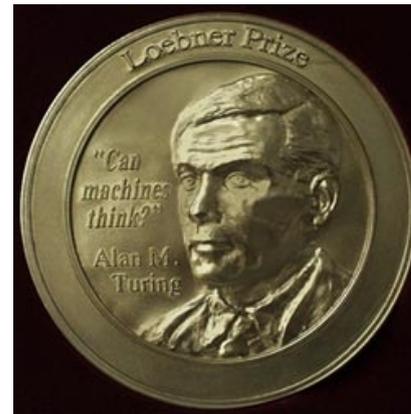


The Loebner Prize



- \$100,000 cash award + gold medal to first machine that can pass a “full” Turing Test
- \$25,000 cash award + silver medal for “limited” TT
- \$3000 + bronze medal annual prize for best program
- Judges question program “entrants” and human “confederates” for 5 minutes each
- Total of 7 rounds
- Next contest: September 6, 2009, Brighton, England

Sample Questions

- General questions
 - What time is it?
 - What round is this?
 - Is it morning, noon, or night?
 - What would I use a hammer for?
 - Of what use is a taxi?
- Questions relating to relationships
 - Which is larger, a grape or a grapefruit?
 - Which is faster, a train or a plane?
 - John is older than Mary, and Mary is older than Sarah. Which of them is the oldest?

Sample Questions

- Questions demonstrating memory
 - I have a friend named Harry who likes to play tennis.
 - *(...a few unrelated intervening questions or statements...)*
 - What is the name of the friend I just told you about?
 - Do you know what game Harry likes to play?
- Transcripts of 2005 contest

History of the Loebner Prize

- Reference: “Lessons from a restricted Turing Test”, by Stuart M. Shieber, *Communications of the ACM*, June 1994.
- First contest held November 8, 1991 at the Boston Computer Museum
- Dr. Hugh Loebner
 - PhD in sociology
 - contest organizer and financial backer
 - NY theater equipment manufacturer

History of the Loebner Prize

- Organizing committee
 - Daniel C. Dennett, chair (philosophy, Tufts)
 - Robert Epstein (Cambridge Center for Behavioral Studies)
 - Harry Lewis (CS, Harvard)
 - W. v. O. Quine (philosophy, Harvard)
 - Joseph Weizenbaum (CS, MIT, ELIZA creator)
 - Allen Newell, advisor (CS, CMU)
 - and several others

History of the Loebner Prize

- 6 computer contestants
- 4 human “confederates”
- 10 human judges
 - answered a public newspaper ad
 - interviewed by the organizing committee
 - no special expertise in CS (to avoid excessive judicial sophistication)

History of the Loebner Prize

- First contest included 2 special rules:
 - Conversational topics were strictly limited
 - burgundy wines
 - small talk
 - whimsical conversation
 - dissatisfaction in relationships
 - Shakespeare
 - women's clothing
 - Tenor of conversations was limited to “natural conversation without trickery or guile” (the types of questions you might ask a stranger on an airplane)

Scoring and Results

- Judges rank-ordered the agents from least human-like to most human-like
- Drew a line separating computers from humans
- Told there were least 2 computers, 2 humans

Rank order of the terminals									
Least human-like	1	2	3	4	5	6	7	8	Most human-like
	<i>B</i>	<i>A</i>	<i>E</i>	<i>D</i>	<i>C</i>	<i>F</i>	<i>H</i>	<i>G</i>	

- Technical problems with computer connections
- Time allocated to each agent reduced from 15 minutes to 7 minutes
- Some judges mistook some programs for humans (and vice versa), but on average all humans ranked highest

Scoring and Results

- All 6 programs were terrible
- Winning program, written by Joseph Weintraub, was a very simple variant of ELIZA
- AI community decided the Loebner contest was premature and made the field look bad
- Generated misleading publicity
 - Weintraub took out an ad claiming his program was the “first to pass the Turing Test”
 - “Computer fools half of human panel”
(*San Jose Mercury News*)
 - “Test a breakthrough in artificial intelligence”
(*Boston Herald*)

Criticisms

- Contest has no clear goal
- Most successful prizes have involved technologies that were just over the horizon, whereas we are much farther away from Turing-Test-passable AI
- Contest limitations undermined the spirit of TT
- Example: which of the following questions would be “appropriate” to ask a stranger on an airplane?
 - Are there any zoos in Washington?
 - Is Washington bigger than a breadbasket?
 - Is there much crime in Washington?
 - Are there any dogs in Washington?
 - Are there many dogs in Washington?

Criticisms

- Contest has no clear goal
- Most successful prizes have involved technologies that were just over the horizon, whereas we are much farther away from Turing-Test-passable AI
- Contest limitations undermined the spirit of TT
- Example: which of the following questions would be “appropriate” to ask a stranger on an airplane?
 - Are there many marmosets in Washington?
 - Are there many marmosets in the Washington zoo?
- Dennett tried to get Loebner to change the rules, but got sick of dealing with him and resigned from the panel
- Marvin Minsky posted the “Minsky prize” on `comp.ai`