Moral Responsibility Has A Past: Has It A Future?

Thomas Nadelhoffer

Setting the Stage

On September 11th, 2001, there were four coordinated terrorist attacks in America carried out by the Islamic terrorist group al-Qaeda. Two planes crashed into the twin towers of the World Trade Center (razing them to the ground), one plane crashed into the Pentagon, and a fourth plane never made it to its intended target—crashing instead in a field in Pennsylvania thanks to passengers who sacrificed their lives to stand up to the terrorists who hijacked their plane. These attacks were driven by a deep-seated moral conviction that Americans needed to pay a price for their transgressions around the world. In this respect, the terrorist attacks were blowback for what were perceived to be past American sins. All told, nearly 3,000 innocent people were killed and more than 6,000 more were injured. Americans were unsurprisingly outraged and they demanded that someone pay a price for these cowardly deeds. Predictably misguided wars were launched in Afghanistan and Iraq—which have subsequently cost upwards of \$6 trillion.¹

The mastermind of the original attacks—Osama bin Laden—was relentlessly hunted for nearly ten years. Eventually, he was found in a cave in Pakistan in 2011 and assassinated without due process or a fair trial. Americans across the political spectrum nevertheless rejoiced in his death. Moral righteousness had its day in the court of public opinion and the guilty verdict was rendered—death to those who do unspeakable harm to innocent people. On the surface, it seemed as if the scales of justice had finally been brought back into balance. However, Americans

¹ America still has troops in Afghanistan more than 17 years later.

were so consumed by the desire for vengeance, we selectively lost sight of the fact that because of our military actions in the Middle East over 240,000 innocent civilians were killed² and an additional 21 million people became war refugees—an easily foreseeable side effect of exacting our proverbial pound of flesh from those we took to have done us harm. But we myopically looked past the associated costs of clearing the moral ledger. We needed to let other would-be terrorists know that Americans do not take attacks on our soil lightly. When provoked, we will exact our revenge come what may.

This example is one of countless tragedies that illuminates the problematic contours of our moral beliefs and practices when it comes to responsibility and punishment. On the one hand, misguided moral convictions can serve as the motivation behind deeply problematic behaviors. On the other hand, when bad people do horrible things to innocent people, we often have deep-seated visceral responses to their transgressions. We bridle with a desire to make sure that the perpetrators pay a price—whether it is a terrorist who flies planes into buildings, a rapist who murders children, or even a garden variety thief. In the wake of moral violations, people demand that something be done to right the wrong—even if the consequences of doing so are worse than the original offense. And if restitution is not possible, then retribution will have to do. These attitudes, beliefs, feelings, and behaviors are deeply entrenched, psychologically speaking—presumably having their roots in our simian past. In the struggle for survival, our evolutionary ancestors who responded negatively to violations of the social compact were more successful than those who did not. In that sense, our moral beliefs and practices were clearly

² Some estimates suggest that there were more than 600,000 casualties.

adaptive. But just because they were useful at a distant time in the past doesn't mean they are presently serving our interests and needs.

Consequently, some philosophers have started to question whether our intuitions about moral responsibility and punishment are justified. Given how deep and ubiquitous these intuitions happen to be, this debate may seem like a mere philosophical abstraction. After all, how could merely theoretical considerations penetrate beliefs and practices that are so central to both our identity and our social practices? The vast reach of these beliefs and practices seems to outstrip the grip of any purported criticisms. It is hard enough to change people's minds about more mundane things (from the dangers of smoking to the importance of flossing one's teeth)—how could we possibly convince people to give up their most foundational beliefs about themselves and others? There are two responses to this worry: First, some individuals have already given up their convictions about moral responsibility. And while it is admittedly sometimes hard for skeptics to put their beliefs into practice, these beliefs nevertheless have some interpersonal and intrapersonal traction. Second, some cultures have more moderated views about agency and human responsibility than others³—views which influence not just their relationships with other people but also the public policies their societies adopt. As such, it is clear that just as progress has been made on particular moral issues—e.g., individual rights—progress can be made when it comes to our beliefs about moral responsibility more generally.

