

Claremont Colleges

## Scholarship @ Claremont

---

Scripps Senior Theses

Scripps Student Scholarship

---

2022

### A Criticism of Retributivism: Identifying the Unavoidable and Undesirable Consequences of Accepting of Not Accepting “Ought Implies Can”

Samantha Bloomfield

Follow this and additional works at: [https://scholarship.claremont.edu/scripps\\_theses](https://scholarship.claremont.edu/scripps_theses)



Part of the [Philosophy Commons](#)

---

#### Recommended Citation

Bloomfield, Samantha, "A Criticism of Retributivism: Identifying the Unavoidable and Undesirable Consequences of Accepting of Not Accepting “Ought Implies Can”" (2022). *Scripps Senior Theses*. 1854. [https://scholarship.claremont.edu/scripps\\_theses/1854](https://scholarship.claremont.edu/scripps_theses/1854)

This Open Access Senior Thesis is brought to you for free and open access by the Scripps Student Scholarship at Scholarship @ Claremont. It has been accepted for inclusion in Scripps Senior Theses by an authorized administrator of Scholarship @ Claremont. For more information, please contact [scholarship@cuc.claremont.edu](mailto:scholarship@cuc.claremont.edu).

**A CRITICISM OF RETRIBUTIVISM:  
IDENTIFYING THE UNAVOIDABLE AND UNDESIRABLE CONSEQUENCES OF  
ACCEPTING OR NOT ACCEPTING “OUGHT IMPLIES CAN”**

by

**SAMANTHA BLOOMFIELD**

**PROFESSOR DION SCOTT-KAKURES**

**PROFESSOR YUVAL AVNUR**

**APRIL 29, 2022**

## Introduction

Retributivism as a moral theory makes two claims: “(1) the innocent deserve not to be punished” and “(2) The guilty deserve to be punished proportionately” (Zimmerman 1951, 67). In this essay, I aim to identify retributivism as an impotent or callous moral theory, depending on one’s stance regarding the principle Ought Implies Can. I begin with a discussion and defining of the terms involved: retribution, regulation, punishment, attributability, and accountability, identifying accountability as the form of blameworthiness we concern ourselves with in a discussion of retributivism. I then introduce OIC, a principle which, if accepted, would assert ability as a necessary condition for accountability. It is logically necessary that either retributivism accepts OIC (and ability is a necessary condition for accountability) or it is not the case that retributivism accepts OIC (and ability is not a necessary condition for accountability). Thus, the functionality of retributivism can be thoroughly explored using these two potential cases.

In exploring these cases, I identify luck and the ineliminable fallibility of our doxastic practices (i.e., the practices through which we come to believe) as nonvolitional forces that may lead an agent to wrongdoing. In the case where retributivism accepts Ought Implies Can (this version I call OIC retributivism), I find the retributivism to be impotent. I argue that it is impossible for OIC retributivism to warrant the punishing of any agent. In the case where retributivism does not accept Ought Implies Can (this version I call non-OIC retributivism), I find retributivism to be callous. I argue that non-OIC retributivism would mandate the punishing of some agents who put forth good-faith efforts at behaving morally, and that this implication should be universally unappealing. Ultimately, I conclude that this exploration into the implications of retributivism reveals retributivism to be an impractical theory.

Michael Zimmerman, in his book *The Immorality of Punishment*, expresses two common and distinct incentives for implementing punishment: regulation and retribution. The first incentive, regulation, employs punishment as it functions to “protect the general public from the harm caused by crime” as “accomplished by various means- rehabilitation, incapacitation, and deterrence” (Zimmerman 1951, 78). The second incentive, retribution, employs punishment because of the belief that “the culpable deserve to be punished ... because, more generally, they deserve to suffer” (Zimmerman 1951, 84). In this paper, I use the term ‘punishment’ to refer exclusively for measures taken for retributive purposes (i.e., actions taken with the ultimate end of causing the wrongdoer to suffer on the basis of said wrongdoer’s culpability; actions taken for the ultimate end of vengeance). I will not use the word ‘punishment’ to describe measures that are put into place for regulative purposes (i.e., measures without suffering as their ultimate end). This narrow definition of the word ‘punishment’ is employed to ensure the conceptual distinction between regulation and retribution.

Note, however, that regulative measures can resemble retributive measures when they take the form of adverse treatment. For example, the regulative measure of mandating the quarantining of a person due to her being infected with a dangerous and contagious illness closely resembles the potentially retributive measure of imprisonment (assuming imprisonment is implemented for the ultimate end of causing the imprisoned to suffer, not as a means of incapacitation or deterrence). In both measures, the full force of some coercive authority functions to prevent an agent from circulating freely as she otherwise would. The resemblance of this regulative measure to some retributive measure is insufficient to rightfully consider the regulative measure to be punishment. Since quarantine does not have the ultimate end of causing

the quarantined agent to suffer, quarantine is not punishment.<sup>1</sup> This applies even to regulative measures implemented with the explicit intention of causing the wrongdoer to suffer as a means of some other end, such as deterring the wrongdoer (or other individuals) from committing future wrongs. The ultimate end of deterrence makes this measure regulative.

Also note that for an instance of punishment to occur, there must be more than just a culpable agent; punishment requires a punisher. The punisher takes the form of some authority authorized and able to facilitate measures in response to the agent's act. Authorized authorities may implement either regulative or retributive measures.<sup>2</sup> Since this paper defines punishment as measures implemented with the ultimate end of causing the wrongdoer to suffer on the basis of said wrongdoer's guilt, it must be said that, in order for some authorized authority's action to be properly considered punishment, the authority, in acting, aims to bring about the wrongdoer's suffering as her ultimate end.

Gary Watson, in his paper "Two Faces of Responsibility" provides two distinct notions of moral responsibility: accountability and attributability. Both are forms of culpability, though they differ in that holding an agent accountable implies the imposition of sanctions (punishment) on the agent that commits the moral wrongdoing, whereas finding an agent attributable does not imply the imposition of sanctions (punishment) on the wrongdoer (Watson 231). On holding an agent accountable, Watson writes, "holding here is not to be confused with the attitude of believing in;" holding "involves a readiness to respond to them in certain ways" (Watson 235). Dividing moral responsibility into these two notions allows us to maintain judgements about moral responsibility without obligating ourselves to act punitively on account of these

---

<sup>1</sup> While some regulative measures may be considered objectionably coercive and paternalistic, so long as the measures are not punitive, they are not rightfully considered punishment within the linguistic confines of this paper

<sup>2</sup> Note that these authorities are referred to as authorized authorities, not justified authorities. Authorities may still take regulative or retributive measures despite their not being justified in doing so.

judgements. Equipped with this distinction, we are able to find some agents attributable, though not accountable, to some harms. Neil Levy, in his essay “The Good, the Bad, and the Blameworthy,” writes, “we ought to be able to say that something is bad without saying that it is blameworthy” (Levy 5). Either an act’s tendency to produce an undesirable result or an act’s being unchoiceworthy are not, on their own, sufficient conditions for the agent who performed the act to deserve punishment.

The distinction between accountability and attributability is a useful one, as individuals can play a causal role in the bringing about a harm without being rightfully considered blameworthy in the accountability sense. Watson offers the example of a person that stays out late drinking the night before an important presentation, fails to get a sufficient amount of sleep, and who performs poorly as a result of these actions (Watson 231). In this case, the agent is certainly found to be attributable for the harm she has caused herself, though we do not hold her accountable. We do not feel the need to punish this agent, and so we do not find her blameworthy in the accountability sense. Watson explains this verdict by claiming that she has failed to meet some necessary condition(s) for accountability. He claims that accountability requires there to be 1) an individual or group responsible to; 2) another individual or group for 3) complying with demands (Watson 235). It is important to note that missing one of these three components is not the only marker of non-accountability.

In cases where we feel an agent to be culpable despite their not meeting the necessary conditions to be held accountable, we may consider them blameworthy in the attributability sense. The concerns of attributability are “what it is to lead a life” and “the quality and character of that life” (Watson 229). Attributability offers an “aretaic” perspective on responsibility, that is, one concerned with “what activities and ways of life are most choiceworthy” and what

“expresses and constitutes [an agent’s] practical identity, what he stands for, what he has made of his life as he found it” (Watson 240). Aretaic appraisals are permissible when an agent is attributable, though these aretaic appraisals are independent from the moral considerations for an agent’s accountable accountability. It is also important to note that an agent’s being merely attributable (and not accountable) is sufficient to warrant action in response to the wrongdoing for which the agent is found attributable. Consider again the case of quarantine, where an authorized authority implements protective, regulative measures without thinking the agent to be deserving of punishment. In cases of non-accountability, the measures put into place must not be implemented for the ultimate purpose of causing the wrongdoer to suffer. When an agent is merely attributable, the measures taken in response must be merely regulative.

Accountability, conversely, is the quality that makes an agent deserving of punishment. When an agent is blameworthy in the accountability sense, their guilt is the kind that mandates the implementation of measures intended to cause the agent to suffer. As a result, a discussion of retributivism is one concerned with the accountability sense of blameworthiness and the conditions for determining accountability.

With these definitions in mind, we can develop a more sophisticated understanding of the claims of retributivism: “(1) the innocent deserve not to be punished” and “(2) The guilty deserve to be punished proportionately” (Zimmerman 1951, 67). The first claim of retributivism may be misinterpreted as a claim of weak immunitivism, which merely asserts that the innocent *do not deserve* to be punished (Zimmerman 1951, 69) Retributivism claims that “the innocent *deserve not to* be punished” (Zimmerman 1951, 80). Both weak immunitivism and retributivism hold that there is no moral obligation to punish the innocent, but retributivism’s claim is stronger than the claim made by weak immunitivism. Weak immunitivism says nothing about what the

innocent *do* deserve; it makes no claim about the moral wrongness of the innocent being punished. Retributivism does. Retributivism asserts that the innocent actively deserve not to be punished, and thus, that it is morally wrong for the innocent to be punished.

Retributivism has two potential prescriptions: punish or do not punish. Punishment is required in cases of guilt and prohibited in cases of innocence, but what about in the case of ambiguity? The option to punish or not punish is forced. To do one *is* to not do the other. We cannot avoid choosing; to not act is still to choose. In cases of ambiguity, retributivism falls silent. I propose the following principle: until agents are rightfully found accountable, we should consider them not accountable. Under this principle, when faced with ambiguity regarding an agent's accountability, the agent ought not be punished. The underlying assumption is that protecting the innocent is of greater moral importance than punishing the guilty. I find this underlying assumption to be intuitive and commonly accepted. It echoes commonly accepted legal sentiments such as "innocent until proven guilty" and the requirement to establish guilt "beyond a reasonable doubt" before implementing certain sanctions.

