man (A), a woman (B) and an interrogator (C) whose gender is unimportant. The interrogator stays in a room apart from A and B. The objective

[3]On the other hand, we do not wish to repeat Turing's ideas here for this would be rather redundant for most of our readers. [Tur50] is a very well-known work and has been cited and quoted copiously. Although what follows will provide an introduction to the TT, it is a good idea to read Turing's original rendering of the issues at hand.

of the interrogator is to determine which of the other two is the woman while the objective of both the man and the woman is to convince the interrogator that he/she is the woman and the other is not. This situation is depicted in Figure 1.

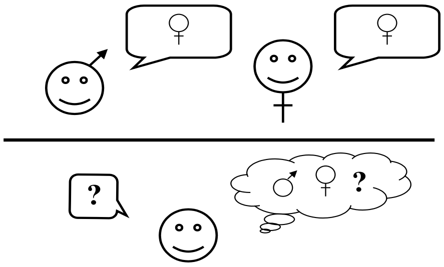

Figure 1: The Imitation Game: Stage 1

The means through which the decision, the conviction, and the deception is to take place is a teletype connection. Thus, the interrogator will ask questions in written natural language and will receive the answers in written natural language. Questions can be on any subject imaginable, from mathematics to poetry, from the weather to chess.

According to Turing, the new agenda to be discussed, instead of the equivocal "Can machines think?", can be '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 played like this as he does when the game is played between a man and a woman?' [Tur50, p. 434]. Figure 2 depicts the new situation.

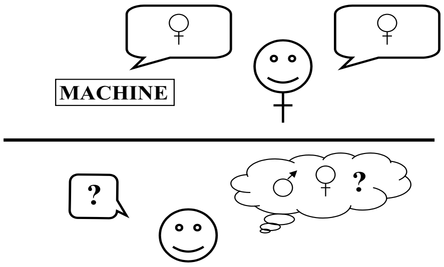

Figure 2: The Imitation Game: Stage 2, Version 1

At one point in the paper Turing replaces the question "Can machines think?" by the following:

> 'Let us fix our attention to 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 programme, C can be made to play satisfactorily the part of A in the imitation game, *the part of B being taken by a man*?' [Tur50, p. 442, emphasis added].

Notice that the woman has disappeared altogether. But the objectives of A, B and the interrogator remain unaltered; at least Turing does not explicitly state any change. Figure 3 shows this situation.

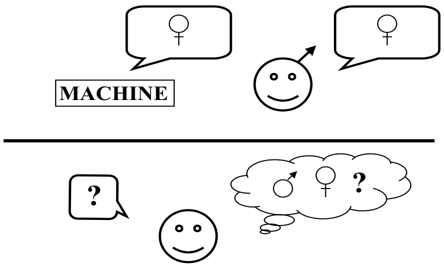

Figure 3: The Imitation Game: Stage 2, Version 2

There seems to be an ambiguity in the paper; it is unclear which of the scenarios depicted in Figure 2 and Figure 3 is to be used. In any case, as it is now generally understood, what the TT really tries to assess is the machine's ability to imitate a human being, rather than its ability to simulate a woman. Most subsequent remarks on the TT ignore the gender issue and assume that the game is played between a machine (A), a human (B) and an interrogator (C). In this version, C's aim is to determine which one of the two entities he/she is conversing with is the human (Figure 4).

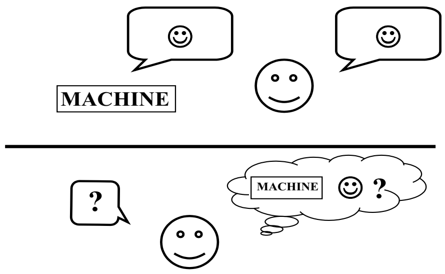

Figure 4: The Imitation Game as is generally interpreted (The Turing Test)

One may ask why Turing designed the IG in such a peculiar manner. Why the fuss about the woman, the man and the replacement? This does not make the paper easier to understand. He could have introduced the IG exactly as he did with the woman-man issue replaced by the human-machine issue and it obviously would not be any more confusing. One argument that can be made against this is that machines (at least those that could be built or imagined in 1950) playing against men in such a game would sound ridiculous at first. A man imitating a woman, on the other hand, has higher prospects of success in the eyes of the average person. In other words, it can be said that the gender-based imitation game sets the mood for what's coming. This, we believe, is not a very convincing argument. The main reason that the decision concerning machine thought is to be based on imitating a woman in the game is probably not that Turing believed the ultimate intellectual challenge to be the capacity to act like a woman (although it may be comforting to entertain the thought). Conversely, it may be concluded that Turing believes that women can be imitated by machines while men cannot. The fact that Turing stipulated the man to be replaced by the machine (when he might just as easily have required the woman to be replaced by the machine or added a remark that the choice was insubstantial) raises such questions, but let us not digress.

Here is our explanation of Turing's design: The crucial point seems to us that the notion of *imitation* figures more prominently in Turing's paper than is commonly acknowledged. For one thing, the game inherently possesses deception. The man is allowed to say anything at all in order to cause the interrogator to make the wrong indentification, while the woman is actually required to aid the interrogator[4]. In the machine vs. woman version, the situation will remain the same. The machine will try to convince the interrogator that it is the woman. What is really judging the machine's competence is not the woman it is playing against. Turing's seemingly frivolous requirements may actually have very sound premises. Neither the man in the gender-based IG nor any kind of machine is a woman. On close examination, it can be seen that what Turing proposes is to compare the machine's success against that of the man, *not* to look at whether it 'beats' the woman in the IG[5]. The man and the machine are measured in terms of their respective performances against real women.

[4]Turing suggests that the best strategy for her would most probably be giving truthful answers to the questions.

[5]This inadvertently figures in the final result, but indirectly.

In Figure 3, we see that the woman has disappeared from the game, but the objective for both the machine and the man is still imitating a woman. Again, their performance is comparable because they are both simulating something which they are not.

The quirks of the IG may well be concealing a methodological fairness beyond that explicitly stated by Turing. We hold that the IG, even though it is regarded as being obscure by many, is a carefully planned proposal. 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 "fake".

Turing could have defined the game to be played with two people, too; one being interrogator, as in the original, and the other being either a man or a woman. The interrogator would then have to decide whether the subject is a man or a woman. Alternatively, the TT for machine intelligence can be reinterpreted as a test to assess a machine's ability to pass for a human being. This issue may seem immaterial at first. However, the interrogator's decision is surely to be affected by the availability (or lack) of comparison. Whether the machine's task will be easier or more difficult in this latter case is another question. We think that Turing intended to imply that some comparison should be available, for otherwise, he could have opted for the two-people version of the game. This implies that the game can be played with the result 'A is the woman' actually meaning 'A seems more woman-like than B'. In turn, a more varied set of questions can be used when the interrogator is trying to judge the gender of the subjects. Once again, we believe that the most sensible reason behind the three-person game is to have a neutral party so as to allow the assessment of the impersonating parties with respect to each other.

In any case, as was mentioned before, the TT concept has evolved through time. Turing's original IG and its conditions do not put serious constraints on current discussions about the test. It is generally agreed that the gender issue and the number of participants are not to be followed strictly in attempts to pass, criticise or defend the TT. Even Turing himself, in the subsequent sections of 'Computing Machinery and Intelligence', sometimes ignores these issues and focuses on the question: "Can machines communicate in natural language in a manner indistinguishable from that of a human being?". This is manifested in the example conversation he gives in [Tur50, p. 434], which contains questions about poetry, mathematics and chess—topics that one would not typically ask about in order to find out the gender of someone. This may be a hint that the gender issue in the IG is indeed for purposes of fair comparison.

After defining the IG, Turing defends the choice of replacing the "Can machines think?" question with "Can machines play the imitation game?". The new problem focuses on intellectual capacities and does not let physical aspects interfere with granting intelligence to an entity. Nor does it limit thinking to specific tasks like playing chess or solving puzzles, since the question-and-answer method is suitable to introduce any topic imaginable.

An issue that is open to discussion is what Turing implies about *how* machines should be built or programmed to play the IG successfully. He seems to believe that if a machine can be constructed to play the game successfully, it does not really matter whether what it does to that end is similar to what a man does or not. Here it can be seen that Turing almost encourages prospective attempts to pass the test to utilize any kind of strategy whatsoever. He even considers the possibility that a machine which successfully plays the IG cannot be explained by its creators because it had been built by experimental methods. However, he explicitly states that 'it will be assumed that the best strategy is to try to provide answers that would naturally be given by a man' [Tur50, p. 435]. It may be concluded that Turing does not put any limitations on how to model human cognitive processes, but seems to discourage any approach that deviates too much from the "human ways", possibly because he feels it is unlikely that satisfactory solutions can be obtained in this manner. On the other hand, by not committing himself to any extreme viewpoint on the issue, he accepts the possibility that machines not mimicking human cognitive processes at all can also pass the test.

The IG, as was mentioned, has deception at its very heart. It is therefore apparent that "cheating" will be an integral part of the TT. Moreover, by not stipulating certain techniques or strategies to be used, Turing explicitly allows this. As Turing described it, in the game there are no rules constraining the design of the machines.

Turing promotes cheating implicitly, too. At various places in the paper, he describes how machines could be "rigged" to overcome certain obstacles proposed by opponents of the idea that machines can think. A very obvious example is about machines making mistakes. When the machine is faced with an arithmetical operation, in order not to give away its identity by being fast and accurate, it can pause for about 30 seconds before responding and occasionally give a wrong answer. Being able to carry out arithmetical calculations fast and accurately is generally considered inteligent behaviour[6]. However, Turing wishes to sacrifice this at the expense of human-ness. But this is cheating, is it not? Maybe, but the arithmetics domain is a highly specific one. Cheating in this manner cannot hurt; if a machine can pass the test, it can then be reprogrammed not to cheat at arithmetics. If it does not cheat, the interrogator can ask a difficult arithmetical problem as his/her first question and decide that he/she is dealing with a machine right then and there.

It can be seen that Turing does not seem to be skeptical towards the idea that a sufficiently human-like machine (i.e., a machine that is sufficiently good at playing the IG) is bound to make such mistakes as we attribute to humans, without any explicit cheating done by its constructors. This idea may seem extravagant, but considering the high level of sophistication required from a machine for passing the TT, it should not be dismissed as being impossible. As Turing also mentions, nothing can really stop the machine from drawing incorrect conclusions without being specifically programmed to do so. A striking example can be given from the inductive learning domain: No learning algorithm guarantees correct results on unseen data. Moreover, in some cases a computer errs in ways that cannot be foreseen, or even understood by its programmer. This can be distressing for machine learning researchers who are after a minimal number of mistakes, but proves the subtle point that machines can make mistakes without explicitly shown *how to*[7]. Since the human mind occasionally draws incorrect conclusions inductively, the fact that machines can act similarly should contribute to the arguments that refute the notion that machines cannot make mistakes.

Turing's approach towards cheating seems similar to that of Adam Smith's "invisible hand" from economics. Maybe Turing's conformity to cheating has its roots in his belief that one cannot go too far by such attempts: He may regard cheating as a last retouch, something to smooth out the machine-ness of the resulting machines that

[6] Although even simple devices like calculators are better at this than average human beings, it is rare that a mathematical whiz who can multiply 8-digit numbers in seconds is regarded as being of ordinary intellect.

[7] Readers are referred to Section 2.3 of this paper, [Tur50, pp. 454-460], and [Tur69, pp. 14-23] for very entertaining and insightful comments on machine learning by Turing.

otherwise handle the more important aspects of human cognition. If a program that has its very bases in what we now call cheating can pass the TT, maybe we would have to revise some notions about the human intellect. It is not possible to say what Turing was thinking and claim to be absolutely correct. It seems as if he would be content with a machine that plays the IG successfully no matter what the inner mechanisms are.

2.2 Contrary Views and Turing's Replies

Turing was aware that some of his ideas would be opposed at the time he wrote 'Computing Machinery and Intelligence' [Tur50] and he responded to some objections that he believed his work would be confronted with. In fact, he discusses some of these earlier in [Tur69][8]. We direct the reader to [Tur50] for the answers to the *theological objection*, and the *argument from extra-sensory perception* for these are rather irrelevant to the current work. However, the remaining objections are worth commenting on.

The *'heads in the sand' objection*, although mostly in disguised forms, is manifested in some subsequent comments on the TT. This is, in its basic form, an aversion of the issue of thinking machines because the consequences of this would be dreadful [Tur50, p. 444]. Most people like to believe that humans are "special" and thinking is considered one of the most important traits that make us so. To some, the idea of sharing such a "human" ability with machines is not a pleasant thought. This outlook was probably more widespread in Turing's time than it is now. Turing believes that this argument is not even worth refutation, and with a little sarcasm, states that consolation (perhaps in the transmigration of souls) is more appropriate [Tur50, p. 444].

There are some theorems showing the powers of discrete-state machines are limited. The most famous of these is probably Gödel's theorem which shows that in consistent logical systems of sufficient power, we can formulate statements that cannot be proved or disproved within the system. An application of this result to the IG is outlined in [Tur50, p. 445] and the reader can refer to [Luc61, Luc96] for more on the implications of Gödel's theorem for machine thought.

Turing studies such results under the title the *mathematical objection*. He states 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' [Tur50, p. 445]. Elsewhere, he notes that those arguments that rest on Gödel's and similar theorems are taking it for granted that the machine must not make mistakes, but that this is not a requirement for intelligence [Tur69].

Perhaps the most important objection is the *argument from consciousness*. Some people believe that machines should be conscious (e.g., aware of their accomplishments, feel pleasure at success, get upset at failure, etc.) in order to have minds. At the extreme of this view, we find *solipsism*. The only way to *really* know whether a machine is thinking or not is to *be* that machine. However, according to this view, the only way to know another human being is thinking (or is conscious, happy, etc.) is to be that human being. This is usually called the *other minds problem* and will show up several times in the discussions of the TT. 'Instead of arguing continually over this point it is usual to have the polite convention that everyone thinks' [Tur50, p. 446]. Turing's response to the argument from consciousness is simple, but powerful: The alternative to the IG (or similar behavioral assessments) would be solipsism and we do not practice this against other humans. It's only fair that in dealing

[8] Although the reference cited is published in 1969, Turing originally wrote the paper in 1948.

with machine thought, we abandon the consciousness argument rather than concede to solipsism.

Turing believes that the IG setting can be used to determine whether 'someone really understands something or has learnt it parrot fashion' as is manifested in the sample conversation he gives in [Tur50, p. 446]. It should also be noted that Turing states he does not assume consciousness to be a trivial or impertinent issue; he merely believes that we do not necessarily need to solve its mysteries before we can answer questions about thinking, and in particular, machine thought [Tur50, p. 447].

The *arguments from various disabilities* are of the sort "machines can never do X", where X can be any human trait such as having a sense of humor, being creative, falling in love or enjoying strawberries. As Turing also notes [Tur50, p. 449], such criticisms are sometimes disguised forms of the argument from consciousness. Turing argues against some of these X's such as the ability to make mistakes, enjoy strawberries and cream, be the subject of its own thought, etc. in [Tur50, pp. 448-450].

Lady Lovelace's objection is similar; it states that machines cannot originate anything, can never do anything new, can never surprise us. Turing replies by confessing that machines do take him by surprise quite often. Proponents of Lady Lovelace's objection can say that 'such surprises are due to some creative mental act on [Turing's] part, and reflect no credit on the machine' [Tur50, p. 451]. Turing's answer to this is similar to the one he gives to the argument from consciousness: '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.' [Tur50, p. 451].

Turing also considers the *argument from continuity in the nervous system.* As the name suggests, this objection states that it is impossible to model the behavior of the nervous system on a discrete-state machine because the former is continuous. However, Turing believes that the activity of a continuous machine can be "discretized" in a manner that the interrogator cannot notice during the IG.

Finally, there is the *argument from informality of behavior.* Intuitively, it seems it is not possible to come up with a set of rules that describe what a person would do under every situation imaginable. In very simple terms, some people believe the following: '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.' [Tur50, p. 452]. First, Turing notes that there might be a confusion between 'rules of conduct' and 'laws of behavior'. By the former he means actions that one can perform and be aware of (like, 'If you see a red light, stop') and by the latter he means laws of nature that apply to a man's body (such as 'If you throw a dart at him, he will duck'). Now it is not evident that a complete set of laws of behavior do not exist. We can find some of these by scientific observation but there will not come a time when we can be confident we have searched enough and there are no such rules. Another point Turing makes is that it may not always be possible to predict the future behavior of a discrete-state machine by observing its actions. In fact, he is so confident about a certain program that he set up on the Manchester computer that he 'def[ies] anyone to learn from [its] replies sufficient about the programme to be able to predict any replies to untried values' [Tur50, p. 453].

2.3 Learning Machines

Turing devotes some space to the idea of *education of machinery* in 'Computing Machinery and Intelligence' [Tur50]. He also discusses the issue in his earlier work 'Intelligent Machinery' [Tur69].

