Proceedings of ICIHCS 2025 Symposium: Voices of Action: Narratives of Faith, Ethics, and Social Practice
DOI: 10.54254/2753-7064/2025.ND28262

Free Will and Determinism: Can Humans Be Morally Responsible Agents?

Zihe Wang

The Hill School, Pennsylvania, USA
cwang26@thehill.org

Abstract. This paper aims to examine whether agents are to be held morally responsible for their actions in a deterministic world. In order to do so, the paper puts the most demanding species of moral responsibility, accountability, at its center and argues for a compatibilist framework grounded in guidance control. Where actions issue from an owned, adequately reasons-responsive mechanism. Main central concepts are explicitly clarified, such as leeway vs. source-hood interpretations of what is necessary to be free, and introducing different perspectives on the topic at hand. It then evaluates the Compatibilism argument against the Incompatibilism, including notable classical arguments and critiques. Building on Fischer and Ravizza's model, the paper defends reasons-responsiveness and historical ownership as sufficient for responsibility in the absence of "freewill." This paper seeks to align theory with intuition, such that the intuitive sense of morality does not diminish and preserve the desert-based evaluation, the integrity of blame, praise, and sanction within a causally ordered world.

Keywords: free will, determinism, moral responsibility, compatibilism

1. Introduction

The problem of free will and moral responsibility asks whether agents can be justly held responsible for their actions if humans are governed by determinism. This problem is not an abstract discussion distant from the ordinary world. It is intertwined with the validity of legal systems, ethical beliefs, and how society views interpersonal relationships. Whether people can logically convince themselves that they can be truly responsible affects how they justify punishments, forgiveness, and praise. If determinism ultimately undermines moral responsibility, then many of social practices will be viewed differently. This is a metaphysical speculation that is crucially relevant.

Philosophers developed many views on this debate. Incompatibilism argue that determinism rules out genuine freedom and responsibility. Their central claim is that the world is made from certain chemical structures and governed by certain physical laws. Humans are part of this world, and if the past cannot be changed, then there could only be one possible future under such laws. Although not all believe that the ability to do otherwise is a prerequisite to moral responsibility, they contend that this leeway is indeed necessary [1,2]. Libertarians agree that if determinism were true, humans would not have free will. However, they argue that there are these indeterministic "self-forming actions" in which agents face alternatives [3]. There are many ways to argue that the world is not

DOI: 10.54254/2753-7064/2025.ND28262

deterministic, or at least, to be strict, "reality" is not deterministic: using perspectives of philosophers such as Descartes' "cogito, ergo sum" to illustrate that there could be dimensions to reality. Compatibilists, by contrast, deny that alternative possibilities are necessary for responsibility. What matters is not whether agents could have done otherwise in some metaphysical sense, but whether they acted through their own decision-making. On this view, responsibility requires only the right kind of control, often described in terms of reasons-responsiveness and ownership of one's actions [4-6]. Finally, skeptics contend that neither determinism nor indeterminism can ground moral responsibility. If determinism is true, people's actions are fixed by factors beyond their control; if indeterminism is true, then people's choices risk being matters of luck. In either case, skeptics conclude, people lack the sort of freedom or autonomy required for desert-based praise or blame [7,8].

This paper will focus on accountability (differences between species will be explained later on in the paper) and examine whether this form of responsibility can survive in a deterministic world. It argues that although compatibilist theories, which ground responsibility in guidance control, reasons-responsiveness, and ownership of action, do not putatively defeat the incompatibilism, it is the most convincing view to confirm the natural intuition that grounds society under some moral basis. By analyzing the Consequence Argument, Frankfurt-style cases, and objections such as manipulation and moral luck, the paper will show how compatibilism can withstand incompatibilism and skeptical challenges while preserving the fairness of moral practices.