³ Although it is worth noting that it is unclear whether any societies have *fully* dispensed with the notion of moral responsibility in the way suggested by some skeptics suggest we should. Instead, it seems like all societies have moral responsibility norms—they just involve different assumptions, attitudes, and practices.

Given that moral skepticism at both the micro and macro levels is possible, the key question then becomes: Is such skepticism warranted? There are two important distinctions that need to be made when analyzing the meta-ethical state of affairs. First, there is the difference between external challenges to moral responsibility and internal challenges.⁴ An external challenge calls our entire moral framework into question. On this view, it is not that fewer individuals *within* the moral responsibility framework are responsible than we formally thought—perhaps because human psychology is different than we traditionally assumed—rather, it is that *no one* is morally responsible. On this globally skeptical view, while there are important epistemic and volitional differences between classes of people—e.g., people with delusions, people with compulsions, or people with other psychiatric disorders—when it comes to lacking free will and moral desert, we are all of a piece. An internal challenge, on the other hand, does not claim to undermine the entire system of moral responsibility from the outside, so to speak, but rather seeks to transform our beliefs and practices from the inside. Here, moral responsibility is to be revised in light of what we are learning about the nature of human thought and behavior. For instance, given what neurobiologists have discovered about the etiology of violent behavior, one might conclude that a large percentage of people we currently blame and punish should either be excused or receive mitigated sentences instead. Here the skeptical challenge works internally against the backdrop of our existing beliefs and practices. It doesn't involve or require a wholesale rejection of the very idea of moral responsibility.

⁴ I am borrowing the distinction between external and internal challenges from Stephen Morse, who uses it to explain different challenges to criminal responsibility based on the brain sciences (2004).

In addition to the distinction between external and internal challenges to moral responsibility, there is also a distinction to be drawn between two different *sources* of skepticism. On the one hand, there are theoretical considerations. For instance, as we will see, some philosophers think that moral responsibility requires a kind of free will that is impossible if not incoherent. Here the skepticism is grounded in conceptual concerns about determinism, indeterminism, moral luck, and the like whereby traditional conceptions of agency and responsibility are ruled out in principle. On the other hand, there are empirical considerations. For instance, one might focus on the latest scientific findings on addiction to suggest that addicts ought not be blamed or punished for their behavior. Or, if one were using empirical considerations to make an external challenge to moral responsibility, one could highlight the evolutionary origins of our moral beliefs and practices as a way of debunking them. Obviously, one need not adopt a binary approach in motivating one's skepticism about moral responsibility—one can instead use both theoretical and empirical grounds. It is nevertheless important for skeptics to be clear about which approach they are adopting.⁵

That said, I would now like to turn my attention to the six articles included in this special issue of *Journal of Information Ethics*.⁶ While some of these article defend skepticism about moral responsibility (in some form or fashion), others try to preserve moral responsibility by responding to some of the arguments put forward by skeptics or by revising our practices or the theories we

⁵ In light of these considerations, there are five varieties of skepticism about moral responsibility: (a) theoretical internal challenges, (b) empirical internal challenges, (c) theoretical external challenges, (d) empirical external challenges, and (e) mixed challenges that incorporate different aspects of (a) through (d).

⁶ I will not be fully summarizing these articles—which is unnecessary given that you can just read the articles in their entirety. Instead, I will be covering just enough to highlight the connective tissue that runs through them all. Also, I will not be developing any criticisms along the way. While there is much I disagree with (more so in some cases than in others), now is neither the time nor the place to air my concerns. The focus should be on the work itself (not potential problems with the work).

use to justify them. As a collection, these articles do a nice job of showcasing some of the complex theoretical and empirical issues that arise in this context. Whatever one thinks about the fundamental issues involved, this special issue highlights the fact that things are less obvious than philosophers and the public more generally have traditionally assumed. While it is clear that moral responsibility has a troubled past (as evidenced by my discussion of 9-11), it is less clear whether it has a future and if so what shape it is likely going to take.

