Peeking Behind the Screen:
The Unsuspected Power of the Standard Turing Test

To appear in the *Journal of Experimental and Theoretical Artificial Intelligence* (2000).

Robert M. French
Quantitative Psychology and Cognitive Science (B32)
University of Liège
Liège, Belgium
email: rfrench@ulg.ac.be
URL: http://www.fapse.ulg.ac.be/Lab/cogsci/rfrench.html

Abstract

No computer that had not experienced the world as we humans had could pass a rigorously administered standard Turing Test. We show that the use of "subcognitive" questions allows the standard Turing Test to indirectly probe the human subcognitive associative concept network built up over a lifetime of experience with the world. Not only can this probing reveal differences in cognitive abilities, but crucially, even differences in *physical aspects* of the candidates can be detected. Consequently, it is unnecessary to propose even harder versions of the Test in which all physical and behavioral aspects of the two candidates had to be indistinguishable before allowing the machine to pass the Test. Any machine that passed the "simpler" symbols-in/symbols-out test as originally proposed by Turing would be intelligent. The problem is that, even in its original form, the Turing Test is already too hard and too anthropocentric for any machine that was not a physical, social, and behavioral carbon copy of ourselves to actually pass it. Consequently, the Turing Test, even in its standard version, is not a reasonable test for general machine intelligence. There is no need for an even stronger version of the Test.

The Rock Test for Strength

Let us begin by considering the following simple thought experiment. Suppose that a committee of sports professors wants to define exactly what is meant by "being strong." After considerable work on the problem, they develop a sufficient test for strength. The test they come up with — the Rock Test — is simple in the extreme: it certifies someone as being strong if he or she can pick up a 5,000-pound rock with their bare hands. But remember, the committee adds, the Rock Test is not a necessary condition for strength, merely a sufficient one. So, if someone fails the Rock Test, no conclusions can be drawn about their strength. Heated arguments ensue about whether or not someone who managed to lift such a rock would necessarily be strong. Certainly, opponents of the Rock Test argue, a person who can pick up a 5000-pound rock might be strong when it came to rock-lifting, but there is so much more to being strong than is evidenced by merely lifting a two and a half ton rock off the ground. What's more, there might be clever tricks that would allow a person to lift such a rock without being genuinely strong. Seeing the reason in their opponents' arguments, the committee agrees the to raise the weight of the test rock to 25 tons, because, while they agree it might be possible to lift a 5000-pound rock by trickery or fluke, 25 tons is a different matter altogether. Everyone agrees that anyone who could lift 25 tons off the ground is strong.

Of course, what the Rock Test overlooks is that no human being now, or at any time in the future, will ever be able to lift a 5000-pound rock off the ground, let alone one weighing 25 tons. In both cases, the number of Rock-Test certified strong people will remain precisely zero and the Rock Test, while not wrong and certainly a sufficient condition for strength, will be useless as a meaningful criterion of strength.

The same is true of the Turing Test, as Turing originally proposed it. The standard Turing Test, when rigorously administered in a manner that will be presented in the remainder of this article, could only be passed by something that had lived life as we humans had, had a body essentially identical to a human body, etc. We will show why the standard Turing Test — and not some other, more rigorous version of it, in particular, the Total Turing Test (Harnad, 1989, 1991) or the Truly Total Turing Test (Schweizer, 1998) — is already too hard to be of any use as a reasonable test of machine intelligence.

The Imitation Game circa 1950 and Now

Alan Turing first proposed his Imitation Game definition of intelligence — today, called the Turing Test — in a classic paper published in *Mind* in 1950 (Turing, 1950). The game consists of a person in one room, a machine in another, and a human interrogator connected to the two candidates by means of a teletype. Turing incorporated a teletype link to hide physical characteristics that he felt were not necessary to cognition (skin, hands, noses, presence and color of hair, etc.) from the Interrogator. The teletype link serves the same purpose as the screen in music competitions that is often placed between the musician and the jury, ensuring that members of the jury not influenced by the musician's physical appearance. The person in the Turing Test attempts to convince the interrogator (rightly) of his or her personhood; the machine tries to persuade the interrogator (falsely) that it is, in fact, the person. If the interrogator is consistently unable to distinguish the person from the machine, the machine will be said to have passed the Test and will be said to be intelligent.