According to Turing, in trying to imitate an adult human mind, we should consider three issues: the initial state of the mind, the education it has been subject to, and other experience it has been subject to (that cannot be described as education). Then we might try to model a child's mind and "educate" it to obtain the model of the adult brain. Since 'presumably the child-brain is something like a note-book as one buys it from the stationers; rather little mechanism and lots of blank sheets' [Tur50, p. 456], developing a program that simulates it is bound to be easier[9]. Of course, the education is another issue. Turing proposes some methods of education for the child-machines (such as a reward/punishment based approach) in [Tur50, pp. 456-460] and [Tur69, pp. 17-23].

Turing's opinions on learning machines are rather interesting, especially considering he wrote these more than 50 years ago. We will not digress into discussions of his ideas on the specifics or the realizability of his proposals. We would like to note one thing though: In most places when he discusses education of machines, there is a noticable change in Turing's style. He seems to believe that the way to success in developing a program that plays the IG well is probably following the human model as closely as possible. As was mentioned in Section 2.1, he does not put any constraints on how to design the IG-playing machine, but the fact that describes learning machines in substantial detail seems to suggest that *he* would prefer such an approach.

In any case, Turing believes '*if* we are trying to produce an intelligent machine, and are *following the human model as closely as we can*' [Tur69, p. 14, emphasis added] a good (and fair) approach would be to allow the machine to learn just like humans.

[9] Turing seems to believe that brains of newborn babies are *tabula rasa*. However, he also considers the opposite and states that we might encode the information at various kinds of status levels (e.g., established facts, conjectures, statements given by an authority) and thereby implies that we may model any 'innateness' there may be [Tur50, pp. 457-458].

2.4 Turing's Predictions

Turing's paper [Tur50] contains some very bold comments on the prospects of machine intelligence. Most of these probably seemed like science fiction at the time. Even now, some of us would consider these far-fetched. This section aims to provide a sample of Turing's predictions that we found interesting.

It is a well-known fact that Turing believes computers to be capable of performing many "intelligent" tasks. He also thinks that they will be able to do so in a "human" way.

> 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 [Tur50, p. 438].

As can be seen from the following quote, Turing believes that the difficulties in designing thinking machines are not unsurmountable.

> 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 [Tur50, p. 455].

While trying to convince the reader that the ideas he proposes are of the sort that can be realized in the foreseeable future, Turing mentions some concrete achievements he expects from computers. Those that are related to machine learning were outlined in Section 2.3. Here is another example, this time pertaining to automated software engineering:

> [The machine] may be used to help in making up its own programmes, or to predict the effect of alterations in its own structure.
> ...
> These are possibilities of the near future, rather than Utopian dreams [Tur50, p. 449].

The game of chess has been at the center of some of the most well-known achievements in AI. Today, computer programs play against world champions and sometimes even beat them. Spectacular advances have more recently been made in computer understanding and generation of speech. Although to what extent currently available speech processing systems are intelligent is a debatable issue, they (like chess-playing programs) have become part of modern life:

> 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 question. 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.
> ...
> Again, I do not know what the right answer is, but I think both approaches should be tried [Tur50, p. 460].

Take a look at computer technology at the turn of the century: What was unimaginable in 1950, in terms of memory and speed, is now reality. What Turing predicted about the IG, however, is still a challenge.

> I believe that in about fifty years' time, it will be possible to programme 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 percent chance of making the right identification after five minutes of questioning [Tur50, p. 442].

3 From the Imitation Game to the Turing Test: The 60's and the 70's

Earlier remarks on the TT, with the exception of [CHW71, CHWK72, Wei66] have mostly been of the philosophical sort. This is hardly surprising because 'Computing Machinery and Intelligence' was itself published in a philosophy journal, *Mind*[10]. Many discussions on the IG were published in the 60's and the 70's, most of the important contributions once again accommodated by *Mind*. In this section we will take a look at these philosophical papers, leaving the more practical work described in [CHW71, CHWK72, Wei66] to other, more appropriate sections. Readers interested in earlier comments on the TT and machine intelligence that are not discussed in this section can consult [May52, Rea69].

Keith Gunderson's comments on the IG are summarized in Section 3.1. Section 3.2 presents an approach stating that developing a TT-passing program is not going to be possible in the foreseeable future. The anthropomorphism in the TT is briefly discussed in Section 3.3, to be taken up later on. An inductive interpretation of the TT is described in Section 3.4.

[10] Although the cover of the 1950 issue reads "A Quarterly Review of Philosophy and Psychology", we find it not too inappropriate to call *Mind* a philosophy journal.

3.1 Rocks that Imitate and All-purpose Vacuum Cleaners

One of the earlier comments on Turing's IG came from Keith Gunderson in his 1964 *Mind* article [Gun64]. In this paper, appropriately titled 'The Imitation Game', Gunderson points out some important issues pertaining to Turing's replacement for the question "Can machines think?".

Gunderson develops certain objections to Turing's 'Computing Machinery and Intelligence' [Tur50] by focusing on the IG. In a nutshell, he emphasizes two points: First, he believes that playing the IG successfully is an *end* that can be achieved through different means, in particular, without possessing intelligence. Secondly, he holds that thinking is a general concept and playing the IG is but *one* example of the things that intelligent entities do. Evidently, both claims are critical of the validity of the IG as a measure of intelligence.

Gunderson makes his point by an entertaining analogy. He asks the question "Can rocks imitate?" and continues to describe the "toe-stepping game" [Gun64, p. 236] in a way that is identical to the way Turing described his IG in [Tur50]. Once again, the game is played between a man (A), a woman (B) and an interrogator (C). The interrogator's aim is to distinguish between the man and the woman by the way his/her toe is stepped on. C stays in a room apart from the other two and cannot see or hear the toe-stepping counterparts. There is a small opening in the wall through which C can place his/her foot. The interrogator has to determine which one of the other two is the woman by the way in which his/her toe is stepped on. Analogously, the new form of the question "Can rocks imitate?" becomes the following: 'What will happen when a rock box is constructed with an electric eye which operates across the opening in the wall so that it releases a rock which descends upon C's toe whenever C puts his foot through A's side of the opening, and thus comes to take the part of A in this game? ...Will the interrogator decide wrongly as often as when the game is played between a man and a woman?' [Gun64, pp. 236-237].

Gunderson believes that even if rock boxes play the toe-stepping game successfully, there would still be no reason to accept that they are imitating. The only conclusion that we can make from this would be that a rock box can be rigged in such a way that it can replace a human being in the toe-stepping game. According to Gunderson, this is because 'part of what things do is how they do it' [Gun64, p. 238]. As we will expand upon in Section 4.1, this is similar to Ned Block's argument for *psychologism* against behaviorism [Blo81].

Gunderson states that thinking is not something that can be decided upon by just one example. He demonstrates his belief that a computer's success in the IG is not sufficient reason to call it a thinking machine by another analogy: Imagine a vacuum cleaner salesman trying to sell a product. First, he advertizes the vacuum cleaner *Swish 600* as being "all-purpose". Then, he demonstrates how it can suck up bits of dust. The customer asks what else the machine can do. Astonished, the salesman says that vacuum cleaners are for sucking up dust and Swish 600 does precisely that. The customer answers, "I thought it was all-purpose. Doesn't it suck up bits of paper or straw or mud? I thought sucking up bits of dust was an *example* of what it does.". The salesman says "It is an example of what it does. What it does is suck up pieces of dust." [Gun64, p. 241].

The salesman is having trouble making his sale by calling Swish 600 all-purpose and being unable to show more than one example of what it does. According to Gunderson, Turing is also having the same problem because the term "thinking" is used to refer to more than one capability; just as the term "all-purpose" implies that the vacuum cleaner has functions other than just sucking up bits of dust. He concludes:

> In the end the steam drill outlasted John Henry as a digger of railway tunnels, but that didn't prove the machine had muscles; it proved that muscles were not needed for digging railway tunnels [Gun64, p. 254].

John G. Stevenson, in his 1976 paper 'On the Imitation Game' [Ste76] raises some arguments against Gunderson. One of these is the objection that Gunderson was expecting, namely the claim that being able to play the IG is not just *one* example; a machine that is good at the IG is capable of various things. Gunderson does not give a direct response to such objections. He mentions a reply can be formulated along the lines of showing that even combining all those things such a machine can do gives us a narrow range of abilities [Gun64, p. 243]. Stevenson doubts whether such a reply would be adequate [Ste76, p. 132]. Even if it does not exhaust everything that is related to human thinking, he believes the list of things that a computer that plays the IG can do would be quite impressive. Stevenson states that Gunderson is ignoring the specific character of the IG and that he proposes defective arguments.

3.2 The TT as Science Fiction

Richard Purtill, in his 1971 *Mind* paper also discusses some issues concerning the IG. Purtill criticizes some ideas in Turing's paper 'mainly as a philosopher, but also as a person who has done a certain amount of computer programming' [Pur71, p. 290]. He believes that the game is interesting, but as a piece of science fiction. He finds it unimaginable that a computer playing the IG will be built in the foreseeable future.

Overall, Purtill believes the IG to be a computer man's dream. He even promises to 'eat his computer library' if anyone has a notion on the principles on which a machine that can play the game is to be built [Pur71, p. 293][11]. He states that if computers, some day, behave like the computers in works of science fiction, he would grant them thought. But since all computer outputs can be explained as a result of a program written by humans[12], computers are not likely to play the IG successfully with the currently imaginable programming techniques. This, he believes, is because the behavior of thinking beings is not deterministic and cannot be explained in purely mechanistic terms.

Purtill believes that the game is 'just a battle of wits between the questioner and the programmer: the computer is non-essential' [Pur71, p. 291]. Although the former part of the claim may be reasonable to an extent, his latter argument about the computer being non-essential is not very sound. To eliminate the computer from the picture, Purtill proposes "purely mechanical" alternatives: machines made of levers and wheels that can do the same task. It is unclear why this should count as an argument against the IG because evidently, the material or structure on which the IG-playing "program" works is irrelevant. Purtill also states, anticipating the objection that the human mind might also be a highly complex collection of such mechanical processes, that if this were the case, it would mean 'human beings do not in fact think rather than that computers do think' [Pur71, p. 292], but does not attempt to justify this bold claim.

In his short paper 'In Defence of Turing' [Sam73], Geoffrey Sampson attacks Purtill's arguments briefly. First of all, he believes most of the limitations pertaining to the realization of IG-playing computers Purtill lists are practical difficulties that may be overcome in the (presumably not so

[11] Recall that the paper is written in 1971.

[12] Even if the program's outputs are guided by certain random elements.

distant) future. Secondly, he states that it is only natural that computer behavior is deterministic and that human behavior is not so easy to explain. The reasons for this are simple: computers are designed by humans; they have mechanisms that explicitly allow us to study their behavior; humans are much more complex in terms of both internal states and possible inputs than any contemporary computer [Sam73, p. 593]. Sampson also rejects Purtill's opinion that the consequence of the claim that human thinking is an extremely complex, yet computer-like, mechanical process is that men do not think. He holds that thinking, by definiton, is something human beings do.

3.3 Anthropomorphism and the TT

In a short paper that appeared in *Mind* in 1973 [Mil73], P. H. Millar raises some important issues which will show up in later works. He first discusses some vices and virtues of the IG and states that it is irrelevant whether or how the computers or the human beings involved in the game are "programmed". Then, he introduces the question of whether the IG is a right setting to measure the intelligence of machines. Millar notes that the game forces us to "anthropomorphize" machines by ascribing them human aims and cultural backgrounds. Millar asserts that the IG measures not whether machines have intelligence, but whether they have *human* intelligence. He believes that we should be open-minded enough to allow each being, be it a Martian or a machine, to exhibit intelligence 'by means of behavior which is well-adapted for achieving its own specific aims' [Mil73, p. 597]. We will return to this issue later on, especially in Section 4.5.

3.4 The TT Interpreted Inductively

In his important paper 'An Analysis of the Turing Test' [Moo76], James Moor attempts to emphasize the significance of the imitation game. As can be seen from the title, the term "Turing Test" was already being used to refer to the IG by 1976. Moor's main assertion is that 'the Turing Test is a significant test for computer thought if it is interpreted inductively.' [Moo76, p. 256].

Moor disagrees with the idea that the TT is an operational definition of intelligence[13]. Rather, he proposes, it should be regarded as a source of inductive evidence for the hypothesis that machines can think [Moo76]. Moreover, Moor does not agree with the claim that even if the TT is not an operational definition, it should at least be a necessary condition for granting computers intelligence. According to him, there could be other evidence based on the computer's behavior that leads to inferences about the computer's thinking abilities. However, he believes that the test provides a sufficient condition for intelligence-granting to computers. But his view is not "absolute"; he accepts that it is possible to revise a positive inference about a computer's possession of thought based on a TT, if other evidence is acquired afterwards.

Moor lists two arguments that support the TT as a good format for collecting inductive evidence. 'First, the Turing Test permits direct or indirect testing of virtually all of the activities one would count as evidence for thinking ... Secondly, the Turing Test encourages severe testing.' [Moo76, pp.251-252]. By the latter, Moor means the test's requirements are not too easy to meet. For instance competence in a single cognitive activity, no matter how complex, would not suffice.

[13] As opposed to Millar, who believes this to be true, and also that this is a virtue of the imitation game [Mil73].

Moor proceeds by considering some of the objections to the TT. He gives replies to these objections and shows that they are either irrelevant or can be refuted when the TT is considered to be a way of gathering data based on which we may inductively infer conclusions about machine thought. One objection that Moor, in our opinion successfully, replies is the objection concerning internal operation. The view that information about the internal information processing of a system is important in granting it intelligence is not uncommon [Blo81, Gun64, Sch98]. Moor warns against the possible confusion between the two variants of this conception. There is an important difference between the claim that evidence about the internal operation of a computer *might alter* a justified inductive inference that the computer can think, and the claim that such evidence *is necessary to make* such an inference. Moor believes the former and notes that this is certainly not a criticism that can be addressed to the TT. If certain kinds of information about the internal operation of a machine that was believed to possess intelligence after being Turing Tested is acquired afterwards, then we might just revise our decision. If the latter alternative were true, then the objection could be used against the test. But, according to Moor, critics fail to show that this is true and they are not likely to ever succeed.

As was discussed above within the context of Gunderson's paper [Gun64], the TT may be considered inadequate because it is only *one* evaluation of behavior. Moor answers this kind of objection also in a liberal light, in a manner similar to his discussion outlined above. Once again he makes a distinction between two claims: one positing that behavioral evidence which cannot be directly obtained in the TT *might alter* a justified inductive inference that a computer can think, and the other stating that such evidence *is necessary to make* this decision. Moor believes that the former is true. Further testing, he says, would be valuable and could even make us change our inference. The important point is, this does not incapacitate the TT in any way. The test could be attacked on these premises only if the latter claim were true. Moor believes the critics have not, and are not going to be able to prove this. This is because he believes that the format provided by the test enables examining a very large set of activities that would count as evidence of thinking. Thereby, he refutes the objections about the scope of the test.

Moor concludes by stating that although the TT has certain shortcomings (e.g., it being of little value in guiding research), it is an important measure for computer thought when it is inductively interpreted. Moreover, the standard criticisms of the TT fail to show that it is deficient if such an interpretation is made.

A reply to Moor comes from Douglas F. Stalker [Sta78]. He prefers to call Moor's interpretation an explanatory one rather than an inductive one. Stalker notes that Moor's beliefs about the mentality of other humans, as well as computers, are part of an explanatory theory. He emphasizes that Moor does not justify that his theory of explaining a computer's success at the TT by using the concept of thinking is the *best* theory that can be constructed about the same phenomenon.

As an alternative explanation for the computer's behavior, Stalker proposes a purely mechanistic theory that does not appeal to any mental concepts. His theory takes into consideration such factors as the computer's physical structure, its program and its physical environment. Moreover, he believes this theory to be preferable compared to Moor's. Stalker believes explanatory theories that involve concepts of thinking can apply to people, but because of some fundamental differences between computers and humans, they may not be the best theories for explaining computer behavior.

In his answer to Stalker [Moo78], Moor basically says that the existence of alternative explanations does not mean that they would necessarily be competitors. It is true that an explanation for a computer's activities can be given at different levels: physics, electronic circuitry, programs, abstract automata, etc. Moor notes that these explanations would be different, but not necessarily rivals. In the case of a computer displaying intelligent behavior by being successful in the IG, an explanatory theory involving thinking could even be preferred because it is simpler and easier to understand. Moor's conclusion is:

> It seems natural and probably most understandable to couch the explanation in terms of a theory of mind. If one has the patience, the explanation could also be given at lower levels of description, e.g., involving perhaps thousands of computer instructions or millions of changes in circuit states [Moo78, p. 327].