The structure of the paper is as follows. Section 2 clarifies the central concepts of free will, determinism, and species of responsibility. Section 3 presents incompatibilism, with emphasis on the Consequence Argument. Section 4 examines Frankfurt-style cases and their critics. Section 5 develops the compatibilist defense. Section 6 addresses major objections, including manipulation cases and moral luck. Section 7 concludes by clarifying why compatibilist responsibility best answers the question.

2. Central concepts and distinctions

The debate over free will and moral responsibility is often ambiguous under esoteric terminology. To make things explicit, this section will sort out the relevant concepts and specify the framework within which this paper will operate. Three distinctions are particularly important: (i) between different notions of freedom, (ii) between different "species" of moral responsibility, and (iii) between metaphysical and practical stakes.

2.1. Freedom and control

Generally, one way to open the free will discussion is to analyze whether an agent of autonomy over their choices. To be specific, could the agent have done otherwise? This will be referred to as leeway freedom. This approach is very intuitive: if agents have free will, and they can be deliberate about their actions, they certainly should have the power to choose between different options. However, this approach appears to run in a dead end when it encounters Peter Van Inwagen's classical Consequence Argument [1].

Which is why compatibilists turn to an alternative way of understanding freedom, shifting the focus from leeway to source-hood. The theory that an agent doesn't necessarily need to prove that they had alternatives; at the same time, philosophers might demand an even more rigorous definition of free will, where source-hood and leeway are necessary. Source-hood theorists state that an agent is free if they are the genuine origin of their actions. Meaning that at least one condition necessary

DOI: 10.54254/2753-7064/2025.ND28262

for an action originates within the agent themself. However, determinism undermines such claims as well. Since every condition necessary for an action is already entailed by events and laws prior to the agent's existence, leaving no ultimate source within the agent.

2.2. Species of moral responsibility

Responsibility itself requires further distinctions. Gary Watson and David Shoemaker emphasize that practices of holding responsible operate at multiple dimensions: (1) Attributability concerns whether an action can be an expression of the agent's character or values. When people describe an agent's action as cruel or generous, can it be said that this is a fair attribution? (2) Answerability emphasizes the capacity to provide reasons. Whether or not the agent is able to distinguish themselves clearly and provide justification for their actions. (3) Accountability concerns liability to the "reactive attitudes" (Strawson): resentment, indignation, guilt, gratitude, and the formalized counterparts of these attitudes in systems of praise and blame [9-11].

These three dimensions are conceptually distinct. Someone might be attributable (their action reflects their character) without being answerable (if, say, they lack rational capacities) or accountable (if they are excused from blame). They also differ in how vulnerable they are to determinism. Even in a deterministic world, attributability may remain intact: a cruel remark still reveals cruelty. Answerability, too, might survive: people could still assess whether reasons were available to an agent. Accountability, however, is the most demanding species. It ties directly to fairness and desert. To hold someone accountable is not only to describe their action but to claim they deserve moral praise or condemnation. This paper, therefore, places accountability at the center of the debate.

2.3. From historical accounts to contemporary stakes

Historically, classical compatibilists such as Hobbes and Hume defined freedom in terms of unimpeded desire satisfaction: one acts freely if one does what one wants without external constraints. Clearly, it is a very generous definition since this conception allows them to affirm both determinism and responsibility. Yet its inadequacy became clear in the twentieth century. Critics pointed out that agents whose desires were themselves manipulated (by brainwashing, coercion, or psychological compulsion) could meet the Hobbes–Hume definition while intuitively lacking freedom. More recent compatibilists have therefore refined the account of control to include second-order desires, reasons-responsiveness [5], or capacities for moral conversation [12,13].

This refinement reflects the practical stakes of the free will problem. If people lacked attributability, moral evaluation would collapse; if people lacked answerability, rational dialogue would falter. But if accountability collapses, the entire structure of praise, blame, punishment, and reward becomes unjustified. That is why this paper will focus on accountability. The central question is whether, in a deterministic world, it is fair to hold people liable in ways that presuppose they deserve moral responses.