This operational definition of intelligence was meant to sidestep the quagmire of philosophical and psychological difficulties associated with attempting to establish a set of conditions that would define intelligence in general and machine intelligence in particular. Ironically, instead of avoiding these difficulties as Turing had hoped, the paper has generated more comments and controversy than any single paper in artificial intelligence. It is arguably one of the most widely discussed scientific papers ever written.

For the purposes of the present discussion, we are going to "update" the Imitation Game in a very minor way, which Turing would certainly have agreed did not violate the

spirit of his original Imitation Game. Instead of a clunky teletype link, we will install a computer link. Each contestant will have a computer screen and a keyboard. However, pictures or free-form art will not be allowed to be sent over the link, only character-based text, as in Turing's original game. Significantly, this slight modification respects what Harnad (1989) has called the "symbols-in/symbols-out" (SISO) nature of the Test as Turing proposed it.

Changing perception of the Turing Test

Over the past forty years the emphasis of the comments on the Turing Test has shifted significantly. From 1950 through the 1970's the overwhelming majority of the papers took for granted the fact that it would be possible for a machine to pass the Turing Test. This served as the springboard for the arguments that followed. Most authors argued that it was possible for a machine to pass the Test and still not be intelligent; far fewer believed that the Test would indeed constitute a sufficient guarantee of intelligence. But this was in the early days of artificial intelligence (AI), when there was an unbridled optimism among most of the researchers in the field that the creation of machine intelligence was just around the corner. But as the initial promises of AI remained unfulfilled, the realization of just how hard it would be to achieve AI gradually began to sink in. Along with this realization, there was a parallel shift in the general perception of the Turing Test. By the mid-1980's a number of papers began questioning, not the sufficiency of the Test as a test of machine intelligence, but, rather, whether any machine could ever actually pass such a test. Dennett (1985) stressed the extreme severity of the Turing Test and asked people to think about what it would actually take to pass it.

French (1988, 1990) introduced a technique involving "subcognitive" questions that indirectly probe the subcognitive associative concept network that humans build up over a lifetime of experiencing the world and made the claim that the Turing Test was, in fact, not a test of intelligence, but rather a test of "culturally oriented human intelligence." He claimed that no machine that had not lived life as a human would ever be able to pass it. Crockett (1994) made similar claims about the Test's anthropocentrism and extreme difficulty that a machine would have to actually pass it.

The technique of "subcognitive questions" will serve as the basis for the present paper. There are, however, a number of significant differences in the manner in which this technique was used in earlier papers and the use to which it will be put in the present paper. In the earlier articles, the technique of subcognitive probing was essentially used to reveal *cognitive* differences in the two candidates, thereby unmasking any machine that had not experienced the world as we humans had. In the present paper, we will show that subcognitive questions can also be used to probe even deeper, revealing not only subtle cognitive differences in the two candidates, but *physical* differences as well. We do this to show that, in fact, there is no need (as suggested by Harnad, 1989, 1991, etc.) to "remove the screen" between the interrogator and the two candidates created by the teletype.

Further, the fact that subcognitive questions can also probe physical structural differences shows not only the role that physical attributes play in cognition, but the extent to which they can be revealed by subtle high-level ("symbolic") questioning. A corollary is, of course, just how tightly cognition is associated with physical attributes, a conclusion that has been reached by numerous authors, especially in the last decade (Harnad, 1990; Lakoff, 198x; Lakoff and Johnson, 1987; Varela, etc.).

A Hierarchy of Turing Tests

Harnad (1989, 1991) introduced the Total Turing Test. He claimed that the level at which Turing had specified his test was inappropriate since it was, basically, an ungrounded (i.e., disembodied, see Harnad, 1990) SISO level that necessarily missed a lot of what we commonly call cognition. He insisted that at this level, which he referred to as T2, the

symbols were not *grounded*, i.e., had no experiential contact with their referent in the real world. This contact, he claimed, was necessary for a fully cognitive machine to pass the Turing Test. T3, the level of the Total Turing Test (he also called this the Robotic Turing Test), meant taking away the teletype screen. The interrogator's communication with the candidates would now be infinitely richer at this new level: now both candidates could be seen, heard, touched, smelled, and spoken to directly. If the machine did not *look like* the human candidate, for example, it would immediately fail the Total Turing Test. In other words, Harnad seems to be arguing, in apparent agreement with Searle (1980), that "mere" SISO questioning could allow a non-intelligent entity to slip through the cracks of the Turing Test, hence the need for the Total Turing Test.

