Imagination machines, Dartmouth-based Turing tests, & a potted history of responses

Chen, Melvin

2018

Chen, M. (2020). Imagination machines, Dartmouth-based Turing tests, & a potted history of responses. AI and Society, 35(1), 283-287. doi:10.1007/s00146-018-0855-3

https://hdl.handle.net/10356/137308

https://doi.org/10.1007/s00146-018-0855-3

Imagination Machines, Dartmouth-based Turing Tests, & A Potted History of Responses

Do androids dream? That is the question that Philip K. Dick's protagonist Rick Deckard asks himself. Human beings (unless they are aphantasic) are able to conjure up mental images of the sheep that they count before sleeping. Can machines or programs imagine, daydream, and dream? Mahadevan (2018) proposes that we are at the cusp of imagination science, one of whose primary concerns will be the design of imagination machines.[1]

Programs have been written that are capable of generating jokes (Kim Binsted's JAPE), producing line-drawings that have been exhibited at such galleries as the Tate (Harold Cohen's AARON), composing music in several styles reminiscent of such greats as Vivaldi and Mozart (David Cope's Emmy), proving geometry theorems (Herb Gelernter's IBM program), and inducing quantitative laws from empirical data (Pat Langley, Gary Bradshaw, Jan Zytkow, and Herbert Simon's BACON). In recent years, Dartmouth has been hosting Turing Tests in creativity in three categories: short stories, sonnets, and dance music DJ sets.[2] This question no longer seems as far-fetched as it might have been to Philip K. Dick's 1968 readership: are we entering the age of the imagination machine? In this article, I will provide a potted history of some responses to these imagination machines and their implications for our understanding of the imagination (creative or otherwise).

Before articulating the various responses to imagination machines, we must first engage with two problems that relate to creativity and the creative uses we make of our

[1] According to Mahadevan, this imagination science will be concerned in addition with the extension of data science beyond its current realm of learning probability distributions from samples.

[2] This should come as no surprise to us: Dartmouth College could well lay claim to being the spiritual home of AI. The term 'AI' had originally been coined to distinguish the subject matter of a Dartmouth-based summer 1956 research workshop from automata studies and cybernetic research. This Dartmouth-based 'Summer Research Project on Artificial Intelligence' workshop would in turn set the agenda for AI research and usher in the classical AI era.

imagination. The first is the definitional problem: how do we define creativity as a mental state and the creative uses that we make of our imagination? The second is the attributional problem: how do we attribute the mental state of creativity or the creative use of the imagination to putatively minded entities other than ourselves? When we encounter the creative output of human beings, it is relatively uncontroversial in these natural instances for us to make a causal inference in favour of the existence of an underlying mechanism for this creativity (typically involving creative uses of the imagination) in these individuals.

One makes the inference that these human beings have minds much like one's own, notwithstanding the problem (otherwise known as the problem of other minds) of how one can be certain or justified that anyone else apart from oneself has a mind, thinks, and makes creative uses of the imagination. Therefore, if the creative output (viz. Shakespearean or Petrarchan sonnets, limericks, children's stories, music deliberately imitating the style of a human composer, improvised music with a human performer) produced by the machine is totally indistinguishable from the creative output produced by human beings, and if we infer that the human producers of creative output in natural instances make creative uses of their imagination, then by the same token we must be allowed to make the same inference about machines in the Dartmouth-hosted Turing Tests for creativity.[3]

The Dartmouth-hosted Turing Tests for creativity may be thought of as setting the (empirical) goal of AI as generating human-scale creative performance (or creative output-generating) capacity, in the same way that the Turing Test for conversational intelligence sets the (empirical) goal of AI as generating human-scale verbal performance capacity. When the

---

[3] The Dartmouth-based Turing Tests for creativity stipulate that machines have to generate category-specific creative output: Shakespearean or Petrarchan sonnets for Poetix, limericks for LimeriX, children's stories for DigiKidLit, music in the style of a human composer for Style or Free Composition, and improvised music with a human performer for Improvisation. The creative output of these machines will be mixed among human creative output in the same category and human judges will be asked to label each creative output as generated by human or machine. Any machine-produced creative output that is indistinguishable from the human-produced creative output in the same category would have passed the Turing Test for creativity.

creative performance of the machine is totally indistinguishable from that of a human being's, the (empirical) goal will be met. To ask for a rigorous definition of mental state of creativity and the creative uses of the imagination is to miss the entire point of these Turing Tests, which address the attributional problem rather than the definitional problem. We do not need a definition of creativity or the creative uses of the imagination to see that once we can no longer tell machine-produced creative output apart from human-produced creative output, we will not have a basis for denying of one (viz. the imagination machine) what we typically affirm of the other (viz. creative human individuals).[4] With this methodological approach to creativity in mind, we are better positioned to trace the potted history of some responses to these imagination machines.