3. Incompatibilism

Incompatibilism argues that determinism and moral responsibility cannot coexist. Two of the most influential incompatibilism strategies appeal to alternative possibilities and source-hood.

DOI: 10.54254/2753-7064/2025.ND28262

3.1. The consequence argument

The most intuitive formulation of incompatibilism is van Inwagen's Consequence Argument [1]. This classical argument proceeds from the claim that if determinism is true, then the conjunction of the past and the laws of nature would surely entail every fact about the future. Thus, it follows that since no agent has power over the past or the laws, no agent has power over their future actions. In short formal terms: (1) No one has control over the past or the laws of nature. (2) If determinism is true, the past and laws entail every future event. (3) Therefore, no one has control over the future, including their own actions [14]. This conclusion suggests that if determinism holds, agents could never have acted otherwise [15,16].

Compatibilists have devoted themselves to undermining the Consequence Argument, often by contesting its central claim that determinism precludes the ability to do otherwise. One early strategy was the conditional analysis of ability [17,18]. According to this view, to say that an agent "could have done otherwise" is simply to say that they had chosen or willed otherwise, they would have acted otherwise. On this analysis, determinism poses no threat: even if their actual choice was determined, the counterfactual statement about what they would have done given a different choice remains true. Thus, alternative possibilities are preserved in a conditional sense.

Incompatibilism has long argued that this move is inadequate. As van Inwagen and Ginet point out, conditional analyses confuse the question of whether an agent had the power to choose differently with the hypothetical consequences of already having chosen differently [1,15]. If determinism is true, then at the actual moment of decision, the agent could not have formed a different choice, so the conditional gloss provides only the illusion of freedom.

3.2. Source-hood and ultimacy

A second strand of incompatibilism reasoning emphasizes source-hood rather than leeway. The Source Incompatibilism Argument holds that an agent acts freely only if they are the ultimate source of their actions. To be the ultimate source requires that at least one necessary condition for their action originates within them, and is not entirely traceable to events and laws predating their existence [19-21]. If determinism is true, then all conditions sufficient for action are already set by the past and natural laws, leaving the agent without genuine origination.

Robert Kane defends a libertarian version of this view, grounding ultimacy in self-forming actions [3]. By contrast, defenders of incompatibilism such as Ginet and Widerker stress that compatibilist theories fail to account for source-hood. If characters and deliberative mechanisms are themselves products of deterministic causes, then appeals to reasons-responsiveness or higher-order desires cannot establish ultimate origination [2,22]. It makes sense, since it is hard to clearly define where these indeterministic moments come from, and how the agent is able to regain freedom in these moments when they make choices that influence their lives.

Skeptics push the objection further. Pereboom contends that indeterminism itself cannot secure source-hood, since if a decision is undetermined, its outcome risks being a matter of chance rather than an agent's authorship [7,23]. Neil Levy develops this into the luck problem, arguing that indeterministic accounts collapse responsibility into randomness [8]. For skeptics, therefore, neither determinism nor indeterminism delivers genuine source-hood, leaving moral responsibility unsupported.

DOI: 10.54254/2753-7064/2025.ND28262

4. Frankfurt-style cases and their critics

An important turning point for the compatibilist was the argument provided by Frankfurt against the consequence argument. If the Consequence Argument secures incompatibilism by showing that determinism rules out the ability to do otherwise, then one natural compatibilist strategy is to deny that such an ability is required in the first place. This is the thrust of Harry Frankfurt's famous 1969 paper [4]. Frankfurt challenges the Principle of Alternative Possibilities (PAP), which claims that a person is morally responsible for what they have done only if they could have done otherwise. If PAP is false, then a part of incompatibilism loses its force.