In later papers Harnad (1994, 1999) goes beyond the Total Turing Test and defines a hierarchy of Turing levels, ranging from t1 (the "t" is in lower case, to indicate "toy level" — the level of all restricted Turing Tests), T2 (the level Turing proposed), T3 (the Robotic Turing Test), T4 (everything is indistinguishable down to the level of blood color, fingernail growth rate, neural wiring, etc.) and, finally, T5, in which everything "down to the last electron" is identical in the two candidates, but they do not share a "Grand United Theory of Everything" (e.g., one may employ undetectable-in-principle quarks, the other not, but the two candidates are otherwise perfectly indistinguishable). Harnad repeatedly claims that the correct level of interpretation of the Turing Test is T3, i.e., robotic indistinguishablity.

The Standard Turing Test is "good enough."

The central claim of the present paper is that Turing's original level is already too strong for any machine that has not lived life as we humans have and, therefore, there is no point in making the test even harder. The power of the SISO test originally proposed by Turing is far greater than most people, including most people who have commented on the Turing Test, realize. In fact, the originally proposed Test is so hard that no machine that wasn't essentially identical to us in virtually all physical respects could pass it (French 1988, 1990). The reason for very close physical resemblance is simple: experiencing life as we humans have — necessary in order to pass the Turing Test —requires a body, arms in the right place, hands in the right place, eyes with a certain degree of precision, located in a particular place, etc. This point will become clearer once the notion of subcognitive questioning has been introduced.

Subcognitive questioning

"Subcognitive questions" (French 1988, 1990, 1996) are the means by which we will peek behind Turing's screen. These questions can be produced on a computer keyboard (i.e., no pictures, objects, etc. are allowed to be part of the question) and will allow us to probe cognitive and even physical characteristics far below the "symbolic" level at which the questions are asked. The idea is as follows. The answers to many questions (e.g., "In what city is the Eiffel Tower located?" "When was Winston Churchill born?" etc.) can be drawn from a database of declarative facts about the world. But a different class of questions, which we will call "subcognitive," pose an entirely different problem for any entity that had not experienced the world as we humans had. And, crucially, to experience the world as we have would require sense organs very similar to our own, almost identical reactions to all manner of cues in the environment, very similar patterns of associations, etc.

Consider, for example, the question: "Does freshly baked bread smell nicer than a freshly mowed lawn?" A machine that had never smelled either baking bread or a newly mowed lawn would have a great deal of trouble answering this question, unless, of course, it had been specifically programmed to answer that particular question. But there are infinitely many such questions that humans can answer immediately because they can make a judgment, based on actual physical experience, about the degree of pleasure associated with

each. Further, on average, most people within a particular culture will respond to this type of question in a similar manner. It is this fact that we will use to infallibly trip up the computer.

To reiterate, the underlying idea of subcognitive questions is that they tap into those things which are associated with our uniquely human manner of interacting with the world, which, among other things, is a product of the presence, precision and location of our sense organs, as well as our lifetime of cultural and social interactions.

A primer on Subcognitive Questions and the Subcognitive Human Profile

Here is how the Well-Prepared Interrogator in a standard Turing Test would conduct herself in order to unmask any computer that had not experienced life as we humans had (which means, in addition, "did not have the same body as we have"). She will first prepare a long list of subcognitive questions that look like this:

"On a scale of 1 (awful) to 10 (excellent), please rate:

- How good is the name *Flugly* for a glamorous Hollywood actress?
- How good is the name *Flugly* for an accountant in a W.C. Fields movie?
- How good is the name *Flugly* for a child's teddy bear?
- etc.

On a scale of 1 (terrible) to 10 (excellent), please rate:

- banana peels as musical instruments
- coconut shells as musical instruments
- radios as musical instruments
- dry leaves as hiding places
- banana splits as medicine
- marbles as eyes
- newspapers as fly swatters
- etc.

Please rate the following smells (1=very bad, 10=very nice).

