Massimo Passamonti **(winner: Trinity term 2023)**

From Schopenhauer to Turing: a novel approach to compatibilism

Introduction

Arthur Schopenhauer and Alan Turing approach the topic of free will from different perspectives. Whilst Schopenhauer investigates the existence of the ultimate origin of free will, Turing discusses the topic indirectly by arguing that for machines to 'think', they must exhibit behaviours indicative of free will. This essay aims to show how Turing takes the foundational elements of Schopenhauer's free will argument and further elaborates on them, adding depth and breadth. Deviating from Schopenhauer's hard deterministic stance, Turing proposes a response-dependent theory in which free will emerges from the observer's perception as opposed to an ultimate 'uncaused effect'. Inspired by a line of reasoning proposed by Diane Proudfoot, I will argue that Turing's interpretation does not rule out the existence of alternative forms of free will that are fundamentally different from human free will and, by doing so, reaches a new account of compatibilism.

Schopenhauer's view of free will

Schopenhauer (1999) begins the 'Prize Essay of the Freedom of the Will' with a comprehensive deterministic examination of the world. According to him, the laws of causality are an immutable a priori principle that serves as the universal norm to which all objects in the world are subject. These laws of causality materialise themselves in human beings in the form of motivations. These motives, that cause us to act, are not different from any other event in nature and must, therefore, follow the same laws of causality. He concludes his essays by arguing that, despite the special kind of freedom that humans possess through rationality, and which allows for a relative freedom compared to other beings such as animals, we do not have a complete freedom of the will. His final position is therefore clear: we do not possess free will because everything is submitted to the laws of causality. Schopenhauer's argument may be formulated in the following way:

1. *All events in nature are fully determined by causes.*
2. *My acts of will are events in nature.*
3. *Anything that is determined by causes cannot be free.*

_ *Therefore, my acts of will are not free.*

The traditional compatibilist argument challenges premise (3), arguing that we are free in some sense of the word even if determinism is true. I will argue that Turing describes an alternative to the traditional compatibilist account not by challenging premise (3) but by building upon Schopenhauer's argument and going beyond his analysis.

Turing's theory of free will

Turing approaches the issue of free will indirectly while investigating whether machines can think or not. To address this question, Turing reformulates the problem in terms of a game which he calls the 'imitation game'. He suggests that if a machine could pass the 'imitation game', it would be reasonable to consider it as being able to 'think' (Turing, 1950). The 'imitation game' relies on another idea that Turing introduces in one of his earlier works. In 'Intelligent Machinery' he claims that the concept of 'intelligence' is "emotional rather than mathematical", in the sense that our perception of a machine's intelligent behaviour is as much influenced by our own state of mind as it is by the characteristics of the entity being observed (Turing, 1948, p. 108). This led Diane Proudfoot (2013) to a response-dependent

interpretation of Turing's theory. A concept is response-dependent if it has a conceptual relationship with the ideas that arise from our reactions in certain delineated circumstances. To address the 'thinking machine' problem, Turing observes that a machine, to replicate a brain-like behaviour, must function as if it inherently has free will (Turing, 1951). According to Proudfoot, Turing uses the same response-dependent concept to address the issue of free will by rephrasing the original question 'Can a machine possess free will?' with the question 'Can a machine take us by surprise?' (Proudfoot, 2017). In 'Can Digital Computers Think?', Turing (1951) enhances and refines, in my view, Schopenhauer's original argument by adding additional layers of reasoning. His line of thinking may be formulated as follows:

1. *Determinism and free will are incompatible (Schopenhauer's premises (1), (2) and (3)).*
2. *There are two solutions: either:*
   a) *x believes to possess free will, but it is an illusion; or*
   b) *x possesses free will, but this can't be proven from x's behaviour.*
3. *It does not matter whether (2.a) or (2.b) is correct.*

