French, R. (2007). If it walks like a duck and quacks like a duck...The Turing Test, Intelligence and Consciousness. In P. Wilken, T. Bayne, A. Cleeremans (eds.). *Oxford Companion to Consciousness*, Oxford, UK: Oxford Univ. Press.

If it walks like a duck and quacks like a duck... The Turing Test, Intelligence and Consciousness

Robert M. French
LEAD-CNRS, Dijon, France
robert.french@u-bourgogne.fr

"If it walks like a duck and quacks like a duck, it probably is a duck." So goes a pithy observation, undoubtedly as old as English itself, that provides a quick-and-dirty way to recognize ducks. This little maxim constitutes an operational definition of a duck that, for better or for worse, sidesteps all the thorny issues associated with actually explicitly defining one (e.g., has feathers, can fly, weighs less than 10 pounds, has webbed feet, swims well, has nucleated red blood cells, has a four-chambered heart, has a flat bill, etc.). What folk wisdom did for ducks, Alan Turing did for intelligence. He was the first to suggest an operational definition of intelligence that has come to be called the Turing Test (Turing, 1950). The underlying idea of his Test is the same as our folk means of duck identification, viz., whatever *acts* sufficiently intelligent *is* intelligent. Translated into the vernacular of modern electronic communication, the Turing Test says that if, by means of an extended e-chat alone, you cannot tell whether you are chatting with a machine or a person, then whomever or whatever you are chatting with is intelligent.

Since it first appeared nearly six decades ago, Turing's article has become the single most cited article in artificial intelligence. In fact, few articles in any field have caused so much ink to flow. References to the Turing Test still appear regularly in artificial intelligence journals, philosophy journals, technical treatises, novels and the popular press. Type "Turing Test" into any internet search engine and there will be, literally, thousands of hits.

How the Turing Test Works

Turing's original description of his "Imitation Game" was somewhat more complicated than the simpler version that we describe below. However, there is essentially universal agreement that the additional complexity of the original version adds nothing of substance to its slightly simplified reformulation that we refer to today as the Turing Test.

The Turing Test starts by supposing that there are two rooms, one of which contains a person, the other a computer. Both rooms are linked by means of text-only communication to an Interrogator whose job it is, by means of questioning the entities in each room, to determine which room contains the computer and which the person. Any and all questions that can be transmitted via typed text are fair game. If after a lengthy session of no-holds-barred questions, the Interrogator cannot distinguish the computer from the person, then the computer is declared to be intelligent. It is important to note that *failing the Turing Test proves nothing*. It is designed to be a sufficient, but not necessary, condition for intelligence.

Commentary on the Turing Test: two lines of argument

There are two major approaches to commenting on the Turing Test. (See French, 2000, for a review.) The first, and by far the most frequent, set of commentaries on the Turing Test attempt to show that if a machine did, indeed, succeed in passing it, that this alone would not necessarily imply that the machine was intelligent. The second accepts the Turing Test as a sufficient condition for intelligence, but says that it would be impossibly hard for any machine to pass that was not embodied as we humans were and had not lived life as we had, and, therefore, the Test is of little value as a real test for intelligence.

The best known paper representing the first approach was written by John Searle (1980). It is a "translation" of Turing's original test into a subtle argument about someone who knows no Chinese and, yet, who can behave for all the world as if he understood Chinese. In Searle's so-called Chinese Room thought experiment, instead of having an Interrogator who must distinguish between an unseen person and an unseen computer based on a series of typed questions, Searle imagines a person who knows no Chinese in a room filled with vast stacks of rules, written in Chinese only, relating groups of Chinese characters to other groups of Chinese characters. The person in the room then receives a question written in Chinese from a Chinese person outside the room (corresponding to the "Interrogator" in the original Turing Test). The person in the room then searches through the stacks of rules and is finally able to construct a reply to the question that is perfectly satisfactory to the Chinese person who submitted it to him. However, even though his behavior (i.e., blindly producing strings of Chinese characters that constitute perfectly correct answers to the questions he has been asked) completely fools the Chinese Interrogator into believing that he actually understands Chinese, clearly he does not. The analogy with the Turing Test is clear: if behavior that perfectly imitates the comprehension of Chinese does not necessarily imply real comprehension of Chinese, then in a like manner, behavior that perfectly imitates general intelligence does not necessarily imply real intelligence.