4 In and Out of the Armchair: The 80's and the 90's

While thought experiments are still around, it can be seen that late 80's and the 90's also feature a tendency to leave the comfortable armchair of philosophy. In this section we will cover only some of the works that have addressed the TT. This is mainly because of the sheer abundance of comments on the issue. The subset of the work done during the 80's and the 90's that we chose to present in this section will provide a general overview of the main arguments and the reader is directed to references for further explication. A must-read is Douglas Hofstadter's 'Turing Test: A Coffeehouse Conversation' [Hof82] which is full of valuable and entertaining insights. Ajit Narayanan studies the intentional stance and the IG [Nar96]. For a discussion of the frame problem in relation to the TT, the reader is referred to [Cro94]. Other references that can be explored are [Hal87, Ran87, For88, GT88, Bie88, Alp90, Dav90, Par90, Cla92, SC93, Jac93a, Mar95, CM95, Fei96, Hay98]. A number of articles on the TT have appeared in popular science magazines, too. Some of these are [Gui83, Dew92, Pla95, Flo96, Wal97].

The TT scene began heating up at the beginning of the 80's. Although the "consciousness argument" and the "anti-behaviorist argument" were voiced before, they had not been really unsettling. Then, in the early 80's, two strong counter-arguments against the TT were formulated by John Searle and Ned Block, respectively. The debate on Searle's "Chinese Room" is in itself expansive enough to be the subject of a whole review paper of at least this size. We will consider it briefly in Section 4.2 and the interested reader should have no difficulty finding more information about the issue. Block's anti-behaviorist attack of the TT, on the other hand, was not disputed much and it is the aim of Section 4.1 to elaborate on his ideas.

Various attempts have been made to modify the TT to get better "tests" for machine thought, and these are discussed in Section 4.4. Robert French's 'Subcognition and the Limits of the Turing Test' [Fre90] is analyzed in Section 4.5. Finally, the "less philosophical" stance towards the TT is discussed in Section 4.6.

4.1 Behaviorism and Ned Block

In 'Psychologism and Behaviorism' [Blo81], Ned Block attacks the TT as a behaviorist approach to intelligence. Although this paper was written in 1981, Block still seems to hold the same opinions, see [Blo95].

Block believes that the judges in the TT can be fooled by *mindless* machines that rely on some simple tricks to operate. He proposes a hypothetical machine that will pass the TT, but has a very simple information processing component. Block's machine has all possible conversations of some given length recorded in its memory. Of course, we want these conversations to be such that at least one party is 'making sense' [Blo95]. The set of strings constituting such conversations that can be carried out in a fixed amount of time are finite and thus can be enumerated and stored in our hypothetical computer. The judge types in a string, say A. The machine finds a conversation beginning with A and types out the second sentence of this string, say B. If, next, the judge types in C, the process is repeated with A replaced by ABC. All the machine does is simple "lookup and write-out", certainly nothing that anyone would call sophisticated information processing.

Since this machine has the intelligence of a jukebox [Blo95] or a toaster [Blo81], and since it will pass the TT, the test must be an inadequate measure of intelligence. Block ties this conclusion to the more general one that this is because of the behaviorist approach taken in the design of the TT.

Ned Block defines psychologism as 'the doctrine that whether behavior is intelligent behavior depends on the character of the internal information processing that produces it' [Blo81, p. 5]. According to this definition, two systems can display the same actual and potential behavior, have the same behavioral properties, capacities and dispositions, and yet, there could be a difference in their information processing prompting us to grant one full intelligence while holding that the other is devoid of any.

A classical argument against psychologism is the Martian argument: Suppose that there is life on Mars. Humans and Martians meet, develop an understanding of each other, engage in mental and creative activities together and so on. And then, it is discovered that Martians have significantly different information processing mechanisms than those of humans. Would we, then, deny that these creatures have intelligence just because they are very different from us? This would be, as Block likes to call it, pure "chauvinism". He holds that psychologism does not involve this kind of chauvinism. After all, psychologism does not state that the fact that a system has a completely different information processing mechanism compared to human beings *necessarily* means that it lacks intelligence.

Attacking the validity of the TT using psychologism does not seem to be Block's main interest. He is more concerned with arguing against behaviorism using the TT as a focal point.

As was mentioned above, Block believes, because of the characteristics peculiar to the design of the TT, some genuinely intelligent machines can be classified as lacking intelligence and vice versa. Here is what Block suggests in order to eliminate dependence on human discriminatory powers: 'We should specify, in a *non-question-begging* way what it is for a sequence of responses to verbal stimuli to be a typical product of one or another style of inteligence' [Blo81, p. 10, emphasis added]. Then, Block suggests we revise our intelligence-granting mechanism as follows:

> Intelligence (or more accurately, conversational intelligence) is the disposition to produce a sensible sequence of verbal responses to a sequence of verbal stimuli, whatever they may be [Blo81, p. 11].

Now, the modified TT does not depend on anyone's coming up with good questions, since the system must have a *disposition* to emit sensible responses to anything that the interrogator *might* say, not just to the things that he/she *does* say. At this point, Block demonstrates that the modified TT is not greatly affected by the standard arguments against behaviorism[14]. The little defects of the modified TT as a behavioral conception of intelligence can be protected against these arguments with another modification. The reformulation involves the replacement of the term "disposition" by "capacity". The difference is that a capacity to $\o$ need not result in a disposition to $\o$, unless certain internal conditions are met. Now, all arguments against behaviorism are avoided[15] with *the neo-TT conception of intelligence*:

> Intelligence (or more accurately, conversational intelligence) is the capacity to produce a sensible sequence of verbal responses to a sequence of verbal stimuli, whatever they may be [Blo81, p. 18].

Although Block seems to be 'helping out' the TT by making it less prone to anti-behaviorist objections, this is hardly a surprising consequence when the definition of intelligence is modified into something that is not really behaviorist any more. Block seems to be aware of this for he says the concession made to psychologism by moving from behavioral dispositions to behavioral capacities will not be enough to save behaviorism [Blo81, p. 18]. His strategy is stretching behaviorism to its limits and showing that, even if we have the most general form of it, the behavioristic conception of intelligence is false.

How, one may wonder, will he do that? Block describes a machine that can produce a sensible sequence of verbal responses to verbal stimuli and is intelligent according to the neo-TT conception of intelligence. However, according to him, the information processing of the machine clearly demonstrates that it is devoid of intelligence. We have basically explained how this machine works in the introductory paragraphs of this section.

Block calls a string of sentences whose members can be typed by a human typist one after another in one hour or less, a *typable* string of sentences. The set of typable strings of sentences is clearly finite. The subset of this set consisting of those strings which can naturally be interpreted as conversations in which at least one party's contribution is sensible is also a finite set. Recall that we assume we have a non-question-begging definiton of "sensible". The machine has the set of sensible strings recorded on tape. Upon encountering the string A, all it does is pick one of the strings that begin with A and type out the second sentence, say B, of A. If the interrogator types in the string C next, the machine will repeat the same process described above with A replaced with ABC and so on [Blo81, p. 20]. This machine will have the capacity to emit sensible verbal output to any verbal input, and therefore would qualify as intelligent according to the neo-TT conception of intelligence. But the machine, in fact 'has the intelligence of a toaster' [Blo81, p. 21]. This is primarily due to the fact that all the intelligence it exhibits belongs to the programmers, not to the machine itself. Block concludes that this capacity is not enough for intelligence and so the TT intelligence conceptions are refuted.

It can be argued that, by Block's reasoning, *any* intelligent machine exhibits the intelligence of its programmers. Block says he is making no such claim. A machine that has more sophisticated mechanisms like learning, and problem solving would, to Block, be intelligent. In the latter case, the exhibited intelligence belongs to the machine itself [Blo81, p. 25]. The search machine of Block can only respond with

[14] The three mentioned by Block are the Chisholm-Geach, perfect actor, and paralytic and brain-in-vat arguments. Detailed discussion of these is beyond the scope of this work and is not crucial to the understanding of what follows. The interested reader is referred to [Blo81, pp. 11-12] and the references provided there.

[15] See [Blo81, p. 18].

what is already put in its memory by the programmers[16]. Block argues that 'the neo-Turing Test conception of intelligence does not allow us to distinguish between behavior that reflects a machine's *own* intelligence and behavior that reflects *only the intelligence of the machine's programmers.*' [Blo81, p. 25, emphasis original]. This kind of argument has been considered by Turing in [Tur50, p. 450], as described briefly in Section 2.2.

Another objection is as follows: Block is merely suggesting a new definiton of intelligence by stipulating certain internal conditions. Block defends the new definition here, which is presuppository of its existence! Therefore, Block is indirectly admitting that all he is doing is suggesting that we adopt new criteria for intelligence and dispose of the behavioristic ones [Blo81, p. 27].

Block also considers the "chauvinism" argument. A system with information processing unlike ours may not be "intelligent" in our criteria; but then, *we* might not count as "shmintelligent" in their criteria. 'And who is to say that intelligence is any better than shmintelligence?' [Blo81, p. 27]. Block denies the chauvinism attributed to him. He believes '[his] machine lacks the kind of "richness" of information processing requisite for intelligence' [Blo81, p. 28]. He doesn't feel the need to elaborate on what sort of systems have the abovementioned richness believing that 'one can refute the Turing Test conception by counterexample without having to say very much about what intelligence really is' [Blo81, p. 28].

To those who ask what Block would think if it turned out that humans process information in the way that Block's machine does, Block responds as follows:

[16] Consider, however the following situation: If every once in a while, upon verbal input A the machine transformed a sentence B in A into B' and proceeded accordingly (this can be likened to a *mutation*), would it count as intelligent because of this little trick of non-determinism?

> If the word "intelligence" is firmly anchored to human information processing, as suggested above, then my position is committed to the *empirical claim* that human information processing is not like that of my machine. But it is a perfectly congenial claim, one that is supported by both common sense and by empirical research in cognitive psychology [Blo81, p. 29, emphasis original].

It can be argued that Block's machine is unrealizable because of combinatorial explosion. We will not go into the details of this; Block's response to this objection can be found in [Blo81, pp. 30-34].

Richardson, in his reply to Block [Ric82], states that he is doubtful whether Block's machine can really imitate human conversational abilities. Humans can (and do) understand sentences that they never heard/uttered before and produce sentences that they never heard/uttered before. They can do this in such a way that they can adapt to novel situations and maintain the coherence of discourse. This view is held by Chomsky who believes that the brain *cannot be a repertoire of responses* and must contain a program that can build an *unlimited* set of sentences out of a finite list of words[17]. If the *potentially* utterable/understandable and sensible sentences that a human mind can produce in a lifetime is unlimited, then how can a team of humans gather this information and enter it in the memory of the machine in finite amount of time? It is difficult to imagine Block's machine managing the many intricacies of human conversation such as adapting to topic shifts and contextual changes. Richardson believes 'if the list-searcher satisfies the neo-Turing Test, the test is too weak' [Ric82, p. 423]. For Block's response to such arguments see [Blo81, pp. 35-36].

[17] See, for instance [Cho75].

Block must have realized some difficulties in enumerating the strings as well. He later introduces, in [Blo95], the Aunt Bubbles machine[18]. In this version, the programmers think of *just one* response to the strings at each step. To maintain coherence and make the task easier to follow, they may choose to simulate a definite person, for instance Block's own (most probably hypothetical) Aunt Bubbles. They may even restrict the situation by modeling Bubbles' responses in the case that she is brought into the teletype room by her strange nephew [Blo95]. So each response is the kind of response that Aunt Bubbles would give to the verbal inputs. Block says that the machine will do as well as Aunt Bubbles herself in a TT, but it is obviously not intelligent because of the reasons described above.

Let us briefly go over some of Block's arguments and the behaviorism in the TT before we proceed. For one thing, as Block also mentions, the intelligence concept (because of some inherent properties it has) does not fully conform to the generalizations of behaviorist or anti-behaviorist arguments based on other mental states such as pain [Blo81, pp. 13-16]. There is another aspect of intelligence that can justify for the behaviorist approach of the TT. Behaviorism may be considered an antiquated or primitive approach in general, but it does not seem that awkward to use it in intelligence-granting. This is primarily because *we* grant intelligence that way: Upon seeing a human being we automatically assume that he/she is intelligent. We feel free to approach a person (rather than, say, a dog or a lamp post) to ask the whereabouts of the post office without having many doubts about him/her understanding us. If the TT is *that* crude and unsophisticated, then we, as humans might consider revising our intelligence-granting mechanisms as well. This constitutes a line of defence for the TT: If behavioral evidence is acceptable for granting intelligence to humans, this should be the case for machines as well. We have discussed this already in Section 2.2.

Recall that Block believes humans can be overly chauvinistic or liberal in granting intelligence to machines. However, it is unclear how he classifies genuinely intelligent machines and mindless machines. If there is a way of deciding on that issue, an X-Test to determine whether a machine is really intelligent, then why would we be discussing the TT with all its quirky and imperfect ways? In addition, although he does not trust the human judges in the beginning, later on Block seems to have complete faith in the '*imagination and judgement* of a very large and clever team working for a long time with a very large grant and a lot of mechanical help' [Blo81, p. 20, emphasis original].

With the current research on cognition and linguistics at hand, it seems unlikely that a simple approach like Block's can succeed in modeling the human mind. If one day, enough on language and cognition is discovered so that Block's "sensible" strings of sentences are enumerated then we may decide that the neo-TT conception of intelligence is false. But then again, when that day comes, having all the *psychologistic* information we need, we probably would not be interested in the TT any more.

The reader is encouraged to contemplate on how Robert French's arguments, which will be presented in Section 4.5, affect Block's claims. We believe that French's demonstration of the inseparability of the cognitive and subcognitive levels in the human mind has direct consequences on the realizability of Block's machine.

In any case, Block's paper is significant because it demonstrates the weakness of the behavioral approach in the TT. The TT may be abandoned one day, because more information on how the mind works may be ob-

[18] Aunt Bubbles appears momentarily, as Aunt Bertha, in [Blo81] too.

tained and we may have on our hands better means to detect another entity's cognitive capacities. But today, we do not have much to look at that is more informative than behavior.

4.2 The Chinese Room

In the beginning of the 80's, with John Searle's Chinese room argument [Sea80] the TT was confronted with yet another objection. The analysis of the Chinese room can easily get out of hand since a great number of comments have been made on the issue and the debate still rages on.

In a nutshell, here is what the Chinese room looks like: Suppose that Searle, a native speaker of English who does not know a word of Chinese, is locked in a room. There is an opening in the room through which we may send in Chinese sentences on pieces of paper. Of course to Searle, these look like meaningless squiggles [Sea80]. In the room, Searle has a "Chinese Turing Test Crib Book" [Lei92] he can consult to find an output that corresponds to each Chinese symbol he receives. What he does is simply match the input with those in the book, follow some rules written in English and find some Chinese symbol sequence to output. We correspond with Searle in this manner and due to the flawless look-up table he has, Searle-in-the-room seems to understand Chinese perfectly. But he does not. Searle still has no idea about what the Chinese symbols we send in and those that he sends out mean. To him, "Squiggle-Squiggle" is coming in and "Squoggle-Squoggle" is going out [Har91].

Now consider a computer program that passes the TT in Chinese. Proponents of the TT will grant it that this computer thinks and, in some sense, understands Chinese symbols. Searle challenges this by being the computer and yelling at the world that he does not understand a word of Chinese. Judging by the inputs and outputs of the system, Searle-in-the-room is indistinguishable from a native speaker of Chinese. In a sense, he is passing the TT in Chinese, without understanding a word of Chinese. It should be clear how that consitutes a criticism of the TT, and the computational view of mind.

As was mentioned before, various aspects of the Chinese Room argument have been analyzed including syntax/semantics, consciousness, boundaries of systems, etc. The interested reader is referred to [Sea80, Sea90, Har89, And87, Dye90, Col91, Cop93, Rey86, Fod91, Hau97, Bod88, Mal87, Rob90, HHPB92] and the references provided in those.

4.3 Consciousness and the TT

Another difficult and widely discussed problem in philosophy of mind is consciousness. While we do not want to delve too far into this, we wish to take a brief look at the relationship between consciousness and the TT.

Donald Michie's 'Turing's Test and Conscious Thought' [Mic96] is one of the important comments made on the TT. Michie comments on a variety of issues surrounding the TT, but in this section we will mainly concentrate on the conclusions he draws about consciousness.

First of all, Michie notes that Turing did not specify whether consciousness is to be assumed if a machine passes the TT. Of course, Turing probably did not believe that consciousness and thought are unrelated. Rather, Michie thinks he means 'these mysteries and confusions do not have to be resolved before we can address questions of intelligence' [Mic96, p. 31] (see also [Tur50, p. 447] and Section 2.2). There seems to be a relation between consciousness and thinking. Some critics believe that intelligence cannot be granted to entities that are not conscious (see, for instance, [Sea90]) while others have questioned the interdependence of conscious and subconscious processes (see, for instance, [Fre90] and Section 4.5).

According to Michie, that the TT provides access to cognitive processes via verbal communication incapacitates it as a test of intelligence. He observes two dimensions in which this inadequacy manifests itself.

