Are we morally responsible in a deterministic world?

Akanksha Kevalramani

In this essay, I will argue that determinism challenges the sort of free will and moral responsibility central to philosophical debate but does not significantly undermine our sense of freedom or practices of moral responsibility. To support my argument, I first introduce basic desert-based retrospective moral responsibility as the prevalent understanding of moral responsibility in both everyday life and philosophical debate. Then, I explore two key conceptions of free will and demonstrate how determinism is incompatible with both conceptions, thus challenging moral responsibility in the basic desert sense. Finally, I address the societal concerns arising from this incompatibilist conclusion by demonstrating how we can uphold our sense of agency and moral responsibility practices, especially in the criminal justice system.

The burgeoning literature on moral responsibility has led to verbal disputes among philosophers debating the definitions and conditions necessary or sufficient for having it (Chalmers, 2011, cited in Pereboom, 2014; Zimmerman, 2015). In this essay, *moral responsibility* will henceforth refer to basic desert-based retrospective moral responsibility. Central to the determinism and free will debate, this kind of moral responsibility is desert-based in the sense that an agent would "*deserve* to be blamed if she understood that [her action] was morally wrong, and she would *deserve* to be praised if she understood that it was morally exemplary" (Pereboom, 2014, p. 2). This is also the basis of how we understand moral responsibility in our everyday lives, as P.F. Strawson famously brought to light in his 1962 in his paper "Freedom and Resentment." For example, the intuitive response to being pushed with the intention to harm is different from the response to being pushed accidentally in a crowded space. In the former situation, we may experience a heightened feeling of

anger or resentment that is absent from the latter scenario. Strawson posited that our emotional responses and reactive attitudes toward an individual's actions as well as the quality of their will form the basis of moral responsibility (Cushing, 2022). These reactive attitudes, such as resentment and indignation, imply that "its target *deserves* to be the recipient of the expression of such emotions" (Caruso and Pereboom, 2022, p. 8). Additionally, this desert-based moral responsibility is basic in that the desert is based solely on the action's inherent moral nature and not any consequential or contractual considerations. Lastly, the moral responsibility here is retrospective, i.e., it pertains to the agent's responsibility for actions that have already occurred. Most philosophical debates on moral responsibility revolve around this basic desert-based retrospective moral responsibility, and in this essay, I will show that this notion of moral responsibility is incompatible with determinism. (Henceforth, when I mention *moral responsibility* without qualification, it refers to basic desert-based moral retrospective responsibility).

Most philosophers agree that moral responsibility requires the existence of free will. This means that for someone to be deserving of blame or praise, they must possess a certain level of control or free will over their actions. However, the concept of free will itself is subject to debate, and there appear to be two fundamental notions of it. The first is a leeway-based approach, which posits that for an individual to possess free will, they must have the ability to do otherwise (Timpe, 2016). This idea that "a person is morally responsible for what he has done only if he could have done otherwise" is termed "the principle of alternate possibilities" (Frankfurt, 1969, p. 829). The second notion of free will is based on the concept of sourcehood, which states that an individual has free will if they are the source of her actions in a specific way (Timpe, 2016). Philosophers often debate moral responsibility by considering these two notions of free will, with some claiming one notion is more relevant than the other. To provide a strong case for the incompatibility of moral

responsibility with determinism, I will demonstrate how both notions of free will are incompatible with determinism.

Determinism posits that all events, including human deliberations, choices, and actions, are causally determined by prior events and the laws of nature. According to van Inwagen (1983, cited in Caruso and Pereboom, 2022), this implies that only one future is physically possible at any given time, with no room for alternate possibilities. By definition, this contradicts the principle of alternate possibilities and, therefore, makes determinism incompatible with the leeway-based conception of free will.

However, prominent semi-compatibilist philosopher John Martin Fischer makes a compelling argument that the presence of alternate possibilities is irrelevant to the grounding of moral responsibility. To support this claim, Fischer presents the following Frankfurt-type case:

> Jones is in a voting booth deliberating about whether to vote for the Democrat or the Republican. After weighing reasons and deliberating in the "normal" way, he chooses to vote for the Democrat. Unbeknownst to him, Black, a neurosurgeon with Democratic sympathies, has implanted a device in Jones's brain that monitors Jones's brain activities. If he is about to choose to vote Democratic, the device does not intervene. If, however, Jones were about to choose to vote Republican, the device would trigger an intervention that would involve electronic stimulation of the brain sufficient to produce a choice to vote for the Democrat and an actual vote for the Democrat (Fischer, 2007, p. 335).