Additionally in cases where we are forced to prioritize either punishing the guilty or protecting the innocent, we prioritize the latter, even when we find the guilty person to possess immense guilt. Otherwise, it would be ineffective for a criminal to take a hostage or employ the tactic of using a 'human shield.' When someone we consider worthy of punishment manipulates a situation such that it would be impossible to punish them without also harming an innocent person, we intuit that it is right to prioritize protecting the innocent.

Finding an agent accountable and thus, finding them deserving of punishment, has potentially grave repercussions: "reputation, liberty, and even life can be at stake" (Watson 242). For this reason, we must have strict and specific conditions for finding an agent accountable that



must be met before an agent can be rightfully punished for their actions. In cases where it is ambiguous as to whether an agent meets the conditions for accountability, we must consider them to be not accountable. If we decide an agent is accountable in an ambiguous case, we run the risk of intentionally bringing about a person's suffering without an appropriate justification. Despite the misguided punisher's moral intentions, a wrongful punishing is simply a harming of a person out of moral ignorance. By most moral theories, this is something one ought not do. Wrongful punishing is a wrongdoing.<sup>3</sup>

I make no theoretical objection to the claims of retributivism. While the notion that some agents can be deserving of suffering for suffering's sake as a result of their wrongdoings is potentially controversial on its face, I look beyond this to explore the implications of the theory. I argue that, if we grant the premise that some actions warrant punishment, retributivism still fails as a functional theory. Since it is not mere proximity to the causation of a harm that marks an agent as deserving punishment for said harm, more than a mere instance of wrongdoing is necessary for retributive action to be warranted. The conditions for accountability must be strict and met unambiguously such to prevent the punishing of the innocent, who, according to retributivism, are deserving of not being punished. Thus, this discussion also warrants an exploration of the epistemic conditions an authorized agent must meet if they are to be considered rightful in their assessments of accountability and in their punishing of alleged wrongdoers. However, our ineliminable susceptibility to the influence of luck and the

---

<sup>3</sup> One could argue that, in the case of ambiguity we may have reason to err in either direction, even while maintaining the primary objective of protecting the innocent. While it might be the case that, in an effort to protect the innocent, ambiguity would make us wary of punishment, one could argue that that ambiguity provides a justification for being more liberal with our assignments of punishment in hopes that the threat of punishment will loom in the background, serving as a deterrent against wrongdoing. This objection confuses regulative action with retributive action. When it comes to taking measures done with the intention of promoting the suffering of the wrongdoer, we must, in cases of ambiguity, err by being conservative with our dispensing of punishment.

unattainability of infallible belief prohibit us from attaining the unambiguity necessary to rightfully find agents accountable. This paper culminates in a criticism of retributivism, claiming that no version of retributivism results in a sufficiently apt dispensing of punishment. All versions of retributivism either display impotence or possess deeply unattractive implications (in that these implications are personally worrisome for all agents). In this way, retributivism is an inept theory.

## Explanation of Ought Implies Can

Now that we have established explicit distinctions between the concepts and terms involved in retributivism, we can begin to lay out the potential conditions for accountability. “Ought Implies Can” (sometimes abbreviated to OIC) is a principle often attributed to Kant that claims that for some action or omission to be rightfully considered an obligation, it must be the case that the agent obligated to said action or omission can do it (in the case of action) or omit it (in the case of omission). As a result, “Ought Implies Can” also holds that it cannot be the case that an agent is obligated to actions or omissions that lie beyond her ability.<sup>4</sup> Under OIC retributivism, if an agent is responsible for violating a moral obligation, it must be the case that said agent had the ability to not violate it. Without ability, there can be no moral obligation. As such, if we accept OIC, ability becomes a necessary condition for accountability. In all cases where an agent is found accountable, it must be said that the agent is guilty of violating a moral obligation that she had the ability to not violate at the time of the violation in question.<sup>5</sup>

Ought Implies Can requires both physical and epistemic ability. Consider the following intuitive cases, both of which serve as gross oversimplifications of the physical and epistemic elements of OIC. First, consider a case where Erin, a person of average strength, encounters a person stuck under a two-ton truck. According to OIC, because Erin lacks the ability to lift a

---

<sup>4</sup> From here, I begin to equate action and omission. Although action and omission are commonly considered as distinct, Randolph Clarke, in his book *Omissions: agency, metaphysics, and responsibility*, draws attention to the often-arbitrary distinction between actions and some kinds of omissions in cases like “holding one’s body still” or “not moving” (Clarke 116). In cases where Clarke argues that there is a significant difference such to warrant a distinct characterization of an omission as separate from action, this difference is often marked by an appeal to some kind of moral standard or norm. In his discussion of these standards, he references a view of moral responsibility that is incompatible with the view I present in this paper. For this reason, for the duration of this paper, I continue to equate actions and omissions and neglect to clarify that the principles I lay out apply to omission in the same way as they pertain to action.

<sup>5</sup> The interpretation of OIC I present here is somewhat contested (Kurthy, Lawford-Smith, Sousa 2017). I do not hold that all interpretations of OIC ought to include these stipulations, nor do I claim that the version of OIC I present here is an unobjectionably accurate characterization of the principle. I am simply employing the term “Ought Implies Can” (sometimes as OIC) in this paper to refer to some principle that has these implications.

two-ton truck, it cannot be said that she ought to do so. Secondly, consider a case where Ayla, a trained EMT, is sitting in her apartment when, unbeknownst to her, someone on the other side of her apartment complex suffers from some timely medical emergency. The neighbor could only survive the emergency unharmed if treated quickly by a trained professional. Due to the distance between Ayla's apartment and that of the person in need, Ayla is unable to hear her neighbor's calls for help. Under OIC retributivism, because Ayla lacked the epistemic ability necessary to help her neighbor, it cannot be said that Ayla ought to have provided such help. In a similar way to Levy's claim that things can be bad without being blameworthy, things can be good without being morally obligatory. While it might be good for Erin to lift the truck or for Ayla to help her neighbor, it is not their obligation to do so under OIC retributivism, given the circumstances that limit their ability.

The notion of the will and volition will also appear in the discussion of ability and OIC. An agent is unable to choose to perform actions that are not from her volition. Choice itself is a matter of volition. When an agent acts contrary to her volition, it must be said that the agent was unable to avoid doing said act. If it were the case that an agent willed against her own action and that she possessed the ability to not do said act, then the agent would not do the act. Accordingly, when an agent acts against her own volition, it must be said this act is done out of the agent's inability to not do it. Action contrary to volition serves as an indicator of an agent's lacking ability.

Ought Implies Can is a controversial principle. I do not defend it here, and so I do not expect the reader to be persuaded to accept it. I will discuss the implications for retributivism were it to include Ought Implies Can (i.e., were retributivism to include ability as a necessary condition for accountability, a version of retributivism I refer to as OIC retributivism). I will also

discuss the implications for retributivism were it to reject Ought Implies Can (i.e., were retributivism to not to include ability as a necessary condition for accountability, a version of retributivism I refer to as non-OIC retributivism). Logically, it is either the case that retributivism accepts or does not accept OIC. In either case (which, because of the binary nature of accepting the principle or not accepting the principle, includes every possible case), retributivism is inept at upholding its own principles. I will demonstrate this conclusion by thoroughly outlining the implications of each version of retributivism (OIC retributivism and non-OIC retributivism) and identifying the necessary and undesirable outcomes for each.<sup>6</sup>

---

<sup>6</sup> There may be more necessary conditions for finding an agent accountable, though I feel it unnecessary to explore these other potential conditions. I believe OIC will mark a significant problem for retributivism's functionality as a moral theory, whether we accept this principle or not.

## **Moral Luck**

In his essay “Moral Luck,” Bernard Williams identifies that traditionally, morality is understood as a sphere immune to the influence of luck (Williams 115). He identifies that this conception of morality is mistaken and that luck permeates our moral lives. Williams uses the term luck “generously, undefinedly” yet “comprehensibly” (Williams 117). I employ the same usage of the term here. Vaguely and imprecisely, luck has to do with external factors beyond an agent’s control. Luck, I argue, has a pervasive influence in our moral lives in a way that compromises our ability to rightfully hold wrongdoers accountable for their wrongdoings under Ought Implies Can.

As a reminder, under OIC retributivism, if an agent is accountable, then the agent must have had the ability, at the time of the wrongdoing, to have not committed the wrongdoing in question. The degree to which an agent ought to be held accountable varies in direct proportion to the degree to which the agent had the ability to comply with moral obligation. A conclusion regarding accountability derived from factors independent from the concerns of accountability is an instance of theory misapplied. Because, under OIC retributivism, agents cannot be more or less accountable on the basis of luck, agents ought not face charges of accountability on the basis of luck.

To demonstrate the plausibility of the claim that luck ought not impact our assessments of accountability, I will first offer an intuitive case before introducing cases that do not align with standard intuition. Consider two possible scenarios. Both begin with Robert letting Emma hold his brand-new phone and Emma dropping it. In one case, the phone screen shatters. In the second scenario, the phone remains totally unchanged. In these two scenarios, Emma may be differently blameworthy in the attributability sense, as accepting OIC does not impact the conditions for

attributability. However, in terms of accountability, Emma cannot be found differently blameworthy in these scenarios, as the only differing factor between these scenarios is luck, which lies outside the realm of her ability. Thus, in the case where Robert's phone breaks, Emma is not more or less deserving of punishment than in the case where Robert's phone does not, according to OIC.

To demonstrate luck's ubiquitous influence in our moral lives, I will now discuss a few of the forms luck may take. Thomas Nagel, in his essay "Moral Luck," outlines four kinds of luck that impact our moral lives. The first, constitutive luck, has to do with "the kind of person you are, where this is not just a question of what you deliberately do, but of your inclinations, capacities, and temperament" (Nagel 140). The second is luck in one's circumstances, or "the kind of problems and situations one faces" (Nagel 140). The third and fourth have to do with the causes and effects of one's actions, which can be roughly summarized as "the way things turn out" and "decision under uncertainty" (Nagel 140-141). Developing an awareness of the manifestations of these various forms of luck reveals that "the natural objects of moral assessment are disturbingly subject to luck" (Nagel 140).

