Abstract

In *MIND,* Victoria University of Manchester’s quarterly review of psychology and philosophy, Alan Turing considers the following question: “Can machines think?” (Turing, 13). In defending the standard for computer intelligence Turing set himself in developing his ubiquitous test, known commonly as the “Turing Test”, Turing makes the claim that the method by which machines “think” or the method by which machines appear to be “conscious”, or exhibit “a diversity of behavior” altogether should not burden our considerations concerning their legitimacy (Turing, 13). Our inclination to a phenomenological argument, that elements of human experience are unique and unreproducible in virtue of some special component to the human not shared in other animals and not found elsewhere in nature, seems to spoil our ability to observe these experiences rationally, and as Turing put it, can often place us in the “solipsist” point of view—believing that we are capable of some things because we actively experience these things but denying that others can because we do not experience them, and cannot conceive of how they could (Turing, 11). Turing breathes humility into this discussion by suggesting that what matters in these considerations is ultimately what is observable to all of us as human beings. This established, certain tests could be designed, based solely on observable qualities, to determine computer intelligence, and consciousness. With this in mind, it is the goal this paper to repurpose Turing’s method for the use of providing a suitable argumentative framework to defend a theory of varying levels of personhood for artificial agents of varying capabilities. In reviewing relevant debates concerning the requisites of personhood through history and its application to non-persons and non-human entities I hope to identify some common driving forces present in those arguments and develop a unified theory of a spectrum of personhood.

Cited:

Turing, Alan M. "Computing Machinery and Intelligence." In *The*

*Philosophy of Artificial Intelligence.* Margaret A. Boden, ed. New York:

Oxford Press, 1990.

On July 9th, 2015 Selmer Bringsjord, chair of New York’s Rensselaer Polytechnic Institute’s cognitive science department, assembled three small *NAO* robots in her robotics lab, each similarly programed and each tasked with solving a puzzle which at its core required an element of self-awareness to solve. The ubiquitous wise men puzzle was being simulated among these robots, each one given the information that they had received a one of two kinds of pills, a “dumbing” pill, which would prevent them from speaking, or a placebo, which would have no effect (techxplore). In reality, the effects of these “pills” would be triggered by pressing a button atop the robot’s heads which when triggered would prompt the beginning of the puzzle--by having the robots believe they were given the one of the two pills. From here, the robots are asked verbally to deduce which pill they received. The simulation begins quietly as each robot attempts to respond to the question. With the other two robots unable to respond the “placebo” robot gets up and reports that it does not know what pill it received, but then, after hearing its own voice it retracts its earlier response and states: “sorry, I know now. I was able to prove that I was not given the dumbing pill.” Not many AI have been capable of passing this test, which relies not only on the agent’s ability to think critically and inductively but also relies on the agent’s ability to recognize itself distinctly from its environment. On the subject of this project’s significance, Bringsjord states that “we…[are] talking about a logical and a mathematical correlate to self-consciousness” (techxplore).

is my goal to develop this framework by extending its uses to the field of programming and Artificial Intelligence. Specifically, this paper aims to provide a suitable argumentative framework to defend a claim to varying levels of personal rights for artificial agents of varying capabilities. In reviewing relevant debates concerning the requisites of personhood through history and its application to non-persons and non-human entities I hope to identify some common driving forces present in those arguments and develop a unified theory of a spectrum of personhood.