

Reasons and Causality: Critiquing the Principle of Alternate Possibilities

Patrick Van Hoven '24
Woodside High School
California, USA

Abstract

When discussing methods of evaluating moral responsibility, one criterion is referenced by both laypeople and academics with undeniable frequency: the Principle of Alternate Responsibilities (PAP). This principle holds that in instances where agents cannot refrain from doing an action, or cannot do any other action, they cannot be considered morally responsible for that action. As moral philosophers such as Daniel Dennett and Harry Frankfurt have pointed out, PAP is often taken to be trivially obvious in cases where people act due to obvious coercion. However, PAP is not always so useful. This paper presents a critique of PAP and proceeds in two parts. First, a definition of moral responsibility, namely a *reasons-responsiveness definition*, is presented and lightly defended. Second, PAP is shown to fail to adequately account for moral responsibility in relation to the reasons-responsiveness definition in three categories of situations. These situations reveal that there is a plenitude of instances in which agents are precluded from acting otherwise, or from preventing certain consequences from occurring, and yet by virtue of their acting on account of internal reasons they earn moral responsibility. This leads to the conclusion that criteria of moral responsibility that make evaluations based on external circumstances have the structural flaw of not paying due attention to the specific moral agent in question.

Introduction

The moral evaluation of agents is a practice endemic to the human experience. In politics, in family, in friendships, and in most other areas of life, we make moral judgments all the time. Moral judgments form the core of our stances on certain political issues, assist us in determining whom to avoid and whom to associate with in a social context, and allow us to evaluate the merits of certain actions that we are considering. But these judgments, these moral evaluations, all depend on a key assumption, one taken to be trivially obvious in many instances: that whoever or whatever is being judged is somehow deserving of that judgment. This key assumption, necessary to justify moral evaluation, is the concept of moral responsibility.

On a base level, moral responsibility is the property that a person can justifiably be held morally accountable for an action that they performed. That is, if someone is morally responsible for an action, it makes sense to apply moral labels to that person based on their action.¹ If a person acts and it is determined that they are not morally responsible, it cannot then be said that *the person* acted morally or immorally. The question of whether or not *the action itself* was moral or immoral, or whether actions separated from agents can have moral value at all, are separate questions altogether but are not the question of moral responsibility. Moral responsibility concerns the degree to which we can fairly morally evaluate *an agent* in light of an action they performed.

One of the first criteria that philosophers reach for when discussing moral responsibility is known as the Principle of Alternative Possibilities (PAP). As stated most influentially by Harry Frankfurt, PAP holds that a person is morally responsible for an action only if they could have refrained from doing that action, or could have done another action.² This paper argues that while PAP may make intuitive sense on a surface level, it does not hold up under closer examination as a viable criterion for moral responsibility.

PAP is often casually invoked in discussions of both moral responsibility and free will, which are related, but different, topics. Importantly, this paper will not lay out an alternate or supplemental list of appropriate criteria for moral responsibility, and will also not discuss

¹ T. M. Scanlon, *What We Owe to Each Other* (Harvard University Press, 1998), 248.

² Harry G. Frankfurt, "Alternate Possibilities and Moral Responsibility," *The Journal of Philosophy* 66, no. 23 (December 4, 1969): 829–39, <https://doi.org/10.2307/2023833>, 829; PAP is not a definition of moral responsibility, but a condition used in its evaluation.

the relationship between PAP and free will. It will only critique the single PAP criterion as it pertains specifically to moral responsibility.