Fischer suggests that in the actual sequence, i.e., when Jones chooses to vote for the Democrat based on his own deliberation, Jones is *choosing freely* and can be held morally responsible for his

choice and action. According to Fischer, this demonstrates that, even though Jones lacks alternate possibilities due to the presence of Black's device, he has the necessary freedom for moral responsibility (Fischer, 2007). However, philosophers, such as Stewart Goetz and Derek Pereboom, raise objections to this example (Fischer, 2007). The primary objection is that the factor eliminating alternate possibilities (in this case, Black's device) exists outside of the actual sequence, as opposed to causal determinism, in which the factor flows through the actual sequence. In other words, Jones' deliberative process, which Fischer regards as *choosing freely*, is itself part of the causal chain. Fischer acknowledges this weakness and concedes that Frankfurt-type examples do not conclusively establish moral responsibility in a causally deterministic world (Fischer, 2007). Therefore, I maintain the position that the leeway-based approach to free will is incompatible with determinism.

In considering the sourcehood conception of free will, I turn to a notable argument presented by Galen Strawson in "The Impossibility of Moral Responsibility" (1994). The argument can be summarized as follows:

a. You do what you do because of the way you are.

b. To be truly morally responsible for what you do, you must be truly responsible for the way you are, at least in certain crucial mental respects.

c. You cannot be truly responsible for the way you are.

d. Hence, you cannot be truly morally responsible for what you do.

An individual may wish to argue against premise c, but doing so would result in a regress, indicating that true self-determination is impossible for finite human beings like us. Given that we lack *ultimate responsibility* for our actions, we do not have the agency required for moral

responsibility in the basic desert sense. Hence, the sourcehood conception of free will and moral responsibility are incompatible with determinism.

Compatibilists predominantly criticize Strawson's argument, claiming that total control as a requirement for moral responsibility is too stringent. Instead, they propose scaling back our demands regarding sourcehood (Fischer, 2012). Fischer argues that compatibilists have never denied our subjection to "factors entirely outside our control" but instead believe that despite these external forces, we possess a "meaningful and robust sort of control" (Fischer, 2012, p. 182). In my view, these contentions fail to present robust counterarguments to Strawson's thesis; instead, they simply dismiss it by bargaining the conditions required for moral responsibility. Therefore, I maintain that Strawson's argument remains unchallenged and provides a thorough and convincing rationale for the incompatibility of sourcehood and determinism.

To address this unresolved disagreement between compatibilists and incompatibilists regarding the conditions required for moral responsibility, Derek Pereboom provides a compelling thought experiment that demonstrates our lack of free will even when compatibilist conditions are met. He presents a four-case argument such that in each case, the agent satisfies compatibilist conditions and is causally determined by factors beyond his control. The cases are designed to ensure there are no responsibility-relevant differences between any two consecutive cases, leading to the conclusion that if the agent is not morally responsible in the first case, then he is not morally responsible in the last case either. To summarize his four-case argument:

> In Case 1, a team of neuroscientists manipulates the agent's neural states remotely. Before the agent begins his deliberation process, the neuroscientists intervene by manipulating his neural states such that his thoughts and actions are determined by the

neuroscientists. This is an explicit case of causally determining manipulation, which most compatibilists would agree absolves the agent of moral responsibility.

In Case 2, instead of manipulating the agent's brain at the time of decision-making, the neuroscientists program the agent's brain at the beginning of his life such that his deliberations and actions are always according to the neuroscientists' programming. Since the only difference between Case 1 and Case 2 is *when* the deterministic manipulation happens, it is intuitive that there are no responsibility-relevant differences between the two cases. Hence, even in Case 2, the agent is absolved of moral responsibility.

In Case 3, instead of being manipulated by neuroscientists, the agent's deliberations and actions are a result of the training practices of his community. The resulting deliberations and actions are the same as in the previous cases, with the only difference being *who* performed the deterministic manipulation. Since this difference between Cases 2 and 3 is not a responsibility-relevant difference, then in this case, too, the agent cannot be held morally responsible.

