The Turing Test is Useful but Non-Decisive

John Allard (1437547)
Phil 11 - Essay #2 - Section G
May 21th, 2015

The Turing Test is perhaps the oldest and certainly one of the most famous proposed tests for machine-intelligence. The test is set-up so that potential intelligent-machines attempt to trick a human into believing that the machine is also human over a certain testing time-span. The implications of this test, in the minds of some, is that if we could make a machine that could reliably trick a human often enough, it should be considered intelligent. While I believe that the Turing test is a useful and interesting thought-experiment, it is not a plausible criterion for intelligence because it is too susceptible to giving false-positive results. While there are situations where a Turing test could correctly identify machine-intelligence, its ability to be fooled by machines that most would consider to be non-intelligent renders the test non-decisive on the issue and thus not a plausible criterion for testing intelligence. In this paper I'll start by going over a famous thought-experiment meant to demonstrate the fragility of strong-AI theory and by extension the Turing test. I'll then go over why the test is susceptible to false-positives, and finish with a possible objection to my argument.

One of the most convincing (in my opinion) thought experiments that counters the idea that the Turing test is useful comes from John Searle, in his 1980 paper "*Minds, Brains, and Programs*". The paper is Searle's response to proponents of strong AI, those that believe the formal manipulation of symbols done by computer programs is actually a lower but similar form of intelligence to that of humans. To counter this, he introduces the Chinese room thought experiment, which abstracts the formal processes of a computer program into an example of an English speaker in a room who is manipulating Chinese symbols so correctly as to be indistinguishable from that of a native-speaker without actually knowing what any of the symbols mean. Here, the room can be

thought of as the computer, and since its ability to communicate is indistinguishable from that of a real, intelligent, Chinese speaker, the room would be able to pass a Turing test (a Chinese one). Searle concludes that "what the example suggests is that [formal symbol manipulations] by themselves have no interesting connection with understanding. They are certainly not sufficient conditions, and not the slightest reason has been given to suppose that they are necessary conditions or even that they make a significant contribution to understanding .." (Searle, pg. 4). I agree completely with Searle here, no-where in the room is any true understanding of the Chinese language contained, even though it can be replicated perfectly. The example points out a flaw in the Turing test, in changing the question from 'can machines think?' to 'can a machine play the imitation game (well)?', the notion of understanding and intention is removed from the equation. The example given by Searle is meant to elucidate that even if a machine could be programmed so well as to be indistinguishable from a human at certain tasks, as long as the internals rely only on pre-defined symbol manipulation there is no intent or understanding present, and by extension no intelligence.

Before continuing any further, I need to clarify but what I mean by intelligence. Intelligence to me is symbol manipulation with belief and intention, which in my opinion requires understanding. My definition would exclude the Chinese room from being considered intelligent because, as discussed in the last paragraph, it does not possess intention or understanding of the symbols it is manipulating. In addition to the arguments of others, I believe the Turing test does not qualify as a proper criterion for intelligence because it is too susceptible to giving false-positives. To be considered a proper test for some trait, the answers that the test gives should be relatively definitive. If there exists a way of consistently fooling the test into giving an incorrect result, then the test should not be considered useful as a measure of that trait. I believe this to be the problem

with the Turing test; although a truly intelligent machine should be able to pass the Turing test, much simpler machines that I would not consider intelligent could also pass. These simpler machines I will call dictionaries. They simply contain enough information to pass a Turing test over a given finite amount of time (like an hour or so), by having pre-programmed outputs that are looked up when requested. The problem is that, no matter how long or how difficult you make the interview with the computer to be, there is always the possibility of just adding more and more responses the computer's memory bank. With enough programming put into the computer, it would be able to pass a Turing test at least some of the time, with the frequency of passing only limited by the amount of work spent programming the computer. As pointed out by Ned Block is *Psychologism and Behaviorism*, all of these machines would be the equivalent of a toaster. They are only programmed to respond to abstract symbols that mean nothing in particular to the actual computer., thus regardless of their ability to pass ever such larger and more complex Turing tests, they never have any more understanding then the man in the room understood Chinese, or your watch understands time (Block). The computer in this fashion is simply parroting information directly, with no grasp on the layers of abstraction associated with the topics it is discussing during the test. The computer in this sense has literally became a parroter of information from the programmers, without actually grasping or having any understanding of the symbols or the relations that exist between them. I am of the opinion that a more formal test will be needed than the one set out by Turing. He thought the original question ('can machines think?') too difficult, so he rephrased it in terms of a game, and assumed that the original question was isometric with respect to this rephrasal. I think it is not, in changing the questions we lose the intricacies of what it means to actually understand something.

Some assert that "A plausible criterion for attributing intelligence to humans is their verbal behavior". While we certainly show our intelligence through verbal communication, there is nothing special about the chosen medium being sound. Communicating through writing, sign-language, or smoke-signals are also just as valid means of demonstrating intelligence as verbal communication. However, none of these succeed in demonstrating that the communicator actually possesses an understand of what they're communicating, so they are alone not enough to establish intelligence. As a possible rebuttal to my arguments, many in the AI field would argue that my distinction between the parroting of information and understanding it is a false one. That because the computer contains all of the information to appear wise and knowledgeable on a topic, this is no different from having true intelligence. I will counter that, at least in the case of programming responses in a dictionary like fashion, the capability to the system to parrot true intelligent behavior is not even analogous to true intelligence because the system has no understanding of the symbols it is manipulating. When a human states something about something, it can regress ever deeper in its knowledge about what is talking. We can espew meaning that is only implied between our chosen symbols, and our possession of an understanding of what is implied by our statements is what sets our intelligence apart from the pseudo-intelligence of a parroting machine. I do believe that if you could program the actual experiences and senses of topics and items that are known by intelligent creatures into the computer, than it would have opportunity to be truly intelligent because it could be said to have an understanding of what it is talking about. The Turing test however, does not accurately distinguish between machines that can understand and produce with intent and belief, and those that are simple dictionary-machines. It is because of this that I believe the Turing test is not an adequate test of machine-intelligence. (citations on back)

Searle, John. 'Minds, Brains, And Programs'. 1980: n. pag. 4 Print.

Block, N. 1981. Psychologism and behaviorism. Philosophical Review XC(1):3–50.