It is worth discussing why PAP is such a common, and intuitive, principle. First, PAP has been an element of conceptions of freedom of action throughout history. PAP has appeared in some form in ancient and medieval philosophy, and the work of early-modern thinkers like John Locke,³ and is rarely critiqued by philosophers.⁴ This reveals how ingrained it is in discussions of moral responsibility. Outside of academic philosophy, PAP is also casually invoked in everyday discussions of moral guilt and responsibility,⁵ typically regarding a specific type of scenario that Daniel Dennett calls a “local fatalism.”⁶ When analyzing an agent’s moral responsibility, it is common, in both juridical and common life, to search for circumstances that controlled that agent’s action by ensuring that they could not do otherwise, regardless of intention or desire. Circumstances such as being locked in a room or having a gun held to your head are common examples of local fatalisms that are assumed to disavow someone of moral responsibility.⁷ Importantly, though, typical searches for and analyses of local fatalisms of this nature not only overlook other types of situations in which an agent might not be able to do otherwise but also ignore key aspects of the moral and decision-making process of the agent, aspects that have a substantive bearing on moral responsibility. This paper will unpack the limitations of such local-fatalism analysis.

Defining Moral Responsibility

Definitions are often some of the most debated and analyzed elements of philosophy, and this is for good reason. Meaningful and productive discourse is almost impossible without well-defined terms. Interestingly, there has been only relatively limited debate over what defines the condition of being morally responsible, at least compared to

³ Robb David, “Moral Responsibility and the Principle of Alternative Possibilities,” ed. Edward N. Zalta, *The Stanford Encyclopedia of Philosophy*, fall 2020, <https://plato.stanford.edu/archives/fall2020/entries/alternative-possibilities/>; William L. Rowe, “Two Concepts of Freedom,” *The American Philosophical Association Centennial Series* 61, no. 1 (September 1987): 43–64, <https://doi.org/10.5840/apapa2013170>.

⁴ Daniel C. Dennett, “I Could Not Have Done Otherwise – So What?,” *The Journal of Philosophy* 81, no. 10 (October 1984): 553–65, <https://doi.org/10.5840/jphil1984811022>, 553.

⁵ *Guilt* and *responsibility* will be used interchangeably. *Guilt* will be used not in the legal sense of conviction by a jury but in the same sense of *responsibility*, referring to everyday interpersonal evaluations of moral responsibility.

⁶ Dennett, “I Could Not Have Done Otherwise – So What?,” 554.

⁷ *Ibid.*

other philosophical definitions of topics such as free will or knowledge. Of course, there has been some debate, debate that has resulted in two primary definitions of moral responsibility. Fortunately, they are fairly intuitive.⁸

One popular approach to moral responsibility was first formulated by P. F. Strawson in his influential paper “Freedom and Resentment.”⁹ Strawson’s approach is essentially inductive; for Strawson, moral responsibility is not normatively deductible *a priori* but is rather based on the justified reactions of others to action. Strawson labeled these justified reactions as “reactive attitudes.” For example, if doing an action warrants the reactive attitude of disgust directed *towards the agent*,¹⁰ then the agent who performed the action would be morally responsible for that action. This responsibility emerges from a historical process of *taking responsibility*; as children grow up, they observe the effects of their words and actions on others, and, provided they are of relatively sound mind,¹¹ children begin to view those reactions as justified, due to the child’s own similar reactions to the words and actions of others. It is through this process that human agents *take* moral responsibility for their words and actions; in this sense, moral responsibility is somewhat of an emergent property. Strawson’s conception of moral responsibility is useful for two reasons. One, Strawson’s account avoids metaphysical or normatively ethical claims about responsibility, and the endless complexity that comes with such claims.¹² Two, it is intuitive, based on phenomenological reactions to actions, reactions that most people have experienced. Moral responsibility, unlike related topics like free will, is a quality that we frequently externally gauge in each other. It makes natural sense that moral responsibility could be defined in relation to that external evaluation.

The second notable conception of moral responsibility is a family of related conceptions known as *reasons-responsiveness views*,¹³ most notably articulated by John Fischer and Mark Ravizza.¹⁴ The reasons-

⁸ David, “Moral Responsibility and the Principle of Alternative Possibilities.”

⁹ Peter F. Strawson, “Freedom and Resentment,” *Proceedings of the British Academy* 48 (1962): 187–211.

¹⁰ It is important that in Strawson’s definition, only reactions directed at the agent himself matter; reactions at the agent’s action are irrelevant here.

¹¹ Psychopathy, sociopathy, etc. may annul moral responsibility under a Strawsonian framework.