Consider first constitutive luck. A person naturally inclined to be patient, attentive, and slow to anger will have far less difficulty avoiding moral wrongdoings than someone impatient, distractable, impulsive, and quick to anger. Were these two individuals of contrasting natures to put the same amount of time and effort into perfecting their moral dispositions, the latter would face far more charges of culpability, not for lack of effort, but for lack of luck, when faced with the same moral challenges. The differences between these two agents, one naturally gifted with virtuous inclinations and one not, exist outside the realm of volition, control, and ability, and

thus, under OIC retributivism, these kinds of considerations ought not impact our assignments of accountability (though they may impact assignments of attributability).

The innate quality of these characteristics may be a topic for debate. Many philosophers claim that one's moral disposition is a matter of volition. There is a pervasive view that an agent can shape her own moral character through intentional action, and that "the capacity for moral agency is supposedly present to any rational agent whatever, to anyone from whom the question can even present itself" (Williams 115). As rational agents, we each have the option to shape our own moral character.

It is tempting to conclude that matters like who we are and what we do are entirely within our ability to control and thus, they are matters for which we may be held accountable. Williams writes, "one is supposed to grasp, not only morality's immunity to luck, but one's own partial immunity to luck through morality" (Williams 116). However, this is an incomplete characterization of the formation of moral disposition. Moral disposition is, at least in part, shaped by factors outside the realm of an agent's control: upbringing, education, neurological variation, hormones, trauma, the community one is brought up in, the availability of information, etc. What external factors one is exposed to and how they shape one's character and actions are matters of luck.

The externally imposed nature of these traits is perhaps most obvious in cases where an individual has a diagnosable mood or attention disorder, or a moral disposition that results from a physical trauma. Consider a case like that of Phineas Gage, who displayed a significant shift in his moral disposition after a premature explosion during railroad construction shot a metal rod through his skull (O'Driscoll & Leach, 1998). While before the accident, friends described Gage as "most efficient and capable," after, he was said to be aggressive, irreverent, impatient, and



impulsive; friends claimed he was “no longer Gage” (O’Driscoll & Leach, 1998). In sum, Gage developed a less virtuous moral disposition as a result of the accident. Regardless of whether the trauma originated due to some neurological or post-traumatic psychological cause, it seems undeniable that the shift in Gage’s character is contingent on the accident. Were Gage luckier, had the explosion had gone off at the proper time, or had the rod shot out at a different angle such that it missed Gage’s head, it is probable that Gage would have maintained his previous moral disposition.

A hypothetical lucky Gage, in his ability to maintain a more agreeable moral disposition, would likely commit less harmful acts than the actual unlucky Gage. While permissible that the lucky and unlucky Gage face different charges of blameworthiness in the attributability sense, it is not appropriate for these different versions of Gage to face different charges of accountability under OIC retributivism, given that the difference between the two versions of Gage is luck. As the lucky Gage is not accountable for the wrongdoings of the unlucky Gage, we must also hold that the unlucky Gage is not accountable for these wrongdoings either. We may generalize from this case: under OIC retributivism, an agent is not accountable for wrongdoings that result from constitutive luck.

Nagel uses the example of Nazi Germany to discuss a second form of luck, luck in one’s circumstances (Nagel 145). He prefaces this example, writing,

The things we are called upon to do, the moral tests we face, are importantly determined by factors beyond our control. It may be true of someone that in a dangerous situation he would behave in a cowardly or heroic fashion, but if the situation never arises, he will never have the chance to distinguish or disgrace himself in this way, and his moral record will be different. (Nagel 145)

Nagel claims that the citizens of Nazi Germany were presented with a moral test not presented to non-citizens of Nazi Germany. It is entirely possible that some individuals, were they to have faced the same moral test as the Nazis, would have behaved badly as the Nazis had. However, by sheer luck, these would-be-Nazis were not placed in Nazi Germany and thus did not have the opportunity to behave badly in this way. In fact, suppose some of these would-be-Nazis were only presented with moral tests through which they easily maintained acceptably virtuous habits. The would-be-Nazis avoid severe charges of culpability entirely, not based on virtue, intent, or effort, but on luck alone.

Similarly, it could be said that some in-fact-Nazis would have behaved virtuously if not presented with the moral test of living in Nazi Germany. Had some of the in-fact-Nazis been placed in the lucky circumstances of the would-be-Nazis, these in-fact-Nazis may have gone on to live apparently virtuous lives. The only difference between the would-be-Nazis and in-fact-Nazis is luck in their circumstances. While assignments for attributability may be different between the in-fact and would-be Nazis, we may not hold the would-be and in-fact Nazis differently accountable, under OIC retributivism. Because we do not hold the would-be Nazis accountable for their would-be actions, we must not find the in-fact Nazis accountable for their in-fact actions. Under OIC retributivism, accountability may not be assigned for wrongdoings that result from luck in one's circumstances.

For the third and fourth kinds of luck, luck in the way things turn out and luck in making decisions under uncertainty, Nagel turns to cases presented by Williams. Williams describes a case where a truck driver, through "a minor degree of negligence, fail[s] to have his brakes checked recently" (Williams 121). This case ends here. Williams then introduces a second case where a child runs into this truck driver's path (Williams 121). Because the brakes had not been

checked recently, despite the driver's best effort to avoid the child, the truck hits the child. Here, the driver is likely thought of as more culpable than in the first case, though he exhibits the same degree of negligence in both cases (Williams 121). Given OIC and that the driver is equally negligent in both cases, the drivers are equally accountable. Under OIC retributivism, agents are not accountable for wrongdoings that result from luck in the way things turn out.

Williams illustrates the fourth kind of luck, luck in making decisions under uncertainty, through a discussion of an artist, Gauguin, who abandons his family in hopes of becoming a great painter (Williams 118). The outcome of Gauguin's choice at the time of his making it is a matter of uncertainty. Williams identifies that confidence in one's creative abilities seems not to be a matter of moral judgement, and notes that, regardless of Gauguin's talent (or his confidence in his talent), Gauguin could have sustained an injury that would have prevented him from succeeding as an artist (Williams 120). Gauguin's success is a matter of uncertainty. In Williams' case, Gauguin happened to be lucky, as he was not injured or thwarted in some other way, and he became a successful and influential artist. It is tempting to consider the lucky Gauguin as less deserving of punishment than an unlucky Gauguin, who abandons his family in vain. However, with luck being the only differentiating factor between the lucky and unlucky Gauguin, our assessment of accountability should be identical between the two cases, as wrongdoings that result from luck in the way things turn out is not something for which agents can be considered rightfully accountable under OIC retributivism.

Luck's influence is pervasive in our moral lives. However, wrongdoings that result from said influence are not wrongdoings for which agents can be rightfully considered accountable under OIC retributivism, as OIC retributivism determines accountability based on an agent's ability, and luck lies outside the realm of this ability. While luck may impact an agent's

attributability for a wrongdoing and thus, our aretaic appraisals of agents and our regulative responses to them, this is the only form of blameworthiness that differs on the basis of luck under OIC retributivism. Wrongdoings based in any form of luck (constitutive luck, luck in one's circumstances, luck in the way things turn out, or luck in decision under uncertainty) are not wrongdoings for which agents are accountable, under OIC retributivism.

## Limits on Epistemic Ability

The purpose of the investigation into epistemology is ultimately twofold. Firstly, this investigation functions to exculpate wrongdoers in the accountability sense in cases where the wrongdoing is sourced in ignorance or some other form of bad believing. This is done by explaining how wrongdoers might come to believe badly as a matter of luck, despite their sincere and competent efforts to believe well, since all doxastic practices are vulnerable to potential unreliability (i.e., all doxastic practices are fallible). Secondly, this investigation is referenced in a later section of the paper to emphasize the fallibility of authorized authorities (i.e., punishers), and to highlight their inability to properly eliminate ambiguity when determining if an agent meets the necessary conditions for accountability. The fallibility of authorized authorities is evidenced by the same arguments that evidence the fallibility of wrongdoers. The investigation into epistemology includes a discussion of epistemic circularity, the fallibility of inductive reasoning, and the influence of doxastic involuntarism.

I aim to identify the potential for unreliability in all doxastic practices (i.e., the practices through which we come to believe). The argument that all doxastic practices are fallible is not to be confused with the claim that all doxastic practices are certainly unreliable. In fact, the notion of certainty is incompatible with my intended conclusion. I am arguing that it is not the case that there is some doxastic practice that can be proven to be certainly reliable, and thus all doxastic practices are fallible, that is, they may be capable of bringing about falsehoods. In his book *Beyond "Justification,"* William Alston investigates what it would take to consider a belief to be completely justified. Ultimately, he concludes that total justification is an unattainable standard due to many problems, one of which he calls "epistemic circularity" (Alston 207). The problem of epistemic circularity goes as follows: for every attempt at showing that any of our basic

sources of belief (e.g., perception, reason, memory, etc.) are reliable, the justification for the source's reliability will eventually depend on premises derived from the sources the argument aims to defend. Fumerton claims:

You cannot use perception to justify the reliability of perception! You cannot use memory to justify the reliability of memory! You cannot use induction to justify the reliability of induction! Such attempts to respond to the skeptic's concerns involve blatant, indeed pathetic, circularity. (Fumerton 177)

This is an analogue of Thomas Reid's case, where he calls us imagine trying to test if a man is honest or not by asking the man himself (Reid 275). Both a liar and a truth teller would claim themselves to be honest, and so the test is useless because it relies on the testimony of the source it aims to test. Justification for belief is circular for the same reason; it relies on the testimony of the source it aims to test. By this account, our beliefs are irreparably circular and thus, not fully justifiable.

Some philosophers advocate that, in cases where we cannot properly justify a belief, we are obligated to withhold said belief.<sup>7</sup> However, since the problem of epistemic circularity applies to all beliefs, this prescription would call us to withhold all belief entirely. Alston writes that "So long as we are alive, we cannot help having beliefs, nor can we avoid using them to guide our thought and action" (Alston 218). It is impossible to suspend all judgment and refrain entirely from belief; even the belief that 'one ought not believe' is a belief itself and is thus born out of circularity.