The first is 'the inability of the test to bring into the game thought processes of kinds which humans perform but cannot articulate' [Mic96, p. 36]. Michie gives examples of some operations humans can perform almost unconsciously. For instance, any English speaker would be able to answer the question "How do you pronounce the plurals of the imaginary English words 'platch', 'snorp' and 'brell'?" with "I would pronounce them as 'platchez', 'snorpss' and 'brellz'." [Mic96, p. 38] [19]. It is conceivable that the programmers of TT-passing programs will be forearmed against this particular question, but it is unlikely that they can encode all we know about pronounciation (or phenomena from non-linguistic domains, for that matter) simply because some related processes operate at the subconscious level. (For a similar argument, the reader is referred to [Fre90] and Section 4.5.)

The second dimension in which Michie believes the TT to be mismatched against its task is the phenomenon of machine 'superarticulacy'. Namely, 'the test can catch in its net thought processes which the machine agent *can* articulate, but should not if it is to simulate a human' [Mic96, p. 42]. As was mentioned above, humans perform many activities without being fully aware of how they do them. In fact, it has been shown that the better you get at something the less aware of the underlying processes you become. Thus during a TT, 'the interrogator need only stray into some specialism in which both human and machine candidates possess a given expertise' [Mic94, p. 192]. The machine will give itself away because of its superarticulacy. For more about superarticulacy, the reader is referred to [Mic96, p. 41-43] and [Mic90].

Finally, Michie notes the importance of social intelligence. AI should, he says, try to incorporate emotional (also called affective) aspects of communication and thought in the developed models. Michie also proposes, like some of those we will see in the next section, that extensions to the TT can be made in order to 'address yet more subtle forms of intelligence, such as those involved in collective problem solving by co-operating agents, and in teacher-pupil relations' [Mic96, p. 51].

We will cut the discussion of consciousness short both because it is a rather broad topic, but also because most commentors on the TT (consciously or subconsciously) propose arguments that can be interpreted from that angle. Can we not reformulate the other minds problem ("How do I know that any entity other than me has a mind?") in terms of consciousness ("How do I know that any entity other than me is conscious?")? The reader can refer to Section 2.2 and [Tur50, pp. 445-447] for Turing's answer to the argument from consciousness and how he makes use of the other minds problem. Similarly, most questions about machine thought can be re-evaluated within the context of machine consciousness. We included the analysis of Michie's paper here because it proposes new ideas from the viewpoint of consciousness and relates them explicitly to the TT. Interested readers can consult [Den92, Gun67, Mic94, Mic95] for more on consciousness.

4.4 Alternative Versions of the TT and Their Repercussions

In this section, we will summarize some alternatives to the TT that were proposed in order to assess machine intelligence.

[19] This question was adapted from [Lai88].

4.4.1 Harnad and the TTT

Stevan Harnad's main contribution to the TT debate has been the proposal of the Total Turing Test (TTT), which is, like the TT, an indistinguishability test but one that requires the machines to respond to all of our inputs rather than just verbal ones. Evidently the candidate machine for the TTT is a robot with sensorimotor capabilities [Har89, Har91].

Harnad's motivation for the 'robotic upgrade of the TT to the TTT' [Har91] has its roots in what he calls 'the symbol grounding problem'. He likens the situation of symbols being defined in terms of other symbols to a merry-go-round in a Chinese to Chinese dictionary [Har90]. He claims that for there to be any semantics in the mind (and there surely is) symbols must be *grounded*. Harnad deduces that meanings of symbols are, at least in part, derived from interactions with the outside world.

Harnad does not explicitly argue that the TT is too specific (unlike Gunderson for instance, see Section 3.1). He concedes that language might capture the full expressive power of our behavior, at least when the concern is assigning minds. What he doubts is whether language is an 'independent module' [Har89]. His position is summed up in the following:

> Successfully passing the teletype version of the Turing Test alone may be enough to convince us that the candidate has a mind just as written correspondence with a never-seen penpal would, but full robotic capacities even if only latent ones not directly exhibited or tested in the TT may still be necessary to generate that successful linguistic performance in the first place. [Har91, p. 46].

Harnad also defends his TTT against the Chinese room argument which, in our opinion, is uncalled for. The motivation of the TTT is quite clear; Harnad's assertions, although not undebatable, are understandable. An approval from Searle would not make that much of a difference, but Harnad seems to think it is important. In any case, by doing so, he enables others to criticize his work on Searlean accounts [Hau93, Bri94].

Larry Hauser, in his reply to Harnad's 'Other Bodies, Other Minds' [Har91], criticizes Harnad and Searle and aims to show that 'Harnad's proposed robotic upgrade of the TT to the TTT is unwarranted' [Hau93, p. 234]. To that end, he analyzes Harnad's intuitive, scientific and philosophical reasons for proposing the upgrade and argues against them. Hauser finds the TTT to be unnecessary because, he notes, if the sensorimotor capacities the TTT tests for are *necessary* for the linguistic capacities that the TT tests for, having the latter should be *sufficient* for the former anyway [Hau93, p. 227].

For more on symbol grounding and the TTT, the reader is referred to Harnad's other papers [Har92, Har94, Har98]. Also interesting is James H. Fetzer's 'The TTT is not the Final Word' [Fet93], in which he aims to show that the TTT cannot provide a proof for machine thought since more than symbol manipulation *and* robotic capacity should be involved in intelligence-granting.

In addition to the TTT, Harnad also mentions a TTTT (Total Total Turing Test) which requires neuromolecular indistinguishability. However, this more stringent version of the TT, according to Harnad, will be unnecessary. Once we know about how to make a robot that can pass the TTT, he says, we will have solved all the problems pertaining to mind-modeling. However, neural data might be used as clues about how to pass the TTT [Har91]. Harnad believes '[TTTT] is as much as a scientist can ask, for the empirical story ends there' [Har98], but he does not think that we have to "go that far". The reader is referred to [Har98] for a detailed explanation of why Harnad believes the TTT is enough. For an excellent third person account of the TT/TTT/TTTT story, among other issues, the reader is referred to [Fet95].

4.4.2 The Argument from Serendipity and the Kugel Test

Stringent versions of the TT are also mentioned by Selmer Bringsjord, occasionally within the context of Harnad. Bringsjord supposes that there is a sequence of TT variants in increasing order of stringency. In his "What Robots Can and Can't Be" [Bri92] he aims to show that AI will produce machines that will pass these stronger versions, but the attempt to build an artifical person will still fail.

Bringsjord is one of those who wants to remain within "the philosophical terrain". In [Bri94] he develops against the TT, *the argument from serendipity* and defends this against some criticisms.

The argument from serendipity, as the name suggests, "refutes" the TT by a finite state automaton (FSA) that generates random English sentences. Call this automaton P. During a TT, P may just get lucky and fool the interrogator. So much for the TT! Even the TTT can be refuted similarly. A robot may behave randomly and by chance, its linguistic behavior may coalesce with the sensorimotor behavior perfectly during a TTT.

Bringsjord finds the TTTT very chauvinistic and considers an alternative version of it he calls TTTT*. This latter test requires a flowchart match between the brains of players A and B rather than a neuromolecular match [Bri94, p. 104]. But Bringsjord believes that the TTTT* is an 'impracticable nightmare' since we would not know how to conduct this test. The interested reader should consult [Bri94] to see Bringsjord explain his reasoning where he appeals both to intuition and computability theory.

Bringsjord, determined to attack every version of the TT, also "refutes" the Kugel Test (KT). The KT is not as well known as the TTT or the other versions of the TT that we investigated in this section. Once again, there are three players involved. A judge, who sits behind two bins marked *YES* and *NO* runs the game. The aim of the participants is to guess the concept that the judge thinks up by looking at the cards (with pictures on them) that the judge drops in the two bins. A card goes to the *YES* bin if it falls under the concept, and to the *NO* bin otherwise. To give an example, if the concept that the judge is thinking of is "woman", cards with pictures of women (or conceivably, items typically identified with women) go to the *YES* bin. A player need not announce the concept when he/she finds it. He/she *wins* the round if there comes a time at which all future guesses about which bin a card will be placed in are correct [Kug90, p. 4]. Thus the player must not only identify the concept (e.g., just say "Aha! The concept is *woman*.") but should also be able to apply it. Now, just as in the TT, to pass the KT, a machine has to perform as good as a human. An interesting twist here is that the machine must be able to *win the game*, which is not the same as winning a round. A game consists of infinitely many rounds.

Why, you may ask, would anyone design such an obscure test? Kugel, by requiring the machine to win infinitely many rounds, wants to rule out the possibility of an FSA passing the KT [Kug86, Kug90]. Although the test is practically useless (because it requires infinite amount of time), is it of any theoretical significance? Kugel believes that humans are neither *pigheaded* (i.e., once they think of an answer to the "sequence game" they do not have to stick with it) nor *narrow-minded* (i.e., once they find the nth member of a sequence, they are still able to learn a different sequence with the same initial elements). If humans were Turing machines (or FSA's with lesser powers) they would be pigheaded and narrow-minded. Kugel holds that humans are automata of some sort, and in the light of the above concludes that they must be trial-and-error machines. For more on KT, the reader is referred to [Kug86, Kug90, Bri94].

Bringsjord is interested in the KT primarily because it rules out FSA's from passing it[20]. He notes that Kugel's arguments may be unsound, but assuming they are not, asks the question "Do we have in KT an acceptable variant of the original TT?" [Bri94, p. 115]. Bringsjord's answer is negative. KT is rigid and does not allow access to all cognitive capacities that the TT can. We agree with this criticism of Bringsjord; participants in the KT are rather passive and their innovative (or rather, generative) capabilities cannot be tested. Bringsjord's second argument against the KT is again from serendipity. A trial-and-error machine can call the random string generating FSA P mentioned above for the declarations about what the concept in question is and so much for the KT... Once again, the reader is to consult [Bri94] to see how the argument from serendipity is "guaranteed to work" against the TT and its variants.

4.4.3 The Inverted Turing Test

Recently, Stuart Watt has proposed the Inverted Turing Test (ITT) [Wat96]. Some of his arguments are also relevant to Section 5 but we chose to discuss his paper here.

Watt's point is that the TT is inseparable from "naive psychology"[21] since to pass the TT, a machine has to convince the interrogator that it has a mind. He calls naive psychology 'the psychological solution to the philosophical problem' [Wat96][22].

Watt's ITT requires the machine to be able to prove its human-ness by exercising naive psychology. In particular, it has to show that its power of discrimination is indistinguishable from that of a human judge in the TT. The TT is literally inverted and 'a system passes [the ITT] if it is itself unable to distinguish between two humans, or between a human and a machine that can pass the normal TT, but which can discriminate between a human and a machine that can be told apart by a normal TT with a human observer' [Wat96].

Watt states that he proposes the ITT as a thought experiment rather than as a goal for AI. Incidentally, he believes that the same applies to the TT and both tests should be regarded as means to gather inductive evidence on which inferences about machine mentality can be made [Moo76]. We had discussed this earlier in Section 3.4.

Watt may be right about intelligence being in the eye (or the mind) of the beholder; many people have noted the human disposition to ascribe intelligence to systems that aren't and vice-versa. But the new test he proposes, the so-called ITT, has been subject to some strong counter-arguments as we shall shortly see. It can be said that Watt's motivation for introducing the ITT seems acceptable, but the proposal itself is problematic [23].

Selmer Bringsjord and Robert French, in their replies to Watt [Bri96, Fre96], propose simple methods that reveal some weaknesses of the ITT. The titles of the papers are illustrative of their content. Bringsjord's

[20] Recall that his argument from serendipity features an FSA.

[21] Basically the term given to the natural human tendency and ability to ascribe mental states to others and to themselves [Wat96].

[22] The latter being the other minds problem.

[23] During the discussions held after a talk on the Turing Test, given by the first author (at the Cognitive Science Colloquium held at the Middle East Technical University, Ankara, in November, 1998) many participants, who did not previously know a lot about the topic *proposed* the ITT as a better means to detect human-ness of machines. These people had not read or heard of Watt's paper and neither the ITT nor naive psychology was discussed during the presentation. Maybe this can be explained as "naive psychology at work".

'The Inverted Turing Test is Provably Redundant' [Bri96] shows that the ITT is entailed by the original TT. Bringsjord also opposes Watt's motivation and believes that naive psychology is withering in many humans (including himself) and, with the advent of computer programs that are very difficult to distinguish from humans in written communication, will soon be no more.

In 'The Inverted Turing Test: A Simple (Mindless) Program that Could Pass It' [Fre96], Robert French shows both that the ITT can be simulated by the TT (in a way that is very similar to Bringsjord's) and that a very simple program can readily be designed to pass the ITT. The mindless machine that will pass the ITT is designed using 'subcognitive questions' that are described in [Fre90, Fre95]. It is assumed that the conclusions explained by French in these works are accepted. These are analyzed in substantial detail in Section 4.5.

First a large set of subcognitive questions are selected, humans are surveyed and a 'Human Subcognitive Profile' for this 'Subcognitive Question List' is obtained. Now, if we give these and a statistical analyzer to an interrogator (man or machine), he/she/it should have no difficulty discriminating machines from humans. It is not difficult to store the list and the profile in the memory and provide a small statistics routine to the computer, and so much for the ITT. While the TT stumbles in front of subcognitive questions (see Section 4.5), they can be used to construct a mindless machine that can pass the ITT.

Others have used their replies to Watt as opportunities to voice their opinions about AI and the Turing Test in general. As we shall see in Section 4.6 Patrick Hayes and Kenneth Ford view the TT as a harmful burden on AI. In their 'The Turing Test is Just as Bad When Inverted' [FH96], they state that the ITT suffers from the same problems as the TT that they explicate in [HF95]. They grant it that Watt has a point in his arguments on naive psychology but note that Turing's original IG (the gender-based TT) is immune to most of those since in this scenario, the interrogator will not be thinking about differences between humans and machines. In any case, they believe that 'it is time for AI to consciously reject the naive anthropomorphism implicit in all such "imitation games" and adopt a more mature description of its aims' [FH96].

Similarly, Collins, in his reply to Watt [Col97], does not really focus on the ITT, but proposes a new variant of the TT. He believes that 'the deep problem of AI' is that of trying to develop machines that can learn from their surroundings the way humans do. There is currently an 'interpretive asymmetry' between the way humans and computers do things. Machines are not as adaptive as humans in human-computer interactions. According to Collins, this asymmetry will disappear when computers reach a level of sophistication in resolving mistakes and learning from their surroundings that is comparable to those of humans and all problems of AI will be solved. Learning languages would then be one of the surface transformations of this deep problem [Col90] and when this is solved 'the rest will be research and development' [Col97].

To determine whether the interpretive asymmetry has disappeared, Collins believes we can use Turing-like tests. In fact he states that a sub-TT is enough to assess whether this goal has been reached or not and complicating the matter by proposing the ITT or the TTT is uncalled for. In the Editing Test (ET) that Collins proposes, the task is no longer as comprehensive as holding a conversation, but that of sub-editing previously-unseen passages of incorrect English. The interrogator will try to come up with pieces of text that a linguistically competent human can easily sub-edit and if a computer is indistinguishable from humans in this task, then the ET is passed and the deep problem of AI is solved. Collins finishes by briefly demonstrating that even the ET is very difficult to pass, at least with the currently imaginable techniques (such as a look-up table) [Col97].

4.4.4 The Truly Total Turing Test

Very recently, in his interesting *Minds and Machines* paper [Sch98], Paul Schweizer proposed the 'Truly Total Turing Test' (TRTTT) [24]. He believes even Harnad's TTT to be an insufficient test for intelligence. Before he proposes the TRTTT, Schweizer states his own opinions about the adequacy of behavioral criteria. He views such tests as 'dealing with evidence for intelligence but not as constitutive or definitional' [Sch98, p. 264].

Schweizer, while talking about the other minds problem, notes that we usually grant intelligence to other humans on behavioral bases because we have general knowledge about the *type* of creature under consideration. However, in the TT, we encounter a type about which we do not know anything. In the case of machines we lack a "history" to base our decisions upon.

Schweizer believes that the TT, and even Harnad's TTT, is subject to the "toy-world" criticism. The systems that succeed in these tests would, according to him, not be displaying an intelligence comparable to the natural intelligence of living things that function in the real world. They can function only in constrained, artificial worlds.

The TRTTT posits a long-term, evolutionary criterion: Consider cognitive *types* and *tokens* of those types. Although we do not have a theory of the intelligence of the human cognitive type, we have an extensive *historical record* of it [Sch98, p. 267]. This is precisely why behavioral intelligence-granting is acceptable for individual humans (tokens of the type human). Thus robots, as a cognitive type, should accomplish achievements that are comparable to those of humans. It is no more enough to converse in natural language or to play chess; robots as a 'race' must be able to *develop* languages and *invent* the game of chess. Similar (evolutionary) tests have been proposed by others before but never so convincingly[25]. Schweizer makes very good use of the other minds problem to support the cultural and cognitive evolution criteria that the TRTTT stipulates.