¹² This isn’t to devalue such normative or metaphysical approaches. However, it passes a sort of Occam’s Razor test due to its relative simplicity, and is as such useful for moral analysis.

¹³ David, “Moral Responsibility and the Principle of Alternative Possibilities.”

¹⁴ John Fischer, *The Metaphysics of Free Will: An Essay on Control* (Wiley-Blackwell, 1994); John Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*

responsiveness definition of moral responsibility is easier stated than Strawson's reactive attitudes approach: if an agent does an action because of internal reasons or motivations, that agent can be held morally responsible for that action. In other words, an agent is morally responsible for an action if "there is a nomologically possible world... in which the same mechanism operates, there is sufficient reason to do otherwise, and the agent recognizes the reason, chooses, and acts on it," the "mechanism" being the set of external, incidental, and possibly limiting circumstances of the situation.¹⁵

As illustrated, reactive-attitude and reasons-responsiveness approaches align with intuitive notions of moral responsibility, and simply approach defining moral responsibility from different directions. Strawson takes a dialectically-external approach to moral responsibility in his reactive attitude definition, defining it in terms of the justified reactions of other agents. Fischer takes a dialectically internal approach, identifying the condition of moral responsibility with the inner conditions and deliberative processes of the agent.

The problem, though, with Strawson's definition is that its social approach to defining moral responsibility makes it difficult to rationally determine the moral responsibility of a person without leaning on pure intuition. Gauging whether someone's action warrants justified reactive attitudes is largely a question of precedent, intuition, and ultimately, moral opinion. Such a definition is not suited for examining the logical components of moral responsibility, as will be done throughout the rest of the paper, and thus Strawson's definition will be of no use in this context. In order to evaluate the extent to which PAP acts as a viable criterion for moral responsibility, we shall proceed in relation to a reasons-responsiveness (RR) definition of moral responsibility, according to which moral responsibility can be evaluated based on facts about the internal deliberative processes of an agent.

A note on methodology: Evaluations of PAP against the RR definition attempt to determine the extent to which an agent acts on their own internal reasons, regardless of external limitations. If scenarios can be found that involve agents that cannot do otherwise, but that act on account of their own reasons, PAP would fail under the RR definition.

(Cambridge: Cambridge University Press, 1998),
<https://doi.org/10.1017/CBO9780511814594>.

¹⁵ Kadri Vihvelin, "Review: The Metaphysics of Free Will by John Martin Fisher," *Nous* 32, no. 3 (September 1998): 406–20, <https://www.jstor.org/stable/2671941>, 408.

PAP's Inconclusiveness

As mentioned earlier, moral responsibility is an external quality as much as an internal one. As such, examples and illustrative scenarios are especially effective in shedding light on aspects of moral responsibility. Note that the following scenarios will be dramatized – this is merely to aid in building intuition. Furthermore, there are three primary varieties of situations in which agents may not be able to do otherwise: cases of historical blame, Frankfurt-style cases, and transfer of non-responsibility cases. These will be examined in turn.

Cases of Historical Blame

This first type of case will not reveal the inconclusiveness of PAP, though further cases will. What this first case will do is shed light on how PAP as typically formulated is too narrowly defined to account for the moral responsibility it is designed to evaluate. To reiterate, PAP holds that an agent is morally responsible for a certain action if and only if they could have refrained from performing that action. This principle, though, fails to account for the effects of our past actions on our future actions over time. Specifically, in cases in which we were morally responsible for an action in the past that is directly causing a present action that we cannot avoid doing, PAP would hold that we are not morally responsible for the present action, at least in a temporally limited interpretation of PAP. This, however, is not true.

Imagine a situation where an agent, Ben, concocts a drug that, when taken, causes the recipient to spontaneously feel an irresistible compulsion to murder someone exactly 24 hours later. Imagine that Ben, knowing the effects of the drug, takes this drug himself and kills someone 24 hours later. In evaluating Ben's level of moral responsibility for the killing, it might be noted that in a certain limited sense, he could not have refrained from the killing, due to the overpowering effects of the drug on his internal reasons, intents, and dispositions. And yet, he did choose to take the drug in the first place with full knowledge of how it would affect him and was thus morally responsible for doing so. Is Ben morally responsible for the killing, even though he could not have refrained from doing so? The answer is clearly yes.