Frankfurt's argument turns on a class of counterexamples now known as Frankfurt-style cases (FSCs). The basic structure is simple. Imagine an agent, Jones, who is deciding whether to perform some action, such as shooting Smith. Unknown to Jones, there is another agent, Black, who is monitoring him. Black has a failsafe mechanism: if Jones shows any sign that he will not shoot Smith, Black will intervene—perhaps by covertly manipulating Jones's brain—to ensure that the action occurs. However, Jones proceeds to carry out his plan to shoot Smith and does not leave a chance for Black to intervene. In such a scenario, Jones acts freely and for his own reasons, but he could not have done otherwise, since Black was prepared to intervene had he wavered [22].

The intuitive force of these cases is that Jones still seems morally responsible for his action, even though alternatives were never genuinely open to him. If this is right, then PAP is false, as agents can be responsible for choices that had no alternatives. Instead, responsibility may only require that the agent act from her own motives or capacities, what John Martin Fischer and Mark Ravizza later call guidance control [5]. This kind of control does not demand genuine alternatives but rather that the agent's action issue from their own reasons-responsive mechanism. Frankfurt's thought experiment thus provided compatibilists with a powerful tool: even if determinism precludes the ability to do otherwise, responsibility might survive.

Despite their influence, Frankfurt-style cases (FSCs) have faced a major line of criticism known as the dilemma objection [2,3,20]. The objection argues that FSCs cannot successfully show that responsibility does not require the ability to do otherwise, because they rest on an unstable middle ground.

The reasoning is as follows: (1) Deterministic interpretation: If the link between Jones's mental states (say, his inclination to shoot Smith) and his final decision is deterministic, then his action is already fixed by prior states and the laws of nature. On this interpretation, the case presupposes determinism and thus simply assumes what incompatibilism denies—that agents can be responsible without alternative possibilities. In other words, the example begs the question against incompatibilism. (2) Nondeterministic interpretation: If instead the link is nondeterministic, then it remains possible, however unlikely, that Jones could have formed a different intention (for instance, to kill Dave instead). If that had happened, Black would have intervened to ensure that Jones carried out the original plan. This residual possibility, sometimes called a "flicker of freedom," means that Jones still had an alternative available, so the Principle of Alternative Possibilities (PAP) is not truly refuted.

The second critique of the FSCs seems rather weak compared to the former. The FSCs, at its core, is trying to create scenarios where the agent can be responsible for an action without the existence of possible alternatives. However, the nondeterministic interpretation does not fully explain how a "flicker of freedom" is linked to Jones's action. There is a difference between trying to have a different outcome and actually succeeding in a different outcome.

Other compatibilists refine FSCs to avoid the dilemma objection. Mele and Robb introduce "buffered" cases in which the intervener is sensitive not to every small flicker but to a more robust

DOI: 10.54254/2753-7064/2025.ND28262

sign that the agent is about to decide otherwise [22]. This eliminates trivial alternatives while keeping the core intuition intact. Similarly, Fischer and Ravizza argue that responsibility should be analyzed in terms of guidance control, which requires that the mechanism leading to action is moderately reasons-responsive and owned by the agent [5]. Frankfurt cases demonstrate that guidance control can exist even in the absence of regulative control (the ability to choose between alternatives).

5. Compatibilist defenses

Compatibilists respond to incompatibilism worries by shifting the focus away from metaphysical leeway and toward the actual basis of action. They argue that what matters for accountability is whether an agent's conduct issues from their own deliberative capacities in the right way, not whether they could have done otherwise in some absolute sense. Three of the most influential approaches are guidance control, reasons-responsiveness, and Strawson's accounts of moral practices.

5.1. Guidance control

John Martin Fischer and Mark Ravizza distinguish between two forms of control: regulative and guidance. Regulative control involves the ability to choose between genuine alternatives, while guidance control involves acting through a mechanism that is reasons-responsive and owned by the agent. Fischer argues that responsibility requires only the latter [5]. Even if determinism rules out alternative possibilities, an agent may still act freely if his conduct flows from his own deliberative processes. In this way, Jones in the Frankfurt case can be responsible; he acts for his own reasons and not from Black's intervention.

