that we must jump over even higher, and still does not help us assess whether or not we are making progress toward it.

Instead of asking the question "Can machines think?," let us ask "Can machines be intelligent?" That, I believe, is the question that artificial intelligence is more interested in. To answer that, we need a definition of intelligence that is applicable to machines as well as humans or even dogs. Further, it would be helpful to have a relative measure of intelligence, that would enable us to judge one program more or less intelligent than another, rather than identify some absolute criterion. Then it will be possible to assess whether progress is being made in finding an answer to the question.

In fact, I believe such a definition and test are within our grasp. The late Allen Newell, in his book *Unified Theories of Cognition,* proposed a definition for intelligence:

> A system is *intelligent* to the degree that it approximates a knowledge-level system...
>
> Thus, intelligence is the ability to bring to bear all the knowledge that one has in the service of one's goals. To describe a system at the knowledge level is to presume that it will use the knowledge it has to reach its goal. Pure knowledge-level creatures cannot be graded by intelligence—they do what they know and they can do no more, that is, no better. But real creatures have difficulties bringing all their knowledge to bear, and intelligence describes how well they can do that. ([4], p. 90).

This definition leaves questions unanswered. What is a goal? How do we know that a creature has knowledge? We are beginning to reach an understanding of these questions as well. Furthermore, this definition can provide the basis for a relative measure of intelligence.

Unfortunately, this definition does not completely dispense with philosophical conundrums such as Searle's Chinese Room. It does make them less compelling, though. We can usually observe directly whether or not an agent is achieving its goals; we cannot directly observe whether the agent is thinking. Furthermore, if Harnad is right, and we must ultimately embody intelligence in a robot in order to prove the existence of artificial intelligence, a good understanding of intelligence will still be necessary for that demonstration.

Perhaps out of the work of Newell and others a precise definition of intelligence can be attained, and with it a means to measure it. That will help clear up public misconceptions about what artificial intelligence is really about. Additionally, by making AI's goals clearer, we may be better able to achieve those goals, and thus perhaps conduct research more intelligently.

Acknowledgements

I wish to thank Alan Frisch, David Powers, and Pat Hayes for constructive comments on an earlier version of this editorial.

References

[1] "Artificial Stupidity," *The Economist*, vol. 324, no. 7770, August 1, 1992

[2] R. Epstein, "The Quest for the Thinking Computer," *AI Magazine*, vol. 13 no. 2, pp. 81–95, Summer 1992.

[3] S. Harnad, "Minds, Machines, and Searle," *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 1, no. 1, 1989.

[4] A. Newell, *Unified Theories of Cognition*, Harvard University Press, 1990.

[5] J.R. Searle, "Minds, Brains, and Programs," *Behavioral and Brain Sciences* vol. 3, pp. 417–424, 1980.

[6] Special Section on Intelligent Cognitive Architectures, *SIGART Bulletin*, vol. 2 no. 4, pp. 12–184, August 1991.

[7] A.M. Turing, "Computing Machinery and Intelligence," *Mind* vol. 59, Oxford University Press, pp. 433–460, 1950.

[8] M.V. Wilkes, "Artificial Intelligence as the Year 2000 Approaches," *Communications of the ACM*, vol. 35, no. 8, pp. 17–20, August 1992.

The Turing Test Is Not A Trick: Turing Indistinguishability Is A Scientific Criterion

Stevan Harnad
Cognitive Science Laboratory
221 Nassau Street
Princeton University
Princeton NJ 08544
harnad@princeton.edu

It is important to understand that the Turing Test (TT) is not, nor was it intended to be, a trick; how well one can fool someone is not a measure of scientific progress. The TT is an empirical criterion: It sets AI's empirical goal to be to generate human scale performance capacity. This goal will be met when the candidate's performance is totally indistinguishable from a human's. Until then, the TT simply represents what it is that AI must endeavor eventually to accomplish scientifically.

