

On Dr. Loebner's Lessons

Stuart M. Shieber

April 8, 1994

How Dr. Loebner spends his money in his private life is his own business. But when he spends it in a public way, when he leverages it with the scarce resources of public funding agencies like the National Science Foundation and private foundations like the Sloan Foundation and the Computer Museum, when he solicits industrial co-sponsors like IBM, Digital Equipment Corporation, and GDE Systems, when he calls in members of the national press as onlookers (and, in the most recent third Loebner competition held on December 8, 1993, as the actual judges), when this public use of his funds has potential deleterious effects on the research community,¹ it becomes everyone's business. With our special expertise, computer scientists have both the right and the responsibility to make our opinions known on matters that affect the research environment. What Dr. Loebner does with these opinions is, once again, his own business.

Consequently, I will take this opportunity to comment on Dr. Loebner's response. Dr. Loebner discusses three topics relating to his prize: the genesis of the prize, his goals in setting it up, and the lessons he himself has learned as they impinge on the future structure of the competition. I discuss each in turn.

History: The history of the prize, though interesting, and in many ways revelatory, is irrelevant to the issues I raise in my paper.

Goals: I find myself in general support of the goals that Dr. Loebner saw for the prize that bears his name, at the level of abstraction at which he presents them. I support the promotion of a better public understanding of the pertinent issues, the advancement of science, and so forth. However, these are goals that even the founder of the patently ill-conceived fictional da Vinci competition could have espoused. Seen in that light, it becomes clear that lofty goals are not at issue here either, but whether the goals are well served by the Loebner Prize.

Dr. Loebner believes that the competition has bred understanding of the Turing test, but in fact it has bred only more media hype. He believes that it will advance the state of

¹Recently, former chair of the Loebner Prize committee Daniel Dennett has stated that "Unless serious competition can be attracted, continuing the contests will not only do no good; it will do positive harm both to the reputation of the Loebner Prize (with both the public and the AI community) and to the public perception of AI, and serious competition will not—and indeed should not—be attracted to the competition until such time as winning it will be a mark of genuine depth of model, genuine progress, in AI." (personal communication) According to Dr. Dennett, he and others on the committee have resigned their positions because of Dr. Loebner's refusal to acknowledge their recommendations.

artificial intelligence research. His own competition shows that he is wrong. In the third running of the Loebner competition, Mr. Weintraub won again, with the same program that won the first two competitions, using the same technology with which Professor Weizenbaum fooled his secretary almost 30 years before. He believes that the competition might have valuable unexpected spinoffs.² Whether this use of funds would be more apt to generate useful spinoffs than some other use (which is the crucial issue) seems highly dubious. Again, the history of the competition itself tells the story.

Dr. Loebner says that if the working hypothesis on which the Loebner Prize was predicated was right, the tremendous “potential social energy” of the field of AI made it overripe for the running of a Turing test, that we were but a breath of wind away from building artificially intelligent machines, and that he had turned on the fan. The experience of the last few competitions shows that he was simply wrong.

Lessons: Dr. Loebner, surprisingly, considers my article a full discussion of the lessons we have learned from the Loebner prize to date. The lessons he gleans from the article, however, seem to be different from the ones I had hoped to present.

Loebner seems to be primarily concerned with minutiae: issues of intellectual property, verification of entrants, requiring of audio and video input, whether to transcribe the interaction at the server end of the conversation or the client end.³ All of these are trivial compared with the basic problem with the structure of the test itself.

Dr. Loebner would now like to “improve” the contest by running a competition unrestricted in topic and tenor. I did argue that a restricted Turing test is meaningless, but if a restricted one is premature, it is hardly imaginable that an unrestricted one is more appropriate. Furthermore, such a modification does not change the fact that the Turing test is not meaningful as a ranking method for comparing things for which is “most human-like in verbal behavior.” (Imagine using a Turing test to force a ranking of, say, a random character generator and a generator of random-length sequences of the letter ‘e’. Which is more “human-like?” Current programs are, of course, much more sophisticated, but comparing them in this way is no more meaningful.) The Turing test, if it makes sense at all, makes sense as a binary determination of “indistinguishable from a human in verbal behavior,” a condition that Turing viewed as a reasonable behavioral replacement for “intelligent.”

I agree with Dr. Loebner on one thing: that there is nobility in the attempt to understand intelligence and thereby some day build an artificial intellect. We disagree only on the right path to take and how far a trip it is.

²For the record, by the way, Dr. Loebner could have chosen a better example in this vein than the story about Mozart’s fundament. The great composer’s discomfort did not owe to the world’s inadequate understanding of stress and strain, but his decision to travel by hard-seat mail coach, rather than by a more expensive class of transportation. Within a few miles of setting out for Vienna, he “upgraded.”

³This last bit of trivia is not only incidental, but Dr. Loebner seems to have chosen the wrong side of the issue on software-engineering grounds: Each server (entrant) runs different code, but all the clients (terminals for judges) run the same code.