French, R. (2009). 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. 461-463.

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 means of duck identification that, for better or for worse, sidesteps all the thorny issues associated with actually explicitly defining a set of features allowing us to identify a duck (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 means of identifying 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 (i.e., to be thinking). 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 have been numerous approaches to discussing the Turing Test. (See French, 2000; Pinyar et al., 2000; Schreiber 2004, for reviews.) 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 (e.g., Scriven, 1953; Gunderson, 1964; Purtill, 1971; and more recently, Searle, 1980; Block, 1981;

Copeland, 2000; who argue against the "behavioristic" – i.e., input/output only – nature of the Turing Test). Numerous authors (Millar, 1973; Moor, 1978; Dennett, 1985; Hofstadter & Dennett, 1981; etc.) argued, on the contrary, that passing the Test would, indeed, constitute a sufficient test for intelligence. Certain authors (Dennett, 1985; French, 1990; Harnad, 1991; etc.) also emphasized the enormous, if not insurmountable, difficulties a machine would have in actually passing the Test.

The input/output-only (I/O) nature of the Turing Test has been the basis of several important criticisms (see, in particular, Block, 1981; Copeland, 2000; and, the Chinese Room argument of Searle, 1980 and Harnad, 1991). Just as critics of behaviorism, arguably beginning with Tolman (1948), have repeatedly demonstrated that there is more to cognition than mere behavioristic I/O, the "behaviorism" critics of the Turing Test contend that there is more to human intelligence than could ever be elicited by an typewritten input-output exchange with a machine. Consequently, they claim, a machine might well pass the I/O-alone subset of human cognition that the Turing Test is capable of testing for, but, since that is only a subset of cognition, passing the Turing Test would be insufficient to demonstrate real human intelligence. In short, the Turing Test tests only for "I/O cognition," not full cognition, and, as such, is not testing for full human intelligence..

Various authors take another tack and emphasize the Test's extreme difficulties. French (1990), for example, takes issue with 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 ask a large number of randomly chosen people 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 and calculates the distribution of answers for each question. Then he puts the same set of questions to the entities in both rooms. The entity whose set of answers is farthest from those of people is the computer.

Crucially, all of the replies to these subcognitive questions are based, not on logic or reasoning or memorized facts, but, rather, 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. In other words, the Turing Test is not actually testing for (general) intelligence, but rather, a test for intelligence in humans, with human bodies, having experienced life as a human being.

The Turing Test as a graded measure of human intelligence and consciousness

The Turing Test is a discrete pass/fail test. Machines that pass a no-holds-barred Turing Test are said to be intelligent; as for those that do not, we withhold judgment. 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 even begin to suspect that the entity in, say, Room 1 might be the computer. It then takes another full hour for him 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 intelligence of the machine. Thus, the further the machine's answers were from average human answers, the less intelligent it would be.

In like manner, the Turing Test could potentially be adapted to provide a graded test for human consciousness. The Interrogator would draw up a list of subcognitive question that explicitly dealt with subjective perceptions, like the question about holding Coca-Cola in one's mouth, about sensations, about subjective perceptions of a wide range of etc.. As before, the Interrogator would pose these questions to a large sample of randomly chosen people . And then, as for the graded Turing Test for intelligence, the divergence of the computer's answers with respect to the average answers of the people in the random sample would constitute a "measure of human consciousness" with respect to our own consciousness. In short, the Turing Test, with an appropriately tailored set of questions, given first to a random sample of people, could be used to provide an operational means of assessing consciousness.

Acknowledgments

This work was supported in part by a grant (FP6-NEST-516542) from the European Commission.

References

Block, N. (1981) Psychologism and behaviourism. *Philosophical Review*, 90:5-43

Copeland, B. J. (2000) "The Turing Test" *Minds and Machines* 10, 519-39.

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

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.

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

Hofstadter, D. and Dennett, D. (1981) Reflections on 'Minds, Brains, and Programs'in *The Mind's I*. Basic Books. 373-382

Moor, J. (1976) An analysis of the Turing Test. *Philosophical Studies, 30*:249-257

Purtill, R., (1971) Beating the imitation game. *Mind*, 80:290-294, 1971

Scriven, M. (1953) The mechanical concept of mind. *Mind*, 62:230-240

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

Shieber, S. M., (2004).. *The Turing Test*. Cambridge, MA: MIT Press

Tolman, E. C. (1948). Cognitive maps in rats and men. *The Psychological Review, 55(4)*, 189-208.

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