5.2. Reasons-responsiveness

Guidance control is specified in terms of reasons-responsiveness. An agent can be morally responsible if the mechanism that produced their action would, to some sufficient extent, respond to logic and reason. For example, Frank Zappa plays the banjo; would he be able to refrain from playing if there are strong reasons not to, such as the arising of urgent events?

An agent is morally responsible if the mechanism that produces their action would, in some range of possible situations, respond to sufficient reasons. For example, if Frank Zappa plays the banjo, it can be asked: if presented with strong reasons not to, would his deliberative system refrain? Similar to previous compatibilist arguments against the Consequence Argument: If the answer is yes in counterfactual scenarios, then his action is responsive to reasons, even if determinism ensures that he acted the way he did in the actual world [5,23]. A model preserving accountability under determinism.

5.3. Strawson's accounts

P. F. Strawson's influential essay reframed the free will debate [11]. This takes on a very similar perspective to this paper's view. Strawson argued that moral responsibility is grounded in the "reactive attitudes"—resentment, gratitude, indignation—that structure interpersonal lives. These practices are so deeply woven into human life that they cannot coherently be suspended, even if determinism is true. For Strawson, the question is not whether determinism is metaphysically

DOI: 10.54254/2753-7064/2025.ND28262

compatible with free will but whether determinism threatens the legitimacy of moral practices. Since these practices are constitutive of responsibility itself, compatibilism follows.

In a sense, this is a consequentialist bias where Strawson disregards the metaphysical proofs of incompatibilism. Instead, he believes that such abstraction is not beneficial for our daily practices. However, there seems to be some circular reasoning in this argument as he uses the existence of moral practices to argue that such practices indeed are valid, despite not being supported on philosophical grounds.

6. Further objections and replies

6.1. Manipulation cases

Critiques of incompatibilism have argued that compatibilist theories fail to reach consistent conclusions when dealing with deterministic forces versus manipulation scenarios. Suppose an agent is captured to a concentration camp at a young age and is programmed to have the very reasons-responsive mechanism Fischer describes. If manipulation undermines responsibility, so too does determinism [7]. Furthermore, if the agent happens to be unfortunate and is brainwashed to conduct immoral deeds, is it fair to hold that agent accountable?

Compatibilists reply by stressing ownership. Fischer and Ravizza argue that an agent is responsible only if they have taken responsibility for their mechanism—internalizing its role in their deliberations. Manipulated agents, unlike ordinary determined agents, have not developed this ownership [5]. Others, like Mele, distinguish the extremes of manipulation and the causation relationships that exist in our natural world, noting that not all causations undermine responsibility [24,25].

6.2. Moral luck

Another challenge comes from the idea of “moral luck.” An argument that seems slightly trivial. Theorists argue that luck is part of a factor in what shapes someone as blameworthy or praiseworthy [8,26]. For instance, two equally reckless drivers may differ in blame if one happens to hit a pedestrian while the other does not. If responsibility is contingent on luck, then is compatibilism simply preserving an unfair practice?

The answer to this is quite simple. Although the outcome may involve luck, accountability tracks the agent’s reasons-responsive quality of view. On this view, both agents are equally blameworthy. Additionally, the fact that moral practices are dependent on the realities of life doesn’t make them illegitimate. People can nevertheless hold each other accountable; it would be even more unfair and counterintuitive if either driver were exempt from blame.

7. Conclusion

The discussion has yet to come to an end for this millennia-long debate. Throughout history, incompatibilism provided powerful arguments. From the consequence argument to the source incompatibilism arguments. Libertarians try to undermine determinism entirely through various methods. Both sides leave no room for morality, nor any justification to assess moral responsibility, which is an integral part of human society.

Compatibilism, however, offers the most appealing framework for accountability. By focusing on reasoning such as guidance, control or reasons-responsiveness, it preserves the integrity of moral practices that cannot simply be erased. Accountability only demands that the agent’s will and

DOI: 10.54254/2753-7064/2025.ND28262

responsiveness to reasons are expressed, not whether they could defy natural laws. This view preserves the fairness of actions while avoiding the excess abstraction of libertarianism and the nihilism of skeptics.