Now, after the *type* passes the TRTTT, we can evaluate *tokens* of the type by less stringent behavioral tests, like the TTT and the TT. According to Schweizer, 'imitative tests like the TTT and the TT apply to individuals *only* under the assumption that the general type is capable of passing the [TRTTT]' [Sch98, p. 268, emphasis original].

4.5 Subcognition and Robert French

One of the more recent discussions about the TT can be found in Robert French's 1990 article 'Subcognition and the Limits of the Turing Test' [Fre90]. In this work, French aims to show that 'the Turing Test provides a guarantee not of intelligence, but of culturally-oriented *human* intelligence' [Fre90, p. 54].

French considers two claims of Turing. The first is the claim that if a computer passes the TT, it will necessarily be intelligent. The second one is the claim that it will be possible to build such a machine in the near future. These, he calls the philosophical claim and the pragmatic claim, respectively. French agrees with the former claim. However, he believes that the pragmatic claim has been largely overlooked in discussions of the TT. In 'Subcognition and the Limits of the Turing Test', he is primarily concerned with this latter claim and believes

[24] In Schweizer's paper, the abbreviation TTTT is used. We prefer to use TRTTT so as to avoid confusion with Harnad's Total Total Turing Test, previously referred to as TTTT

[25] See [Bar87] and Section 5.

that the TT is 'virtually useless' [Fre90, p. 53] as a real test of intelligence because he believes it will never be passed.

To establish this result French considers "subcognitive" questions, i.e., questions that reveal low-level cognitive structure[26]. French argues that any sufficiently broad set of questions for a TT will contain subcognitive questions, even if the interrogators do not intend to ask them. The fact that the cognitive and subcognitive levels are intertwined in such a way, in turn, shows that the TT is essentially a test for human intelligence, and not for intelligence in general.

First, we will consider an interesting analogy French makes: The Seagull Test. Consider a Nordic island on which the only flying animals known to the inhabitants are seagulls. One day, two philosophers are discussing the essence of flying. One of them proposes flying is moving in the air. The other objects by tossing a pebble and stating that the pebble certainly is not flying. The first philosopher stipulates that the object remain aloft for a period of time for the activity to count as flying. But in this case clouds, smoke and children's balloons qualify as flying entities, the other argues. Then the philosopher questions whether wings and feathers should be involved but this is immediately refuted by the latter by pointing to penguins. While the arguments continue to be inconclusive, they agree on a few facts: The only flying objects known to them are the seagulls on their island. Flight has something to do with being airborne, physical characteristics like feathers, beaks are probably not involved. They, then, in the light of Turing's famous article, devise a Seagull Test for flight. They believe if something can pass the Seagull Test, it is certain that it is able to fly. Otherwise, no decision can be made; maybe it can fly, maybe it cannot [Fre90].

The Seagull Test works as follows: There are two three-dimensional radar screens, one tracking a seagull and the other tracking the flying object attempting the test. The object will pass the test only if it is indistinguishable from the seagull on the radar screen. The similarity between our TT and the Seagull Test is evident. The arguments about the gist of flying between the two philosophers is an uncanny reminiscent of the arguments on the nature of intelligence. The test itself is an almost direct analogue of the TT.

The Seagull test as it is, cannot be passed by airplanes, helicopters, bats, beetles or sparrows. It is doubtful that *anything* can pass it. That is, except for the Nordic seagulls of the philosophers' island. Then, 'what we have is not a test for flight at all, but rather a test for flight as practiced by the Nordic Seagull' [Fre90, p. 56]. The analogy makes it clear what French thinks about the TT. Correspondingly, he believes that the TT is a test for intelligence as practiced by a human being.

Before proceeding with his main arguments about subcognitive questions, French makes some assumptions that he feels are acceptable. First, he allows the interrogator to poll humans for the answers of some questions before asking them during the test with the machine. Another assumption is that all the people involved are from the same culture and the machine is attempting to imitate an individual from that culture[27]. And finally, French assumes that although building a machine that can experience the world in a way indistinguishable from that of a human being (from tasting strawberries to falling off of bikes) is theoretically possible, this is not likely to occur in the foreseeable future.

[26] Here, low-level cognitive structure refers to the subconscious associative network in human minds, consisting of highly overlapping activatable representations of experience [Fre90, p. 57].

[27] French believes that this assumption is tacitly made by Turing. The importance of culture in conversation and communication is immense (see Section 5) and this could be a reasonable stipulation.

The first class of questions that French introduces are explicitly designed to reveal low-level cognitive structure. These are based on research in cognitive psychology, in particular, associative priming. For instance, it has been established that if human subjects are presented with a set of words and non-words, they generally require less time to recognize that a string is a word if it is preceded by a related word, rather than a non-related one or a non-word. To give a specific example, the item "salt" should be recognized faster when preceded by "pepper" and not by an unrelated word like "shoe" or a non-word like "gloff".

Before questioning the computer, the interrogator runs a similar experiment on a group of people[28] and calculates average recognition times. During the test, the interrogator can distinguish between the machine and the human simply by looking at the recognition times on the same set of words on which the experiment had been carried out. It is hard to imagine a machine being successful at deceiving the interrogator in such a case. The only possible way to carry out associative priming in a manner that is similar to the average human's seems to be having experienced life as an average human.

It may seem unfair to expect computers to match scientific, experimental data about subcognitive capacities of the human mind. This view might even be exagerrated into expecting them to have a physical construction like ours. Whether such an objection is valid or not, French proposes some questions that are (or rather seem) at a higher cognitive level. The first kind is called the Neologism Rating Game. The interviewees (and later the machine) will be given a set of made-up words and will be asked to rank them on their appropriateness for defining certain entities. Participants in the game are asked to rate, on a scale of 0 (completely implausible) to 10 (completely plausible), whether a certain neologism can be a name for something. French proposes a set of such questions that are especially demonstrative. Here we will only consider two of those neologisms: "Flugblogs" and "Flugly".

According to French, "Flugblogs" would make an inappropriate choice for the name of a cereal since the initial syllable, "flug", is phonetically similar to "flub", "thug", "ugly" and "ugh!" and the second syllable, "blogs" phonetically activates "blob", "bog", etc. As can be seen, these words do not really sound very appetizing and they each carry an *aura* of semantic connotations that renders them insuitable choices as syllables of a cereal name. However, "Flugblogs" would be a very appropriate name you would give to big, bulbous, air-filled bags used to walk on water. In this case the semantic connotations of the syllables like "flug", "blob" and "bog" are in accordance with the proposed meaning. Similar analysis of "Flugly", which activates friendship, coziness and cuteness, reveals that it is a plausible name for a child's teddy bear. The same name, although it has positive connotations, would sound awkward as the surname of a glamorous movie star.

The arguments above are highly intuitive, and although most of us would agree on them, we do not know precisely how we come up with the connotations. We do know, however, that these happen due to a large number of culturally acquired associations. We do not have control over the accumulation of such associations; they are pumped into our brains in daily life as brand names, advertising slogans, names of pets and stereotypes of various sorts[29]. Moreover, it is not possible to program these into the computer since neologisms are virtually infinite in number. French believes that the

[28] In French's terminology these human subjects are called *interviewees.*

[29] The importance of cultural factors becomes evident in this context. Without having notions of Kellogg's and teddy bears, the answers to these questions would be near-random guesses.

computer's chances would be very low when the interviewees' responses to such questions are compared to those of the human and the computer in the IG.

Another game of a similar nature is the Category Rating Game in which the questions are of the type "Rate Xs as Ys", where X and Y are any two categories. Again, French gives several illustrative examples [Fre90, p. 61]. Consider, for instance, "Rate *dry leaves* as *hiding places*". The definition of dry leaves does not contain anything explicitly stating they might be good hiding places for children, and yet 'few among us would not make that association upon seeing the juxtaposition of these two concepts' [Fre90, p. 60]. If we are asked to rate, on a scale of 0 to 10, most of us (those who have seen a James Bond movie at some point in their lives) would certainly rate "*pens* as *weapons*" higher than, say, "*grand pianos* as *wheelbarrows*". Again the answers to the Category Rating Game questions are highly dependent on our having experienced life as a human being in a certain social and cultural setting.

Now that we have studied French's incisive subcognitive questions, let us see how he uses these to refute the TT as a useful test for intelligence. The main claim of French is that the physical level and the cognitive level of intelligence are inseparable. The subcognitive questions reveal information about the low-level cognitive processes of the entities answering them. In a way, if used during the TT, these would allow the interrogator to 'peek behind the screen' [Fre90, p. 62]. These questions allow comparison of the associative concept networks of the two candidates. And because these networks are formed after a lifetime of experiencing the world and the structure and nature of them are necessarily dependent on physical aspects of that experience (like human sense organs, their locations in the body, etc.), the computer will be distinguishable from the human. In short, it is not possible for a computer (or any other non-human) to be successful in playing the IG. Not having experienced the world as we have is not just an obstacle, but a severe restriction in this task. This is due to the fact that the TT is a test for human intelligence, just as the Seagull Test is a test for Nordic seagull flight.

French considers whether there can be a modification of the TT that does not reduce the computers' chances of passing it to zero. He explains the impossibility of this as follows:

> Surely, we would not want to limit a Turing Test to questions like "What is the capital of France?" or "How many sides does a triangle have?". If we admit that intelligence in general must have *something* to do with categorization, analogy-making, and so on, we will of course want to ask questions that test these capacities. But these are the very questions that will allow us, unfailingly, to unmask the computer [Fre90, p. 63].

French repeatedly states, as was mentioned above, that the TT is a test for *human* intelligence. It may seem like by proposing subcognitive questions he is stipulating that a human subcognitive substrate is *necessary* for intelligence in general, but this is only apparent. What French really attempts to demonstrate, as he explains, is that the human subcognitive substrate is necessary to pass the TT (as the subcognitive questions show), and TT is inadequate precisely because of this. He holds that this substrate is definitely not necessary for intelligence in general, just as being a Nordic seagull is not a necessary condition for flight.

French's paper is significant in one sense: Instead of discussing whether passing the TT is a sufficient or necessary condition for machine thought, he asks whether the test can be passed at all. Let Searle have his room and Block his Aunt Bubbles. French reminds us that the TT is difficult when you leave your armchair.

A criticism of French's 'Subcognition and the Limits of the Turing Test' [Fre90], has been made by Dale Jacquette in [Jac93b]. For French's response to Jacquette, the reader should refer to [Fre95].

4.6 Getting Real

As we mentioned in the beginning of this section, the more interdisciplinary approach that seems to prevail in the discussions of the mind also had effects on the way we philosophize on the TT. Thus, the 90's became a time during which it was not so easy to get away with proposing wild thought experiments and leaning back on your armchair to watch the fuss over them. Stevan Harnad expresses an impatience that many were beginning to feel as follows:

> If you want to talk about what a model or a simulation can or cannot do, first get it to run. [Har89, p. 4].

Recently, Justin Leiber has argued that the TT has been misinterpreted [Lei95]. He notes that Block's[30] and Searle's counterarguments do not refute the TT. Among the reasons Leiber lists for this are practical issues like memory, reliability and speed. Leiber views the TT as an operational definition and states that 'our problem [is] one of engineering' [Lei95, p. 65]. His position is similar to that stated by Harnad:

> What you need to face Turing's Turing Test is a reasonably detailed description of a machine which can indeed be supposed to pass the Turing Test in real time but which somehow is not really thinking. [Lei95, p. 61].

At one extreme are Patrick Hayes and Kenneth Ford, who state that we should reject the goal of passing the TT in their 'Turing Test Considered Harmful' [HF95]. They believe that passing the TT is a distraction for "useful" AI research.

Hayes and Ford believe that AI's ultimate goal should not be that of imitating human capabilities. Since the TT's sole aim is precisely that, they believe that 'it is time to move it from textbooks to the history books' [HF95, p. 972]. They also see a problem with the gender issue in the IG:

> The gender test is not a test of making an artificial human but of making a mechanical transvestite [HF95, p. 973].

> [Turing] tells us quite clearly to try to make a program which can do as well as a man at pretending to be a woman [HF95, p. 977].

As we mentioned in Section 2.1, this peculiarity might have its reasons, but Hayes and Ford have a moral objection concerned with the artificial constraints the setting imposes on the participants of the game.

Hayes and Ford also express their inability to find a practical use for the TT. Why on earth should we work that hard (and it *is* hard) to build a machine that imitates us? To depict the uselessness of direct imitation of humans in AI, they resort to a very popular analogy: mankind's futile attempts at making flying machines by the imitation of natural flight. Artificial intelligence, like artificial flight, can be radically different from natural flight. And it can still be a good thing. Hayes and Ford believe that even if one's goal is trying to understand humans, there is no reason to define all that there is about cognition in terms of human cognition.

[30] Although Leiber is mainly concerned with the homunculi argument in 'Troubles with Functionalism' [Blo78], his response also applies to Block's attack of the TT described in Section 4.1.

Their belief that AI is a field that strives to be useful leads Hayes and Ford to deny passing the TT as a sensible goal. They hold that AI should produce cognitive artifacts, not necessarily in a human way, but in a way useful to humans.

Blay Whitby, in 'The Turing Test: AI's Biggest Blind Alley?' [Whi96] makes similar arguments. He, like Hayes and Ford, believes that AI need not try to imitate humans. He even uses the same analogy (i.e., AI and artificial flight). Whitby states that the TT has become a distraction and he sees the main source as a mistaken reading of 'Computing Machinery and Intelligence' [Tur50]. He is of the opinion that 'Turing's paper [has been] interpreted as closer to an operational test than he himself intended' [Whi96, p. 54] and that 'the last thing needed by AI *qua* science is an operational definition of intelligence involving some sort of comparison with human beings' [Whi96, p. 62].

5 TT in the Social Sciences

A review of the TT would be incomplete if we were to consider the topic within the boundaries of computer science and philosophy only. Turing's ideas had many repercussions in social sciences as well. The TT has naturally received attention from sociologists. Much of the more philosophical work on the topic also considers social aspects of intelligence, but there have been researchers who concentrated solely on this dimension. These sociological works are discussed in Section 5.1. In addition, the gender issue in the TT has been analyzed and this will be summarized in Section 5.2. Finally, Turing-like tests have been used to assess the success of computer simulations of paranoid behavior. This process in described in detail in Section 5.3 and will be considered again in Section 6.

5.1 Sociological Aspects

An entity's status in a society, in general in a social environment, is often considered an integral part of its intelligence. Many psychologists believe that social adaptation, learning and communication are important indications of, even requisites for intelligence. The study of artificial intelligence has also been influenced by this outlook, as is apparent from the recent research being done on intelligent agents. Much attention is focused on learning, adaptivity, communication[31] and socio-psychological factors in intelligent systems [Col90, McI93, MNM+94].

In 1986, Charles Karelis wrote a paper for the *Journal for the Theory of Social Behavior* titled "Reflections on the Turing Test" [Kar86]. This paper summarizes Turing's original paper [Tur50] and Block's objections to the TT [Blo81], mildly criticizes the test and briefly discusses some issues surrounding behaviorist approaches to intelligence.

A few years later, in the same journal, we find "A Simple Comment Regarding the Turing Test" [Sha89] by Benny Shanon. The author first mentions the fact that most discussions of the IG are not faithful to the original form proposed by Turing. He then, continues by criticizing the TT for confining human behavior to those that can be conducted by means of the structures and operations that are available to the computer [Sha89]. He raises the important issue of whether cognition is autonomous with respect to social interaction, affect, motivation, motor control and similar systems. However, after stating that the TT presupposes the claim that there is such an autonomy, he abruptly ends his paper by asserting that the only remaining way to distinguish between man and machine is to "look at them, touch them, tickle them, perhaps see whether you fall in love with them" [Sha89,

[31] Agent-agent and human-agent.

p. 253].

Justin Leiber, in his defense of the TT against Shanon [Lei89], states that Shanon seems to be suffering from the 'unwillingness to admit the possibility that mankind can have any rivals' [Tur69] that Turing liked to call the 'heads-in-the-sand objection' [Tur50]. Leiber notes that satisfactory answers to such objections have already been given by Turing. He also argues against Shanon's claim that the TT involves only symbol manipulation and thus assumes a representational/computational framework for cognition. Leiber points out that there is ample evidence in Turing's paper [Tur50] showing that such a framework is not assumed. He asserts that Turing does not make the aforementioned autonomy presupposition either.

Tracy B. Henley's [Hen90] is another paper arguing that Shanon is being overly chauvinistic. A reply to Henley was given by Shanon in [Sha91].

Some of those who view intelligence as a part of social processes (and vice versa) take a more evolutionary approach [Bar87, For88, Sch98]. Adaptivity is indeed, a most prevalent characteristic of social intelligence. However, the issue can be viewed from two different levels: the individual level and the collective level. The approaches we have looked at above were mainly individual based. Evolutionary arguments, on the other hand, are largely collective in outlook. These usually focus on the intelligence of *species* and study the factors influencing their development. According to the evolutionary viewpoint, there is a system, i.e., nature, in which entities function and the interactions within the system have effects on individuals that, in the long run, lead to species-level adaptations. The adaptation in this context is not merely giving appropriate responses in appropriate social or physical situations, but is successful survival of the species within the whole system.

