Trust and Decision Making in Turing's Imitation Game

A

Huma Shah
Coventry University, UK

Kevin Warwick
Coventry University, UK

INTRODUCTION

Trust can be defined as "confidence in relying on another person" and is the basis for "sharing new ideas with others" (Chua et al., 2012). Turing's imitation game provides a platform for human and machine interlocutors to share knowledge and opinions through text-based communication, but more so it can "make oneself vulnerable" (ibid). This is because human participants open-up asking and answering questions, which can lead them to trust naïvely.

The susceptibility of human interrogators is one of the reasons why Turing's imitation game is frequently dismissed as an unsuitable criterion for machine success (Hayes and Ford, 1995). It is also considered a bad idea (McDermott, 2010), and in need of updating for the 21st century (AISB, 2012). Being able to convince a human interrogator that you are human is viewed as too weak a benchmark and "highly game-able" thus a stronger test for machine intelligence is advocated (AAAI, 2015). Alternative notions to Turing's skip around and fail to address what the imitation game, commonly known as the Turing test, actually is. Turing too, in his scholarship on intelligent machinery, bypassed definitions, so whether machines could think or not, he described 'thinking' as a "sort of buzzing" in his head (Turing, 1952: p. 667). Turing did warn that the concept of *intelligence* was an emotional rather than a mathematical one (Shah, 2014; Turing, 1948). The emotional context of human-machine interaction is betrayed through trusting an unseen interlocutor in text-based conversation that they are like *oneself,* another human.

In this chapter we present a study giving the reader an opportunity to examine trust in decision-making by humans reading a transcript of a conversation between a human interrogator questioning a hidden machine and hidden human in parallel. We begin with Turing's idea showing his imitation game is a simple and implementable scientific experiment. We contend the imitation game is a widely applicable method to compare machine performance against a human's. In the human language imitation game, the interaction between human and machine is conducted in interview style through the prism of the latter's capacity to answer any questions in a satisfactory and sustained manner. Additionally the test provides a means to examine the decision-making process, in natural language exchanges, and why a human bestows trust on a stranger.

BACKGROUND

Analyses and opinions on the imitation game's salience have varied (see Shah & Warwick, 2015; Shah, 2013; Shah, 2011; Shah & Warwick, 2010). Turing evolved his ideas on an imitation game posing an interview in which a human interrogator questions a hidden entity to determine whether it is human or machine (Turing 1950; Turing 1952). This was Turing's *viva voce* test (Shah, 2010; Turing, 1950). The 'standard Turing test' is accepted

DOI: 10.4018/978-1-5225-2255-3.ch023

as involving *a human interrogator simultaneously questioning two hidden entities at the same time* (Stanford Encyclopedia of Philosophy, 2011). Designing an experiment to implement both of Turing tests requires setting parameters interpreting Turing's description. These include:

- Adequate duration for a test;
- Number of interrogators; and
- Style of interrogation.

An evaluation is necessary of what it means *exactly* for a machine to *pass as human*: what are the implications of any pass beyond the test? Can it be used to raise awareness of human susceptibility to deception and safeguarding trust in cyberspace interactions?

In the next section the authors present Turing's scholarship on the imitation game.

Turing's Question-Answer Test

Turing derived his natural language test for a machine from a chess game that he first introduced in his 1947 lecture on 'The ACE machine' to the London Mathematical Society (Shah, 2013). In his 1948 paper 'Intelligent Machinery' Turing advanced the possibility of a machine learning from experience and competing against humans in chess. His reason for developing the imitation game, beyond chess to conversational question and answers, was the belief that language learning was one of the most accomplished of human feats (Shah, 2011). In 1948 Turing described a "little experiment" with three participants, A, B, and C, playing chess:

1. A and C are humans located in different rooms;
2. A and C are poor chess players; and
3. B is a machine operated by a mathematician.

Player *C* was invoked to play both *A* and *B*. Turing felt *C* may find it difficult to say which they are playing. In this early version of the imitation game Turing did not say what *C* should be told about the hidden players *A* and *B*: whether *C* should be informed that between *A*, *B* one is the machine and the other is a human (or both *A* and *B* are machine, or both are human). Turing set the ground for a game based on hidden interlocutors answering questions from a human interrogator who cannot see or hear them (Turing, 1950). It should be noted here that Turing was not advocating a machine to simply imitate a human; he was putting forward the idea that it was possible to build machines to answer any question put to it if the machine were designed with a sufficiently sophisticated programme (Turing, 1950).

By the end of Turing's 1950 paper *Computing Machinery and Intelligence* Turing's quest to examine machine thinking could be executed in two different ways:

1. A 3-participant game in which a human interrogator questions two hidden entities *simultaneously* and determines which is human and which is machine based on their respective answers (see Figure 1), or
2. A 2-participant *viva voce* game in which a human interrogator questions one hidden entity and determines whether it is human or machine based on responses received (see Figure 2).

In 1952 Turing detailed his imitation game further elaborating his two participant *viva voce* test (Shah, 2013; Shah, 2011). Table 1 compares the *simultaneous comparison* and *viva voce* tests, both exploring a machine's intellectual capacity to engage in human-like dialogue (Shah, 2010).

The essential features in both Turing's scenarios are:

1. The questions must be put in typewritten form to ensure *fair play* to the machine so that it is not judged on "tone of voice" or "beauty" (Turing, 1950: p. 434);
2. The questions be unrestricted: the interrogator can ask any question, "introduce almost