The Total Turing Test and the Loebner Prize

David M. W. Powers
Department of Computer Science
The Flinders University of South Australia

powers@acm.org

Abstract

The Loebner Prize is the first, and only regular, competition based on the Turing Test, but in order to stage the competition various modifications to the original test have been made. In particular, the Grand Prize has a controversial and as yet undefined Audio-Visual condition attached to it. This paper discusses the value of the test with and without the A/V condition, and makes a proposal about what the general nature of the A/V test should be.

1. Introduction and Motivation

The Turing Test has always been controversial, and has been a continuing topic in AI journals and discussions since well before the Loebner Prize. We, however, will not pick up the story till after the first Loebner Prize competition was held, since this sparked a number of critical articles in both the general and the specialist media, including AI Magazine, The Economist, and multiple treatments in each of SIGART Bulletin (e.g. 3#4, 4#1, 4#4, 6#4) and CACM (Wilkes, 1992; Shieber, 1994; Loebner, 1994; .

As Harnad (in a SIGART Editorial and Commentary, 1992) puts it, Turing's insight is that we can discern the intelligence of a lifelong pen pal without ever meeting him, but the Eliza and Parry experience suggest that it is too easy to trick the computer, and Shapiro points out the importance of knowing that you may be talking to the computer. Our default 'charitable assumption' is that we are talking to 'a person like us'.

Even in a special issue on an apparently unrelated controversy, the connectionism/symbolism debate, the issues relating to the representation of knowledge merge in with the debate about the nature of intellegence and the Turing Test (Powers, 1993). The issue of Symbol Grounding is the bridge, and Harnad in particular has devoted considerable attention to this in bothof these forums and elsewhere (e.g. 1987, 1989, 1990, 1991, 1992), and it is also the issue you I wish to focus on here. I will therefore ignore the questions as to whether the Turing Test promotes dishonesty or the Loebner Prize promotes tricks, and focus on the issue of whether robot-like interaction with the world is necessary to pass the Turing Test, and how the Loebner Prize should implement the Audio-Visual component of the Grand Prize which Loebner added in the face of the criticisms that the Turing Test was too shallow (Shieber, 1994; Loebner, 1994).

2. Symbol Ground and the TTT

The problem with Natural Language Processing today is that it tends to amount to translation between formalisms without understanding. The original criticisms of the Loebner Prize, and earlier the Turing Test, focus on the fact that Eliza-style 'parrying' involves no understanding but can still fool the unitiated for short periods of time in highly constrained circumstances (e.g. where they have been led to believe the program is a doctor or a patient).

The converse of this is the claim that nothing better than this kind of translation and reflection, albeit in increasingly sophisticated forms, can be achieved by a computer that has no connection with reality. Even an encyclopaedia has pictures, and a dictionary assumes a basic vocabulary and understanding based on a lowest common denominator form of basic human experience and language capability. Harnad (1990) calls this the symbol grounding problem, and claims that it will not be possible to pass an unrestricted Turing Test without symbol grounding - that is our symbols, or words, need to have some connection to a sensory-motor experience of the world. Thus a computer system capable of passing the Turing Test would have to be more like a robot capable of passing the Total Turing Test, in which behaviour in interaction with the world must be the same.

This is the reason that Loebner (1994) felt constrained to place an Audio-Visual condition on the Grand Prize. The very first winner of the annual Loebner Prize, fooled half the judges in a restricted Turing Test (and a carefully chosen domain, "whimsical conversation"), but fell far short of exhibiting intelligence or of providing a convincing performance to those familiar with Eliza and the standard tricks. There was no evidence of understanding, and according to the symbol grounding principle there was not possibility of understanding. One way to ensure that systems are grounded is to move to a version of the Total Turing Test by adding in a Audio-Visual requirement which has not yet been fully defined, and has served to create another controversy.

David M. W. Powers (1998) **The Total Turing Test and the Loebner Prize.** In D.M.W. Powers (ed.) *NeMLaP3/CoNLL98 Workshop on Human Computer Conversation,* ACL, pp 279-280.

There are two main reasons for this: one is that not all agree that grounding is necessary, and the other is that it is no longer the Turing Test as defined by Turing. On the other hand we have a solid hypothesis. If Harnad is right, the first program to pass the TT should be able to pass the TTT. If Harnad is wrong, there are likely to be some years between the passing of the TT and the passing of the TTT.

