What does the Turing Test measure?

The Turing test is popular and attractive largely because of humankind's recognition that we are unique. While this is a somewhat narcissistic notion, it is not without merit. We are unique not because of any innate physical characteristics, other than our disproportionate large brains. These brains allow a flexible substrate for the acquisition of useful abilities. However, the ability that sets us most apart is our ability to make use of language. No other species even comes close to having the size of vocabulary that we humans make use of. This large vocabulary allows nuanced and complex communication. This complexity is leveraged in the Turing test to provide an environment of the utmost complexity to test a computer.

At first glance the Turing test seems to be primarily a test of our ability to create a computer that acts in a manner consistent with human behavior. While it does test this, human conversation is only useful insofar that it provides a complex environment. **What the Turing test is really measuring is the ability of an algorithm to act according to a set of complex rules**. The rules governing human conversation are complex and changing over time, but these rules do exist and can be known, even if they are not obvious to us, the developers of the rules.

Humans are a group-oriented species, and, as such, it is advantageous for humans to be able to communicate. A group of highly communicative individuals will out-compete a group of less communicative individuals and, over time, come to dominate the evolutionary tree. This process leads to the gradual development of more and more complex language, such as we use today. However, initially, primitive languages must have been very simple. Imagine, for example, a language called LSUeese. In this hypothetical language there are only three words, "*Grunt*", "*Mmmm",* and *"Blah". "Grunt*" would mean "Is that food good or bad?" And the only two existing replies could be "*Mmmm*" or "*Blah*" for "It's good" and "It's bad", respectively. Now suppose the Turing test is conduct in LSUeese. The judge would "*Grunt*" and the computer and confederate could then reply with either "*Mmmm"* or "*Blah*". What would it say about an algorithm if it could pass such a Turing test?

The limited complexity of such a language would make it trivially easy to trick the judge, so such a success would be insignificant. The only difference between an LSUeese Turing test and the modern English Turing test is the language used. So what is special about today's Turing test (other than that we haven't yet passed it)? Modern English is many orders of magnitude more complex than LSUeese, which lends the modern Turing test its difficulty. Modern English simply represents a complex, abstract, environment in which to test a computer's ability to act. Consider that the Turing test would likely be easier or harder to pass given what language it was conducted in. Given that, and an agreed upon algorithm, the Turing test could become an index for the complexity of a language - such that the easier it is for the judge to tell which player is the algorithm, the more complex a given language would be. According to this index, LSUeese would at the least-complex end of the spectrum.

Since its inception, progress has been made in the quest to pass the Turing test. However, this progress has been gradual. Given such a history of incremental progress, it is likely that progress will continue in this way. So, the year that computer success goes from 48% (not passing) to 50% (passing), will that algorithm suddenly be wildly more powerful than its predecessors? Also, how could any incremental improvements be made to an algorithm that already passes the Turing test? A new test would have to be devised.

Given these issues, does the Turing test measure intelligence? This does not seem to be the case. Rather, it measures some other capacity, which might be related to intelligence. However, is there any one measure of intelligence? Intelligence, as a psychological construct, is conceptualized as a latent variable, one that cannot be directly measure, only inferred from other measures. So, given that intelligence is unmeasurable, the Turing test is doomed from the start to measure intelligence. Is there any one word or concept that entirely captures what the Turing test measures? The safest definition of what the Turing test measures is that it measures is the ability of an algorithm to pass the Turing test. A more aggressive definition, given the sensitivity of the test to the complexity of the language it is performed in, is that it measures the ability of a computer to act in complex environments.

Intelligence cannot be evident unless an organism is in a complex environment. Complexity allows room for intelligent behavior. Take, for example, a bacterium living in your gut. Such an organism has sensors for the suitability of its present location, and, at best, directional sensors indicating the direction the greatest suitability gradient. The actions available to this bacterium are to either stay here or move (and possibly also increase or decrease certain biochemical pathways depending on its location). What test could we possibly administer this bacterium to get some idea of its intelligence, considering that we are limited by the environment observable to the bacterium? Such an organism has no opportunity for intelligent behavior because of the limitations of its environment.

Humans, on the other hand, live in a much more complex environment, and it is this complexity that allows for intelligent behavior. Consider the complexity of the environment of the Turing test. While there is surely great complexity in the physical world, that complexity is dwarfed by the complexity of the world of language, which is the inner world of our consciousness. In this world, the environment is not bounded by physical possibility, increasing the level of complexity many times over. It is this rich, seemingly limitless, inner world that is shared and developed during communication. Language allows us to explore this exceedingly complex environment. To some extent, the complexity of one's inner world is the product of all the complexity that that person has experienced in their entire life. Complexity is built upon existing complexity, in the same sense that learning is a cumulative process over an individual's entire life. Contrast this with the physical world, whose properties, at scales larger than the atomic scale, are well-defined and predictable. Therefore, language, because it operates in this environment of great complexity, is by far our most useful and powerful tool. However, it is not just our ability to wield this tool that makes us intelligent. Rather, we are intelligent, so we wield it. Just because a computer can wield language and navigate such a complex environment doesn't mean it is intelligent. It simply means that that computer can decide, in the face of gross complexity, upon the same action that an intelligent being would have decided upon.

Again, it is this property that is measured by the Turing test: How intelligently can a computer act in a complex environment, one which allows for acting intelligently. The important distinction is that it is a measure of *acting* intelligently, not *being* intelligent. The difference is that an intelligent being is capable of intelligent behavior, but is not defined by such behavior. If Alan Turing, prior to his death, was given a paralyzing serum such that he could no longer move or talk, would he be any less intelligent than he was prior to receiving the serum? Certainly not. Mr. Turing would be able to live just as rich an inner life as he did prior to receiving the serum. The only difference would be that he would be able to share it with anyone or augment his own insights with the insights of others.

While the Turing test is not a test of intelligence itself, it is a test of the ability of a machine to act under conditions of great complexity, and such a capacity should not be underestimated. Complexity is inherent in all great questions before they are properly understood, and computers will hopefully continue increase the ability to allow us to appreciate such complexities.