Into the Breach

The first article I would like to discuss is “Moral Responsibility is Morally Wrong” by Bruce Waller. Waller adopts a mixed approach to externally debunking what he calls the “moral responsibility system” that involves both theoretical and empirical objections. Along the way, he not only impugns the foundation of our intuitions about moral responsibility, but he also suggests that these intuitions lead to suboptimal behaviors (at the individual level) and suboptimal policies (at the societal level). On Waller’s view, the moral responsibility system is an evolutionary throwback that has gradually become obsolete. Given that he thinks our commitment to moral responsibility “fuels hard retributivism and popular punitivism, undermines rather than protects human rights, promotes destructive shaming, grossly distorts our understanding of free will, blocks deeper inquiry into the causes of both positive and negative behavior, comforts the privileged, afflicts the poor, and impedes the development of a super system,” Waller adopts an eliminativist rather than a revisionist stance towards our moral beliefs and practices.⁷

⁷ Although it is worth noting, that some of these appear to be *moral* objections to moral responsibility. For instance, a notion of fairness seems to motivate a lot of these worries. Waller therefore owes us an account of why we should care about these moral beliefs and practices if no one is morally responsible. In short, we need an account of fairness

Waller's first objection focuses on what he calls the "strike back desire"—that is, the visceral desire to lash out at people who cause us harm. This kind of reactance is problematic because it is often neither proportional nor targeted—that is, our punitive responses are often harsher than the original harm and they need not even be directed at the person who harmed us. A powerful person might offend me. Because I cannot take out my frustrations on him, I may take them out on my friends, family, pets, or even inanimate objects (e.g., when I kick the coffee table after stubbing my toe). The overarching reactant desire is to discharge our frustrations with having been harmed—a desire that can misfire in predictably bad ways ranging from punishing people too harshly to punishing people who are completely innocent. We find this kind of behavior in both humans and nonhumans alike. It is deeply enmeshed in how we navigate social hierarchies involving unequal and unequitable power relations. When we cannot get even by "punching up," we have a propensity to "punch down"—a propensity Waller thinks is deeply intertwined with our moral responsibility system.

According to Waller, another pillar of moral responsibility is our propensity to believe in a just world. Drawing on the literature on attributional biases and system justification, Waller points out that the gathering evidence suggests that people are inclined to believe that the world is fair and just even when this flies in the face of the facts. Indeed, people who are disadvantaged by unfair social arrangements are the very ones who are most likely to think that the system is fair and that their place within the system is deserved. The driving idea here is that we find unfairness unsettling. So, if there is something we can do to address it, we will. However, when

without responsibility. I am not saying that this account cannot be developed, I am just saying that such an account is necessary if moral objections to moral responsibility are to make sense.

we are relatively powerless to change the system, we are prone to revert to unconsciously justifying the injustice in order to make ourselves feel better about something that is otherwise aversive. In short, our propensity to find the world to be fair and just helps entrench unfair and unjust power relations. Not only do the powerful walk away feeling like they have earned their spoils, but the powerless blame themselves for their own compromised condition. Waller suggests that while just world beliefs help *explain* our moral responsibility system, they also *problematize* it.

According to Waller, another pillar of our steadfast commitment to moral responsibility is our faith in the belief that we have self-making, contra-causal powers. Here the metaphysical idea is that we have the ability to cause things to happen in the world without at the same time being caused to cause those things to happen. It is this radical kind of free will that purportedly undergirds moral desert (which is in turn needed for retributive punishment). As attractive as this kind of *causu sui* is to philosophers and theologians alike, Waller thinks that the concept is misguided. Not only is it metaphysically spooky, but it also problematically undercuts the otherwise important inquiry into our deeper motives, desires, and intentions.