However, the crucial issue that both Turing and Searle avoid entirely is just how (impossibly) hard it would be for either a computer to pass the Turing Test or for a person with no knowledge of Chinese to fool a Chinese person into believing he had native fluency in Chinese. Both Turing's and Searle's thought experiments assume as a premise the possibility of actually being able to do their respective tasks.

French (1990) takes issue with this premise. He challenges Turing's assumption that a disembodied machine that had not experienced the world as we humans had could ever actually pass the Turing Test. In French's argument the Interrogator relies on the vast web of "subcognitive" (i.e., unconscious) associations that we humans develop over the course of a lifetime of interacting with the world. He has the Interrogator go out prior to the start of the Turing Test and interview a large number of people. The Interrogator asks them questions that derive from their interactions with the world over the course of their lifetime, such as: On a scale of 1 to 10, rate kisses as medicine, billiard balls as Christmas ornaments, credit cards as banana peels, etc. He asks them questions that derive from the fact that humans have bodies that are designed in a very particular way, e.g., Is holding gulp of Coke in your mouth more like having pins and needles in your feet or having cold water poured on your head? The Interrogator then collects all the responses to these subcognitive questions and puts together a Human Subcognitive Response Profile (HSRP), which consists of the distribution of answers for each question. Then armed with the HSRP, he puts the same set of questions to the entities in both rooms. The entity whose set of answers is farthest from the Profile is the computer.

Crucially, all of the replies to these subcognitive questions are based, not on logic or reasoning, but on having a human body and on having experienced the world as we have. Not having either a human body or benefiting from human experience, any computer now, or in the foreseeable future, would have immense — arguably insurmountable — difficulties answering questions of this kind as humans do.

The Turing Test as a graded measure of intelligence and consciousness

French's point is that the Turing Test is not actually testing for (general) intelligence, but rather, intelligence in humans, with human bodies, having experienced life as a human being. The problem lies in its discrete pass/fail nature. Only machines that actually pass a no-

holds-barred Turing Test are said to be intelligent. But what if a machine *almost* passed the Test? Let us assume that only after an hour of intense questioning with subcognitive questions does the Interrogator begin to suspect that the entity in, say, Room 1 might be the computer. It then takes another full hour for the Interrogator to correctly identify the computer. Would we not be willing to grant that this machine has a higher degree of intelligence than one for which the same conclusion (i.e., correctly deciding that it was a machine) had been reached in a single minute of questioning? Most likely.

The idea then is that we could use the Turing Test as a way of providing a *graded* assessment, rather than an all-or-nothing decision, on the performance of the machine, once it is unmasked as being the computer. Thus, how far the machine's answers were from the HSRP would constitute a "measure of intelligence" with respect to our own intelligence.

In like manner, we could potentially adapt the Turing Test to provide a graded test for human consciousness. Instead of the full range of subcognitive questions, the Interrogator would focus on a subset of these questions that explicitly dealt with subjective perceptions, rather like the question about holding Coca-Cola in one's mouth. Then, as above, the Interrogator would draw up a Human Subcognitive "Qualia" Profile (HSQP) for the Turing Test. And then, exactly, as for the graded Turing Test for intelligence, how far the machine's answers were from the HSQP would constitute a "measure of consciousness" with respect to our own consciousness.

Of course, those in the consciousness community who suggest the logical possibility of (philosophical) zombies (e.g., Chalmers, 1996) would undoubtedly remain unconvinced by such a graded measure of consciousness. A philosophical zombie, after all, would, according to the Turing Test for consciousness, score exactly the same as a conscious human on the Turing Test described above. But for those who accept that consciousness admits of degrees with respect to full human consciousness, the Turing Test could be used to provide an approximate measure of the machine's consciousness as compared to human consciousness.

References

Chalmers, D. J., (1996). The Conscious Mind: In Search of a Fundamental Theory. Oxford, UK: Oxford University Press.

French, R. M. (1990). Subcognition and the Limits of the Turing Test. *Mind*, *99*(393), 53-65. Reprinted in: P. Millican & A. Clark (eds.). *Machines and Thought: The Legacy of Alan Turing* Oxford, UK: Clarendon Press, 1996 and in S. M. Shieber. *The Turing Test*. MIT Press, 2004.

French, R. M. (2000). The Turing Test: the first 50 years. *Trends in Cognitive Sciences*, *4*(3), 115-121.

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

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