_ *If x can take us by surprise, then x has free will.*

Proudfoot (2017) points out that if 'machine intelligence' is a response-dependent concept, then it should not be explained by the underlying mechanics that determines it (i.e., the computations that the machine carries out) but rather by the response that it generates in the observer. Similarly, a response-dependent formulation does not allow us to find the ultimate cause of free will by stripping away all causally determined actions, rather it supports us to evaluate whether a machine appears to behave as if it had free will in the eye of an observer. Turing uses the 'skin of an onion' analogy to address the mistaken approaches of exploring 'thinking' (Turing, 1950). When examining how the mind works, we observe processes that can be explained in simple mechanical terms. We are tempted to say that these are not the true essence of the brain, but rather layers that need to be peeled away to disclose the true underlying mind. However, as we continue with this process, we are left with the dilemma of whether we will ever reach the 'real' mind, or whether 'intelligence' is simply a construct whose essence is determined by the whole rather than by its parts. If the second option is true, Turing concludes, we should be inclined to believe that the entire mind is mechanical. I argue that this analogy can also be used in the context of free will. If we remove all causally determined actions, motives, stimuli, and rationality what remains may not be the "effect without cause" that Schopenhauer (1999, p. 66) was seeking. Free will is, based on Turing's analogy, the result of all these elements interacting one with another and, ultimately, with the observer's perception. Proudfoot (2017) further suggests that if we accept an observer-dependent formulation, then this might provide a novel understanding of free will in the context of the connection between primary and secondary qualities. This perspective implies that free will is not an inherent quality existing separately from the agent or its actions. Instead, it emerges from the observer's perception and as a result of the interaction between the agent's actions and the observer's sensory system. Proudfoot argues that the same way an object can be coloured even if its constituent elements (the particles) are not, so an action can be regarded as being free even if its causes are completely determined. This line of reasoning suggests that Turing's argument can be framed in Lockean terms. As a secondary quality, free will bears no "resemblance of [the body] at all"; rather, it is solely the object's capacity to generate the impression of free will in the observer's minds (Guttenplan, S., *et al.,* 2021, p. 324).

Conclusion

In conclusion, Schopenhauer and Turing concur that all actions, in humans and machines, originate from preceding causes, negating the existence of free will that is independent of causality. Turing, contrasting Schopenhauer, transcends the traditional account of compatibilism and introduces a novel approach that is able to reconcile free will and determinism. He suggests that free will is an observer-dependent concept, a secondary quality arising from the interaction between the machine's actions and the observer's perception. His theory implies that the reactions we experience of the whole is the 'real' free will, whilst its parts only contribute to its overall experience. Actions are, therefore, truly free even if they are the result of completely determined chains of causes. Turing's 'skin of an onion' analogy for the investigation of the mind further supports this interpretation. But an observer-dependent formulation makes another, more significant, contribution in the effort to reconcile determinism and free will. While Schopenhauer's investigation seeks the ultimate uncaused effect of human free will, Turing introduces a universal methodology to prove its existence. His criterion aims at providing a *sufficient* condition for free will in an observer-*dependent* world, whereas Schopenhauer's approach intends to define the *necessary* conditions for free will in an observer-*independent* world. In other words, Turing does not rule out the existence of other types of free will that are capable of producing similar experiences in the observer but that are fundamentally distinct from human free will. Turing's distinction between the essence of free will and the observer's experience of it avoids the use of overly rigorous or ambiguous definitions, provides a universal approach to defining free will, and results in a new account of compatibilism.

Reference list

Guttenplan, S., *et al.* (2021) *Reading Philosophy: selected texts with a method for beginners*, 2nd edition. Hoboken: Wiley-Blackwell.

Proudfoot, D. (2013) Rethinking Turing's Test. *Journal of Philosophy,* 110 (7), 391–411.

Proudfoot, D. (2017) Turing and Free Will: A New Take on an Old Debate. In: Floyd, J., Bokulich, A. (eds). *Philosophical Explorations of the Legacy of Alan Turing*. Cham: Springer, 305–321.

Schopenhauer, A. (1999) *Prize essay on the freedom of the will*. New York: Cambridge University Press. Edited by Günter Zöller.

Turing, A. M. (1948) Intelligent Machinery, In: Copeland B. J. (ed.) (2004). *The Essential Turing*. Oxford: Oxford University Press, 395–432.

Turing, A.M. (1950) Computing Machinery and Intelligence. *Mind*, 59(236), 433–460.

Turing, A. M. (1951) Can Digital Computers Think?, In: Copeland B. J. (ed.) (2004). *The Essential Turing*. Oxford: Oxford University Press, 482–486.