This example reveals a flaw in the common formulation of PAP: it doesn't account for cases in which actions were responsibly performed in the past that cause or influence present actions. This is not an end-all-be-all argument against PAP. A defender of PAP might respond that Ben could have done otherwise when he chose to take the drug in the

first place and thus was morally responsible for the killing. This – assuming for a second that PAP is valid – is true. What cases of historical blame illustrate is that for PAP to function well as a criterion for moral responsibility, it must take a wider view than merely the circumstances of the action at hand. For PAP to make sense at all, it must also take into account the past actions of an agent for which that agent *was* morally responsible. If an agent is morally responsible for putting themselves in a situation in which they cannot do otherwise, their moral responsibility evidently carries over, even if, examined locally, they act in that situation without being able to refrain from thus acting. If an agent has the ability to avoid a situation in which they may be unable to do otherwise and does not, then it is clear that the agent assumes moral responsibility for their future actions in that situation. PAP, as typically formulated, does not specify such a wider view as it focuses on incidental present circumstances and thus fails to adequately prove what its adherents desire it to.

Frankfurt-Style Cases

In contrast to cases of historical blame, which illustrate problems with the analytical scope of PAP, Frankfurt-style cases provide evidence for a direct critique of the central normative claim of PAP. Two scenarios will be presented. Observe that the scenarios differ from each other in only a few ways, none of which have any relation to the freedom to do otherwise. Also, observe the probable evaluations of moral responsibility included in each example; the examples were chosen so that those evaluations are intuitively obvious. They will be normatively justified *post hoc*.

Scenario A: John is a vegan, and is abducted by a group of mysterious individuals. He is told that unless he eats a bite of meat, he will be killed. He eats the meat. His vegan friends rightly do not blame him for this; they rightly do not believe he was morally responsible for transgressing on his veganism.

Scenario B: Mike is *not* a vegan, but he is *very* hungry and is abducted by a similar group of mysterious individuals. He is told that unless he eats a bite of meat, he will be killed. He eats the meat. His vegan friends rightly *do* blame him for this,¹⁶ as they believe that eating meat is immoral; they rightly hold him morally responsible for his action.

¹⁶ *Rightly* in the sense that Mike is a philosophically-appropriate target of blame, not that what he did – eating meat – is bad or blameworthy.

First, note that the only difference between the two scenarios is the agent's preordained attitude toward eating meat. In the first scenario, scenario A, the agent, John, is a vegan; his preordained position is to not eat meat. In B, Mike is not a vegan and is quite hungry. In A, removing the coercive threat from the scenario would lead to John not eating the meat. In B, removing the coercive threat would likely do nothing to change the outcome – Mike is hungry and has no aversion to eating meat. It would make sense that he would consume the available food.

These two scenarios are of a kind made prominent by Harry Frankfurt and are commonly referred to as “Frankfurt-style counterexamples.”¹⁷ Such scenarios are compelling arguments against PAP. Note that in the first example, the coercive force of the threat is strong enough that it is fair to say that John could not have done otherwise, not with the threat of certain death as the price for maintaining his vegan piety. It would be right in that instance to hold John morally responsible.

By the RR definition of moral responsibility, it is clear that John ate the meat not because of internal attitudes, dispositions, or reasons, but rather due to the external threat of death. He was not reasons-responsive. If there were to be an alternate but nomologically possible world in which the abductors were absent, it is likely that John would have had internal reasons to not eat the meat due to his vegan morals, would have recognized those reasons, and would have acted on those reasons, not eating the meat.¹⁸

It has been established that in A, John was not morally responsible under the RR definition. This is not the case in B – Mike is guilty under the RR definition in this example, as he is reasons-responsive. It is likely that, should the coercive threat be removed from the scenario, he would still eat the meat, because he is a hungry meat eater. He would recognize his internal reasons to eat the meat, and would likely act on those reasons – in that sense, he has moral responsibility.