In Case 4, instead of being determined by a team of neuroscientists or a community, the agent's deliberations and actions are causally determined by the past and the laws of nature. This is an example of a human in the ordinary deterministic world, and Pereboom contends that there is no responsibility-relevant difference between this case and Case 3 either, leading us to the conclusion that the agent lacks the required freedom for moral responsibility in a deterministic world (Caruso and Pereboom, 2022, pp. 26–27).

Some compatibilists object to this argument by acknowledging there are no responsibility-relevant differences between adjacent cases and consequently assert that the agent must be morally responsible in all four cases. However, this starkly contradicts our intuitive understanding of sourcehood and moral responsibility. No matter how lenient the requirements of sourcehood for moral responsibility may be, the agent in Case 1 evidently lacks any semblance of sourcehood for moral responsibility (Caruso and Pereboom, 2022). Other compatibilists attempt to identify responsibility-relevant differences between the cases such that the agent may be considered morally responsible in some cases and not others. The drawback of this approach is that any difference these philosophers identify would apply solely to the "current manipulation cases but not future cases" (Caruso and Pereboom, 2022, p. 29). Caruso and Pereboom (2022) contend that it is impossible for us to find any kind of freedom or control, even in principle, that would be available to agents in deterministic settings but not in manipulation cases. Given the issues with these objections, I believe the four-case argument provides a robust argument for the incompatibility of sourcehood and determinism.

Thus far, I have demonstrated the incompatibility of different conceptions of free will and determinism, which implies that basic desert-based moral responsibility must also be incompatible with determinism. In other words, if an individual is unable to do otherwise and cannot be an appropriate source of their actions, then they cannot be truly deserving of blame or praise for their actions.

Many philosophers oppose this incompatibilist stance, expressing concerns that a deterministic world could undermine our sense of agency and capacity for deliberation, as well as negatively impact our norms for handling criminal behavior. However, I argue that determinism would not significantly diminish our sense of freedom or moral responsibility practices.

Philosophers such as John Searle (2001, cited in Fischer, 2007) and Peter van Inwagen (1983, cited in Fischer, 2007) argue that if we lived in a deterministic world without free will, then there would be *no point* in engaging in deliberation, and doing so would be contradicting ourselves or holding an inconsistent set of beliefs. In other words, if we believe there is only one option available to us, then it would be futile or contradictory to engage in deliberation as if there were more possibilities available to us. However, a lack of metaphysical possibilities does not deprive us of epistemic possibilities (Fischer, 2007). Even if we believe determinism is true, our inability to know what is determined compels us to engage in deliberation and reasoning. Moreover, the doctrine of causal determinism neither impairs our capacity to deliberate nor affects our desire to choose and act according to our judgment (Fischer, 2007). Renowned hard incompatibilist Galen Strawson states that, regardless of the true nature of our freedom and moral responsibility, our *experience* of freedom and moral responsibility is an inescapable part of being human (Strawson, 1994). Therefore, due to our inescapable *capacity* and *necessity* for deliberation and reasoning, the truth of determinism or its incompatibility with free will would not substantially diminish our sense of agency and responsibility.

Furthermore, while determinism would influence some existing criminal justice procedures, various methods would still remain available for managing offenders. Among the different approaches to criminal justice, determinism would particularly affect retributivism, which is the notion of punishment as an end in itself with no other purpose, such as deterring others or rehabilitating the offender. Since retributivism is rooted in the belief that individuals deserve punishment in proportion to the moral blameworthiness of their actions, it inherently assumes free will and basic desert-based moral responsibility. However, as previously demonstrated, in a deterministic world, an individual cannot be deserving of blame or praise for their actions as they

lack the ability to do otherwise and ultimately is not the source of their actions. Although retributivism loses its justification in a deterministic world, embracing a forward-looking, non-desert-based approach to moral responsibility, as proposed by Derek Pereboom, is not only viable but also advantageous. His approach emphasizes the "protection of potential victims, reconciliation in personal relationships and with the moral community, and formation of moral character" (Pereboom, 2015, p. 285). Several existing criminal justice strategies would be compatible with this forward-looking non-desert account, such as moral education and deterrence; however, the most enduring and compelling approach is incapacitation. This theory "invokes our right to protect ourselves" by drawing an analogy to quarantine (Pereboom, 2014, p. 169). Just as we find it acceptable to quarantine carriers of dangerous diseases, we could justify quarantining dangerous criminals with the intention of protecting potential victims. Therefore, by focusing on the safeguarding of potential victims, Pereboom's forward-looking approach to handling offenders is the most optimistic and ethical approach, irrespective of the truth of determinism.