Even beyond this objection, Alston is right. Living a life requires belief. For example, I believe that certain physical sensations signal thirst. I believe the appropriate response to thirst is

---

<sup>7</sup> Take, for example, the works of W.K. Clifford (Madigan 2009)

to drink. I believe turning the faucet will cause water to come out. Without these beliefs, I would have no reason for turning the faucet or doing anything else for that matter. Living a life requires belief.

In an attempt to provide guidelines for a responsible doxastic practice that accommodates the impossibility of entirely withholding belief, Alston explores alternatives to complete justification. He discusses the reliabilist view of belief justification where, even though the sources for belief formation are not provably reliable a priori, we can determine a belief source to be reliable if it produces results accurately and consistently (Alston 126). This, Alston calls the “track-record argument” (Alston 203). While seemingly plausible upon first impression, the track-record argument and the conclusions derived from it also fall victim to fallibility and circularity.

Alston notes that the track-record argument may lead us to conclude that disreputable methods of belief formation such as “crystal ball gazing” are reliable (Alston 204). Were crystal ball gazing, by the influence of luck, to make a string of predictions that correspond with reality, an agent could appropriately conclude crystal ball gazing to be reliable. It could then be said that when this agent forms beliefs sourced in crystal ball gazing, they are doing so rightly and responsibly. The track-record argument is functional because if crystal-ball gazing were reliable, the track-record argument would claim it to be so. However, reliabilism is dysfunctional because if crystal-ball gazing were unreliable, the track-record argument might still advocate for its reliability, so long as crystal-ball gazing produced consistent results.

Not only might it be the case that the track-record argument would have us base our belief in disreputable sources, it might also be the case that the track-record argument would have us disregard reputable sources of belief. Goldman, in “What Is Justified Belief?” identifies

that according to the track-record argument, we may encounter evidence that supports the belief that sense perception is unreliable (Goldman 11). He offers the example of a congenitally blind man who wakes up from a surgery that gives him 20-20 vision (Goldman 11). Upon waking, the man sees a stuffed cat and forms the false belief that the cat is real (Goldman 11). Due to the man's fear of cats, he goes into cardiac arrest and dies (Goldman 11). Goldman claims that this example provides evidence in support of the belief that forming beliefs based on sense perception can lead to inaccurate beliefs. This evidence also supports the belief that the false beliefs sourced in sense perception are potentially harmful.

More common cases can also evidence the belief that our basic belief sources are unreliable. Consider the following case: the day after getting a new bike lock, I try to recall the numerical code I set. I think back to setting the code, and remember the code to be my birthday. I put in the corresponding digits to no avail. I then re-remember; I set the code as the street address of my childhood home. I put in the code and the bike lock opens. Because memory has produced inaccurate and inconsistent results, the track-record argument approach would have us consider memory unreliable. In a test of internal coherence as a metric of reliability, our most reputable sources of belief may not be found reliable.

Internal coherence does not entail objective accuracy about the external world. Consider the following case. Jake wants to paint his house red. He only has blue and yellow paint, but he remembers hearing that blue and yellow mix to make red. He mixes the paints together and sees them to be red. In this example, Jake has misremembered and is also (unbeknownst to him) colorblind. To Jake, it is internally coherent that blue and yellow mix to make red. Jake continues to see instances of blue and yellow coming together to make red, and thus, Jake forms the belief that red is formed by mixing blue and yellow. Although the information provided by



Jake's sense perception is consistently compatible with itself, his sense perception is not accurately interpreting the objective facts of the extended world. While this echoing of Cartesian skepticism can feel contrived and unpersuasive, consider additional ways the track-record argument may lead to erroneous conclusions.

Although dependent on the reliability of memory and induction, the track-record argument cannot to prove the reliability of memory or induction without circularity. Consider the following thought experiment. Clementine tries out some disreputable source of belief formation, such as crystal ball gazing. Although she knows crystal ball gazing not to be reliable a priori, she wants to test if this method of belief formation can produce accurate results. Clementine's crystal ball gazing predicts Y. Later, Clementine learns that, in fact,  $\sim Y$  is true. However, Clementine's memory is unreliable too. She recalls the crystal ball to have predicted  $\sim Y$ . Clementine thus considers crystal ball gazing a reliable source of belief. She comes to this conclusion on the basis of her memory. Clementine considers her memory a reliable source of belief because she remembers it to be. Again, it is true that if memory were reliable, the track-record argument would prove it to be so. However, it is also still the case that if memory were unreliable, the track-record argument might still advocate for its reliability. The track-record argument method of belief-source verification is insufficiently selective; it cannot protect our beliefs from potential vulnerability to error.

Similarly, if induction were unreliable, the track-record argument would be unable to identify it as such. Induction can only prove a conclusion to be probable; it does not eliminate the possibility of the conclusion's being false. Consider the example from Nassim Taleb's book *The Black Swan*.<sup>8</sup> Taleb writes that it was commonly accepted in 16<sup>th</sup> century Europe that all swans

---

<sup>8</sup> Regardless of the historical accuracy of this example, it can still serve as a thought experiment that demonstrates the fallibility of conclusions made through inductive reasoning

must be white, as 16<sup>th</sup> century Europeans had only seen white swans (Taleb 2007). He demonstrates the confidence people had in this belief by claiming that the phrase “black swan” was colloquially accepted as a stand-in for something impossible or nonexistent (Taleb 2007). The eventual encounter between a European and a black swan exposed the belief, ‘all swans must be white’ to be false (Taleb 2007). The conclusion that ‘all swans must be white’ is one based on limited experience and induction. The unluck of the European’s circumstances led them to a conclusion, through induction, that does not accurately represent reality.<sup>9</sup> Although the premises of an inductive argument do not contradict the conclusion, they do not entail it either. Inconclusive premises yield an inconclusive conclusion. On induction, Hume writes:

it implies no contradiction, that the course of nature may change, and that an object, seemingly like those which we have experienced, may be attended with different or contrary effects. May I not clearly and distinctly conceive, that a body, falling from the clouds, and which, in all other respects, resembles snow, has yet the taste of salt or feeling of fire? ... Now whatever is intelligible, and can be distinctly conceived, implies no contradiction, and can never be proved false by any demonstrative argument or abstract reasoning *à priori*. (Hume 846)

If a conclusion is made through inductive reasoning, and the negation of the conclusion proves to be true, no logical contradiction occurs. The improbable should not be mistaken for the impossible; there is a nonnegligible likelihood that the improbable may be. Only having seen white swans does not logically entail the nonexistence of black swans. Conclusions found by means of induction are fallible, including conclusions made through induction about the

---

<sup>9</sup> One might note that my persuasive efforts against the reliability of induction takes the form of an inductive argument. This is not to be taken as a successful objection. Rather, it shows that induction, even according to itself, is unreliable.

reliability of belief sources. The track-record argument is an inductive process, thus, its conclusions are fallible.

To summarize: complete justification for belief is seemingly unattainable, yet completely withholding belief is impossible. The only viable solution to our epistemic situation is to employ doxastic practices guided by track-record style arguments based on induction. Because the track-record argument is ultimately a test of internal coherence, it does not allow us to conclude anything about the objective goings on in the external world. Even the reliability of this test of internal coherence is still dependent on circular reasoning, as seen through memory and induction. Induction is unreliable in that its conclusions only state what is probable, though the improbable is still within the realm of the possible. Additionally, inductive reasoning is dependent on the evidence one has access to, which is a matter of luck in one's circumstances. Inductive reasoning does not guarantee us immunity from error. Our doxastic practices are irreparably subject to error in many ways. As a result, we may come to adopt erroneous beliefs due to our inability to believe better. For holding erroneous beliefs that we adopt due to our inability to believe better, we cannot be found accountable under OIC retributivism.

Our beliefs are not entirely a matter of our volition. Hume notes that the mind employs inductive reasoning automatically (Hume 841). What we believe is not entirely a matter of our volition. While it may be true that we have some kind of direct, relevant, immediate control over some of our beliefs, this degree of control is not applicable to all kinds of belief. William James writes,

Does it not seem preposterous on the very face of it to talk of our opinions being modifiable at will? Can our will either help or hinder our intellect in its perceptions of truth? Can we, just by wiling it, believe that Abraham Lincoln's existence is a myth, and

that the portraits of him in ‘McClure’s Magazine’ are all of someone else? Can we, by any effort of our will, or by any strength of wish that it were true, believe ourselves well and about when we are roaring with rheumatism in bed, or feel certain that the sum of the two one-dollar bills in our pocket must be 100 dollars? We can say any of these things, but we are absolutely impotent to believe them; and of just such things is the whole fabric of the truths that we do believe in is made up- matters of fact, immediate or remote, as Hume said, and relations between ideas, which are either there or not there for us if we see them so, but which, if not there, cannot be put there by any actions of our own.

(James 4-5)

While we may, in some cases, be able to volitionally switch into doubt and agnosticism, in other cases we are truly unable to let go of some beliefs in a complete and immediate way. When one’s senses present her with evidence of her pain, and her memory recalls her diagnosis of rheumatism, her belief that she is in pain with rheumatism seems to be unconquerable, regardless of what she might will. She lacks the ability to believe otherwise. Other beliefs hold a similar status; they appear entirely or partially resistant to the influence of the will. To clarify, I am not arguing against the existence of some volitional influence belief. Rather, I am arguing for the existence of some nonvolitional influence on belief.