The Machines "R" Us Response. If mechanical systems are capable of these creative feats of invention and discovery on the artistic, musical, mathematical, and scientific fronts, then perhaps human beings are no more than elaborate bits of clockwork (at least on the creative imaginative front). This idea is at least as old as Thomas Hobbes, who had already declared in his Introduction to *Leviathan* (1651, p. 5):

> Nature, the art whereby God hath made and governs the world, is by the *art* of man, as in many other things, so in this also imitated, that it can make an artificial animal. For seeing life is but a motion of limbs, the beginning whereof is in some principal part within; why may we not say, that all *automata* (engines that move themselves by springs and wheels as doth a watch) have an artificial life? For what is the *heart*, but a *spring*; and the *nerves*, but so many *strings*; and the joints, but so many *wheels*, giving motion to the whole body, such as was intended by the artificer?[5]

---

[4] Harnad (1992) makes a similar case in his defense of the Turing Test as an empirical criterion that addresses the attributional rather than the definitional problem of (conversational) intelligence. I have merely extended his argument to the Dartmouth-based Turing Tests for creativity. In short, the governing idea is this: however creativity might (ultimately) be defined, if the creative output of machines and the creative output of humans are indistinguishable by human experts, then we cannot attribute the mental states of creativity and creative uses of the imagination to humans while refraining from attributing these same mental states to machines.

[5] Given these (and other) intellectual inclinations, George Dyson (1997, p. 7) has seen fit to characterize Hobbes as the 'patriarch of artificial intelligence'. One should note that this metaphor of the mechanical body (with automatons and clocks serving as the exemplary machines for Hobbes) is introduced in *Leviathan* with the intention of shedding light on the notion of the body politic.

Can computers do everything the human mind can? This is the question that Penrose poses in *The Emperor's New Mind* (1989). Although Penrose replies in the negative, one might hold that JAPE, AARON, Emmy, and BACON are demonstrating that facets of human thinking having to do with the creative imagination can be successfully emulated by sophisticated machines. If algorithms are appropriately specified and possess the right level of sophistication, creative processes that employ the use of the imagination can be mechanized in machines that are technological implementations of these algorithms.[6] If humans are machines, then they will be subject to whatever limitations (e.g. Gödelian limitations) apply to these machines (Nilsson, 2010). The creative imagination will be de-mystified and the Romantic myth of the creative genius will be dissolved, once and for all.

It may be objected that once JAPE, AARON, Emmy, and BACON are taken off the high mesa of their respective areas of creative specialization, they fall to the flat plain of ignorance (Nilsson, 1995). This is known as the Brittleness Objection. What these brittle machines lack is a certain robustness and a good dose of common sense (McCarthy, 1990). However, the Brittleness Objection is not a decisive one: at least some creatively imaginative human beings similarly behave as idiot savants once taken off the high mesa of their areas of creative specialization. Furthermore, there is every possibility that deliberative processes involving the use of common sense and native good judgment can be mechanized in the same manner that creative processes employing the use of the imagination have been.

The Many-Roads-Lead-to-Rome Response. It is traditionally held that consciousness, intelligence, and the imagination supervene on certain physical and biochemical properties. Only the right kind of stuff (viz. neuroprotein) can support creative imagining. If however silicon-and-metal-based machines are capable of creative output, then we appear to be led to

[6] As a limiting case, one might think of consciousness, intelligence, thinking, and creative uses of the imagination as emergent properties that spring forth from an appropriately sophisticated level of hardware and software organization. See LaChat (1986).

the multiple realizability thesis, according to which the same mental state of creatively imagining can be realized by certain entities that differ from human beings physically, physiologically, or neurologically.[7]

Take the analogous mental state of being in pain. In human beings, the mental state of being in pain is realized by the activation of C-fibers, whereas in other non-human animals the physiological mechanism that realizes pain might be vastly different. Once we take into account (logically possible) non-carbon-and-protein-based biological systems and putatively minded electromechanical systems, we are led to conclude of mental states, properties, and concepts that there are no constraints on the actual physical mechanisms or structures that realize them (Kim, 1996, pp. 74-75). If this applies by analogy to the mental states of creatively imagining, then many roads lead to Rome and mental states involving creative uses of the imagination ought not to be reduced to brain states. Naturalistic approaches to the philosophy of imagination have thus far sought integration with cognitive science (Nichols & Stich, 2003, Weinberg & Meskin, 2006). They might be advised to extend their gaze to salient developments within AI research as well.