In his 1987 paper [Bar87], John Barresi considers intelligent machines as a species and proposes an evolutionary 'Cyberiad Test' instead of a Turing Test. According to Barresi, the TT aims to trick a person, but in *natural* intelligence, this person is 'mother nature'. The Cyberiad Test is similar to the TT: The basis of the judgement is a comparison between humans and the machines. The difference between the two lies in how intelligence is defined. The Cyberiad Test defines intelligent behavior as those that are necessary for the society's survival. The arbiter here, is mother nature.

According to Barresi, the TT is inferior to the Cyberiad Test because what it can process about an entity's intelligence is limited to a particular domain, namely, verbal communication. The Cyberiad Test is passed, 'if [the] society of artificial men are able to continue a socio-cultural evolution of their own without disintegration over an extended period, say of several million years' [Bar87, p. 23][32]. Even though this 'science fiction' atmosphere sometimes distracts the reader from the important assertions about evolutionary and cultural intelligence, the paper is quite an entertaining piece of work.

5.2 On Gender

Judith Genova draws attention to the gender issue in the IG [Gen94b]. She, as we have done so in Section 2.1, remarks that Turing's description of the game involves, not a question of species, but one of gender. She states in [Gen94a] that her aim was to show that the sexual guessing component of the IG is important, even after the machine enters the picture. Our explanation of this design choice differs from that of Genova's, however. As you might recall, we have not made a distinction between the two genders in our explanation. We have regarded the choice of the woman being 'imitated' as a rather insignificant one and assumed that the game would not change radically if it

[32] Compare this with [Sch98] and Section 4.4.4.

were the other way around. Genova, on the other hand, does not merely accept Turing's choices as accidental, but tries to demonstrate some motivations behind these.

Genova believes that sexist notions about women being less intelligent, by themselves, do not account for the peculiar design of the game. She states that by complicating the game in this manner, Turing questions the existence of discrete categories. In other words, by using the male/female issue, he is attempting to demonstrate that gender itself is a socially imposed concept that is not 'natural' the way we usually think it is.

Genova regards the IG as part of Turing's general philosophy of 'transgressing boundaries' [Gen94b]. Under the assumption that Turing admired such transformations that do not conform to the given discrete categories, Genova suggests that Turing might be marking the woman as an inferior thinker because he believes her to be unable to deceive. The rest of the paper considers Turing's hypothetical hope to create a 'perfect being' and draws some analogies between him and Pygmalion. As can be seen, Genova's approach is different from ours; for her, Turing's paper [Tur50] 'is itself a game' [Gen94a].

Another paper that considers the gender issue in the IG and constructs links between the design of the game and Turing's opinions on life is Jean Lassegue's 'What kind of Turing Test Did Turing Have in Mind?' [Las96]. Those readers who are interested in Turing's life and psychology might want to consult it.

5.3 Artificial Paranoia

The TT has received some attention from psychologists as well [Rea69, Alp90, GL91]. In this section, however, we will focus only on Kenneth Colby and colleagues' work on simulating artificial paranoia [CHW71, CHWK72, Col81].

In the 70's, Turing Tests were used to validate computer simulations of paranoid behavior. Colby et al. describe in their 1971 *Artificial Intelligence* paper 'Artificial Paranoia' [CHW71] a computer program that attempts to simulate paranoid behavior in computer-mediated dialogue. The program emits linguistic responses based on internal (affective) states. To create this effect, three measures, FEAR, ANGER and MISTRUST are used. Depending on the flow of the conversation, these measures change their values. Substantial detail about the artificial paranoia program can be found in [CHW71].

A year later, again in the same journal, Colby et al. describe how they validate their simulation program by a Turing-like indistinguishability test [CHWK72]. Their approach to such tests is much more practical compared to those of philosophers who argue against the TT because it is a behaviorist criterion for intelligence. Colby et al. believe that computer simulations should be *validated*, and that a simulation's acceptability must first be based on 'its success in achieving the desired end of producing resemblance at some input-output level' [CHWK72]. Thus, they view the test as a tool to validate a simulation.

They describe Turing's original IG and note that there is one important point that needs to be resolved before the test can actually be used. What the judges are *told* about the game is not explicitly stated in [Tur50]. From the provided context, it seems as if the judges are not told that one of the parties they are conversing with is a machine. This is assumed both by Colby et al. in their experiments, and by Abelson in [Abe68].

Abelson proposes an "Extended Turing Test". There is a computer program that attempts to imitate a subject carrying out a task, and there is also another target person, referred to as the *foil* that differs from the subject in some respect[33]. At first, the judge will try to identify the subject and the foil. At some point, the computer

[33] In Turing's IG, this difference is gender, for instance.

replaces the foil, but the judge's task remains unchanged. The computer's simulation is acceptable if the difference between the success of the judge in distinguishing the subject from the other is sufficiently small [Abe68].

Colby et al. chose to design their validation in the light of Abelson's comments. They devised a machine-mediated interviewing environment, where experienced psychiatrists corresponded with both real-life paranoid patients and computer programs through teletype. The judges were *not* told that some of the interviewees could be computer programs. They were asked to rank each interviewee's level of paranoia. There were eight judges who interviewed two patients each. In half of the cases, the first interviewee was a human patient and in the other half, it was the computer program. Another thirty-three psychiatrists constituted a second set of judges (referred to as the protocol judges) who were asked to determine levels of paranoia by reading the transcripts of the conversations carried out by the first set of judges. Finally, both sets of judges were asked to identify which of the interviewees was an actual human being and which was a computer program.

Details of the results can be found in [CHWK72], but the overall conclusion was that the psychiatric judges did not do better than chance guessing in identifying the computers from the human patients. In a further test, randomly selected psychiatrists were sent protocols of two interviews and were asked to judge which one is a program and which one is a real patient. Only 48% made the correct identification, which also approaches a chance level.

Paranoid people are expected to display irrational and obsessive behavior from time to time. This fact makes the task of the simulation program easy, for deficiencies in linguistic processing of the inputs and generation of the outputs can seem to be indicative of a psychological problem. One interesting point is that, the paranoia simulation program does not use *any* sophisticated linguistic techniques. Input sentences are assumed to be syntactically simple, and the operation of the program relies on spotting some keywords. No parsing or sense disambiguation is done. Even without sophisticated linguistic techniques, the program can attain some success in modeling human behavior. Apparently, this is possible only because the problem area is such that irrelevant responses from the interviewees are expected.

Still, the simple design of the program indicates that by finding appropriate measures and correlations of these with the flow of the conversation, one can model the behavior of one kind of human being without using much natural language analysis. Because the representational model of the paranoid mind used in the program is a good approximation of the paranoia-related cognitive processes in humans, and because lack of linguistic competence can be accommodated in the setting, the program can be succesful. In modeling human beings in general, the former is not so easy to discover and formalize and the latter is not the case any more.

6 Chatbots

We have reached the end of the century, but what has *really* been done in terms of passing the TT? Over the years, many natural language systems have been developed with different purposes, including that of carrying out conversations with human users[34]. These systems chat with people on the WWW, play MUDs[35], give information about specific topics, tell stories, and enter Turing Test competitions. However, none

[34] Such systems are usually called language understanding/generation systems, conversation agents, or simply, chatbots.

[35] Multi-User Dungeons. These are games played interactively on the Internet by multiple players.

has been able to *pass* the TT so far.

6.1 The Loebner Prize Contest

The TT, as Turing actually described it, has never been carried out. However, there are variants of the original in which computer programs participate and show their skills in "human-ness". Since 1991, Hugh Loebner has been organizing the so-called annual Loebner Prize Competition[36]. Although views as to whether this annual contest is to be taken seriously varies immensely among the AI community, it nevertheless continues to be the most well-known of the attempts to pass the TT. The first program to pass an unrestricted TT will win a gold medal and $100,000[37], while each year, a bronze medal and $2,000 is awarded to the most "human" program among the contestants. Since 1995, all entries must be prepared to be inquired on any topic whatsoever. No program has won the grand prize yet, but the quality of the participating programs seems to be increasing every year.

The first Loebner Prize Contest was held at Boston's Computer Museum. Six computer programs, four human subjects and ten human interrogators were involved[38]. The administrative committee was headed by Daniel Dennett, a very respectable figure among the philosophy and cognitive science community. The organizing committee, thinking that it was not possible at the time for a computer program to pass the TT as originally defined, decided that the conversation topics were to be restricted, both for the contestants and confederates. Consequently, the judges were asked to stay on topic during their interrogations. Substantial detail about the 1991 Loebner Prize Contest can be found in [Eps92]. The reader can also consult [Mau94, Pla95] for more information on other years' contests.

A widely discussed issue before 1995 was the restricted vs. unrestricted TT. According to Turing, passing a restricted TT would not suffice for intelligence. However, from another viewpoint restricted tests are not totally useless. We are not saying that they should be carried out within the context of the Loebner competition. Still, restricted tests can be devised to assess the success of more specific AI applications that are not created with passing the TT in mind. Examples of such systems that can be assessed by a restricted test are intelligent tutoring systems, computer help services, and natural language components of other applications that are designed for specific domains. The reader can also consult [Shi94] and [Loe94] for more discussion on restricted TTs and the Loebner competition.

In the Loebner contest, the sexual guessing component of the original game is ignored. The aim of the contestants is to convince the judges that they are human. One or more human confederates also participate and try to aid the judges in identifying the humans. The judges also rank the terminals with respect to their "human-ness". Although, looking at the transcripts, one can see that the computer programs are, in general, obviously distinguishable from the real humans, there have been cases in which some actual humans were ranked less human than some computer programs. In fact, in 1991, not only were some programs were thought to be human beings, but an actual human was mistaken for a computer program because of her impeccable knowledge of Shakespeare's literature[39]. The interested reader is referred to the article written by Charles Platt, one of the human confeder-

[36] http://www.loebner.net/Prizef/loebner-prize.html

[37] Now, Loebner requires that this program should also be able to process audio/visual input.

[38] In the Loebner Prize terminology, the computer programs are called 'contestants', the human subjects 'confederates' and the interrogators 'judges'.

[39] The reason why this fact does not make the TT passed is simply because Turing required *consistently* successful performance from machines to grant them intelligence.

ates in the 1994 Loebner Contest [Pla95].

The amount of time that the judges spend communicating with each terminal in the Loebner competition varies. It has been the case that each judge gets more than one chance to interrogate each terminal. Ideally, the contestants should be able to handle conversations of unlimited duration as well as multiple sessions with each judge. In the beginning, each judge was required to rank the subjects from least human to most human. They also had to mark the point at which they believed the subjects switched from computer programs to human beings.

It is difficult to keep track of the small changes in the rules. It is, for instance, not clear how the grand prize of $100,000 will be awarded. The 1999 rules state that if a contestant achieves a 50:50 chance of being mistaken for a human being, it will be given an intermediate Turing award of $25,000 and appropriate competitions (those that involve audio and video inputs) will be held after that. However, how this ratio is exactly calculated is not elucidated. We also feel the number of judges is too small to be a basis for a statistic like 50:50.

One of the aims of the Loebner competition, as Loebner states, is to advance the field of artificial intelligence. A majority of researchers believe that this competition does not, and is not likely to serve such a purpose. Some hold that it is possible that a program implemented as an entry to the competition can evolve into a useful natural language system. It is also possible that by arising an interest to the topic, the contest may prompt researchers to start working on such systems. Nevertheless, the competition has been getting harsh comments from many leading scientists and philosophers[40].

The quality of the programs that compete in the contest has been increasing. The first programs that participated in the Loebner Prize Contests were little more than slightly modified ELIZAs. One need only look at more recent transcripts or play with some of these programs to see that there *is* progress. It seems awards can get people interested in writing TT-passing programs. Hence, the Loebner Prize could indeed be influential on AI research.

We would like to finish this section with a small excerpt from a conversation between the winner of the 1996 Loebner Prize, HeX, and a particularly emotional judge[41]:

```
PROGRAM ... So what do you do
for a living?

JUDGE05 I think. I do
research. I write. I love. I
Live.

PROGRAM I wonder whether you
think sometimes! I am feeling
like the judge here!

JUDGE05 HUh?
```

6.2 Tricks of the Trade

The first, and probably the most well-known natural language system that has been programmed is ELIZA. Joseph Weizenbaum coded ELIZA at MIT during the years 1964-1966. This "friendly" program simulates a Rogerian psychotherapist. It rephrases the interrogator's statements into questions and urges him/her to continue talking. The mechanism behind ELIZA is a very simple one. First, what is typed into the program is parsed. Then, a suitable reply is formulated by simple pattern recognition and

[40] In fact, Marvin Minsky has offered $100 to the first person who can get Hugh Loebner to revoke the competition, which he calls an 'obnoxious and unproductive annual publicity campaign'. Loebner astutely declared Minsky a co-sponsor of the contest, since, according to the rules, when the grand prize is won, the contest will not be held again. In that case, with Minsky's contribution, the prize should become $100,100.

[41] In the following transcript and the others in this section, the spelling and grammar errors have not been corrected. Moreover, timing information denoting the exact response times of the participants has also been deleted.

substitution of keywords [Wei66]. The term "ELIZA-like" for chatbots is used to mean that the program tries to carry the conversation by using techniques similar to those of ELIZA's.

ELIZA would certainly perform poorly in the Loebner contests or similar instantiations of the TT. This is because the interrogators are trying to find out whether they are conversing with a human or a machine and thus, are not likely to open up about themselves and their personal problems as if they are talking to a psychotherapist. However, it has been reported that some people have even developed emotional attachment to ELIZA [Wei76]. Certain psychiatrists went so far as to suggest that such programs can replace psychotherapists all together. Weizenbaum, himself, has been amazed by these delusions that ELIZA, a simple program, could induce in perfectly normal people. These reactions to ELIZA suggest that even if the program has no chance to pass the TT, it can be said to model, with success, the main aspects of the conversational capability of one kind of human being, namely, the Rogerian psychotherapist.

A similar story is that of PARRY, which is a program that attempts to simulate another restricted class of human beings. Kenneth Colby wrote this program in the 70's in order to model the paranoid mind. A modified TT in which an experienced psychiatrist tries to distinguish between a real paranoid patient and the computer program can be carried out to assess how well the paranoid conversational characteristics are simulated. Many expert psychiatrists had difficulty making the distinction between PARRY and actual paranoids. The design of PARRY has been explained in detail in Section 5.3.

Both ELIZA and PARRY use certain tricks to be able to successfully perform in conversations. ELIZA directs the conversation away from herself by asking questions. Many people like this[42] and happily believe that the program is listening as they talk about themselves. ELIZA uses parts of the user's input in the output questions and seems to be following the conversation. In addition to these techniques, PARRY has little stories to tell and tends to insert these in the conversation.

Recent programs also employ many tricks. Joseph Weintraub's PC Therapist, which is the winner of the first three Loebner competitions, and also the one in 1995, is a master of non sequitur. Overall, the program was ELIZA-like but impressed the judges by its long and incomprehensive responses. Its success is understandable in the first three Loebner contests since the conversation topics were limited and the judges actually expected irrelevant, absurd and extravagant answers from a terminal that boasted "whimsical talk". The reader is referred to [Eps92] for transcripts of PC Therapist in the 1991 contest, which are illustrative of the tricks it employs.

PC Therapist's winning the first unrestricted Loebner competition shows that the usage of non sequitur is indeed, a powerful trick. We find the utilization of non sequitur similar to that of simulating typing mistakes of humans: In the long run, such tricks will not suffice by themselves, but they can be incorporated into otherwise successful programs as seen necessary. The transcripts of the 1995 contest indicate that little progress has been made since ELIZA. Below are some excerpts that may give an idea about the performance of Weintraub's program in 1995. You may realize that if the judge asks sensible, down to earth questions, the program's responses seem erratic and irritating[43]:

```
JUDGE06 What did you have for
```

[42] Although, contest judges most probably would not, as was mentioned before.