Even in cases where the agent believes that the strongest evidence opposes her initial belief, an agent’s belief about the strength of opposing evidence is not always sufficient to fully eliminate one’s initial mistaken impression. Consider optical illusions as an example. The Müller-Lyer illusion presents two lines of the same length that appear to not be the same length (Rock 1997). While the lines are easily shown to be the same length with a ruler, the senses are unresponsive to this evidence, and the lines still appear to be different lengths. However, we

believe measuring the lines with a ruler to be a better method of forming belief about the length of the lines. Since we are most compelled by the evidence in favor of S, the belief that the lines are the same length, we believe it is best to adopt S. If we want to believe the best beliefs, then we want to adopt S. However, even with the will to adopt S, we still cannot fully discount the evidence from the senses, which support  $\sim S$ , the belief that it is not the case that the lines are the same length. An overthinker may stare at the illusion for a while and, despite the evidence, begin to more seriously consider  $\sim S$ . Alternatively, a lapse in attention may cause an agent to, by default, rely on her sense perception and momentarily adopt  $\sim S$ . In this way, we may adopt beliefs contrary to our volition, which indicates that the adoption of some beliefs is beyond our control.<sup>10</sup>

The conflict between belief and the will can occur in more than just seemingly low-level beliefs about the physical layout of the world. Consider Pascal's Wager, where an agent's desire to believe in God is insufficient to instantly bring about her belief in God. Even as a proponent of doxastic voluntarism, Pascal acknowledges the limits on volition's ability to shape belief. Pascal prompts us to participate in religious rituals that he claims will gradually establish a belief in God (Pascal 1670). He acknowledges that there will be a period of time during which one must feign belief in God before the belief becomes legitimate (Pascal 1670). During this transitional period, the agent's belief does not correspond with her volition. The agent wills her belief to be different than it is, but this willing does not bring about the desired belief's immediate actualization. An agent may believe herself to be believing badly, and yet, she may still lack the

---

<sup>10</sup> It should be noted that these kinds of innocuous misperceptions may not be entirely harmless. These misperceptions could prompt an agent to act in such a way that has the potential to result in harm. Although contrived, the following sort of example is possible. Imagine a person guiding another person who is somehow visually impaired. If the sighted person were to mistake the physical layout of the world and provide the visually impaired person an inaccurate description based on this misperception, the person with visual impairment may move according to the inaccurate description, trip, and get injured.

ability to make herself instantly believe better. Lack of ability makes this bad believing something for which an agent cannot be found accountable under OIC retributivism.

The kind of case where an agent's belief is in conflict with her volition is not the only kind of case in which an agent lacks the ability to believe otherwise. Consider James' one of examples once more. I both do not will myself to believe that Abraham Lincoln's existence is a myth, and I think myself unable to bring about the belief that Abraham Lincoln's existence is a myth. There is no conflict in volition and belief, and yet, ability is absent as well. We cannot look to an agent's volition as the sole indicator of ability. Agents may both want to believe X and lack the ability to not believe X.

Our processes of belief formation are irreparably fallible as demonstrated through the unsolved problem of epistemic circularity, the limits of inductive reasoning, and the existence of nonvolitional influence on belief. However, we lack a more reliable alternative for belief formation. Thus, we do not have the ability to believe infallibly. Our unavoidable fallibility may lead us to believe badly, trust in unreliable belief sources, and come to nonveridical conclusions. This is not something we are able to avoid. If our bad believing leads us to wrongdoing, we may not be rightfully held accountable for these wrongdoings under OIC retributivism.

## **Ignorance and Incompetence as Exculpatory Under Ought Implies Can**

If an agent is ignorant to a moral obligation, said agent lacks the ability to comply with said moral obligation. For this reason, under Ought Implies Can, ignorance is exculpating. Under OIC retributivism, if an agent's wrongdoing is rooted in ignorance, they cannot be held accountable for it. After a further defense of these claims, I will provide a collection of thought experiments. These thought experiments are intended to demonstrate how any wrongdoing (regardless of the type of mens rea we traditionally think an agent must possess to commit some wrongdoing) may still be sourced in an agent's ignorance. This is the case even when agents are ignorant as a result of their unique incompetence. Incompetence and ignorance exculpate agents from accountability under OIC retributivism.

Zimmerman argues against the distinction between moral ignorance and factual ignorance (i.e., "ignorance of non-moral fact") (Zimmerman 2018, 628). He writes that when a harm is sourced in ignorance, the distinction between these two kinds of ignorance is not productive (Zimmerman 2018, 628). Any harm committed out of factual ignorance ultimately results in moral ignorance because the agent is ignorant of the moral fact that she is acting wrongly (Zimmerman 2018, 628).

A denial of the wrongdoing's wrongness is not necessary for an agent to be exculpated. In his paper, "A Plea for Excuses," John Austin distinguishes between justifications and excuses, remarking that although both make attempts at exculpating the alleged wrongdoer, justifications do so by denying the overall wrongness of an agent's action, whereas excuses exculpate the wrongdoer while maintaining that the wrongdoer's action was overall morally wrong. Roughly put into the framework of this paper, a justification alleviates assignments of both attributability

and accountability, while excuses maintain assignments of attributability while eliminating assignments of accountability.

Under OIC retributivism, ignorance is a valid moral excuse. In his article “Excuses and ‘Ought’ Implies ‘Can,’” Lawrence Heintz succinctly summarizes H.A. Prichard’s finding that, under Ought Implies Can:

in order for one to be under an obligation one must [believe] that one has the obligation. ...the principle is at work in its contrapositive form: If we do not [believe] that some action of ours will have as one of its consequences the satisfaction of the obligation, then we cannot properly be said to be obligated (Heintz 451).<sup>11</sup>

Think back again to the case of Ayla, the EMT with a neighbor in need of assistance. Despite her hypothetical partial ability to aid her neighbor given her medical knowledge, she did not believe her neighbor to be in need of help and thus, she did not have the ability to help her neighbor. Ayla’s ignorance is a successful moral excuse.

Zimmerman writes, “all blameworthiness has its origin in witting wrongdoing, that is, in wrongly doing something that one did not do from ignorance of the fact that one was wrongly doing it” (Zimmerman 2018, 625). Under OIC retributivism, blame must be assigned relative to the origin of the wrongdoing. We should consider the origin to be the earliest member of the causal sequence that results in the wrongdoing in question.

In cases of moral ignorance, some members of the causal sequence that results in the wrongdoing in question may be considered benighting acts (Smith 547). Holly Smith, in her

---

<sup>11</sup> Here, Heintz uses the word “know” in place of the word “believe.” While the change to the word ‘believe’ may alter the intended meaning of the quote in some applications, for the application of the quote in the context of this paper, I believe the intended meaning of the quote to be preserved despite this change. This change is made because the scope of this paper intentionally does not include an explicit discussion of knowledge. Knowledge is a broad and messy topic in epistemology, and its inclusion is unnecessary for the argument I present.



essay “Culpable Ignorance,” identifies a “benighting act” as an act an agent performs that results in her own moral ignorance regarding the wrongness of her subsequent action (Smith 547). As is the case for all acts under OIC retributivism, agents may only be found accountable for a benighting act if said benighting act has its origin in witting wrongdoing.

Consider the following case from Carl Ginet’s essay, “The Epistemic Requirements for Moral Responsibility”: some agent, Herb, takes a drug that induces lapses in memory (Ginet 274). Herb forgets to keep a promise as a result of his taking the drug; in breaking this promise, Herb fails to comply with a moral obligation (Ginet 274). To investigate Herb’s status regarding accountability, we must trace the wrongdoing back to its origin. The origin of Herb’s wrongdoing is not his forgetting, as this was caused by his benighting act of taking the drug. This benighting act may or may not be the origin of the wrongdoing. Consider the following three variations for Herb’s case.

- 1) Herb was ignorant to the fact that taking the drug would result in lapses in memory. Thus, Herb was ignorant to the fact that taking the drug would result in breaking his promise. Herb’s taking the drug was not a witting wrongdoing. If we possessed more information about Herb’s case, we may continue our investigation to see if Herb’s benighting act of taking the drug may have been the result of some previous benighting act. We would carry out the same investigation for Herb’s previous benighting act. However, we do not possess this information; Herb’s case is ambiguous. Without the information necessary to continue our investigation, under OIC retributivism, we cannot conclude that Herb is accountable.
- 2) Herb believed that taking the drug would result in lapses in memory but was ignorant to the fact that taking the drug would cause him to break his promise. Herb is ignorant to the

fact that his forgetting would result in his wrongdoing. Therefore, Herb's taking the drug was not a witting wrongdoing. Thus, Herb cannot be found accountable for this wrongdoing under OIC retributivism. As before, if we possessed more information about Herb's case, we may continue our investigation. Again, we do not possess this information, and Herb's case is ambiguous. Thus, he cannot be held accountable under OIC retributivism.

- 3) Herb believed that taking the drug would cause him to break his promise. In this case, we might investigate the causal sequence that resulted in Herb's wanting to do such a thing. We would have to conduct a further investigation into Herb's beliefs and motivations (as well as their respective origins) to ensure that the true origin, the very first member of the causal sequence, was one of witting wrongdoing (i.e., not one of ignorance or luck). Without the information necessary to conduct this investigation, Herb cannot rightly be found accountable under OIC retributivism.

Ginet himself responds to Herb's case, claiming that we should find Herb blameworthy because Herb "should have [believed]" that his taking the drug would result in his failing to comply with moral obligation (Ginet 274). The notion that Herb 'should have believed' is an appeal to normative competence (a concept separate from an agent's actual ability) and is therefore not a valid consideration for accountability under OIC retributivism.

Zimmerman discusses notions similar to that of 'should have believed' in his response to Rik Peels' *Responsible Belief: A Theory in Ethics and Epistemology*. Peels describes the following case on moral ignorance:

Imagine that I am babysitting and that, before leaving, my friend tells me that her daughter Sarah is seriously allergic to milk. If, an hour later, I completely forget about

that and give Sarah a glass of milk, I am clearly blameworthy for that, simply because I had the dormant<sup>12</sup>, but unactivated (non-occurrent)<sup>13</sup> belief that I should not give her any milk. I could and should have thought about it or remembered it. (Peels 580)

Peels determines that the babysitter is blameworthy for giving Sarah a glass of milk, though Zimmerman argues that this verdict is premature. Zimmerman describes the propositions that the babysitter may believe:

- (1) Sarah is allergic to milk.
- (2) It is morally wrong (*ceteris paribus*) to give someone something to which that person is allergic.
- (3) It is morally wrong (*ceteris paribus*) to give Sarah a glass of milk.
- (4) I am giving Sarah a glass of milk.
- (5) I am giving Sarah something to which she is allergic.
- (6) I am wrongly giving Sarah something to which she is allergic.
- (7) I am wrongly giving Sarah a glass of milk. (Zimmerman 629)

Zimmerman then identifies that at the time the babysitter gave Sarah the milk, even though the babysitter (perhaps dormant) believed propositions 1-3, it is also true that the babysitter was ignorant to propositions 5-7. Zimmerman argues that the babysitter's ignorance to propositions 5-7 is evidenced "because it seems natural to say that [the babysitter] wouldn't have done this had he [believed these propositions]. He was acting *from* ignorance of this fact" (Zimmerman 630). Zimmerman anticipates that Peels may object by claiming that the babysitter was not ignorant to these propositions, but rather, the babysitter was inadvertent to them (Zimmerman

---

<sup>12</sup> Dormant belief is one of three kinds of belief Peels describes in his paper. Zimmerman roughly summarizes dormant belief writing, "S dormantly believes that p iff S is not considering whether p but has done so in the past, and the last time he did so he thought that p" (Zimmerman 626).

