

Panel Communication

Some Misconceptions Regarding the Turing Test

Hugh Loebner

Abstract. Several misconceptions are presented regarding Alan Turing's "Imitation Game" These relate to (a) using other than three entities, (b) using an incorrect criterion for passing the test and (c) unnecessarily restricting the content of the test.

1 INTRODUCTION

I have directed or otherwise been associated with 19 annual Loebner Prize Competitions. During this time I have been the recipient of a multitude of communications regarding the Turing Test. Most of these excoriated the competition, detailing in what manners the Loebner Prize fails as a Turing Test. My association with the Loebner Prize and my reflections on these communications lead me to recognize that there are several misconceptions regarding the test, or, at least, a failure to appreciate salient characteristics.

The Turing Test is a paired comparison between a human and a computer program judged by a human. The Method of Paired Comparisons is a well studied research tool that permits the development of an interval scale and can test the transitivity of an underlying dimension.

2 IMPROPER NUMBER OF PARTICANTS

Turing wrote "It [the imitation game] is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex" [1] and "We now ask the question, "What will happen when a machine takes the part of A in this game?" Note that the judge is presented with precisely two entities which are to be compared against each other on how "human" they seem. In order to conduct a proper Turing Test one must be able to count to three. One must gather a judge, a human, a computer and then Halt. One could say that this is another type of "Halting Problem."

- One error occurs when the creator of a chatbot claims that his or her entry "has passed the Turing Test" if someone interacting with an unpaired chatbot mistakes it for a human. This is not passing the Turing Test because the chatbot was not chosen as the human when compared with a human. I would say this is the most frequent error made by the lay community.

Consider also the "CAPTCHA" device used by many websites to insure that a human and not a web bot is sending an email or otherwise using the site. CAPTCHA is an acronym for "Completely Automated Public Turing Test To

Tell Computers and Humans Apart"¹ However, it is not a Turing Test at all.

- A more subtle error happens when a judge is required to judge more than two entities without being told how many are human and how many are computers. This, in fact, was the design used in the Loebner Prize until 2004, and the contest was directed by some very eminent scholars during that period. The initial design of the Loebner Prize was under the direction of a distinguished committee headed by Daniel Dennett, and including, among others W.O. Quine, Joseph Weizenbaum, and I. Bernard Cohen. Nevertheless, the first contest, held in 1991 at The Computer Museum in Boston, MA, did not conform to Turing's prescription of paired comparisons.

Consider Stuart Shieber's critique of the 1991 Loebner Prize Contest [2]. Despite mentioning that the 1991 Competition did not present binary choices, Shieber's complaints revolved around limiting the "tenor" of the comments and limiting "topics" of content. (His main complaint, however, was that the competition was premature.)

When I (along with several others) judged the Loebner Prize in 2001 at The Science Museum in London, the same general design was used. At the time, even I did not realize that I was not participating in a Turing Test. In point of fact, the 2004 competition was the first Loebner Prize wherein the judges were actually faced with a paired comparison. It was only *after* the 2004 competition, while I was reflecting upon it, that it occurred to me that this was actually Turing's design. To be honest, I specified the paired comparison design not because it was the correct design for a Turing Test, but to facilitate holding the competition. In 2004 I held the competition in my apartment in New York City, in part to support my lawsuit against the Cambridge Center for Behavioral Studies². Rather than develop the necessary communications protocols, I required each contestant to enter his own programs to allow interactions between judge, human, and computer³. Rule 5 stated:

- (5) Each Computer Entry must be accompanied by a matching communications program.. The Com Program must (a) run as two programs on two

¹ <http://www.captcha.net/>

² Loebner v. Cambridge Center, et al., (03 CV 10959 RCL Fed. Dist. Ct., Mass.)

³ I thought: "Why should I have to do the work of developing a communication protocol?" It turns out the answer is: to prevent contestants from entering a communication program that garbles the human-judge interaction, which actually occurred in the 2004 competition.

computers (b) communicate character by character with each other via easily implemented communications (c) mimic the appearance and behavior of the matching Computer Entity. "Easily implemented communications" means directly two computers with a cable having standard plugs or connecting each to an Ethernet hub.

I do not claim that requiring one or more judges to select one or more computer programs from among more than two unknown entities is not a test of "intelligence." I am merely stating that it is not the Turing Test. Nor am I claiming that academicians do not know what is the Turing Test. I am just pointing out that (1) the general public does not understand the test and (2) that academicians, at least those who have been associated with the Loebner Prize, seem not to appreciate the central fact that the Turing Test is a Paired Comparison.

2 PASSING THE TEST

I believe that there are two other errors regarding the interpretation of the Turing Test, although these are open to different interpretations of the design.

One error involves what constitutes "passing" the test. Turing wrote "I believe that in about fifty years' time it will be possible, to programme computers, with a storage capacity of about 10^9 , to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning."⁴ This is sometimes interpreted to mean that mistaking a computer program for a human for five minutes constitutes passing the test. However, Turing did not write words to the effect "...and that would mean passing the test, or, explicitly, "I believe that in 50 years it will be possible to have a computer program pass the test." In my opinion, Turing was simply making a prediction about what would be the state of the art in 50 years, although he did continue "... I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted," which statement obscures the point. My personal belief is that Turing would have said that a true "thinking machine" should be able to fool the judge for any given length of time.

3 CONTENTS OF THE TEST

Another misinterpretation, I believe, is the view that the Turing Test should be restricted to textual material only. If we adhere to this limitation, we are faced with the limitation that

⁴ Turing uses 70% chance of success. In a paired comparison, we would expect that a "thinking" computer capable of perfectly imitating a human would only have a 50% chance of success. Also, there is some ambiguity about the five minutes of questioning. Are the five minutes to be divided between the two entities for a five minute test, or is each entity to be questioned for five minutes for a total test time of 10 minutes?

the Turing Test can only be played with a literate person. Illiterates need not apply.

I received a great deal of criticism for requiring that the "Grand Prize" winning program be able to deal with audio/visual input. However, in his article, Turing discusses the memory requirement for a successful program. He wrote "I should be surprised if more than 10^9 was required for satisfactory playing of the imitation game, at any rate against a blind man." Now, we may ask, why should the storage capacity of the computer depend on whether it were played with a blind or sighted person? If we permit visual images when played with the sighted person the answer becomes obvious. And, in the penultimate paragraph of his article, Turing writes:

"We may hope that machines will eventually compete with men in all purely intellectual fields....It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English. This process could follow the normal teaching of a child. Things would be pointed out and named, etc."

REFERENCES

- [1] Alan Turing. Computing machinery and intelligence *Mind*, 59:433-490, (1950)
- [2] Stuart M. Shieber, Lessons from a Restricted Turing Test, Communications of the Association for Computing Machinery, volume 37, number 6, pages 70-78, 1994, also available at <http://www.eecs.harvard.edu/shieber/Biblio/Papers/loebner-rev-html/loebner-rev-html.html>