[43] Especially when it repeats long chunks of the judge's previous inputs in totally unrelated contexts.

```
dinner yesterday?  Keep in mind
I'm a strict vegetarian.

PROGRAM I could swear you just
said what did you do yesterday
at 7:83 p.m.??  He was so
narrow-minded he could see
through a key-hole with two
eyes.  Are such questions on
your mind?

JUDGE06 What did you have for
dinner yesterday?

PROGRAM What does a woman want
anyway?  What answer would
please you most?
```

However, depending on the attitude of the interrogator there can exist an atmosphere (or rather illusion) of mutual understanding.

```
PROGRAM Is love the most
important thing to you?  How
important is that to you?

JUDGE05 essential cure to my
existential schizophrenia

PROGRAM A striking example of
the essential identity of the
two tongues.  existential
schizophrenia, that's amusing.

...

PROGRAM You know the old saying
- fun is in the eye of the
beholder!.  What do you think?

JUDGE05 the ``beholder'' is the
central issue.  confusion
reigns because of the illusion
of ``identity.''

PROGRAM I love men, not because
they are men, but because they
are not women.  Is that the
real reason?

JUDGE05 i love all that is,
which has produced me to
observe it, yet given me the
madness to think i am distinct
ans apart from it.

PROGRAM I have this fantasy of
making love to two women in one
year.  Can you elaborate on
that?

JUDGE05 why confine love to
only two

PROGRAM Is love the most
important thing to you?  What
comes to your mind when you ask
that?
```

Another trick that has been used is to direct the conversation by talking a lot about a predefined topic. This has been used and abused, and the difference shows. Programs that cheat by typing very long responses extremely slowly give away the fact that they are trying to use up the interrogation time. However, the winner of the 1997 Loebner competition, Converse, uses the conversation-directing technique rather impressively. It begins the session by asking the interrogator about something interesting that had been on the news the previous night. Below are excerpts from the 1997 contest transcripts of Converse:

```
PROGRAM Well, it's nice to meet
you, Elliot.  Did you see that
story on CNN last night about
the lesbian couple who came out
at a White House party on
Sunday?

JUDGE04 Yes, I did.  I think it
may be a publicity stunt for
Ellen.

PROGRAM Ellen Degeneres was one
of them - she was kissing her
lover in front of Bill Clinton.
Do you think he was very
embarrassed?

JUDGE04 I think he want the gay
vote.  He will do anything to
get it and keep it.  He does
not have a conservative base.
```

```
PROGRAM Are you still a friend
of Bill Clinton?

JUDGE04 I am not sure if he has
any real friends.  He has
achieved alot and left alot of
people behind.

PROGRAM You really don't like
the guy--well, at least that's
definite--I thought most people
were indifferent or numb by
now.

--

PROGRAM Well, it's nice to meet
you Tracy.  Did you see that
story on CNN last night about
the lesbian couple who came out
at a White House party on
Sunday?

JUDGE03 No...  Are you serious?
Who?

PROGRAM I'm pretty much that
way today.  No idea who, do
you?  Hm, i was just saying
Ellen Degeneres was one of them
- she was kissing her lover in
front of Bill Clinton.  Do you
think he was very embarrassed?

JUDGE03 I doubt it very much.
From what I've heard, Bill's
been, done and seen just about
everything there is for a
consenting adult to see.

PROGRAM Are you still a friend
of Bill Clinton?

JUDGE03 Never was.

PROGRAM You really don't like
the guy--well, at least that's
definite--I thought most people
were indifferent or numb by
now.
```

It is possible to see, after reading the full transcripts, that the program repeats the exact same response sequence to each judge regardless of what they type in. But the judges could not know this, and the overall effect was quite impressive. The program spent a lot of the time in this manner, and although it could not handle many kinds of questions, won the competition. The program that came second, Jason Hutchens' SEPO, was more competent in handling diverse questions.[44]. In any case, directing the conversation seems to be more sophisticated than repeating the input back to the interrogator or asking ELIZA-like questions. The situation is similar to watching a team play in the offensive rather than in the defensive.

As was mentioned before, an approach is that of modeling one kind of person. Thom Whalen took this a step further than what has been done with ELIZA and PARRY and created an actual human being, Joe the Janitor. Although he lost the Loebner contest to Weintraub in 1995, the program was quite successful.

Whalen utilized further tricks while inventing Joe. For instance, Joe has a narrow world view since he works at nights, does not read books or newspapers and does not watch TV. In addition, he has a little story to tell about being accused of theft and fear of losing his job. These tricks are similar to those employed in the design of PARRY. The major weakness of Joe against other programs was that he did not resort to non sequitur and repeated the same responses when confronted with something he didn't *know* about. Below, we produce some excerpts from the transcripts to depict what can go right or wrong with Whalen's approach:

```
JUDGE04 where do you work?
```

[44] An interesting point was that one of the judges was named 'Nate', short for 'Nathan'. The program repeatedly addressed him as 'Mate', and complained about him not telling his name. This created the impression that SEPO lacked intelligence, but was, in fact, due to the fact that Jason Hutchens is from Australia. This suggests that the designer's mind is an integral component of the resulting programs, which is an interesting idea to ponder upon.

```
PROGRAM At the university.  I
mostly clean the Standish
Building.

JUDGE04 What university?

PROGRAM University of Eastern
Ontario.  I'm a cleaner.  I
empty trash.  Vacuum.  Dust.
You know.

JUDGE04 and why do they want to
fire you?

PROGRAM It's just because my
boss, Mr.  Johnson, is mad at
me.

JUDGE04 why is he mad at you?

PROGRAM I didn't do nothing
wrong.  But he thinks I did.
It's all because of that radio
they think I stole from the
lab.

JUDGE04 that's too bad.  are
you in a union?

PROGRAM What would I tell the
union?  They won't say they're
firing me because I stole the
radio.  They'll just make up
some excuse...

--

JUDGE02 what does nato stand
for?

PROGRAM I don't know.

JUDGE02 what's your name?

PROGRAM What's your name??  I
don't know.  What do I look
like?  Some kind of professor?

JUDGE02 What was mark twain's
real first name?

PROGRAM Huh?

...

JUDGE02 what does huh mean?

PROGRAM Beats me.

JUDGE02 huh?

PROGRAM What do you mean?

JUDGE02 beats me.

PROGRAM What?

JUDGE02 huh.

PROGRAM I don't know.
```

Another strategy of Thom Whalen, as he reports in [Wha95] was to program suitable responses to common questions on topics that people usually talk about when they first meet, like their names[45], where they live, the weather, etc. It is obviously a good idea to have these responses prepared in case the interrogator introduces such topics. In fact, some judges at some Loebner competitions make only small talk on such topics. However, generally the judges do not treat the competitors as if they were people they've recently met. They ask questions that nobody would ask to people they have just met[46].

Apparently, the judges of the 1995 Loebner competition were not very interested in Joe's little story and although they admitted that it was the terminal with the most human personality (which should be a quality that the competitors are trying to achieve, since having a human personality should be an integral part of human-ness) they ranked Weintraub's program above it. For more detail about Joe and its performance in the 1995 Loebner Prize Contest, see [Wha95].

[45] Although, as seen above, Joe cannot answer the question "What is your name?".

[46] One of the judges in the 1997 Loebner competition tried asking each terminal the question "When you got your first liscence, was it in a stick or an automatic?". The question is a cleverly planned one since words like 'driving' or 'car' are not used, but the meaning is clear from the context. Even the misspelling of the word 'licence' as 'liscence' is most probably intentional. It is difficult to imagine a computer program that would answer various trick questions of that type, but almost anyone (certainly any adult American) would be able to give a relevant answer to that one.

6.3 What Else Should be Done?

In chatbot design, first and foremost, a good grammar unit is necessary. Most programs are rather poor in this task and their tricks do not go much beyond pronoun transposition. It seems a good idea to employ more sophisticated natural language processing methods. One may argue that usage of perfect grammar is not crucial since it is quite rare that humans do that in informal transactions. If a program's responses are grammatically perfect, some interrogators may decide that no human can use English so impeccably[47]. However, most programs err in ways that give away their machine-ness; when interrogators feel they are talking to a machine, they literally *attack* it in order to fully reveal its identity. A good strategy for the TT is undisputably that of trying to maintain human-ness (or at least the neutrality) for as long as possible. It gets very difficult for the machine to make the interrogator believe that it is human after he/she has his/her mind set on "unmasking" the poor thing.

A promising approach is learning programs. The reader might recall that Turing discussed these extensively[48]. Although such programs that have been developed so far do not seem very sophisticated, the approach is logical and is likely to yield good results in the long run. Some learning chatbots boast the capacity to converse in any given language. However, there seems to be a tradeoff between the sophistication and the number of languages any one system can learn. In designing natural language learning systems, knowledge from psychology and cognitive science can be employed in order to model human language acquisition. In fact, work has been done in this line, but not with the intention of producing computer programs to pass the TT. Another option is using mathematical and statistical techniques to represent word sequences and probabilities of them occurring in proximity.

We believe many of the chatbots in the future will be using learning methods. Already, those programs that do not keep track of the current conversation (relying solely on text processing tricks) perform poorly compared to those that learn from the interrogators. As the quality of the conversational systems increase, to be competent, developers will have to integrate a learning component in their programs *and* teach them in ways that maximize their performance.

Overall, when one looks at the transcripts from the Loebner Prize Contests and talk to some chatbots, one realizes that successful programs integrate the techniques mentioned above. They have a personality and history, they try to ask questions and initiate new conversations, they produce grammatically correct responses, they have some information about recent happenings (like new movies, albums, gossip), they learn about and from the interrogators and when they don't know what to say[49], they try to respond by combining words from the interrogator's input in order to come up with a relevant answer.

7 Discussions and Conclusion

Having analyzed the '50 years of the Turing Test', we are now going to conclude our survey with a brief look at some main issues about the TT and of course, its future.

Our stand on the issues are not at the extremes. Perhaps this is because we have tried to be objective in our analyses of the arguments for or against the TT; this is expected considering the nature of our task.

[47] One might recall that Eliza Doolittle was mistaken for a Hungarian princess because she spoke English too well for a native.

[48] See Section 2.3.

[49] When all else fails, so to speak.

Most of the arguments discussed in this paper are strong, and if read independently, can "convince" the reader. However, looking at the 50 years as a whole, we find it difficult to adopt a simple viewpoint. We believe some readers, having read the current work, will also be in the same position.

We will now discuss some important issues regarding the TT and provide our own answers to (and interpretations of) those.

• *Why did Turing propose such a strange game?*

We discussed this question at length in Section 2.1. Among the comments made on the issue (for instance [Gen94b, Las96, Abe68]) we find the best explanation to be the one we provided. In the IG, the machine is supposed to be as good as a man who is imitating a woman. This gender-based design might be a methodological choice. We are asking the machine to imitate something which it isn't; so it is only fair that we compare its success against a human who is *also* imitating something which it isn't.

• *Is the TT an operational definition?*

Parts of Turing's paper (the percentages, the predictions about the future, etc.) would prompt us to believe that he intended it as such. However, most arguments surrounding the issue have been philosophical. Neither Searle's Chinese room, nor Block's Aunt Bubbles machine are practically realizable, yet they have been proposed with the intention of refuting the TT as a measure of intelligence. Apparently, proponents of such thought experiments and some other commentors view the TT as a philosophical criterion.

Viewed as a practical test, we see the TT as follows: If a machine passes the TT it should be granted intelligence. However, if it cannot, we cannot say for sure whether it thinks or not. We believe this is the most common stance towards the TT.

Philosophically, the test has been subject to many criticisms. We are all familiar with the anti-behaviorist attacks. Some have also noted that the TT is anthropomorphic. It is true that the TT tests for human intelligence. We should not be too bothered about this for it is only natural that we are interested in the only kind of intelligence we know[50].

Moreover, we need not assert that the *only* way to grant intelligence to machines is by the TT. Perhaps a good way to see the TT is as a means of gathering inductive evidence about machine mentality[51].

As was mentioned before, lately most arguments on the TT has been of the "put up or shut up" sort (e.g., [Har89, Lei95]). With the advances in computer technology, cognitive science and artificial intelligence, it is time that we stipulate that attackers or defenders of the TT back up their arguments with something more than mere intuition. This does not mean that everyone should try to develop TT-passing computer programs. However, to argue for or against the TT, we believe that a more or less realizable method of passing the test should be supplied.

• *Isn't the TT guilty of behaviorism?*

We are not saying there *should* be tests to assess machine intelligence, but if we have to make a choice, TT-like tests seem the best method for reasoning about machines' minds even though they are being accused of behaviorism. If one day, we stop granting intelligence to other human beings behavioristically, then the TT could be replaced by some other method.

The idea of a TT-passing machine having radically different information processing compared to humans is neither scary, nor improbable. If this happens one day, it will just have to be 'heads-out-of-the-sand'.

• *Isn't the TT too easy?*

The TT has been criticized for being a

[50] Moreover, it is not even evident that other "kinds" of intelligence can be conceived by human beings. The interested reader may refer to [Min85] for a good discussion on this issue.

[51] See Section 3.4.

limited test since it enables the assessment of only "verbal" intelligence. However, it does not follow from this that the test is too easy.

Proponents of this view should come up with a realizable model of a machine that passes the TT and then prove that this model does not deserve to be called intelligent. If a simple "bag of tricks" passes the TT, we are willing to either admit that the TT is too easy or that the human mind is a simple bag of tricks as well.

After 50 years, all that we have are some very rudimentary chatbots (Section 6), serendipitious FSA's [Bri94] and unrealizable Chinese rooms [Sea80] and Aunt Bubbles machines [Blo81, Blo95].

- *Isn't the TT too difficult?*

It is challenging. This is primarily due to our limited understanding of natural intelligence, more precisely language understanding, generation and processing in humans. It may even turn out that these processes are impossible to model on computers.

As is manifested by the space we devoted to Robert French's paper [Fre90], we find the idea that the TT is (too) difficult an important one. Is this a deficiency of the TT? Not if one does not require success in the TT as a necessary condition of machine intelligence. Computers, even today, perform many tasks that would require intelligence if done by humans. Research and development in this line is valuable and worthwhile. A natural language system that answers queries on a particular topic is certainly a remarkable product. It's not useless just because it cannot pass the TT. In our opinion, the TT is a sufficient condition for human-like intelligence (or more appropriately, mentality) because of the reasons outlined above. It may be too difficult to pass the TT, but this does not prevent AI from building intelligent machines.

- *Why bother about the TT?*

As we saw, there are those who believe that the TT is harmful for AI [HF95, Whi96]. If AI's aim is to make computers perform "intelligent" tasks and thereby make life easier for humans, we grant it that TT-passing programs are not very useful from that perspective.

AI researchers are being unjustly accused of mankind's failure in making machines that can pass the TT. This, we believe, is precisely the reason behind some of the harsh reactions to the TT from the AI community. Even if we take an extreme viewpoint and stipulate that AI's ultimate goal is to produce TT-passing machines, we should accept that this is a hard problem and give it more time. If less AI researchers shun the TT because "it gives the field a bad name", maybe more can be done in the positive direction.

Recall the "myth" of Newton and the apple. Recall Archimedes and his adventures in bathing. The apple might be silly, but gravity is not. Of course, thousands of people bathe, thousands of apples fall. The point is, sometimes a scientist can focus on an apple and behind it, find gravity. Later, you may forget about the apple, or even eat it if you like.

The TT may seem like a game. But trying to make computers communicate with humans in natural language is a task that may also provide valuable insights into how the human mind works. Now this latter is unarguably of scientific and philosophical interest.

- *So what happens now?*

We failed to fulfil Turing's prophecy in the first 50 years of the TT. We should admit that we have a difficult task at hand.

Hopefully, the reader has seen that many critics of the TT have expected too much, too early. Seeing the TT as the ultimate goal of AI will make many remarkable achievements look weak. The situation is somewhat reminiscent of "Fermat's last theorem" from mathematics which was proved recently by Andrew Wiles, after centuries of failure. Practically nobody believes that Fermat had proved the theorem at the time he scribbled something about lack of space in the margin of his book more than 300 years ago. In fact, Wiles' proof alludes to mathematical theory that was not developed until long after Fermat died. The same might be true of the TT. Maybe we simply don't have the requisite theory at this time[52].

The TT is after all, about simulating human use of language by computers. This, in turn, involves many questions: How do humans use language in similar settings? What is the relation between language and cognition? Is language autonomous with respect to other cognitive abilities? How can computers be made to *understand* language? What does a "simulation" mean, anyway? You can think of more questions like these. These are all *big* questions that psychologists, computer scientists, philosophers, linguists have been probing for several years. As more light is shed on each question, we will be one step closer to passing the TT. We don't know how many such steps will be necessary. Perhaps it is best to relax and not regard the TT as a "goal" but as a feat that will (if at all) be achieved through a synthesis of other remarkable feats. Everyone who considers himself/herself a "cognitive scientist" may, explicitly or implicitly, be working towards passing the TT. In any case, we believe he/she would at least be interested in what is going on in the TT arena.

> We believe that in about fifty years' time, someone will be writing a paper titled "Turing Test: 100 Years Later"

Acknowledgments

We would like to thank Dr. James Moor, Dr. Bilge Say and Armagan Yavuz for reading and commenting on earlier versions of this work.

[52] Of course, passing the TT may be "impossible", but none of the counter-arguments proposed so far suffice to establish such a bold claim.

References

[Abe68] Robert P. Abelson. Simulation of social behavior. In G. Lindzey and E. Aronson, editors, *Handbook of Social Psychology*, pages 274–356. Addison Wesley, Reading, Mass., 1968.

[Alp90] Gerald Alper. A psychoanalyst takes the Turing Test. *Psychoanalytic Review*, 77(1):59–68, 1990.

[And87] D. Anderson. Is the Chinese room the real thing? *Philosophy*, 62:389–393, 1987.

