A paradox related to the Turing Test

Samuel A. Alexander*
Department of Mathematics, the Ohio State University

Published in *The Reasoner* vol. 5, issue 6, p. 90.

I will describe a paradox which arises assuming it is possible to distinguish machines from non-machines. In the "fly on the wall" version of the Turing Test, player A passively observes the dialog of players B and C. Player A's goal is to determine whether B is a machine and whether C is a machine. For simplicity, remove C from the game. Let A observe B as B recites a monologue, and let A try to determine whether or not that recitation is computable.

I further modify the Test as follows. Player A guesses the nature of B after every new line. The Test runs forever and A wins if his guesses are *eventually* always correct: he is allowed finitely many wrong guesses. This is justified because every finite string is computable, so no finite speech can rule out a machine; any finite enunciation can be canned. Only an infinite one has a chance of non-computability. Further, the Test is run by an Operator who delivers B's lines to A.

Suppose Player B is a human trying to appear human and that A can distinguish machines from non-machines. Without further caveats, A will eventually detect B is non-machine and A's guesses will converge to the correct answer.

But suppose the Operator is mischievous. If A most recently guessed that B is non-machine, the Operator will lie and tell A that B said "Wait," storing what B really said. Only if A incorrectly guessed B was a machine does the Operator let the real monologue go on. What will happen to A's guesses?

When A believes B is a machine, the Operator presents the correct lines from B. Eventually, A will realize from these lines that B is non-machine, and will correct himself. This causes the operator to begin lying, and as far as A knows, B begins saying "Wait" repeatedly. A Turing Machine can produce any finite speech followed by "Wait" forever, so A will eventually think B is a machine. This process continues, causing A to change his mind infinitely often.

But no machine can generate the lines A sees: if they are computable, they remain computable with all "Wait"s removed, meaning B's genuine lines are computable, contradicting that they're supposed to be distinguishable from a machine. The lines which A is told are not mechanical and A eventually realizes so, and stops changing his mind. This contradicts the previous paragraph.

This is a special case of a more general paper. In Alexander (2011: On Guessing Whether a Sequence has a Certain Property, http://www.cs.uwaterloo.ca/journals/JIS/VOL14/Alexander/alex2.pdf, J. of Integer Sequences, 1–11) I show that a set S of sequences of naturals is "guessable" (in a sense like the above) if and only if S can be defined in a $\forall x\,\exists y$ way and also in an $\exists x\,\forall y$ way. If S is the set of Turing computable sequences, then S can be defined in an $\exists x\,\forall y$ way: $f:\mathbb{N}\to\mathbb{N}$ is computable iff $\exists x\,\forall y\ f(y)=\phi_x(y)$. But S cannot be defined in any $\forall x\,\exists y$ way, so S is nonguessable.

*Email: alexander@math.ohio-state.edu