3. Gold, Silver and Bronze Medals

To address this, Loebner has agreed to provide three levels of prizes, initially $100,000, $25,000 and $2,000, and corresponding Gold, Silver and Bronze Medals for the Grand Prize Winner, Turing Test Winner and Annual Prize Winners, respectively. The Silver Medal will be awarded for the straight Turing Test, whilst the Gold Medal and Grand Prize are for a version of the Total Turing Test.

Note that the prize is now awarded by a panel of judges of which 50% are experts in AI - which will mean Eliza-like tricks will not suffice to win the Silver Medal. The A/V version of the test will be run separately, with only programs capable of fooling 50% of the judges being eligible. The remainder of this paper discusses a proposal for the A/V condition to be used for the Grand Prize competition.

4. Show and Tell

Let's go back to kindergarten for a moment. Remember when you had to bring in a favorite object, show it to your classmates, and talk about what was so special about it. This is the basic concept.

Now let's consider some possible Loebner Prize entries. Some may have robot arms and cameras; some may have speech recognition and synthesis capabilities along with a broader range of auditory capabilities.

Now let's consider some possible Loebner Prize confederates. Some intelligent people have disabilities and lack arms or vision; some may be deaf or have no musical ability, but they will have a range of other senses open to them.

The aim of the Loebner Prize for Artificial Intelligence is not to test entrants' auditory or visual acuity, but rather their ability to take in information from and interact with their environment, to deal with it in an intelligent way, and to discuss it in a way that is indistinguishable from a human participant — allowing for the fact that some humans may be deaf or blind.

Part of the purpose of the Loebner Prize is to provide an understanding of what it is that distinguishes a human from a computer, natural intelligence from artificial intelligence. Another aspect is to understand what it is that leads us to decide that a person, or a computer, is intelligent, and to wean ourselves away from the superficial accidental features which might distract from this focus on intelligence. For this reason, the current organizers of the Loebner Prize competition aim to include a confederate or a judge with some kind of disability each year. We aim for it to be representative of a broader class of people, both in the general population and in the scientific community. If the computer can't see, does that mean that a blind person may be confused more easily with a computer, or that a computer may be more easily confused with a blind person?

Giving the participant an object allows them to make use of whatever sensory-motor capabilities they have, and shouldn't bias against any particular disabilities. If you hand a deaf-person a musical instrument, or a CD and CD-player, they can tell you they are deaf, but still appreciate the objects per se and demonstrate their intelligence in terms of their understanding of the use or function of the object(s).

5. Discussion

This paper is intended to lead to discussion of this and other possibilities for the implementation of the Loebner Prize A/V condition. We have shown that it is closely related to the Total Turing Test, and that there is a significant empirical question to be answered in relation to whether it is or is not effectively a stronger test than the standard Turing Test.

6. References

Editorial (1992), Artificial Stupidity, *The Economist* **324**#770:14.

Editorial+Commentary (1994) SIGART Bulletin **3**#4:7-11

Epstein, R. (1992), The Quest for the Thinking Computer, AI Magazine **13**#2:80-95.

Harnad, S. (1987, ed.) *Categorical Perception*, CUP

Harnad, S. (1989) Minds, Machines & Searle, *JETAI* **1**:5-25

Harnad, S. (1990) The Symbol Grounding Problem, *Physica D* **42**:335-346

Harnad, S. (1991) Other Bodies, Other Minds, *Minds and Machines* **1**:43-54

Harnad, S. (1992) Connecting Object to Symbol in Modeling Cognition, in Clarke A. & Lutz R. (Eds) *Connectionism in Context,* Springer-Verlag

Loebner, H.G. (1994) Response to Lessons from a Restricted Turing Test, *CACM* **37**#6:79-82

Powers, D.M.W. (1993) Special Issue on Connectionism versus Symbolism, *THINK* **2**:1

Shieber, S.M. (1994) Lessons from a Restricted Turing Test, *CACM* **37**#6:70-78

Turing, A.M. (1950) Computing Machinery and Intelligence, *Mind*, **59**#236:433-460

Wilkes, M.V. (1992) Artificial Intelligence as the Year 2000 Approaches