<sup>13</sup> "S occurrently believes that p iff, while considering whether p, S thinks that p" (Zimmerman 626)

630). Zimmerman responds by saying that so long as the babysitter cannot be found blameworthy for his inadvertence, his inadvertence to the moral facts of the situation is sufficient to constitute a valid moral excuse for the babysitter's wrongdoing.

George Sher, in his book *Who Knew?: Responsibility Without Awareness*, emphasizes the nonvolitional nature of inadvertence:

We all know what it is to be assaulted by an urgent problem that drives all other thoughts from our minds; to emerge from a reverie into which we have no recollection of choosing to enter; and to have our defenses against drowsiness infiltrated by a momentary lapse in our awareness of the need to sustain them. When such things happen, there is simply no point at which we are conscious of choosing to allow them to happen. Thus, if someone were to assert that all such failures of attention are nevertheless voluntary, his claim would be implausible on its face. That claim, if not backed by some powerful independent argument—and I know of none that supports it—will not so much advance our understanding of the problem as simply define it out of existence. (Sher 25)

Ignorance and lapses of attention can result in a failure to comply with moral obligation. The benighting acts that bring about ignorance or inadvertence must originate in witting wrongdoing if the agent is to be found accountable for them. In the case of Sarah's allergy, the babysitter does not intend for the outcome of his actions to be Sarah's having an allergic reaction. The babysitter's wrongdoing is unwitting; he is ignorant to the moral facts of his situation.

The notion that the babysitter 'should have believed' under Ought Implies Can is mistaken. 'Should' is synonymous with 'ought.' To say that the babysitter ought to have believed (at the time of his giving Sarah a glass of milk) that he should not give Sarah the glass of milk is to either violate OIC or to mistake the babysitter's believing this as within his ability. Notions

like ‘should have’ that lie outside of an agent’s actual ability are often appeals to normative competence. Sometimes, agents are abnormally incompetent. Unless one’s incompetence is originally sourced in witting wrongdoing, then incompetence is not something for which an agent can be found accountable. ‘Should have believed’ and any other appeals to normative competence are arguments independent from the agent’s actual ability and are thus independent from assignments of accountability under OIC retributivism.<sup>14</sup>

Ignorance to moral fact can exist even in cases where agents are not ignorant of the immediate consequences of their actions. Consider another case: Nicky gets word of a new vaccine that purportedly protects against infection from the virus causing a global pandemic. She looks first to her trusted news sources. She has come to trust these news sources not only due to their predictive accuracy, but also, because her friends and family (whom she has also been led to trust) regularly advocate for the legitimacy of these news sources. These trusted news sources not only claim that the vaccine is ineffective, but they also warn that the vaccine has dangerous side effects, even arguing that the vaccine is counterproductive towards ending the pandemic. Nicky decides to go beyond these secondary sources and do her own research. She spends hours online, reading more information from sources she has come to trust through her doxastic practices. She reads on not only about the potential dangers of the vaccine, but also learns that some individuals, with whom she shares many beliefs, suspect there to be a political agenda that is antithetical to her beliefs associated with the vaccine. Nicky continues to get information from a variety of sources. She looks at news sources with differing perspectives and some scientific studies (though at times, she lacks the experience necessary to fully comprehend these academic, scientific sources). Nicky does not find the news articles that advocate for the safety of the

---

<sup>14</sup> The sentiment of ‘should have believed,’ is more commonly expressed as ‘should have known’

vaccine compelling; the news sources that write these articles often release information she finds counterfactual. Nicky notes that no studies discuss the long-term effects of the vaccine, and that many of the major studies overlook the way the vaccine may impact recipients differently based on sex and gender.

In fact, Nicky is skeptical about the reliability of the major medical and political institutions that support the vaccine. She recalls learning that these reputable medical institutions conducted a study in which they intentionally lied at the cost of hundreds of lives, and another instance where reputable medical institutions withheld information from a patient about their usage of her cells. Nicky recalls studies that identify the prevalence of bigotry in the medical world and formulates the belief that these medical authorities may not be entirely reliable.<sup>15</sup> She notes that these instances of unreliability often result in harm to individuals on the margins of society, where she considers herself to be.

Nicky has a great deal of evidence that supports the belief that the vaccine is unsafe. Although she has also had exposure to a great deal of evidence that could have supported the belief that the vaccine is safe, she has reason to doubt the validity of this evidence. Nicky, as a result, believes the vaccine to be unsafe and acts accordingly. She does not get vaccinated, shares the evidence that led to her decision with the people she encounters, and encourages others to abstain from the vaccine, as she believes this to be the safest option.

Suppose that the truth of the matter is that the vaccine is safe, effective, and essential for protecting Nicky and her community from the virus. How is Nicky to know this? Given the

---

<sup>15</sup> Although Nicky's case is fictitious, there are (perhaps very obviously) real-world events that inspire this thought experiment. Specifically, I draw attention to the USPHS Syphilis Study at Tuskegee, the use of Henrietta Lacks' stem cells without her knowledge or consent, the withholding of information regarding COVID-19 by the CDC, and a study by PubMed on "racial bias in pain assessment and treatment recommendations and false beliefs about biological differences between blacks and whites" ("Syphilis Study at Tuskegee" 2021; Cramer 2021; Mandivilli 2022; Hoffman 2016).

evidence at her disposal and her ability to interpret this evidence, Nicky lacks the ability to believe contrary to the notion that the vaccine is unsafe. For this reason, Nicky cannot be held accountable for this belief under OIC retributivism. Also, under OIC retributivism, Nicky cannot be held accountable for any action she takes as a result of this belief, regardless of the harm such actions may cause. Because these actions are motivated by a belief that originates in ignorance, these actions are originated in ignorance themselves. Under OIC retributivism, Nicky has a successful moral excuse for the harm she causes as a result of her ignorance.

Note that Nicky's exculpation is due solely to her inability. Her nonaccountability has nothing to do with the fact that her sources of belief somewhat resemble our own in their thoroughness and methodology. Suppose Nicky's doxastic practices, in their fallibility, had led her to believe that testimony from friends alone would sufficiently evidence a belief against vaccines, or that crystal-ball gazing or prayer would sufficiently evidence this kind of belief. In these cases, Nicky's ignorance still comes from inability, and her ignorance would still exculpate her under OIC retributivism. In these cases, we may struggle to believe that Nicky truly lacked the ability to believe otherwise. We may be tempted to make claims about what Nicky 'should have' done to avoid this belief, in our own disbelief of her unluck or incompetence. However, our inability to believe in Nicky's inability does not impact the truth of the matter.

OIC can be taken further in more extreme cases of moral ignorance. Consider again the case of Phineas Gage. While he may be exculpated under OIC retributivism for his wrongdoings immediately following his injury, it may be said that eventually, he might be held accountable for his failure to rehabilitate his moral disposition. Perhaps Gage may have a moral obligation to re-learn patience, kindness, civility, and cooperation. It is not obvious that Gage has the ability to do such a thing. Even if Gage found himself lucky enough to access competent psychiatric and

psychological care, he may put up resistance to his rehabilitation as a result of his newfound impulsive and obstinate disposition. As a reminder, this disposition, he acquired through luck. The acquisition of this disposition is not something for which Gage can be held accountable. It is not within Gage's ability to develop the belief that he ought to develop a moral disposition compatible with the norms of his society. Thus it is not within his ability to rehabilitate his disposition. Under OIC retributivism, Gage is exculpated for his wrongdoing on the basis of his ignorance.

I conclude this discussion with a case of seemingly archetypal malice to illustrate the way that truly any harm might be considered to have originated in ignorance and unluck.<sup>16</sup> Consider the case of Robert Alton Harris, who murdered two 16-year-old boys outside a McDonalds in San Diego in 1978 (Corwin 456). After Harris shot the boys multiple times, he returned to the boys' car, ate what was left of their food, and began laughing as he joked to his brother about the murders (Corwin 457). Harris' sister claims that Harris had told her that "he had his chance, he took the road to hell, and there's nothing more to say" (Corwin 259).

Here are some of my personal aretaic assessments, made on the basis of Harris' uncontested attributability for his wrongdoings.<sup>17</sup> Robert Alton Harris was a man who did not respect human life, and who did incredibly repulsive and unchoiceworthy things. His actions were wrong and have brought about unimaginable amounts of undeserved suffering to innocent people. Additionally, it is surely right for regulative measures to be taken to ensure further such harm does not occur.<sup>18</sup> Accountability, however, the matter of whether Harris deserves to suffer for the sake of his suffering as a result of his guilt, is a separate issue.

---

<sup>16</sup> This is another case where the actual facts are secondary to the ideas they demonstrate.

<sup>17</sup> I offer these to partially clarify what is and is not permitted according to mere findings of attributability, and partially to ensure the reader that the conclusions I draw are not intended to justify or condone Harris' actions.

<sup>18</sup> What form this regulative measures should take or should have taken, I feel myself unqualified to say.



Under OIC retributivism, we must investigate if Harris' actions originate in witting wrongdoing. He himself seems to believe so. Harris' claim that he 'had his chance' functionally expresses the belief that he had the opportunity to have done otherwise. Harris' belief that the origin of his crimes was his witting wrongdoing is subject to error, as all beliefs are. His testimony is insufficient on its own to determine the origin of his wrongdoings.

Before we determine whether Harris' crimes come from witting wrongdoing, I feel it necessary to provide more information regarding Harris' life and upbringing. Miles Corwin, in a report on Harris, writes that Harris' life "had been dominated by incessant cruelty and profound suffering" (Corwin 258). Harris' birth was induced in a hospital after his father kicked his mother in the stomach in a drunken rage (Corwin 258). The years that followed were no less tragic. Harris' life was marked by alcoholism, poverty, emotional neglect, physical abuse to the point of serious injury, the arrests of his parents, the death of his mother, unaided learning disabilities, violent stays in youth detention centers, subjection to sexual violence, and multiple suicide attempts (Corwin 258). His sister claimed that when Harris was 10 years old, before he was hardened by years of abuse, out of his nine siblings, "he was the most sensitive of all" (Corwin 258). Despite Harris' believing himself to have chosen his path, it does not seem obvious that this is the case. It seems improbable that Harris became the person he was solely through his volition. Harris was shaped by his remarkably unlucky upbringing. It is possible that, had Harris' childhood been even marginally less horrific, he may not have become a murderer.