- Newly cut grass.
- Freshly baked bread
- A wet bath towel
- The ocean
- A hospital corridor
- The interior of a new car.
- Ground pepper
- Fried garlic
- Burning pine needles
- Burning rubber
- Burning paper
- Green oak leaves
- Yogurt

All of these questions attempt to elicit information from the vast, largely unconscious associative concept network that we have all built up over a lifetime of interacting with our environment. Furthermore, there is nothing "tricky" about these questions – for a human being, that is.

It is worth considering this point in detail. Consider how good "Ethel Flugly" would be for the name for a glamorous Hollywood actress. It just doesn't work. (Anymore than "Archibald Leach" worked for a handsome male movie star . . . which is precisely why Hollywood movie moguls rechristened him "Cary Grant.") On the other hand, it works

perfectly for an accountant in a W.C. Fields movie. Why? Because, in your mind's ear, you can hear a cantankerous W. C. Fields saying, "Flugly, get my gloves and let us pay a little visit to Miss Whipsnade." It also works for a child's teddy bear, because it partially activates words like "fluffy," "cuddly" (similar sounds), etc. Of course, "ugly" will become active but most likely in the sense of the Ugly Duckling, with all the connotations surrounding the loveable little duckling in the children's story, etc. In any event, even if we aren't sure exactly *why* it works, most people would agree that *Flugly* would be a downright awful name for a sexy actress, a good name for a character in a W.C. Fields movie, and a perfectly appropriate name for a child's teddy bear. But why? Notice that no explicit rules determine these choices. They emerge from a lifetime of experience with the world, with teddy bears, with hearing names, with watching movies, etc. It would be absurd to think one could explicitly program in all of the answers to all possible questions of this type, especially since the words are made up. And the list really is endless: "Is *Flugblogs* a good name for a startup computer company?" (Answer: ghastly) "Is *Flugblogs* a good name for air-filled bags that you tie on your feet and walk across swamps with?" (Pretty good!), etc., *ad infinitum*.

The same is true when we rate one concept as an instance of another (e.g., "Rate dry leaves as hiding places"); experience with the world is required. No dictionary definition will ever include the fact that piles of dry leaves in the autumn are marvelous places for little children to hide in, but who among us, especially those of us who have actually crawled under piles of dry autumn leaves as children, can help but make this association?

Using Subcognitive Questions to Unmask the Computer

How will the Interrogator use this type of question to unmask the computer?

She will go out into the population from which the human contestant in the upcoming Turing Test will be drawn. From that population she will select a fairly large, random sample of people and ask them all of the questions on her subcognitive question list and record their answers. The distribution of their answers will constitute the Human Subcognitive Profile. She will then come to the Turing Test armed with her list of Subcognitive Questions and the corresponding Human Subcognitive Profile. She will ask both candidates all of the questions on her list. The candidate statistically closest to the Human Subcognitive Profile will be the human. Given a large enough set of questions, she will always be able to unmask a computer that hadn't experienced life as we had.

The point of subcognitive questioning is that it probes our underlying set of associations built up over many years of experiencing the world. If the set of associations for one of the candidates is significantly different from the Human Subcognitive Profile, then that candidate will be eliminated. Now, of course, this is an advantage and a problem. The Turing Test that includes, as it must, subcognitive questions turns out to be so strong that it will correctly eliminate computers that haven't lived life like we have, but it will also eliminate *anything*, however, intelligent, that hasn't lived our culturally-oriented human life. For example, suppose that someone was born with eyes on his knees, but was, in every other respect, like a human being. This physical fact would make him Turing-test detectably different from the other (normal) human candidate. His profile would be skewed when it came to questions involving wearing long pants, falling off bicycles, skinning his knees in the playground, etc.

Of course, this also means that lots of perfectly intelligent people would also fail to pass the Test. For example, people from other cultures, with other lifestyles, with disabilities, etc. would fail the Test. But we must keep in mind that our only goal is to unmask the computer and so the Test's extreme difficulty and excessive anthropocentricity is, in some sense, irrelevant (even though one could certainly argue that this fact means that the Test not a particularly useful measure of general intelligence). Failing the Test is not a demonstration of anything, and certainly not a demonstration of non-intelligence. We are only interested in

exploring how difficult it would be for a machine to pass the Test, accepting that *if* it succeeded in passing the Test, it would be considered intelligent.