Pen-Pals Versus Robots

In my own papers I have tried to explain how trickery, deception and impersonation have nothing at all to do with the scientific import of Turing's criterion [2,4]. AI is not a party game. The game was just a metaphor. The real point of the TT is that if we had a pen-pal whom we had corresponded with for a lifetime, we would never need to have seen him to infer that he had a mind. So if a machine pen-pal could do the same thing, it would be arbitrary to deny it had a mind just because it was a machine. That's all there is to it!

This entirely valid methodological point of Turing's is based on the "other minds" problem (the problem of how I can know that anyone else but me actually has a mind, actually thinks, actually has intelligence or knowledge—these all come to the same thing): It is arbitrary to ask for more from a machine than I ask from a person, just because it's a machine (especially since no one knows yet what either a person or a machine *really* is). So if the pen-pal TT is enough to allow us to correctly infer that a real person has a mind, then it must by the same token be enough to allow us to make the same inference about a computer, given that the two are totally indistinguishable to us (not just for a 5-minute party trick or an annual contest, but, in principle, for a lifetime). Neither the appearance of the candidate nor any facts about biology play any role in my judgment about my human pen pal, so there is no reason the same should not be true of my TT-indistinguishable machine pen-pal.

Now, although I too am critical of the TT, I think it is important that its logic—which was only implicit in Turing's actual writing —should be made explicit, as I have tried to make it here and in my other writings, so we can see clearly the methodological basis for his proposed criterion. Elsewhere I have gone on to take issue with the TT on the basis of the fact that humans also happen to have a good deal more performance capacity over and above their pen-pal capacity. It is hence arbitrary and equivocal to focus only on pen-pal capacity; but Turing's basic intuition is still correct that the only available basis for inferring a mind is Turing-indistinguishable performance capacity. For *total* performance indistinguishability, however, one needs *total*, not partial, per-

formance capacity, and that happens to call for all of our robotic performance capacities too: the Total Turing Test (TTT). And, as a bonus, the robotic capacities can be used to *ground* the pen-pal (symbolic) capacities, thereby solving the "symbol grounding problem" [3], which afflicts the pen-pal version of the TT, but not the robotic TTT.[1]

In fact, one of the reasons no computer has yet passed the TT may be that even successful TT capacity has to draw upon robotic capacity. A TT computer pen-pal alone could not even tell you the color of the flower you had enclosed with its birthday letter—or indeed that you had enclosed a flower at all, unless you mention it in your letter. An infinity of possible interactions with the real world, interactions of which each of us is capable, is completely missing from the TT (and again, "tricks" have nothing to do with it).

Is the Total Turing Test Total Enough?

Note that all talk about "percentages" in judging TT performance is just numerology. Designing a machine to exhibit 100% Turing indistinguishable performance capacity is an empirical goal, like designing a plane with the capacity to fly. Nothing short of the TTT or "total" flight, respectively, meets the goal. For once we recognize that Turing-indistinguishable performance capacity is our mandate, the Totality criterion comes with the territory. Subtotal "toy" efforts are interesting only insofar as they contain the means to scale up to life-size. A "plane" that can only fall, jump, or taxi on the ground is no plane at all; and gliding is pertinent only if it can scale up to autonomous flight.

The Loebner Prize Competition is accordingly trivial from a scientific standpoint. The scientific point is not to fool some judges, some of the time, but to design a candidate that *really* has indistinguishable performance capacities (respectively, pen-pal performance [TT] or pen-pal + robotic performance [TTT]); indistinguishable to any judge, and for a lifetime, just as yours and mine are. No tricks! The real thing!

The only open questions are (1) whether there is more than one way to design a candidate to pass the TTT, and if so, (2) do we then need a stronger test, the TTTT (neuromolecular indistinguishability), to pick out the one with the mind? My guess is that the constraints on the TTT are tight enough, being roughly the same ones that guided the Blind Watchmaker who designed us (evolutionary adaptations—survival and reproduction —are largely performance matters; Darwinian selection can no more read minds than we can).