The What-You-See-Is-Not-What-You-Get Response. While the output of JAPE, AARON, Emmy, and BACON might be ostensibly creative, what you see need not be what you get. For starters, we could still distinguish between the creative output and the creative process (as is typically done in creative cognition research). These machines remain the product of human ingenuity, and the algorithmic processes that are implemented in these machines cannot properly count as creative imaginative processes. After all, since these machines can only do what they are programmed to do, they fail to take us by surprise and cannot achieve

[7] The concept of multiple realizability, first introduced by Putnam (1967) and later receiving its current expression from Lewis (1972), has vexed many philosophers of mind. Attempts have been made, *vis-à-vis* the multiple realizability thesis, to demonstrate that some version of a mind-brain identity theory is still viable. For a relatively recent discussion, see Polger (2002).

any novelty (Lovelace, 1953).[8] The What-You-See-Is-Not-What-You-Get Response is a highly popular response, supported *inter alia* by a number of distinct objections:

The Easy Dupe Objection. One could also deny that the ostensibly creative output generated by these programs is really creative. If machines are able to pass the Dartmouth-based Turing Tests in creativity, it is not because their output meets some creativity threshold.[9] Rather, human beings are in general easily duped.[10] The Turing Tests in creativity and the more standard Turing Test, according to this skeptical line of reasoning, confirm neither the creativity of the output generated by programs nor the conversational intelligence of programs, but merely the more general fact that human beings are in general easily deceived into making conclusions that are unwarranted.[11]

The Lovelace Objection. As machines can only do as they have been programmed to do, the relevant creative process is derivative in these machines. If the ostensibly creative output is indeed creative, then credit ought to be properly assigned to the human algorithm designers and AI researchers instead of the machines. The Lovelace Objection is more plausible if AI continues to be understood in the classical computationalist sense, wherein algorithms are designed with decision-making rules that are handwritten. In a machine-learning context, however, algorithms have the capacity to define or modify their decision-making rules autonomously. As algorithms increasingly rely on machine learning capacities, human input will be minimized.

---

[8] For anecdotal evidence of how problem-solving algorithms have creatively subverted the expectations and intentions of their human AI designers and researchers and produced unexpected results, see Lehman *et al* (2018).

[9] Besides (and as things stand), there remains a worrying lack of consensus about assessment criteria for creativity. Having addressed this elsewhere, it is not my intention to revisit my arguments here.

[10] Defenders of the Easy Dupe Objection are effectively denying that the Turing Tests for creativity provide an adequate empirical criterion for generating human-scale creative performance (or creative output-generating) capacity. What is further implied is that only human beings remain capable of possessing mental states, whether we might have in mind the mental state of creativity or the mental state of being easily duped.

[11] Weizenbaum (1966) provides a similar objection to the Turing Test. More specifically, Weizenbaum's ELIZA program has deceived people into mistaking it for a human being, despite the fact that ELIZA is able to engage a person in a conversation despite not understanding anything about what is being said.

The Vagueness Objection. Turing (1950) and Minsky (2006) have notoriously given short shrift to the questions ‘Can machines think?’ and ‘Are machines capable of consciousness?’. They would have dismissed Deckard’s question on behaviorist grounds. Thinking, consciousness, dreaming, and imagining are vague, ambiguous, mysterious notions, and it simply will not do to pose questions that include these notions. If JAPE, AARON, Emmy, and BACON are capable of generating creative output, we can only conclude that they exhibit at least some of the behavioral dispositions of which human producers of creative output are capable. Whether or not these machines and programs can exercise creative processes that employ the use of the imagination is a question we should ultimately refrain from answering, given the gappiness of the notion of the creative imagination.

The Further-Missing-Parts Objection. Machines might share the relevant behavioral dispositions of which human producers of creative output are capable. However, they crucially lack the further conditions (viz. motivations and emotions) for the algorithmic processes to properly count as creative imaginative processes. According to this objection, the relevant behavioral dispositions, motivations, and emotions are necessary but insufficient on their own. Rather, they form a suite of conditions that are jointly sufficient when taken together. Nonetheless, nothing in principle prevents us from furnishing machines or programs with human-like bodies, in order for these machines or programs to perceive, act, and attain human-like motivations, emotions, and experiences (*cf.* Brooks, 1990). This would constitute a postclassical enactivist response to the Further-Missing-Parts Objection. Alternatively, while it could be readily conceded that we still understand precious little about the computational architecture of the mind, nothing in principle prevents us from understanding motivation and emotion in computational terms and more adequately in the future. This would constitute a classical computationalist response to the Further-Missing-Parts Objection.

A few final thoughts in this potted history: as aforementioned, when encountering the creative output of human beings, it is relatively uncontroversial for us to make a causal inference in favour of the existence of an underlying mechanism for this creativity (typically involving creative uses of the imagination) in these individuals. If we resist making the same inferential move when encountering the ostensibly creative output of machines, do we do so simply because of our species chauvinism or anthropocentric bias? How do we avoid being invidiously discriminatory in our responses to these imagination machines? To what extent will these imagination machines alter our more general understanding of the imagination (whether by their heroic failure or successful emulation of the uses we make of our creative imagination)? While I do not presume to possess the answers to all these questions, I do hope at least to have stimulated further lines of enquiry about imagination machines and their implications for our understanding of the imagination (creative or otherwise).

