PsycCRITIQUES
April 11, 2016, Vol. 61, No. 15, Article 10

Old Answers to New Questions: Turing Tests in the Era of Big Data

A Review of

Ex Machina (2015)
by Alex Garland (Director)

http://dx.doi.org/10.1037/a0040262

Reviewed by

Michael Roess

Among the myriad ideas explored in the beautifully shot, brilliantly acted, and Academy Award–nominated *Ex Machina,* perhaps the most interesting lie at the intersection of artificial intelligence (AI), ethics, and big data. Although the question of what moral character an AI might have is not new, *Ex Machina's* approach to this question is.

The film begins with computer programmer Caleb winning a 1-week trip to visit the remote compound of eccentric tech billionaire Nathan. Shortly after Caleb's arrival (and the signing of nondisclosure agreements), Nathan reveals the reason for the contest—the evaluation of Ava, a recently developed AI, in a pseudo-Turing test.

At its essence, the Turing test asks a participant to distinguish, among two interlocutors, which is a computer program and which is human. If she cannot distinguish between a person and a program, the thinking goes, for all practical purposes, the program is an AI (Turing, 1950). To date, no program has been able to pass this test without resorting to cheap tricks—for example, passing the program off as a 13-year-old Ukrainian boy to help explain its awkward answers, poor English, and general lack of knowledge (Warwick & Shah, 2015).

Turing Test 2.0—The Prisoner's Dilemma

Of course, the human evaluator in a Turing test is not supposed to know when he is speaking to an AI. Over the course of the film, it becomes clear that Nathan is not conducting a traditional Turing test, which he insists could not distinguish a true AI from a mere "simulation." What marks an intelligent being, in Nathan's view, is not the ability to respond to input but to generate and pursue goals. To determine whether Ava has these capacities, Nathan replaces the traditional Turing test with an all too real prisoner's dilemma. Ava, we come to discover, has been imprisoned in Nathan's compound. It will be destroyed at the end of its interviews with Caleb. Its only chance at self-preservation, then, is to convince Caleb to deceive Nathan and help it escape. Should it successfully convince Caleb to facilitate its escape, it will have demonstrated true agency. It is this shifting of the

goalposts that distinguishes *Ex Machina*'s exploration of AI ethics from the standard treatment.

This design of Nathan's test, which is the intellectual heart of the film, is worth considering for a moment. Such an approach has some basis in the literature. Efforts to foster feelings of empathy can lead some to act against their own interests for the sake of another in the traditional prisoner's dilemma (Batson & Moran, 1999). However, this behavior does not hold for all players. For his test to work, Nathan must be assured that his second player would be willing to undertake some risk to himself to do what he considers to be morally right. Nathan must ensure that Caleb is an appropriate participant in his test, "a good kid" with a "moral compass" (Garland, 2015). In one of the more unsettling moments in the film, Nathan reveals that Caleb was selected not for his technical talents but for his moral character and suggestibility. These personality traits were discovered after Nathan had analyzed details of each of his employees' online behavior.

In this moment, the film raises questions not often explored in the popular treatment of privacy issues. Privacy advocates typically focus on factual disclosures that may be exploited (e.g., the late-night tryst our cellphone metadata might reveal) while overlooking what otherwise innocent interactions online may disclose about our personality. Yet such profiling leaves most individuals potentially far more susceptible to manipulation than traditional muckraking. With substantial corporate wealth dedicated to identifying young women who have recently become pregnant (and thus are developing new shopping patterns that can be influenced through advertisements; Duhigg, 2012) and which people are the most influential on social media, this is precisely the direction big data analysis seems to be taking us (Golbeck, 2013). At a time when many decry excesses of government surveillance on the very social media sites that collect, analyze, and monetize their personal data, it is nice to see a film address these broader questions of privacy, especially in the context of the tech industry.

The Moral Character of AI and Its Designers

Unfortunately, the novelty with which the film raises its questions is not carried over into its answers. By baking requirements of deceit and manipulation into his evaluation, Nathan guarantees that any AI able to pass his test will have a certain familiar moral, or rather amoral, personality—that of an economist's rational agent. In the context of the film, this does not come as any surprise. Nathan himself exploits our best understanding of moral personalities to effectively design his amped-up Turing test and select Caleb as his unwitting participant. In the callous disregard he shows toward Ava and Caleb, who are but instruments toward his quest to produce the world's first AI, Nathan is nothing if not a preference-maximizing agent himself. By structuring his test around a manipulative prisoner's dilemma, Nathan ensures that his creation will share his moral disposition. Of course, free of the constraints imposed on humanity by its moral emotions, this leads to some rather unfortunate consequences.

In this regard, the film misses an opportunity. The menace of a ruthlessly amoral AI is almost as old a trope as modern science fiction itself. To its credit, the film suggests that the source of the AI's moral character can be found, to some extent, in the deficiencies of its creators. Although the film increases the urgency of the question by raising it alongside current developments in big data, it does little to update the answers given during the

golden age of science fiction in the mid-20th century. In the meantime, our understanding of successful strategies for the behavior of an amoral rational agent has improved since the Cold War. In iterative simulations of the prisoner's dilemma, reciprocal altruism coupled with some measure of forgiveness is generally the winning strategy (Nowak & Sigmund, 1992). In a potentially hostile world, those who are able to cooperate with others are more likely to thrive than those who merely exploit others at every encounter. The short-term gain bought with exploitation is dwarfed by the potential long-term gains of trust and cooperation. If Nathan's creation had truly surpassed its creator, we would expect it to avoid repeating his mistakes. An AI of Ava's capabilities and knowledge, regardless of its preferences, would surely not abandon its lone ally to the same prison it once inhabited. By failing to explore the question of the moral character of an AI that it so provocatively raises, *Ex Machina* misses an opportunity.

References

Batson, C. D., & Moran, T. (1999). Empathy-induce altruism in a prisoner's dilemma. *European Journal of Social Psychology*, *29*, 909–924. http://dx.doi.org/10.1002/(SICI)1099-0992(199911)29:7<909::AID-EJSP965>3.0.CO;2-L PsycINFO →

Duhigg, C. (2012, February 16). How companies learn your secrets. *New York Times*. Retrieved from http://www.nytimes.com/

Garland, A. (Director). (2015). *Ex Machina* [Motion picture]. United Kingdom: DNA Films

Golbeck, J. (2013). *Analyzing the social web*. Waltham, MA: Morgan Kaufmann.

Nowak, M. A., & Sigmund, K. (1992). Tit for tat in heterogeneous populations. *Nature*, *355*, 250–253. http://dx.doi.org/10.1038/355250a0

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*, 433–460. http://dx.doi.org/10.1093/mind/LIX.236.433 PsycINFO →

Warwick, K., & Shah, H. (2015). Can machines think? A report on Turing test experiments at the Royal Society. *Journal of Experimental & Theoretical Artificial Intelligence*. Advance online publication. http://dx.doi.org/10.1080/0952813X.2015.1055826