Sher discusses the unconquerable ambiguity of the origin of wrongdoing, emphasizing the influence of moral luck in shaping one's moral disposition:

...we are generally not in a position to appreciate the cumulative effects of the innumerable smaller decisions of which our daily lives consist. Even if we grant that

[some agent was] remiss in not taking what in retrospect can be recognized as the steps that were necessary to prevent the development of their current bad traits, it will remain implausible to suppose that their wrongful failures to take those steps were themselves witting. (Sher 37)

It is possible that many people would have become wrongdoers, if only they were born into circumstances conducive to their becoming a wrongdoer. Similarly, some wrongdoers may have avoided their fate of becoming wrongdoers, if only they were they born in to circumstances conducive to that. It is a matter of unluck in one's circumstances that some wrongdoers become wrongdoers, and there is no perfectly reliable way to distinguish between these unlucky agents and those that are truly evil.<sup>19</sup> It is possible that Robert Alton Harris came to commit his crimes because of his moral unluck and that his belief that he had a choice in the matter is the product of his epistemic fallibility. If this is the case, it cannot be said that Harris is accountable for his wrongdoings under OIC retributivism. Because the origin of Harris' wrongdoing is too ambiguous for us to rightfully arrive at a verdict, we cannot find him accountable for his wrongdoings under OIC retributivism.

Ignorance exculpates under OIC retributivism, and it is possible for any sort of wrongdoing to have originated in ignorance (i.e., to have not originated in witting wrongdoing). The appearance of witting wrongdoing can be traced back to benighting circumstances that lie beyond the realm of an agent's control. Also outside the realm of an agent's control are appeals to moral standards like normative competence, as the actual ability of an agent is indifferent to objective standards. Appeals to normative competence are thus not grounds for determining accountability under OIC retributivism. For these reasons, ignorance and incompetence are

---

<sup>19</sup> If such a thing exists.

successful moral excuses under OIC retributivism, and any harm has the potential to have originated in ignorance or incompetence. Thus, under OIC retributivism, we cannot rightfully hold wrongdoers accountable unless we are rightfully certain that their actions originate in witting wrongdoing.

## **The Fallibility of Authorized Authority**

I now draw our attention away from the actual fact of an agent's being accountable and redirect it to our ability to recognize the fact of an agent's being accountable. There are notable philosophical traditions that hold that every harm is rooted in ignorance, and that people never knowingly commit wrongdoings. We need not go so far. I leave open the possibility for some harms to be sincerely rooted in malice, in known and intentional wrongdoing. I only aim to argue that we are too fallible to assign accountability in a way that would prohibit erroneous assignments of accountability on the innocent under OIC retributivism. We ourselves, as well as any other authorized authority, are vulnerable to fallibility. This is demonstrated by the same arguments and evidence that support the irreparable fallibility of all belief.

Ability can be compromised in many hard to detect ways. We do not have reason to think ourselves aware of all relevant pieces of evidence at any given time. We do not have reason to assume we will not err in our assessments of the evidence to which we are exposed. In fact, we have reason to believe otherwise. In the event that we, or any authorized authority, believe an agent's action to be sourced in witting wrongdoing, we have reason to doubt this belief. Just as unwitting ignorance can have an effect on wrongdoers, the same applies to authorized authorities.

One may claim that, despite the fallibility of our assessments of accountability, this fallibility should not inhibit our assessments of accountability. I am fallible in my belief that turning the faucet will cause water to come out, and yet, when I am thirsty, I turn the faucet. Despite the fallibility of our beliefs, we allow them to govern our action. This is necessary to lead a life, and thus, our assessments of accountability should be no different.

However, assignments of accountability have far higher stakes than other kinds of action. “Reputation, liberty, and even life can be at stake” (Watson 242). When we determine agents to be accountable in cases of ambiguity, we risk harming the innocent. One may object by saying that some sacrifice is appropriate to warrant the punishing of the guilty. This claim on its own is highly controversial, but suppose it were true. Suppose there were some ratio we could achieve, of innocent punished compared to guilty punished, such that we could consider our assignments of accountability in cases of ambiguity to be ethical. We have no way of determining that we are punishing within the acceptable ratio. One may even believe it highly probable that we would, in fact, meet this ratio. Again, the improbable is not impossible. We cannot guarantee our punishing of the innocent would be in the appropriate proportion to our punishing of the guilty, were such a proportion to exist.

No matter how confident we feel in some belief, all beliefs are fallible. To act according to a fallible belief is a gamble, and it seems morally questionable to gamble with the wellbeing of nonconsenting agents who very well may be innocent. For one to set the precedent that they may assign accountability despite their own fallibility is to allow for others to do the same. Not only do overeager assignments of accountability run the risk of the harming the innocent, but one risks becoming a victim themselves when adopting a culpability-independent means for determining worthiness of punishment.

Any case of wrongdoing has the potential to be sourced in ignorance or unluck, both of which are matters independent of ability. When we come to believe that a wrongdoing is rooted in witting wrongdoing, we have reason to doubt this conclusion. Our own fallibility results in perpetual ambiguity. In cases of ambiguity, we cannot responsibly determine accountability. Under OIC retributivism, we must withhold assignments of accountability as a recognition of our

own fallibility. We must do this in order to protect the innocent from wrongful punishment. OIC retributivism thus cannot rightfully prescribe assignments of accountability. OIC retributivism fails to punish the guilty in that it fails to punish anyone. A less conservative approach to assignments of accountability, due to our own fallibility, still would not guarantee that the guilty are the ones who receive punishment. This approach may result the innocent's being punished without solving the problem of the guilty going unpunished. For these reasons, OIC retributivism is impotent in its application.

## **Addressing an Anticipated Misunderstanding Regarding the Prescriptive Nature of These Claims**

One may mistake my argument regarding OIC retributivism's impotence as merely descriptive and not prescriptive. One may claim that, if regulative action remains an option despite an agent's not being accountable, and regulative measures can take the same form as retributive measures, then the argument as discussed above does not influence our conduct when it comes to responding to agents who commit wrongdoings. We may act as we always have and implement whatever measures we intuitively see fit, so long as we call them 'regulative.' This is not necessarily the case. When retribution is removed as an incentive and only regulation remains, our response becomes one only concerned with most effectively ensuring security and minimizing further harm. We need not compromise the efficiency of our regulative efforts to ensure the wrongdoer 'gets what they deserve.'

In cases where we have failed to find an agent accountable, we must treat them as though they are undeserving of retribution. Although perhaps guilty in the attributability sense, the wrongdoers maintain their innocence in the accountability sense and thus, retributivism claims they ought not be punished. We must consider the wellbeing of the wrongdoer along with the wellbeing of the innocent. In cases where the wrongdoer is not found accountable, the wrongdoer *is* innocent, in the relevant sense; the wrongdoer deserves not to be punished.

Sometimes, the most effective form of regulation can take the form of offering resources and aid, as well as taking preventative measures. I believe that we, both as individuals and as a society, can have the tendency to hesitate in employing these kinds of regulative measures, despite their proven efficiency. While perhaps these forms of regulation may be interpreted as a reward, it is not the case that retributivism makes a claim about one's deservingness of reward.

We cannot justify our withholding of the most efficient forms of regulation without appealing to the notion that it is morally right for the wrongdoer to suffer. As a result, when assignments of accountability and retribution are unwarranted, we are liberated to take the most effective course of action in terms of regulation. In this way, the argument I present regarding the impotence of OIC retributivism is not merely prescriptive.



## **Implications of Non-OIC Retributivism**

Thus begins our discussion of retributivism that does not include Ought Implies Can as a necessary condition for accountability (non-OIC retributivism). Without OIC as a necessary condition for accountability, murderers, bad believers, and those with poor moral dispositions may very well be found accountable. We may rightfully punish them for their actions. However, this consolation does not obviously outweigh the potential dangers of a moral theory that does not take into account the ability of the agents it punishes. Any form of non-OIC retributivism may mandate the punishing of some agents who put forth a good-faith effort in acting morally. This claim may not be particularly alarming on its face, but it becomes more so when we realize that we may find ourselves to be this agent. Under non-OIC retributivism, we may find ourselves in a position where the bringing about of our suffering (for the ultimate end of our suffering), is morally required, despite our belief that we are acting rightly. This possibility makes non-OIC retributivism hugely unappealing, as we risk our own wellbeing in exchange for the moral freedom to bring about the suffering of others.

In some cases, the prescriptions of OIC feel intuitive. Recall again the examples of Erin and Ayla. Erin, who could not lift the two-ton truck as a result of her lack of physical ability to do so, and Ayla, who neglected to aid her neighbor in need of urgent medical assistance as a result of her lack of belief that such a need existed. Although the consequences of Erin's and Ayla's actions (or lack thereof) are bad, we do not intuitively consider Erin or Ayla to be blameworthy in any sense that would warrant their being punished. In some cases of wrongdoing brought about by inability, we consider this inability to be exculpatory. This point is succinctly argued by Levy:

Suppose, for instance, that plants can be harmed, and that this harm is a moral reason against killing or treading on them. In that case, many of us are causally responsible for a great many moral harms. Are we morally responsible for them? Do we flout a moral requirement, and challenge plants' standing as objects to which some moral consideration is owed? No to all these questions: If we do not grasp the moral requirement, and this ignorance is not culpable, we do nothing blameworthy. (Levy 9)

How is non-OIC retributivism to account for this kind of intuitive conclusion? Without identifying a linkage between ability and responsibility, what is to stop individuals from being punished for committing wrongdoings they lacked the ability to not commit? Intuitively, it seems wrong to punish Erin, Ayla, or the plant-killer, despite their actions being bad (or at least, not good). Within non-OIC retributivism, we must develop a means for distinguishing between cases where an agent's action is merely bad and cases where an agent should be held accountable for their actions.

Like in OIC retributivism, the necessary conditions for accountability must be strict under non-OIC retributivism, as to prevent ambiguity that may lead to the punishing of the innocent. As a reminder, protecting the innocent is not a mere preference; it is mandated by retributivism.

One may propose a less strict implementation of some ability principle, one that includes a 'should have believed' or normative competence standard, where agents can be held accountable in the event that they fail to comply with said standard. Note that any normative-competence standard, where the standard is established independent from the actual ability of the agent to whom said standard applies, may result in a situation where an agent is punished despite their good-faith effort at behaving morally.