BIBLIOGRAPHY:

Brooks, Rodney A. 1990. 'Elephants Don't Play Chess,' in *Robotics & Autonomous Systems*, Vol. 6, pp. 3-15

Dyson, George. 1997. *Darwin Among the Machines: The Evolution of Global Intelligence*, Helix Books

Harnad, Stevan. 1992. 'The Turing Test is not a Trick: Turing Indistinguishability is a Scientific Criterion,' in *SIGART Bulletin*, Vol. 3 No. 4, pp. 9-10

Hobbes, Thomas. 1651. *Leviathan*, ed. & intro. Michael Oakeshott, Oxford: Basil Blackwell

Kim, Jaegwon. 1996. *Philosophy of Mind*, Boulder: Westview Press

LaChat, Michael. 1986. 'Artificial Intelligence & Ethics: An Exercise in the Moral Imagination,' in *AI Magazine*, Vol. 7 No. 2, pp. 70-79

Lehman, Joel, Jeff Clune, Dusan Misevic, Christoph Adami, Julie Beaulieu, Peter J. Bentley, Samuel Bernard, Guillaume Beslon, David M. Bryson, Patryk Chrabaszcz, Nick Cheney, Antoine Cully, Stephane Doncieux, Fred C. Dyer, Kai Olav Ellefsen, Robert Feldt, Stephan Fischer, Stephanie Forrest, Antoine Frénoy, Christian Gagné, Leni Le Goff, Laura Grabowski, Babak Hodjat, Frank Hutter, Laurent Keller, Carole Knibbe, Peter Krcah, Richard E.Lenski, Hod Lipson, Robert MacCurdy, Carlos Maestre, Risto Miikkulainen, Sara Mitri, David E. Moriarty, Jean-Baptiste Mouret, Anh Nguyen, Charles Ofria, Marc Parizeau, David Parsons, Robert T. Pennock, William F. Punch, Thomas S. Ray, Marc Schoenauer, Eric Shulte, Karl Sims, Kenneth O. Stanley, François Taddei, Danesh Tarapore, Simon Thibault, Westley Weimer, Richard Watson & Jason Yosinksi. 2018. 'The Surprising Creativity of Digital Evolution: A Collection of Anecdotes from the Evolutionary Computation and Artificial Life Research Communities,' cite arxiv:1803.03453

Lewis, David. 1972. 'Psychophysical & Theoretical Identifications,' in *The Australasian Journal of Philosophy*, Vol. 50, pp. 249-58

Lovelace, Ada. 1953. 'Notes on Manabrea's Sketch of the Analytical Engine Invented by Charles Babbage,' in *Faster than Thought*, ed. Bertram Vivian Bowden, London: Sir Isaac Pitman & Sons

Mahadevan, Sridhar. 2018. 'Imagination Machines: A New Challenge for Artificial Intelligence,' *AAAI Conference*, link available at <https://people.cs.umass.edu/~mahadeva/papers/aaai2018-imagination.pdf>

Minsky, Marvin. 2006. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, & the Future of the Human Mind*, New York: Simon & Schuster

Nichols, Shaun & Stephen Stich. 2003. *Mindreading: An Integrated Account of Pretense, Self-awareness & Understanding Other Minds*, Oxford: Oxford University Press

Nilsson, Nils. 1995. 'Eye on the Prize,' in *AI Magazine*, Vol. 16 No. 2, pp. 9-17

Nilsson, Nils. 2010. *The Quest for Artificial Intelligence*, Cambridge University Press

Penrose, Roger. 1989. *The Emperor's New Mind*, Oxford University Press

Polger, Thomas. 2002. 'Putnam's Intuition,' in *Philosophical Studies,* Vol. 109 No. 2, pp. 143-70

Putnam, Hilary, 1967. 'Psychological Predicates,' in *Art, Mind, & Religion*, ed.s W.H. Capitan & D.D. Merrill, Pittsburgh: University of Pittsburgh Press, pp. 37–48

Turing, Alan M. 1950. 'Computing Machinery & Intelligence,' in *Mind,* Vol. 59 No. 236, pp. 433-60

Weinberg, Jonathan & Aaron Meskin. 2006. 'Puzzling over the Imagination: Philosophical Problems, Architectural Solutions,' in *The Architecture of the Imagination: New Essays on Pretence, Possibility, & Fiction*, ed. Shaun Nichols, Oxford University Press, pp. 175-202

Weizenbaum, Joseph. 1966. 'ELIZA – A computer program for the study of natural language communication between man & machine,' in *Communications of the ACM*, Vol. 9 No. 1, pp. 36-45