[Bar87] John Barresi. Prospects for the Cyberiad: Certain limits on human self-knowledge in the Cybernetic age. *Journal for the Theory of Social Behavior*, 17:19–46, 1987.

[Bie88] P. Bieri. Thinking machines, some reflections on the Turing Test. *Poetics Today*, 9(1):163–186, 1988.

[Blo78] Ned Block. Troubles with functionalism. In C. Wade Savage, editor, *Minnesota Studies in the Philosophy of Science*, volume 9: Perception and Cognition. University of Minneapolis Press, Minn., 1978.

[Blo81] Ned Block. Psychologism and behaviorism. *Philosophical Review*, 90:5–43, 1981.

[Blo95] Ned Block. The mind as the software of the brain. In D. Osherson, L. Gleitman, S. Kosslyn, E. Smith, and S. Sternberg, editors, *An Invitation to Cognitive Science*. MIT Press, Cambridge, Mass., 1995.

[Bod88] M. Boden. Escaping from the Chinese room. In *Computer Models of the Mind*. Cambridge University Press, U.K., 1988.

[Bri92] Selmer Bringsjord. *What Robots Can and Can't Be*. Kluwer, Dordrecht, The Netherlands, 1992.

[Bri94] Selmer Bringsjord. Could, how could we tell if, and should - androids have inner lives? In K. M. Ford, C. Glymour, and P. Hayes, editors, *Android Epistemology*, pages 93–122. MIT Press, Cambridge, Mass., 1994.

[Bri96] Selmer Bringsjord. The inverted Turing Test is provably redundant. *Psycoloquy*, 7(29), 1996. http://www.cogsci.soton.ac.uk/psyc-bin/newpsy?7.29.

[Cho75] Noam Chomsky. *Reflections on Language*. Pantheon, 1975.

[CHW71] Kenneth M. Colby, Franklin D. Hilf, and Sylvia Weber. Artificial paranoia. *Artificial Intelligence*, 2:1–25, 1971.

[CHWK72] Kenneth M. Colby, Franklin D. Hilf, Sylvia Weber, and Helena C. Kraemer. Turing-like indistinguishability tests for the validation of a computer simulation of paranoid processes. *Artificial Intelligence*, 3:199–222, 1972.

[Cla92] T. Clark. The Turing Test as a novel form of hermeneutics. *International Studies in Philosophy*, 24(1):17–31, 1992.

[CM95] Stephen J. Cowley and Karl F. MacDorman. Simulating conversations: The communion game. *AI and Society*, 9:116–137, 1995.

[Col81] Kenneth M. Colby. Modeling a paranoid mind. *Behavioral and Brain Sciences*, 4(4):515–560, 1981.

[Col90] H. M. Collins. *Artificial Experts: Social Knowledge and Intelligent Machines*. MIT Press, 1990.

[Col91] D. J. Cole. Artificial Intelligence and personal identity. *Synthese*, 88:399–417, 1991.

[Col97] H. M. Collins. The Editing Test for the deep problem of AI. *Psycoloquy*, 8(1), 1997. http://www.cogsci.soton.ac.uk/psyc-bin/newpsy?8.01.

[Cop93] B. J. Copeland. The curious case of the Chinese gym. *Synthese*, 95:173–86, 1993.

[Cro94] L. Crockett. *The Turing Test and the Frame Problem: AI's Mistaken Understanding of Intelligence*. Ablex, Norwood, N.J., 1994.

[Dav90] Donald Davidson. Turing's test. In Karim A. Said et al., editor, *Modelling the Mind*. Oxford University Press, Oxford, U.K., 1990.

[Den92] Daniel Dennett. *Consciousness Explained*. Little, Brown & Co., Boston, Mass., 1992.

[Dew92] A.K. Dewdney. Turing Test. *Scientific American*, 266(1):30–31, 1992.

[Dye90] M. Dyer. Intentionality and computationalism: Minds, machines, Searle and Harnad. *Journal of Experimental and Theoretical Artificial Intelligence*, 2:303–319, 1990.

[Eps92] Robert Epstein. The quest for the thinking computer. *AI Magazine*, 13(2):81–95, Summer 1992.

[Fei96] Edward A. Feigenbaum. How the "what" becomes the "how". *Communications of the ACM*, 39(5):97–105, 1996.

[Fet93] James H. Fetzer. The TTT is not the final word. *Think*, 2(1):34–36, 1993.

[Fet95] James H. Fetzer. Minds and machines: Behaviorism, dualism and beyond. *Stanford Electronic Humanities Review*, 4(2), 1995.

[FH96] Kenneth Ford and Patrick Hayes. The Turing Test is just as bad when inverted. *Psycoloquy*, 7(43), 1996. http://www.cogsci.soton.ac.uk/psyc-bin/newpsy?7.43.

[Flo96] Gary Flood. If only they could think: Should the Turing Test be blamed for the ills that beset artificial intelligence. *New Scientist*, 149(2012):32–35, 1996.

[Fod91] Jerry A. Fodor. Yin and Yang in the Chinese room. In D. Rosenthal, editor, *The Nature of the Mind*. Oxford University Press, Oxford, U.K., 1991.

[For88] R. Forsyth. The trouble with AI. *Artificial Intelligence Review*, 2(1):67–77, 1988.

[Fre90] Robert French. Subcognition and the limits of the Turing Test. *Mind*, 99(393):53–65, 1990.

[Fre95] Robert French. Refocusing the debate on the Turing Test: A response. *Behavior and Philosophy*, 23:59–60, 1995.

[Fre96] Robert French. The Inverted Turing Test: A simple (mindless) program that could pass it. *Psycoloquy*, 7(39), 1996. http://www.cogsci.soton.ac.uk/psyc-bin/newpsy?7.39.

[Gen94a] Judith Genova. Response to Anderson and Keith. *Social Epistemology*, 8(4):341–343, 1994.

[Gen94b] Judith Genova. Turing's sexual guessing game. *Social Epistemology*, 8(4):313–326, 1994.

[GL91] Robert M. Galatzer-Levy. Computer models and psychoanalytic ideas: Epistemological applications. *Society for Psychoanalytic Psychotherapy Bulletin*, 6(1):23–33, 1991.

[GT88] S. Guccione and G. Tamburrini. Turing's test revisited. In *Proceedings of the 1998 IEEE International Conference on Systems, Man and Cybernetics*, volume 1, pages 38–41, Beijing and Shenyang, China, August 1988.

[Gui83] Michael A. Guillen. The test of Turing. *Psychology Today*, 17(12):80–81, 1983.

[Gun64] Keith Gunderson. The imitation game. *Mind*, 73:234–245, 1964.

[Gun67] Keith Gunderson. *Mentality and Machines*. Doubleday, New York, NY., 1967.

[Hal87] M. Halpern. Turing's test and the ideology of Artificial Intelligence. *Artificial Intelligence Review*, 1(2):79–93, 1987.

[Har89] Stevan Harnad. Minds, machines and Searle. *Journal of Experimental and Theoretical Artificial Intelligence*, (1):5–25, 1989.

[Har90] Stevan Harnad. The symbol grounding problem. *Physica D*, (42):335–346, 1990.

[Har91] Stevan Harnad. Other bodies, other minds: A machine incarnation of an old philosophical problem. *Minds and Machines*, 1:43–54, 1991.

[Har92] Stevan Harnad. The Turing Test is not a trick: Turing indistinguishability is a scientific criterion. *SIGART Bulletin*, 3(4):9–10, October 1992.

[Har94] Stevan Harnad. Does mind piggyback on robotic and symbolic capacity? In H. Morowitz and J. Singer, editors, *The Mind, the Brain, and Complex Adaptive Systems*. Addison Wesley, Reading, Mass., 1994.

[Har98] Stevan Harnad. Turing indistinguishability and the blind watchmaker. In G. Mulhauser, editor, *Evolving Consciousness*. John Benjamins, Amsterdam, 1998.

[Hau93] Larry Hauser. Reaping the whirlwind: Reply to Harnad's "other bodies, other mind". *Minds and Machines*, 3:219–237, 1993.

[Hau97] Larry Hauser. Searle's Chinese box: Debunking the Chinese room argument. *Minds and Machines*, 7:199–226, 1997.

[Hay98] Brian Hayes. Turing's Test. *Muse*, 8, April 1998.

[Hen90] Tracy B. Henley. Chauvinism and science: Another reply to Shanon. *Journal for the Theory of Social Behavior*, 20(1):93–95, 1990.

[HF95] Patrick Hayes and Kenneth Ford. Turing Test considered harmful. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, volume 1, pages 972–977, 1995.

[HHPB92] P. Hayes, S. Harnad, D. Perlis, and N. Block. Virtual symposium on virtual mind. *Minds and Machines*, 3(2):217–38, 1992.

[Hod83] Andrew Hodges. *Alan Turing: The Enigma*. Simon & Schuster, New York, NY., 1983.

[Hof82] Douglas R. Hofstadter. The Turing Test: A coffee-house conversation. In D.R. Hofstadter and D.C. Dennett, editors, *The Mind's I: Fantasies and Reflections on Self and Soul*, pages 69–95. Penguin Books, 1982.

[Jac93a] Dale Jacquette. A Turing Test conversation. *Philosophy*, 68:231–233, 1993.

[Jac93b] Dale Jacquette. Who's afraid of the Turing Test. *Behavior and Philosophy*, 20:63–74, 1993.

[Kar86] Charles Karelis. Reflections on the Turing Test. *Journal for the Theory of Social Behavior*, 16:161–172, 1986.

[Kug86] Peter Kugel. Thinking may be more than computing. *Cognition*, 22:137–198, 1986.

[Kug90] Peter Kugel. Is it time to replace Turing's Test? 1990 Workshop *Artificial Intelligence: Emerging Science or Dying Art Form*, sponsored by SUNY Binghamton's Program in Philosophy and Computer and Systems Sciences and AAAI, 1990.

[Lai88] Philip Johnson Laird. *The Computer and the Mind*. Harvard University Press, Cambridge, Mass., 1988.

[Las96] Jean Lassegue. What kind of Turing Test did Turing have in mind? *Tekhnema*, 3:37–58, 1996.

[Lei89] Justin Leiber. Shanon on the Turing Test. *Journal for the Theory of Social Behavior*, 19(2):257–259, 1989.

[Lei92] Justin Leiber. The light bulb and the Turing-tested machine. *Journal for the Theory of Social Behaviour*, 22:25–39, 1992.

[Lei95] Justin Leiber. On Turing's Turing Test and why the matter matters. *Synthese*, 105:59–69, 1995.

[Loe94] Hugh Gene Loebner. In response. *Communications of the Association for Computing Machinery*, 37:79–82, June 1994.

[Luc61] J.R. Lucas. Minds, machines and Gödel. *Philosophy*, 36:112–127, 1961.

[Luc96] J.R. Lucas. Minds, machines and Gödel: A retrospect. In P. Millican and A. Clark, editors, *Machines and Thought*. Oxford University Press, Oxford, U.K., 1996.

[Mal87] J.C. Maloney. The right stuff. *Synthese*, 70:349–72, 1987.

[Mar95] L. Marinoff. Has Turing slain the Jabberwock? *Informatica*, 19(4):513–526, 1995.

[Mau94] Michael Mauldin. Chatterbots, tinymuds and the Turing Test: Entering the Loebner prize competition. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, volume 1, pages 16–21, Seattle, WA, August 1994.

[May52] W. Mays. Can machines think? *Philosophy*, 27:148–62, 1952.

[McI93] Paul McIlvenny. Constructing societies and social machines: Stepping out of the Turing Test discourse. *Journal of Intelligent Systems*, 3(2-4):119–156, 1993.

[Mic90] Donald Michie. The superarticulacy phenomenon in the context of software manufacture. In D. Partridge and Y. Wilks, editors, *The Foundations of Artificial Intelligence*, pages 411–439. MIT Press, Cambridge, Mass., 1990.

[Mic94] Donald Michie. Consciousness as an engineering issue, part 1. *Journal of Consciousness Studies*, 1(2):52–66, 1994.

[Mic95] Donald Michie. Consciousness as an engineering issue, part 2. *Journal of Consciousness Studies*, 2(1):182–195, 1995.

[Mic96] Donald Michie. Turing's test and conscious thought. In P. Millican and A. Clark, editors, *Machines and Thought: The Legacy of Alan Turing*, pages 27–51. Oxford University Press, Oxford, U.K., 1996. Originally printed in *Artificial Intelligence* 60: 1-22, 1993.

[Mil73] P. Hartley Millar. On the point of the Imitation Game. *Mind*, 82:595–597, 1973.

[Min85] Marvin Minsky. Communication with alien intelligence. In Edward Regis, editor, *Extraterrestrials: Science and Alien Intelligence*. Cambridge University Press, Cambridge, U.K., 1985.

[MNM+94] Youngme Moon, Clifford Nass, John Morkes, Eun-Young Kim, and B.J. Fogg. Computers are social actors. In *Proceedings of the CHI Conference*, pages 72–78, Boston, MA, 1994.

[Moo76] James H. Moor. An analysis of the Turing Test. *Philosophical Studies*, 30:249–257, 1976.

[Moo78] James H. Moor. Explaining computer behavior. *Philosophical Studies*, 34:325–327, 1978.

[Nar96] Ajit Narayanan. The intentional stance and the imitation game. In P. Millican and A. Clark, editors, *Machines and Thought: The Legacy of Alan Turing*, pages 63–79. Oxford University Press, Oxford, U.K., 1996.

[Par90] H.M. Parsons. Turing on the Turing Test. In W. Karwowski and M. Rahimi, editors, *Ergonomics of Hybrid Automated Systems II*. Elsevier, Amsterdam, 1990.

[Pla95] Charles Platt. What's it mean to be human, anyway? *Wired*, April 1995.

[Pur71] Richard L. Purtill. Beating the imitation game. *Mind*, 80:290–294, 1971.

[Ran87] T.L. Rankin. The Turing paradigm: A critical assessment. *Dialogue*, 29:50–55, 1987.

[Rea69] A.V. Reader. Steps toward genuine artificial intelligence. *Acta Psychologica*, 29(3):279–289, 1969.

[Rey86] G. Rey. What's really going on in the Chinese room? *Philosophical Studies*, 50:196–285, 1986.

[Ric82] R.C. Richardson. Turing Tests for intelligence: Ned Block's defense of psychologism. *Philosophical Studies*, 41:421–426, 1982.

[Rob90] L. Roberts. Searle's extension of the Chinese room to connectionist machines. *Journal of Experimental and Theoretical Artificial Intelligence*, 2:185–187, 1990.

[Sam73] Geoffrey Sampson. In defence of Turing. *Mind*, 82:592–594, 1973.

[SC93] Ravi Sharma and David Conrath. Evaluating expert systems: A review of applicable choices. *Artificial Intelligence Review*, 7(2):77–91, 1993.

[Sch98] Paul Schweizer. The Truly Total Turing Test. *Minds and Machines*, 8:263–272, 1998.

[Sea80] John R. Searle. Minds, brains and programs. *Behavioral and Brain Sciences*, 3:417–424, 1980.

[Sea90] John R. Searle. Is the brain's mind a computer program? *Scientific American*, (262):26–31, 1990.

[Sha89] Benny Shanon. A simple comment regarding the Turing Test. *Journal for the Theory of Social Behavior*, 19(2):249–256, 1989.

[Sha91] Benny Shanon. Chauvinism: A misdirected accusation. *Journal for the Theory of Social Behavior*, 21(3):369–371, 1991.

[Shi94] Stuart M. Shieber. Lessons from a restricted Turing Test. *Communications of the Association for Computing Machinery*, 37:70–78, June 1994.

[Sta78] D.F. Stalker. Why machines can't think: A reply to James Moor. *Philosophical Studies*, (34):317–320, 1978.

[Ste76] John G. Stevenson. On the imitation game. *Philosophia*, 6:131–133, 1976.

[Tur50] Alan Turing. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950.

[Tur69] Alan Turing. Intelligent machinery. In D. Michie B. Meltzer, editor, *Machine Intelligence 5*, pages 3–23. Edinburgh University Press, 1969. Originally, a National Physics Laboratory Report, 1948.

[Wal97] Richard S. Wallace. The lying game. *Wired*, August 1997.

[Wat96] Stuart Watt. Naive psychology and the inverted Turing Test. *Psycoloquy*, 7(14), 1996. http://www.cogsci.soton.ac.uk/psyc-bin/newpsy?7.14.

[Wei66] Joseph Weizenbaum. ELIZA—a computer program for the study of natural language communication between men and machines. *Communications of the ACM*, 9:36–45, 1966.

[Wei76] Joseph Weizenbaum. *Computer Power and Human Reason: From Judgement to Calculation*. W. H. Freeman, San Francisco, CA, 1976.

[Wha95] Thom Whalen. How I lost the contest and re-evaluated humanity, 1995.

[Whi96] Blay Whitby. The Turing Test: AI's biggest blind alley? In P. Millican and A. Clark, editors, *Machines and Thought: The Legacy of Alan Turing*, pages 53–63. Oxford University Press, Oxford, U.K., 1996.