Consider a variation on Nicky's case regarding the vaccine. Imagine a different agent, Annie, faces the same situation as Nicky. Annie gets word of a new vaccine that purportedly protects against infection from the virus causing a global pandemic. For the sake of example, consider Annie to be equally as competent, truth-seeking, and well-intentioned as Nicky. Assume that Nicky and Annie have equally fallible doxastic practices (i.e., they should be considered equally likely to develop a true or false belief). Annie, through luck or unluck, comes to believe that the vaccine is safe, effective, and essential for protecting her community. Annie acts accordingly. She gets vaccinated, shares the evidence that led to her decision with the people she encounters, and encourages other to get vaccinated, as she believes this to be the safest option.

Suppose the normative-competence standard for accountability is employed, and Annie's society grows to adopt the normative standard that one ought to believe the vaccine is unsafe and act in accordance with this belief. In this case, Annie, despite her best efforts, violates the normative standard. Under non-OIC retributivism, it is morally right for Annie to suffer for the ultimate end of her suffering; her society would be rightful in their choice to take vengeance on Annie. Although Annie acted to the best of her ability and with good intentions, her being punished would be morally required under non-OIC retributivism with a normative competence standard for accountability. Under this kind of non-OIC retributivism, guilt is not established relative to truth or ability, but relative to normative standards.

In this case, I intentionally do not specify whether the vaccine is, in fact, safe or not. The normative standards are ones made by the individuals in Annie's society, whose beliefs are fallible. The ideals promoted by normative standards are not a matter of truth, rather, they are a

matter of what the prominent voices in a society perceive as the truth.<sup>20</sup> Normative standards operate independent of the actual truth and thus, normative standards have the potential to not correspond to truth. As a reminder, retributivism mandates punishment beyond regulation; the punishment must have the ultimate end of bringing about suffering. Non-OIC retributivism with a normative competence standard for accountability dispenses this kind of punishment to those that fail to meet a fallible and potentially subjectively unmeetable moral standard.

No person should think it impossible that they might end up in a position like Annie's. The agents referenced here are well meaning, and they arrive at their conclusions hugely due to circumstantial luck, constitutive luck, luck in the way things turn out, and luck in decision under uncertainty. Presumably, this is true of us as well; we have no valid reason to believe ourselves to be more resistant to the influences of luck than anyone else. Still, we may struggle to imagine ourselves in Annie's position. We may mistakenly think ourselves to be exceptional. We may assume that we, by virtue of being ourselves, are epistemically superior. We have no reason to think that we would be aware of our own ignorance, were we to possess it. More likely, we would be ignorant to our own ignorance. The commitment to a belief in one's superiority, especially a kind of superiority that can (even partially) overcome luck, is a display of hubris and chauvinism.<sup>21</sup>

---

<sup>20</sup> Like previous discussions of fallibility, this is not to be mistaken with advocacy for relativism. I maintain that there is an objective truth in this situation. My point here is merely that agents are presented with limited evidence and have unreliable doxastic practices that have the potential to lead away from this truth. History is full of examples where the moral standards set by a society are seemingly not in alignment with truth. That said, this thought experiment can be read either as taking place in a world where the vaccine is safe, or as taking place in a world where the vaccine is unsafe. One may even read the thought experiment with an indifference to this fact. However, it should not be read as though both truths can exist simultaneously and subjectively. For this example, there must be an objective truth regarding the vaccine.

<sup>21</sup> Not only is this belief fallible, it is contradictory. The argument that some agent is especially resistant to the influences of luck by virtue of her having been born herself is dependent on the premise that this agent's moral status is significantly impacted by luck, in that being born as herself is a matter of luck

Normative-competence-non-OIC retributivism and any other non-OIC retributivism requires the drawing of boundaries around which kinds of lacking ability should exculpate an agent and which kinds of lacking ability should not. This is an arbitrary practice; the assumption is that some instances of lacking ability are less legitimate than others. There is no basis for this distinction, besides that of normative competence.<sup>22</sup> Other conditions for assigning accountability may be proposed, though anything short of OIC will allow for cases where an agent is rightfully found to be deserving punishment for an action she was unable to avoid, despite her good-faith effort to behave morally. Thus, non-OIC retributivism will always fall victim to some counterexample that demonstrates how we ourselves, despite our sincere best efforts, may end up deserving punishment.<sup>23</sup> To think oneself exempt from this possibility is to be mistaken. For this reason, non-OIC retributivism remains an unattractive moral theory in its callousness.

---

<sup>22</sup> Which, above, was shown to be a problematic metric

<sup>23</sup> I suppose this would not be true in a case where some version of non-OIC retributivism included some principle that specifically immunized some person or peoples from punishment. I do not see this as a legitimately viable or defensible moral theory, and so I do not feel the need to address it or object to it further.

## Conclusion

In this paper, I have argued against the functionality of retributivism. Through an exploration of the implications of retributivism were it to either accept or reject Ought Implies Can, I have identified retributivism as having unavoidable implications that challenge its viability as a moral theory. Due to moral luck, epistemic circularity, the limits on inductive reasoning, and doxastic involuntarism, our beliefs are irreparably fallible. These influences may lead us to error despite our best efforts to believe responsibly and act morally. Additionally, due to the ubiquity of these same nonvolitional factors that influence action and belief, the authorities that intend to make assignments of accountability are fallible as well. This results in an unavoidable ambiguity regarding an agent's status of deserving or not deserving to be held accountable for her actions.

As a result, under OIC retributivism, no agent can ever be rightfully found accountable, and thus no agent can ever be punished. This type of retributivism is impotent. Non-OIC retributivism, in necessarily falling short of OIC, will always include cases that mandate the punishing of well-meaning agents acting to the best of their ability. These agents might very well be ourselves. The notion that a moral theory may mandate action with the ultimate end of bringing about our suffering, despite our belief that we are acting rightly and according to our best efforts is hugely unappealing. This type of retributivism is callous. In that all versions of retributivism would fall under one of these two versions, retributivism is unavoidably unappealing in either its unavoidable impotence or callousness.

### Works Cited

- Alston, W. P. (2018). Beyond “Justification”: Dimensions of Epistemic Evaluation (Vol. 1–1 online resource). Cornell University Press.
- Bartha, P. F. A., & Pasternack, L. (Eds.). (2018). Pascal's wager (Ser. Classic philosophical arguments). Cambridge University Press.
- Clarke, R. K. (2014). Omissions : agency, metaphysics, and responsibility. Oxford University Press.
- Cramer, M. (2021). “Henrietta Lacks, Whose Cells Were Taken Without Her Consent, Is Honored by W.H.O.” The New York Times.
- Miles Corwin, Icy Killer's Life Steeped in Violence, L.A. TIMES, May 16, 1982, reprinted in Watson, supra note 13, at 268-270.
- Fumerton, R., & Brueckner, A. (1997). Metaepistemology and skepticism. *Mind*, 106(424), 782.
- Goldman, A. (n.d.). Goldman, A. 1979. “What Is Justified Belief?” in G. Pappas, ed. *Justification and Knowledge* (Dordrecht: D. Reidel), 1-23.
- Ginet, C. (2000). The Epistemic Requirements for Moral Responsibility. *Philosophical Perspectives*, 14, 267.
- Heintz, L. L. (1975). Excuses and “‘Ought’ Implies ‘Can.’” *Canadian Journal of Philosophy*, 5(3), 449–462.
- Hoffman, K. M., Trawalter, S., Axt, J. R., & Oliver, M. N. (2016). Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. *Proceedings of the National Academy of Sciences of the United States of America*, 113(16), 4296–4301.

- Hume, D. Hume: *An Enquiry Concerning Human Understanding*. In Cahn, S. M. (1985).  
Classics of Western philosophy (2nd ed). Hackett Pub. Co.
- James, W. (2012). *The will to believe: and other essays in popular philosophy*. Duke Classics.  
Retrieved April 29, 2022
- Kurthy, M., Lawford-Smith, H., & Sousa, P. (2017). Does ought imply can? *Plos One*, 12(4),  
0175206
- Levy, N. (2017). The good, the bad, and the blameworthy. *Journal of Ethics and Social  
Philosophy*, 1(2), 1–16.
- Madigan, T. (2009). *W.k. clifford and "the ethics of belief"*. Cambridge Scholars Pub.  
Retrieved April 29, 2022, from INSERT-MISSING-URL.
- Mandivilli, A. (2022). “The C.D.C. Isn’t Publishing Large Portions of the Covid Data It  
Collects” *The New York Times*.
- O’Driscoll, K., & Leach, J. P. (1998). “No longer Gage”: An iron bar through the head. *BMJ :  
British Medical Journal*, 317(7174), 1673–1674.
- Peels, Rik (2011). Tracing Culpable Ignorance. *Logos and Episteme* 2 (4):575-582.
- Plato. (1987). *Gorgias*. (D. J. Zeyl, Trans.). Hackett Pub.
- Prichard, H. A., & MacAdam, J. (2002). *Moral writings* (Ser. British moral philosophers).  
Clarendon Press.
- Reid, T. (1983) *Inquiry and Essays*. Eds. Ronald E. Beanblossom & Keith Lehrer. Indianapolis:  
Hackett.
- Rock, I. (1997). *Indirect perception* (Ser. Mit press/bradford books series in cognitive  
psychology). MIT Press. Retrieved April 29, 2022



Sher, G. (2009-11-01). *Who Knew?: Responsibility Without Awareness*. : Oxford University Press. Retrieved 29 Apr. 2022

Smith, H. (1983). Culpable Ignorance. *The Philosophical Review*, 92(4), 543–571.

Taleb, N. N. (2007). *The black swan : the impact of the highly improbable* (First). Random House.

“The U.S. Public Health Service Syphilis Study at Tuskegee,” (2021). Center for Disease Control

Watson, G. (1996). Two Faces of Responsibility. *Philosophical Topics*, 24(2), 227–248.

Williams, B. A. O., & Nagel, T. (1976). Moral Luck. *Proceedings of the Aristotelian Society*, Supplementary Volumes, 50, 115–151.

Zimmerman, M. J. (2011). *The immorality of punishment*. Broadview Press. (Original work published in 1951)

Michael J. Zimmerman (2018) Peels on Ignorance as a Moral Excuse, *International Journal of Philosophical Studies*, 26:4, 624-632.