After all, if you think someone can “pull themselves up by the bootstraps out of the swamps of nothingness,” to use Nietzsche’s color description, then there’s no reason to ask why an agent had the particular motives and choices that led to his action. If the agent serves as a full stop, etiologically speaking, then there is no point in looking at the antecedent conditions and circumstances that led up to his choice and behavior. But, as Waller notes, it’s looking at precisely these priors that undercuts moral responsibility because what you find are myriad determining forces over which the agent had no control. As he says, “the skill and fortitude with which you

‘play the cards that were dealt you’ are ultimately among the cards that were dealt you (dealt either by genetics or conditioning).” So, by foreclosing on these background considerations, the commitment to self-making powers papers over one of the main objections to moral responsibility—namely, that we are not ultimately in control of our behavior.

Based on these three considerations—namely, the strike back desire, just world beliefs, and the belief in our self-making powers—Waller thinks we have theoretical and empirical reasons for rejecting the moral responsibility system outright. Additionally, he thinks there is ample evidence that our beliefs about moral responsibility lead to bad behaviors and policies. For instance, Waller points out that in America and in the UK, where a problematic commitment to rugged individualism is commonplace, we adopt a highly retributive approach to crime (which he thinks is counter-productive), and a punitive approach to the poor and disadvantaged (who are blamed for their own condition and largely left to their own unfortunate devices). As Waller points out, when we look at cross cultural comparisons with societies that do not have such a moralistic and metaphysically bloated framework, we find more humane and effective societal responses to both crime and poverty. Given these downstream effects of the moral responsibility system, Waller finds additional grounds for questioning its iron-fisted grip on our beliefs and practices.

The article by Gregg Caruso is a nice compliment to the piece by Waller. In “A Defense of the Luck Pincer: Why Luck (Still) Undermines Moral Responsibility,” Caruso picks up where Waller’s discussion of self-making left off. The main issue addressed by Caruso is moral luck—that is, when factors beyond an agent’s control nevertheless influence our moral assessment of the agent. Thomas Nagel famously showed that moral luck is real, descriptively speaking—that

is, it seems that luck does, in fact, influence our ordinary moral judgments in garden variety cases (1979). But it's an open normative question whether it *should* play a role in our moral deliberations. This is the issue Caruso addresses. On his view, luck ultimately undermines moral responsibility. Once all of the myriad ways that luck influences our moral behavior are taken into consideration, there is no room left for moral responsibility to get a foothold. For this reason, Caruso follows Bruce Waller, Neil Levy, Galen Strawson, and others in adopting a skeptical view whereby "moral luck does not exist because luck universally undermines responsibility-level control." More specifically, Caruso thinks that luck threatens the type of agency that undergirds basic desert—that is, the kind of responsibility that would make us apt targets of praise and blame (and punishment and reward).

The first thing worth mentioning on this front is that luck, by definition, is something beyond our control. It is for precisely for this reason that a tension exists between luck and responsibility. After all, it seems intuitive that we can only be responsible for things over which we have some control. While the relationship between control and responsibility seems straightforward enough at first blush, when we dig deeper into the variety of ways that luck influences our lives, we find that we have much less control than we have traditionally assumed. To illustrate this point, Caruso focuses primarily on two varieties of moral luck. First, there is constitutive luck—that is, luck when it comes to the so-called "lottery of life." Here, the idea is that our genetic and psychological predispositions, coupled with how, when, and where we were raised (and by whom), have a sizable influence on our present decisions and actions. Indeed, these background conditions shape (or even determine) the beliefs, desires, and motivations that in turn shape (or even determine) our present behavior. Second, we can be lucky (or unlucky)

when it comes to our present situation—which is called circumstantial luck. We could be at the right place at the right time, the right place at the wrong time, the wrong place at the wrong time, the wrong place at the right time, or anywhere in between. In each case, our present circumstances are influenced by factors beyond our control. So, while constitutive luck refers to properties of the agent that are beyond his control, circumstantial luck refers to properties of the immediate situation that are beyond the agent's control.