Physical differences can be brought to light with the standard Turing Test

In addition, the Interrogator will include certain questions that would not be part of the Human Subcognitive Profile. These are questions that will test physical aspects of the two candidates. The important thing is that they would probe for physical characteristics that are present but clearly *irrelevant* to cognition.

Here is an example of a question that would indirectly test for the physical attributes of the two candidates:

> "Please bring your two hands together, palms pressed together, as if you were praying, touching the fingertips of your left hand with the corresponding fingertips of your right hand. Fold down your two middle fingers — and only your two middle fingers — so that the middle knuckles of both come together. (The tips of your thumbs, index, ring and pinky fingers should still be together.) Now, move your other fingers one at a time and report what happens."

Now, of course, for a machine without hands like ours to know what would happen in this case would require a complete knowledge of human hand muscle positions, strengths and a theory of their movement. For people, on the other hand, all that is required is for them to bring their hands together and try it. Go ahead and try it: You cannot separate your ring fingers. This curious fact about human hand muscle structure that is revealed is a completely irrelevant quirk of our anatomy, of the way our musculature is put together. There would be no earthly reason to include this feature when building a robot hand, and, yet, if you didn't, your robot would be detectably different from the human candidate.

We can invent all sorts of these questions that are, in fact, nothing more than little experiments that test for bodily sensations, sensations that a computer that didn't have bodies essentially identical to ours would not be able to answer. Consider the following question:

> "Does holding a mouthful of Coca-Cola in your mouth feel more like:
>
> • having a cup of cold water poured on your head?
> • having someone slap your backside?
> • having your foot wake up after being asleep?
> • having a mouthful of cold soup in your mouth?

Everyone to whom I have posed this question agrees that it is most like having pins and needles in your foot. It is a question that they had never been asked before and had never thought about . . . but one that they were able to answer without hesitation because it involves previously experienced bodily sensations. How could a computer without a mouth like ours or without feet like ours — that occasionally fall asleep when we stay too long in the same position — possibly consistently answer this type of question correctly?

Another experiment: Since we have agreed to allow the Interrogator to communicate with the candidates by means of a computer monitor instead of a teletype, she could ask each candidate to place his/her/its nose at a location on the screen marked by an X. If a human candidate did this, the exact location of his eyes would then be known, which would mean that the exact location of the blind spots in each eye (i.e., the spot where the optic nerve connects to the eye) would be known. Small randomly chosen letters would be flashed briefly at various locations on the screen. *Some of these letters would be intentionally located at a spot exactly corresponding to a human being's blind spot*. The subjects would be asked to identify the letters that had been flashed on the computer screen. The human would not be able to identify those letters that fell exactly in his blind spot. But what about the computer? In designing a computer eye (or if there were no eye at all), there would be no reason whatsoever to include a blind spot, or more precisely a blind spot at the exact location where it was found in the human eye, when the only reason that humans have such a blind spot is because the optic nerve is connected to the eye at that point. Without a blind spot of its own,

the computer would have to have such a highly developed theory of the physionomy of the human eye that it could not only predict where the blind spot should be, but then would reply that it could not see the letters in that spot. Perhaps this is possible, but it would be a very strange design for a mechanical eye to intentionally include dysfunctional characteristics of the human eye that are nothing more than a by-product of how the human eye happens to be connected to the brain.

It can be seen from the above example that the computer screen can become the source of a wide variety of experiments that will allow the Interrogator to "perceive" many physical features of the candidates, much as a particle physicist indirectly infers facts about atoms and sub-atomic particles by bouncing other particles off of them. For example, we can imagine experiments where the Interrogator asks the candidates to hold their open hand vertically between their eyes (i.e., a vertical "salute" along the ridge of the nose), thereby forming a vertical screen between the left and right eye. Then the candidates are asked to move up towards the computer screen until their vertical hand on their nose comes in contact with the computer screen. Now the right eye will not see what the left eye sees and the Interrogator can conduct a series of experiments that will reveal the existence of a right-left brain structure in one of the candidates. Again, unless the computer was explicitly designed with a bi-partite brain structure, then the computer will be unmasked.

