

---

# The Turing Test

by [\*Lynellen D.S. Perry\*](#)

In 1950, Alan Turing wrote about a test that would ``refute anyone who doubts that a computer can really think: if an observer cannot distinguish the responses of a programmed machine from those of a human being, the machine is said to have passed the Turing test" [2]. In 1991, the first Turing test competition was conducted and the Loebner Prize was awarded to the winner. (For information, see <http://acm.org/~loebner/loebner-prize.htmlx>) Charles Platt was a judge at a Turing test competition a few years later, and you can read about his experience at [www.wired.com/wired/3.04/features/turing.html](http://www.wired.com/wired/3.04/features/turing.html). To make these Turing tests practical, the programs had to respond to only a small subset of topics, including women's clothing, Burgundy wine, and romantic relationships [1]. The restriction of discussion topics takes the teeth out the Turing test, for it is much easier to converse about a restricted range of topics than to demonstrate general conversational ability.

The controversy around the Turing test is that it doesn't seem to be very general and it defines intelligence purely in terms of behavior. Thus the Turing test is not an adequate test of intelligence. Conversation is not the ultimate display of intelligence, and real thinking is not indicated by spitting out sentences, because that is all the computer is programmed to do.

In some ways, this test reminds me of the intelligence tests that immigrants had to take when arriving at Ellis Island at the turn of the century. The Ellis Island tests were culturally dependent (like identifying the correct sequence of pictures that depicted an event, or identifying the emotion on drawn faces). Similarly, a Turing test with a restricted domain is like an intelligence test which only accurately measures white males who watched a lot of television in 1970.

The Turing test focuses too much on the behavior of conversation. Just because I am speaking to a student who speaks lousy Pig-Latin does not mean the student is not intelligent. A computer program that sometimes miscues and says something that makes no sense does not necessarily mean its not intelligent. On the other hand, a person who has a very extensive knowledge in a small domain can seem to be a computer due to this intricate knowledge. This doesn't necessarily imply intelligence either, since it says nothing about the person's ability to learn, handle a new situation, or to merely converse about some other subject. In the end, conversational skills are not the ultimate sign of intelligence, even in a world where communication media are pervasive.

Casti [1] writes about philosopher Ned Block's argument against the Turing test. If we were to ``write down a tree structure in which every possible conversation of less than five hours' duration is explicitly mapped out" [1] and then have the computer follow this tree during the conversation, it would have all the appearances of being intelligent. But this ``strongly suggests that the machine has no mental states at

all ...Intelligence is not just the ability to answer questions in a manner indistinguishable from that of an intelligent person; to call a behavior intelligent is to make a statement about how that behavior is produced" [1]. Following a preset structure is not intelligent, but being able to adapt dynamically is intelligent.

Since behavior alone is not a test of intelligence, what exactly is intelligence? How can it be noticed or observed? This alone is a large debate in the artificial intelligence community, just as it is debated in the psychology community and the philosophy community. Henley argues that most AI applications under development today are "pragmatic" in their definition of intelligence. "If it functions in the same capacity as an intelligent, indeed expert, human being then that is that. Intelligent in these situations is defined in the practical terms of cost/benefits and bottom line performances" [4]. This may seem to be a problem until he points out that neither philosophy nor psychology have reached a consensus about the definition of intelligence. So perhaps a pragmatic definition is not so horrible.

Newell and Simon say that intelligence will involve "the use and manipulation of various symbol systems, such as those featured in mathematics or logic" [2]. Others suggest intelligence includes such things as "feelings, creativity, personality, freedom, intuition, morality, and so on" [3]. Before we can definitively test for the presence of intelligence, we must arrive at a consensus about its definition.

Another factor to look at in this quest is the purpose of building artificial intelligence. Are we trying to simulate human minds in order to experiment with hypotheses about how they work? Or are we interested solely in the end result? If we are only interested in the consequences of a program's execution, its output, then perhaps the Turing test is applicable. In this type of situation, it doesn't matter how the program arrived at the response, but merely the fact that the output matched, to some degree, the output that would be expected from a human. The appearance of intelligence could be sustained by a program that had merely a large enough database of pre-programmed responses (or a fast method of generating a response) and a good pattern recognizer that could trigger the appropriate output.

But if we care at all how the output was produced, then this type of simple program will not demonstrate intelligence. Rather, we would say that its output was not the result of any real thinking but a set of rules for matching input to output. This reminds me of Searle's Chinese room (for a summary of this, see <http://sun1.iusb.edu/~lzynda/searle.html> for example). The output appears to be intelligent because it correctly translates the input, but there is no real understanding or intention or purpose behind the output. The output is generated by blindly following a set of rules.

I think that the process of arriving at a result is part of intelligence. Intelligence partly involves sensing the environment, processing that information, and acting to make changes. An intelligent program must be able to handle previously un-encountered situations, be able to operate effectively with inaccurate and incomplete knowledge, be able to learn from past experience, have access to a large amount of common sense knowledge, and be able to predict possible outcomes and prepare for them. A conversational test can not adequately test these requirements. Probably we can not test adequately for intelligence until we have learned to mate machine intelligence with robotics so that the collection can manipulate a physical

environment in order to test its hypotheses. One expression of this idea can be found in the Virtual Turing Test (see [www.sunsite.unc.edu/dbarberi/vr/ultimate-turing](http://www.sunsite.unc.edu/dbarberi/vr/ultimate-turing)).

It seems that this sort of test, though not conversational, is again focused on behavior. It may be that the only way we can measure human intelligence is by evaluating the actions taken in various situations. This is because we do not have access to the inner states of the human mind, so we don't completely understand what sequence of states would constitute intelligence while any other sequence would indicate a lack of it. However, although we could have access to the states a program goes through, simply monitoring these leads to a dangerous assumption. If we did know the exact way a human mind works, and the computer program did not follow the same method (but got the same results), then some might say that the program wasn't intelligent because it didn't do things the way humans do. So we have to ask if the goal of artificial intelligence is to exactly model the way a human does things, or is the goal to solve the same types of problems that humans solve (or problems with which humans have a hard time).

In either case, the Turing test is not a good test for intelligence. If the computer performs better than the human and thus gives itself away, then it has technically failed the test. If a judge is comparing the problem solving methods of the human and the computer and they don't match, then the computer fails again even if it is able to solve the problem when the human was not. While the Turing test is not adequate, it does seem that we will end up with some variety of behavioral test when looking for intelligent programs. However, we still have a ways to go on getting everyone to agree to a definition of intelligence. While waiting around for a definition, why not apply your own Turing test to some of the conversation programs on the net? There's a list at: [www.yahoo.com/Recreation/Games/Internet\\_Games/Interactive\\_Web\\_Games/Artificial\\_Intelligence](http://www.yahoo.com/Recreation/Games/Internet_Games/Interactive_Web_Games/Artificial_Intelligence).

## References

1

Casti, John L. 1994. [\*Complexification\*](#). Harper Collins Publishers.

2

Gardner, Howard, 1987. [\*The mind's new science: A history of the cognitive revolution\*](#). Basic Books.

3

Haugeland, J. 1985. [\*Artificial intelligence: the very idea\*](#). Cambridge, Mass: MIT Press.

4

Henley, Tracy B. 1991. Natural problems and artificial intelligence. *Behavior and Philosophy* 18:2, p.43-55.