To reiterate, in both scenarios, there was a coercive force of sufficient strength to preclude the ability to do otherwise.¹⁹ If the agent

¹⁷ Frankfurt, “Alternate Possibilities and Moral Responsibility,” 831.

¹⁸ Vihvelin, “Review: The Metaphysics of Free Will by John Martin Fisher,” 408.

¹⁹ If the reader is not convinced of this, it should be noted that the thought experiment can be refined to address almost any objection to the strength of the coercive force. See Frankfurt, “Alternate Possibilities and Moral Responsibility,” 835; and Vihvelin, “Review: The Metaphysics of Free Will by John Martin Fisher,” 407.

could not have done otherwise in both scenarios but *was* morally responsible in one of them, scenario B, PAP fails as a criterion for moral responsibility in cases such as this. The problem with PAP is that “[t]he fact that he could not have done otherwise clearly provides no basis for supposing that he *might* have done otherwise if he had been able to do so.”²⁰ The relevant question, the question overlooked by PAP, is not if the agent *could* have done otherwise. It is if the agent *would* have done otherwise.

When examining cases of coercion such as the examples presented above, it is usually implicitly assumed that the coercive impulse was the preeminent, and indeed the only, factor in the agent’s decision-making process. This makes salient the fact that what matters in assessing moral responsibility is not the mere existence of coercion, but the way in which coercion impacts the agent’s eventual decision. If a person was about to get out of their car, and a gunman appeared and told them to get out of their car or face death, it would be unfair to say that that person got out of their car solely because of the gunman’s coercion, because the person was *already* internally motivated to do so. According to Frankfurt, “[s]ituations in which a person who does something cannot do otherwise because he is subject to coercive power are either not instances of coercion at all, or they are situations in which the person may still be morally responsible for what he does if it is not because of the coercion that he does it.”²¹

One possible objection here is that when an agent is coerced to do an action that they already had internal reasons to do and end up doing that action, it is because of the coercive force and not of the internal reason that they do the action, despite the preexistence of the internal reason. In the context of the previous example about eating meat, Mike was predisposed to eat the meat, and then was coerced to eat the meat, and then in fact ate the meat. It could be reasonably proposed that the final reason for Mike eating the meat was coercion, and thus the inability to do otherwise caused by the coercion does in fact relieve him of moral responsibility. There are a few reasons such a view of moral responsibility should be rejected. First, there is little reason to assign moral weight to reasons for action in terms of the order in which those reasons appear, especially when considering a local and narrow timeframe. While it is true that the coercive impulse to eat the meat was the *last* reason Mike was introduced to before eating the meat, there is no reason to suppose that it was the *primary* or most *morally relevant* reason that he acted the way that he did. The fact an agent has internal reason

²⁰ Frankfurt, “Alternate Possibilities and Moral Responsibility,” 837.

²¹ *Ibid.*, 833–834.

and intention to perform an action and is soon after confronted with a coercive impulse to do that same action provides no reason at all to ignore that original internal reason and intention. Second, the irrelevance of an agent's inability to do otherwise in the face of coercion or local fatalism is made salient when the coercive reasons are hidden from the agent's deliberative view, and are thus not present in the agent's decision-making process. Frankfurt offers a refined illustration of the inconclusiveness of PAP by examining an instance in which coercive force remains hidden from the agent.²² He imagines an instance in which a coercer, Black, decides in the following way to ensure that an agent, Jones, does a certain action: Black will wait in hiding until Jones has made up his mind, and decide accordingly what steps to take.²³ If Jones decides to perform the action that Black does not desire, Black will intervene and coerce Jones into acting how Black wants. If, on the other hand, Jones decides to perform the action that Black does desire, Black will do nothing, and Jones will never even know of his existence or conditional intent to coerce him – Jones will never know that he could not have done otherwise. As such, in the case where Jones acts according to Black's wants, he does so solely on the merits of his own internal reasons, as he was never even aware of Black's desires or conditional intent to coerce him. This is an instance where Jones could not have done otherwise – Black would have stopped him – and yet was utterly morally responsible, on an RR account of moral responsibility, as he decided to perform the action solely due to his internal reasons.

