

BOOK REVIEW

Bernardo Gonçalves, *The Turing Test Argument*

**New York: Routledge, 2023. Pp. 238. ISBN 978-1-032-29157-4.
£130.00 (hardback).**

Harry Law 

Cambridge University

The Turing Test Argument seeks to make sense of Alan Turing's 'imitation game', a well-known but frequently misunderstood scripture in the canon of artificial intelligence (AI). The book satisfyingly delivers on its primary objectives: to take to task popular interpretations of the Turing test and make the case for an alternative theory grounded in the intellectual climate in which the imitation game was conceived. In doing so, Gonçalves knits together existing Turing scholarship with new historical work to cleave apart the intention behind, and reception of, an essential experiment in the history of science.

The book begins by proposing that Turing developed the imitation game to substantiate what Gonçalves describes as the 'Turing test argument', the notion that machines are capable of manifesting a sort of thinking that may allow them to pass for humans under certain conditions (p. 1). The significance of this claim, the author argues, is that it provides the conditions under which the boundaries of human-like intelligence (called think¹) might be redrawn amidst contact with an alternative model of cognition (called think²) used by machines. The point of the Turing test was therefore to show that synthetic intelligence, while different in nature to the human variety, may in fact contain the potential for an equally valuable type of intelligence. It is for this reason that the author proposes that the test can be understood as a form of 'propaganda' designed to force Turing's intellectual opponents into recognizing the weaknesses of their arguments (p. 6).

To substantiate this claim, Chapter 2 summarizes the reception of the Turing test since its introduction in 1950. It sorts early respondents into two camps: AI practitioners who saw the test as a legitimate (and soluble) target of the emerging programme of AI research, and philosophers who stressed its neo-behaviourist credentials. Next, Gonçalves considers questions related to minds, machines and consciousness, describing how Ernst Nagel and James Newman's *Gödel's Proof* (1958) influenced pockets of philosophers to believe that 'Gödel's theorems had established that the power of the human mind exceeds that of machines' (p. 31). By the 1980s, the chapter argues, the charge of behaviourism drove John Searle to formulate his popular 'Chinese room' thought experiment based on a faulty reading of the imitation game that took it as a test to determine the mental states of machines. Our potted history of the Turing test's reception concludes with the game's rejection by AI researchers and its defence by philosophers from the 1990s to the 2010s.

Chapter 3, which centres Turing's principle of imitation, begins by dressing down two popular conceptions of the test. The 'reductionist' view posits that the imitation game was developed to measure intelligence. In this telling, intelligence can be directly measured by

the test, and therefore the test itself is a legitimate target of AI research. The ‘constructionist’ interpretation foregrounds the belief that the test constitutes a certain type of intelligence through its design and implementation. This account proposes that the test actively shapes our understanding of intelligence rather than passively measuring it. Describing both approaches as ‘cartoons’, the chapter proposes that the true goal of the imitation game was to establish a new logic of machine intelligence by introducing a task through which learning machines could imitate the human brain (p. 56).

Gonçalves injects a dose of much-needed historical context into debates surrounding the Turing test in Chapter 4. Perhaps the book’s most important contribution, this portion situates the paper within the ‘controversy’ surrounding Turing’s debates with physicist Douglas Hartree, chemist and philosopher Michael Polanyi, and neurosurgeon Geoffrey Jefferson (p. 81). It proposes that the test’s focus on learning and adaptability sought to counter Hartree’s view of computers as calculation engines, that tasks like the composition of poetry aimed to address Jefferson’s demands for creative abilities, and that the introduction of open-ended conversation (rather than the use of rule-based games like chess) was designed to confront Polanyi’s concerns that human knowledge could not be formalized by machines. In an original contribution to the extensive Turing scholarship, Gonçalves tackles Jefferson’s view that hormones were crucial for producing facets of behaviour that machines could not replicate. In doing so, the author proposes that the gender imitation element of the original game aimed to contest the perspective that certain modes of behaviour were dependent on physiological conditions.

Chapters 5 and 6 recover the nature of the imitation game as a thought experiment. Gonçalves begins by proposing that the Turing test corresponds with Ernst Mach’s basic method for thought experiments by centring the continuous variation of experimental conditions, before arguing that the game ‘satisfies’ Popper’s conception of the heuristic uses of imaginary experiments (p. 126). This line of analysis continues as the author introduces an unlikely figure in the form of Galileo Galilei. The book deals with the ‘rhetoric of Turing’s argument’ by drawing parallels between moves by Turing to construct an experiment for his interlocutors and Galileo’s description of an idealized fall in a void (p. 128). With one eye on the rhetorical currency of experimentation, the author draws the section to a close by reminding us that ‘not all experiments are meant to be performed’ (p. 155).

The book’s final chapter of analysis engages the circulation of, and critical response to, Turing’s views about the nature of thinking machines. It produces three images of Turing: as what Jefferson called a ‘scientific Shelley’ closely linked with utopian ideals, a gentle humourist of childlike manners, and an irresponsible scientist whose work risked nothing less than the enfeeblement of humanity. In a single-page conclusion, the author invites us to grapple with both the ‘exaggerated’ and the ‘naive’ perspectives on the imitation game, proposing instead that the test ought to be remembered as a container for Turing’s maverick views about machine intelligence (p. 187). A rich and tightly written little book, *The Turing Test Argument* succeeds in its central aim: to recover the human element in an essential moment in the history of machine intelligence.