Other forward-looking accounts of moral responsibility are criticized for potentially attributing blame to innocent individuals in pursuit of beneficial societal outcomes (Smart, 2017, cited in Talbert, 2023). For example, if a series of serious crimes go unsolved, leaving the actual perpetrator at large, others might come to believe they can get away with similar offenses. Hence, there could be an argument to punish the innocent in order to "maximize utility" (McCloskey, 1965 cited in Pereboom, 2014, p. 164). Another example is the possibility of detaining those who have not yet committed crimes but have the potential to do so. However, Pereboom counters these criticisms by arguing that only agents who have genuinely acted immorally should be subject to forward-aiming blaming practices, and adding that "the right to

liberty must carry weight" in these contexts (Pereboom, 2014, p. 170; Pereboom, 2017, cited in Talbert, 2023). I believe Pereboom's forward-looking approach comprehensively addresses various concerns related to criminal behavior in a deterministic world without invoking basic desert.

In conclusion, the traditional notions of free will and basic desert-based moral responsibility face significant challenges in the face of determinism. However, this does not mean our sense of freedom and moral agency would be completely undermined in a deterministic world. The ability and need to deliberate and reason provide us with a sense of agency and responsibility in our everyday lives, even if our choices may be ultimately determined. Moreover, adopting a forward-looking, non-desert-based approach to moral responsibility can help us navigate our interpersonal relationships and encourage us to refine our justice system in a more morally justifiable manner.

Reference List

Caruso, G.D. and Pereboom, D. (2022) *Moral Responsibility Reconsidered*. 1st edn. Cambridge University Press. Available at: https://doi.org/10.1017/9781009219730.

Cushing, Simon (2022) *P.F. Strawson: 'Freedom and Resentment'*. 20 October. Available at: https://www.youtube.com/watch?v=mYOtIamkenw (Accessed: 26 November 2023).

Fischer, J.M. (2007) 'Free Will and Moral Responsibility', in *The Oxford Handbook of Ethical Theory*. Oxford University Press. Available at: https://doi.org/10.1093/oxfordhb/9780195325911.003.0013.

Fischer, J.M. (2012) 'Sourcehood: Playing the Cards That Are Dealt You', in J.M. Fischer (ed.) *Deep Control: Essays on Free Will and Value*. Oxford University Press, p. 0. Available at: https://doi.org/10.1093/acprof:osobl/9780199742981.003.0010.

Frankfurt, H.G. (1969) 'Alternate Possibilities and Moral Responsibility', *The Journal of Philosophy*, 66(23), p. 829. Available at: https://doi.org/10.2307/2023833.

Pereboom, D. (2014) *Free Will, Agency, and Meaning in Life*. Oxford University Press. Available at: https://doi.org/10.1093/acprof:oso/9780199685516.001.0001.

Pereboom, D. (2015) 'A Notion of Moral Responsibility Immune to the Threat from Causal Determination', in R. Clarke, M. McKenna, and A.M. Smith (eds) *The Nature of Moral Responsibility: New Essays*. Oxford University Press, p. 0. Available at: https://doi.org/10.1093/acprof:oso/9780199998074.003.0013.

Strawson, G. (1994) 'The impossibility of moral responsibility', *Philosophical Studies*, 75(1–2), pp. 5–24. Available at: https://doi.org/10.1007/BF00989879.

Strawson, P. (1962) 'Freedom and Resentment', *Proceedings of the British Academy*, 48, pp. 187–211.

Talbert, M. (2023) 'Moral Responsibility', *The Stanford Encyclopedia of Philosophy.* Fall 2023. Edited by E.N. Zalta and U. Nodelman. Metaphysics Research Lab, Stanford University. Available at: https://plato.stanford.edu/archives/fall2023/entries/moral-responsibility/.

Timpe, K. (2016) 'Leeway Vs. Sourcehood Conceptions of Free Will', in K. Timpe, M. Griffith, and N. Levy (eds) *Routledge Companion to Free Will.* Routledge, pp. 213–224.

Zimmerman, M.J. (2015) 'Varieties of Moral Responsibility', in R. Clarke, M. McKenna, and A.M. Smith (eds) *The Nature of Moral Responsibility: New Essays*. Oxford University Press, p. 0. Available at: https://doi.org/10.1093/acprof:oso/9780199998074.003.0003.