Specifically, in regard to the RR definition of moral responsibility, it can be seen that PAP is not a necessary *or* sufficient condition. RR does not require PAP as a criterion because it is possible to act in the presence of coercion, or in any other circumstance where all alternate possibilities are blocked, and still act primarily under internal reasons. While additional reasons stemming from coercion or local fatalisms may also play a role in deciding how to act, the preexistence of internal reasons to do that action is morally significant and often rises to the level of responsibility. In that light, evaluations of moral responsibility in cases where coercion is present must go further than merely denying moral responsibility based on the presence of coercion. Evaluations in such instances must evaluate the extent to which the coercion directly influenced the agent's decision, whether or not the coercion aligned with the agent's previous internal reasons (or lack thereof), and whether or not the agent's eventual action would be identical in a nomologically-possible world absent of coercion. If it can be reasonably argued that the

²² *Ibid.*, 835–836.

²³ For the sake of illustrative convenience, Black is assumed to have a way of determining Jones' intention before Jones acts on it.

coercion contradicted the internal reasons of the agent, and if it can be reasonably argued that should the coercion not have been present, the agent would have acted differently, then it can be reasonably argued that the coercion was the causal force behind the action and thus the agent would be disavowed of moral responsibility. The mere presence of an inability to do otherwise, though, provides no conclusion in and of itself. The causal force of the coercion on the agent's action must be proven.

Cases of the Transfer of Non-Responsibility

A second variety of PAP, known as the transfer of non-responsibility, or transfer-NR, is commonly invoked as a logical entailment of causal determinism. Causal determinism is the principle that any event is the direct and only result of the antecedent events and conditions along with the governing laws of nature.²⁴ In light of that definition, the following argument is often presented: "If no one is morally responsible for the fact that P, and no one is morally responsible for the fact that if P then Q, then it follows that no one is morally responsible for the fact that Q."²⁵ The key element here is that determinism mandates that if P is a past state, then Q, where Q is a present state, must obtain because determinism holds that present states are completely determined by past states. If P is a state in the past, and Q is a current state where an action is being performed by an agent, determinism principally holds that Q is inevitable as a result of natural law, and thus nobody is morally responsible for Q, as it did not arise as a result of a choice made on the account of an agent's internal reasons. Thus, if Q is a present state that contains an action being performed by an agent, and nobody is morally responsible for the state Q, nobody is morally responsible for any actions occurring within Q, or so it is argued. This is the transfer-NR argument.

First, let us examine why this should be considered a variant of PAP. PAP, recall, is the idea that if an agent could not have refrained from performing an action, they are not responsible for that action. Determinism holds that every action is the direct result of past states and natural laws, and thus every action is outside the real control of agents. Thus, agents cannot refrain from acting in certain ways as those actions are just steps in an inevitable deterministic process. The idea, then, that determinism implies a lack of moral responsibility translates to the idea

²⁴ See Shirley Matile Ogletree and Crystal D Oberle, "The Nature, Common Usage, and Implications of Free Will and Determinism," *Behavior and Philosophy* 36 (2008): 97–111, <https://www.jstor.org/stable/27759547>.

²⁵ Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, 156–157.

that because of determinism, agents have no real choice over their actions, *and thus cannot do otherwise*, and thus are not morally responsible. Transfer-NR is a clear variant of PAP.

Importantly, this paper takes no stance on the validity of determinism. Why? It doesn't need to. The argument in focus in this section, transfer-NR, is the one stated previously, that "if no one is morally responsible for the fact that P and no one is morally responsible for the fact that if P then Q, then it follows that no one is morally responsible for the fact that Q." As some philosophers have argued, such as the aforementioned John Fischer and Mark Ravizza, this argument is not valid, and can be shown to be invalid without impugning the underlying determinist principle.