Building **an intelligent machine versus** ***testing*** **an intelligent machine**

In the literature on the Turing Test, one often sees a general confusion between the notion of *testing* an intelligent machine by means of the Turing Test and actually *building* such an intelligent machine. The two activities are often conflated and should not be. The symbolic level is, indeed, inappropriate and insufficient for building an intelligent machine. To reproduce all of the high-level facets of human cognition that emerge from a vast substrate of subcognitive interactions, a physical symbol system (Newell & Simon, 1976) would run into problems of combinatorial explosion. Put another way, the high-level cognitive phenomena that emerge from our human subcognitive substrate — choosing this or that dress, making this or that analogy, joke, or play on words, relating this situation to another, reacting to a TV advertisement, making this or that slip-of-the-tongue, etc. — are too numerous and complex to be modeled by symbols alone. On the other hand, *this does not imply that the symbolic level is inadequate for testing a machine*, since the symbolic level can be used to indirectly *probe* the lower levels of cognition. This "symbolic" probing will reveal differences in the two candidates far below the symbolic level. Put another way, symbols alone are most likely not powerful enough to build an intelligent machine; symbols alone are, however, powerful enough — in fact, they are ultimately, *too* powerful — to test for machine intelligence.

Conclusion

The main point of the present article is a very simple one, but one that is often overlooked in discussions of the Turing Test — namely, the extraordinarily powerful ability of the standard, SISO Turing Test to detect differences in the two candidates. This point has been made over the years by a number of authors (Dennett, 1985; French, 1988, 1990; Davidson, 1990; Crocker, 1994; etc.) but is rarely taken seriously enough. As a test of general intelligence, the Turing Test is not particularly appropriate precisely because it is so hard: it tests not for intelligence, in general, but, rather, for culturally oriented human intelligence. In order to pass it, a machine would have to experience the world in essentially the same manner as we humans had, and, in order to do this, it would have to have a body and a set of experiences very similar to our own. And it is this that would make it virtually impossible for any machine to actually pass the Turing Test. And this is why it makes no sense to raise the philosophical (or empirical) bar even higher, as Harnad (1989, 1991, 1994, 1999) has done

with his Total Turing Test (and the even harder levels T4 and T5) or as Schweizer (1998) has done with his Truly Total Turing Test. In short, the "symbols-in/symbols-out" level is easily hard enough, so hard that the chances of any machine actually passing it are vanishingly small.

Acknowledgments

This work was supported in part by a research grant PAI 4-19 from the Belgian government.

References

Crockett, L. (1994). *The Turing Test and the Frame Problem: AI's Mistaken Understanding of Intelligence*. Norwood, NJ: Ablex.

Davidson, D. (1990). Turing's test. In Karim A. Said et al. (eds.), *Modelling the Mind*. UK: Oxford University Press, pp. 1-11.

Dennett, D. (1985). Can machines think? In *How We Know*. (ed.) M. Shafto. San Francisco, CA: Harper & Row.

French, R. (1988). Subcognitive Probing: Hard Questions for the Turing Test. *Proceedings of the Tenth Annual Cognitive Science Society Conference*, Hillsdale, NJ: Lawrence Erlbaum. 361-367.

French, R. (1990). Subcognition and the limits of the Turing Test. *Mind*, 99(393):53-65.

French, R. (1996). The Inverted Turing Test: A simple (mindless) program that could pass it. *Psycoloquy*, 7(39).

Harnad, S. (1989). Minds, machines and Searle. *Journal of Experimental and Theoretical Artificial Intelligence*, (1):5-25.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, (42):335-346.

Harnad, S. (1991). Other bodies, other minds: A machine incarnation of an old philosophical problem. *Minds and Machines*, 1:43-54.

Harnad, S. (1994) Levels of functional equivalence in reverse bioengineering: The Darwinian Turing Test for Artificial Life. *Artificial Life* 1(3): 293-301.

Harnad, S. (1999). Turing on reverse-engineering the mind. *Journal of Logic, Language, and Information*. (to appear).

Newell, A. & Simon, H. (1976). Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the Association for Computing Machinery*, *19*, 113-126.

Schweizer, P. (1998). The Truly Total Turing Test. *Minds and Machines*, 8:263-272, 1998.

Searle, J. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, 3:417-424.

Turing, A. (1950). Computing machinery and intelligence. *Mind*, 59(236):433-460.