If determinism is true, agents may not transcend the laws of nature or the causal past. Yet they remain beings who deliberate, respond to reasons, and interact within moral communities. To hold one another accountable is to recognize and sustain these capacities. Compatibilism, therefore, secures the kind of responsibility that matters most: responsibility that grounds praise, blame, and justice in a way that is philosophically defensible.

References

[1] van Inwagen, P. (1983) An Essay on Free Will. Oxford: Oxford University Press.
[2] Ginet, C. (1996) In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing. Philosophical Perspectives, 10, 403–417.
[3] Kane, R. (1996) The Significance of Free Will. New York: Oxford University Press.
[4] Frankfurt, H. G. (1969) Alternate Possibilities and Moral Responsibility. The Journal of Philosophy, 66, 829–839.
[5] Fischer, J. M. and Ravizza, M. (1998) Responsibility and Control: A Theory of Moral Responsibility. Cambridge: Cambridge University Press.
[6] Vihvelin, K. (2013) Causes, Laws, and Free Will: Why Determinism Doesn't Matter. Oxford: Oxford University Press.
[7] Pereboom, D. (2001) Living Without Free Will. Cambridge: Cambridge University Press.
[8] Levy, N. (2011) Hard Luck: How Luck Undermines Free Will and Moral Responsibility. Oxford: Oxford University Press.
[9] Watson, G. (1996) Two Faces of Responsibility. Philosophical Topics, 24(2), 227–248.
[10] Shoemaker, D. (2011) Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility. Ethics, 121(3), 602-632.
[11] Strawson, P. F. (1962) Freedom and Resentment. Proceedings of the British Academy, 48, 187–211.
[12] Frankfurt, H. G. (1971) Freedom of the Will and the Concept of a Person. The Journal of Philosophy, 68, 77-91.
[13] Watson, G. (1987) Responsibility and the Limits of Evil: Variations on a Strawsonian Theme. In F. D. Schoeman (ed.) Responsibility, Character, and the Emotions: New Essays in Moral Psychology. Cambridge: Cambridge University Press, 256–286.
[14] O'Connor, T. (2022) Free Will. The Stanford Encyclopedia of Philosophy. Retrieved from https: //plato.stanford.edu/archives/win2022/entries/freewill/
[15] Ginet, C. (1990) On Action. Cambridge: Cambridge University Press.
[16] van Inwagen, P. (1975) The Incompatibility of Free Will and Determinism. Philosophical Studies, 27, 185–199.
[17] Ayer, A. J. (1954) Freedom and Necessity. In A. J. Ayer. Philosophical Essays. London: Macmillan, 271–284.
[18] Lewis, D. (1981) Are We Free to Break the Laws? Theoria, 47, 113–121.
[19] McKenna, M. and Coates, D. J. (2024) Compatibilism. The Stanford Encyclopedia of Philosophy. Retrieved from https: //plato.stanford.edu/archives/sum2024/entries/compatibilism/
[20] Widerker, D. (1995) Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities. The Philosophical Review, 104, 247–261.
[21] Pereboom, D. (2014) Free Will, Agency, and Meaning in Life. Oxford: Oxford University Press.
[22] Talbert, M. (2025) Moral Responsibility. The Stanford Encyclopedia of Philosophy. Retrieved from https: //plato.stanford.edu/archives/fall2025/entries/moral-responsibility/
[23] Mele, A. and Robb, D. (1998) Rescuing Frankfurt-Style Cases. Philosophy and Phenomenological Research, 58, 162–173.
[24] Fischer, J. M. (1994) The Metaphysics of Free Will: An Essay on Control. Oxford: Blackwell.
[25] Wolf, S. (1990) Freedom Within Reason. New York: Oxford University Press.
[26] List, C. (2019) Why Free Will Is Real. Cambridge: Harvard University Press.