Fischer and Ravizza present the following scenario as a counterexample to transfer-NR:

Betty plants her explosives in the crevices of the glacier and detonates the charge at T1, causing an avalanche that crushes the enemy fortress at T3. Unbeknownst to Betty and her commanding officers, however, the glacier is gradually eroding (in a certain way). Had Betty not placed the dynamite in the crevices, some ice and rocks would have broken free at T2, starting an avalanche that would have crushed the enemy camp at T3.²⁶

First, observe that this scenario is of a transfer-NR form. Nobody is morally responsible for the fact that the glacier is eroding, and nobody is responsible for the fact that the glacial erosion may cause an avalanche, and yet, as Fischer and Ravizza note, "in virtue of Betty's freely detonating the explosives at T1, she is at least in part morally responsible for the enemy camp's being destroyed at T3."²⁷ Betty's detonation did in part contribute to the avalanche – that we know. Betty was (as is assumed in the example) morally responsible for the detonation she caused; she planted and detonated the explosives due to internal reasons. Thus, Betty was morally responsible for an action that partially contributed to the destruction of the camp.

One specification to make here is that the transfer-NR variant of PAP concerns the responsibility of an agent for the *consequences* of an action they perform and not for the action itself. This can be seen in the Betty example; what is under examination is not that Betty could or could not have done otherwise regarding setting and detonating the

²⁶ Ibid., 157.

²⁷ Ibid.

explosives, but whether or not Betty could have prevented (by not planting the explosives) the eventual consequence of the destruction of the camp. The answer is no, of course – remove Betty and the causal force she emanates from the situation altogether and the same consequence would occur. And yet, by virtue of being morally responsible for an action that contributed to that eventual consequence, Betty at least earns partial responsibility for the eventual consequence. Because transfer-NR is about the consequences of actions and not the actions themselves, the flaw in the transfer-NR argument is that it presumes that a consequence must have only one cause. But this is true only of actions, not consequences. Actions are performed by the agent that performed them, and only that agent. Consequences, on the other hand, can be the aggregate result of a large panoply of causes. If an agent is *wholly* morally responsible for one of those many causes, then the agent must be at least *partially* responsible for the end consequence.

This translates to a more general principle: if there is a situation where you were to remove an agent from the situation and the consequence of that agent's actions still occurs, the agent can still be held morally responsible. This principle is, fortunately, consistent with common intuitions about similar occurrences in life. Take, for example, voting. In an election, many people come together to elect a candidate for some office. Each one of the voter's actions – their action being the casting of their ballot for their preferred candidate – contributes directly to the eventual consequence. When a candidate wins, it is commonly felt, and indeed is true, that each voter who voted for that candidate is in some small part responsible for that candidate's victory. And yet, if you were to remove a single voter's ballot, the result would still occur. The transfer-NR argument would lead to the conclusion that if a single specific voter is not responsible for the votes of every other voter, and not responsible for the fact that those votes determine the winner of the election, then they must not be responsible for the result of the election. But this is false. When agents are morally responsible for contributory actions that lead in part to a specific consequence, they are morally responsible in part for that consequence.

One possible determinist challenge to this argument is that what matters about transfer-NR cases is not that non-human occurrences may lead to a certain consequence – such as the glacier causing an avalanche – but that human agents cannot be responsible for their own actions, and thus cannot be responsible for the consequences of those actions. Why can't human agents be responsible for their own actions? The determinist answer is that because their actions are just the result of past states and natural laws impacting their deliberation and internal reasons,

their action is inevitable – this is, after all, the determinist thesis. Because their actions are just the mechanical output of a natural deterministic formula, the agent never could have not done the action, and thus is not responsible.

This argument fails for reasons unpacked in the previous section. As shown, in evaluations of moral responsibility the inability of an agent to do otherwise is an inconclusive factor. If an agent acts on account of internal reasons, they remain morally responsible on an RR account of moral responsibility. Determinism does not affect this conclusion.