Let me close with the suggestion that the problem under discussion is not one of definition. You don't have to be able to define intelligence (knowledge, understanding) in order to see that people have it and today's machines don't. Nor do you need a definition to see that once you can no longer tell them apart, you will no longer have any basis for denying of one what you affirm of the other.

1. In a nutshell, the symbol grounding problem can be stated as follows: Computers manipulate meaningless symbols that are systematically *interpretable* as meaning something. The problem is that the interpretations are not intrinsic to the symbol manipulating system; they are made by the mind of the external interpreter (as when I interpret the letters from my TT pen-pal as meaningful messages). This leads to an infinite regress if we try to assume that what goes on in *my* mind is just symbol manipulation too, because the thoughts in my mind do not mean what they mean merely because they are interpretable by someone *else's* mind: Their meanings are intrinsic. One possible solution would be to ground the meanings of a system's symbols in the system's capacity to discriminate, identify, and manipulate the objects that the symbols are interpretable as standing for [1], in other words, to ground its symbolic capacities in its robotic capacities. Grounding symbol-manipulating capacities in object-manipulating capacities is not just a matter of attaching the latest transducer/effector technologies to a computer, however. Hybrid systems may need to make extensive use of analog components and perhaps also neural nets, in order to connect symbols to their objects [5, 6].

References

[1] Harnad, S., (ed.) *Categorical Perception: The Groundwork of Cognition*. New York: Cambridge University Press. 1987.

[2] Harnad, S., "Minds, Machines and Searle." *Journal of Theoretical and Experimental Artificial Intelligence* 1: 5-25. 1989.

[3] Harnad, S., "The Symbol Grounding Problem." *Physica D* 42: 335-346. 1990.

[4] Harnad, S., "Other bodies, Other minds: A machine incarnation of an old philosophical problem." *Minds and Machines* 1: 43-54. 1991.

[5] Harnad, S., Hanson, S.J. & Lubin, J., "Categorical Perception and the Evolution of Supervised Learning in Neural Nets." In: *Working Papers of the AAAI Spring Symposium on Machine Learning of Natural Language and Ontology* (DW Powers & L Reeker, Eds.) pp. 65-74. Presented at Symposium on Symbol Grounding: Problems and Practice, Stanford University, March 1991; also reprinted as Document D91-09, Deutsches Forschungszentrum für Küenstliche Intelligenz GmbH, Kaiserslautern FRG. 1991.

[6] Harnad, S., "Connecting Object to Symbol in Modeling Cognition." In: A. Clarke and R. Lutz (Eds) *Connectionism in Context*, Springer Verlag. 1992.

The Turing Test and The Economist

Stuart C. Shapiro
Department of Computer Science
and Center for Cognitive Science
SUNY at Buffalo
226 Bell Hall
Buffalo, NY 14260 U.S.A.
(716) 645-3935; FAX (716) 645-3464
shapiro@cs.buffalo.edu

The Turing Test

I have always thought the Turing Test a perfectly good test for AI, and I want to say briefly why.

First of all, I am assuming the version of the Turing Test in which the Interrogator interacts with entities A and B, one of which is a human, and the other of which is a computer, and the Interrogator must guess which is which. Furthermore, I have always assumed that the Interrogator knows these facts, an assumption met neither in the cases where Eliza nor where Parry were said to have passed the Turing Test. I have also assumed a reasonable amount of time—not a mere five minutes—and that it would not be significant to the outcome that the computer intentionally be programmed to make mistakes so that it not look "too smart." (If interrogators guess correctly, and attribute their success to the computer's making fewer mistakes than the human, then certainly the computer has demonstrated thinking, intelligence, etc.)

I believe the above assumptions to be at least in the spirit of what Turing specified for his imitation game. Perhaps they are so different from what Turing said explicitly that you will conclude that this is no longer the Turing Test, but a replacement, but I don't think so.

What I may have missed in my thinking about the Turing Test is that sufficiently unsophisticated Interrogators might provide such a weak test that "obviously unintelligent" programs might