Closing

What this critique of the Principle of Alternate Possibilities reveals is that attempting to use an external principle to gauge moral responsibility is a fool's errand. What matters, in the end, about what makes a person deserving of moral evaluation is the nature of their internal processes and not the external circumstance in which those processes occur. The external circumstances can matter, of course, but they do not matter intrinsically, as PAP holds. They only matter instrumentally, insofar as they directly exert causal influence on the deliberative mechanism and internal environment of the agent in question. However, this causal influence is not always present when agents are simply unable to do otherwise. PAP fails to adequately account for moral responsibility when the grounds for moral responsibility for a given action are found outside the local temporal window of that action; when coercion is present but does not supersede the agent's internal reasons for action; and when an agent's action is the partial cause of a consequence. Such instances reveal that as an evaluative metric, PAP is simply not useful.

In a more abstract sense, it seems only reasonable that moral responsibility be evaluated based on the merits of the individual at hand – the rich inner moral life of the individual deserves to lie at the heart of moral inquiry. Evaluative rules such as PAP attempt to gauge morality – a distinctly human quality made possible by our unique capacity for rational moral analysis – by analyzing external conditions, and fail to adequately focus on the individual moral subject as the center of moral importance.²⁸ What any authentic evaluation of moral responsibility must

²⁸ In regard to humans' unique capacity for moral analysis, see Grace Clement, "Animals and Moral Agency: The Recent Debate and Its Implications," *Journal of Animal Ethics* 3, no. 1 (season-01 2013): 1–14, <https://doi.org/10.5406/janimalethics.3.1.0001> and S.F. Sapontzis, "Are Animals

do is derive its conclusive strength from an analysis of the individual agent in question. To do otherwise would be folly.

Moral Beings?," *American Philosophical Quarterly* 17, no. 1 (January 1980), <https://www.jstor.org/stable/20009783>.

Bibliography

Clement, Grace. "Animals and Moral Agency: The Recent Debate and Its Implications." *Journal of Animal Ethics* 3, no. 1 (season-01 2013): 1–14. doi.org/10.5406/janimaethics.3.1.0001.

David, Robb. "Moral Responsibility and the Principle of Alternative Possibilities." Edited by Edward N. Zalta. *The Stanford Encyclopedia of Philosophy*, Fall 2020. <https://plato.stanford.edu/archives/fall2020/entries/alternative-possibilities/>.

Dennett, Daniel C. "I Could Not Have Done Otherwise – So What?" *The Journal of Philosophy* 81, no. 10 (October 1984): 553–65. <https://doi.org/10.5840/jphil1984811022>.

Fischer, John. *The Metaphysics of Free Will: An Essay on Control*. Wiley-Blackwell, 1994.

———. "Ought-Implies-Can', Causal Determinism and Moral Responsibility." *Analysis* 63, no. 3 (July 2003): 244–50. <https://doi.org/10.1093/analys/63.3.244>.

———, and Mark Ravizza. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press, 1998. doi.org/10.1017/CBO9780511814594.

Frankfurt, Harry G. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66, no. 23 (December 4, 1969): 829–39. <https://doi.org/10.2307/2023833>.

Ogletree, Shirley Matile, and Crystal D. Oberle. "The Nature, Common Usage, and Implications of Free Will and Determinism." *Behavior and Philosophy* 36 (2008): 97–111. <https://www.jstor.org/stable/27759547>.

Rowe, William L. "Two Concepts of Freedom." *The American Philosophical Association Centennial Series* 61, no. 1 (September 1987): 43–64. <https://doi.org/10.5840/apapa2013170>.

Sapontzis, S.F. "Are Animals Moral Beings?" *American Philosophical Quarterly* 17, no. 1 (January 1980). <https://www.jstor.org/stable/20009783>.

Scanlon, T. M. *What We Owe to Each Other*. Harvard University Press, 1998.

Strawson, Peter F. "Freedom and Resentment." *Proceedings of the British Academy* 48 (1962): 187–211.

Vihvelin, Kadri. "Review: The Metaphysics of Free Will by John Martin Fisher." *Nous* 32, no. 3 (September 1998): 406–20.
<https://www.jstor.org/stable/2671941>.