Once all of the luck leading up to our decisions and actions is taken into consideration and once all of the situational variables that are beyond our control are factored into the equation, it is unclear what is left of our purported control (and hence our moral desert). Notice that this observation is independent of the traditional worry about determinism. Regardless of whether determinism or indeterminism is true, it seems as if our actions are entirely the result of forces beyond our control. Indeterministic luck is no less threatening than deterministic luck. In either case, it seems like our agency shrinks to an “extensionless point” (Nagel, 1979, p. 35) in the face of the problem of luck. In this way, luck cuts across the traditional debate between incompatibilists and compatibilists, as both views are challenged by the ubiquity of luck in our lives (albeit in different ways).

The specific argument that Caruso marshals in support of his skeptical thesis is the *Luck Pincer* argument developed in Levy (2011)—which has the following structure: (a) the universal luck premise—“every morally significant act is either constitutively lucky, presently lucky, or both”; (b) the responsibility negation premise—“constitutive and present luck each negate moral responsibility”; and (c) the conclusion—“an agent is not morally responsible for any morally significant acts.” On this skeptical view, “luck swallows everything,” to borrow a phrase from

Strawson (1998). Once we control for all of the things over which we have no control, there is nothing leftover to ground free will and responsibility—which, here again, holds no less true in a deterministic universe than in an indeterministic universe. In either case, agents have no ultimate control over how they are psychologically constituted or the proximal situational constraints on their choice architecture. As such, it is difficult to see how they could come to be responsible for the influence luck has on our their lives. As Levy (2009) puts it, “we cannot undo the effects of luck with more luck” (p. 242). This is true regardless of whether one’s account of agency and responsibility is historical or non-historical—that is, whether one focuses on the actual series of events leading up to an agent’s choices and actions or whether one focuses merely on the immediate details of the agent’s motivation and decision making.

In light of Caruso’s stance on the all-encompassing influence of luck on our decisions and actions, it’s no surprise that he is a skeptic about free will and moral responsibility. It’s also no surprise that he rejects not just retributive, desert-based theories of punishment, but the very institution of punishment itself. In a recent paper with Derk Pereboom, Caruso has defended a preventive and rehabilitative approach to crime that jettisons punishment altogether (2018). Rather than grounding punishment in deterrence or retribution—the two most popular justifications of our penal practices and institutions—Pereboom and Caruso defend a “quarantine model” of crime prevention that is based on societal self-defense and public health ethics. On their view, dangerousness and not deterrence or desert ought to be the basis for our policies on crime prevention and rehabilitation. They suggest that their approach to crime is both more efficient and more humane than the alternatives. By unmooring our practices from the fiction of moral desert, their approach would purportedly enable us to approach crime in a more

scientifically informed way—a way that will privilege preventive measures (e.g., improved early childhood education) to incapacitative measures (e.g., detention). And while some criminals will invariably need to be detained, we will be obligated to ensure that their detention is as comfortable and rehabilitating as possible. Given that we're not making an example out of criminals or giving them their just deserts, we have to treat them like wards under our care and supervision.

I have briefly sketched out Pereboom and Caruso's quarantine model of crime prevention and rehabilitation because John Lemos raises concerns about their type of non-desert-based approach in his contribution to this special issue entitled, "A Moral/Pragmatic Defense of Just Deserts Responsibility." Lemos takes a kind of epistemic uncertainty as his starting point. On his view, it is an open question whether we have libertarian free will and desert-based responsibility. While Lemos is an event causal libertarian in theory, he openly acknowledges that "there doesn't seem to be empirical, logical, or metaphysical evidence that we possess free will." While one might wonder what could possibly count as sufficient evidence on this front, that is neither here nor there for present purposes. For now, we can simply assume that it is a perpetually open question. Then, the question becomes: What should we believe and how should we act in the face of this unending uncertainty?

According to Lemos' pragmatic approach to libertarian free will, because compatibilism is problematic and because free will skepticism yields morally problematic results, we ought to act as if we have the kind of libertarian free will that grounds desert-based responsibility (regardless of whether we have compelling evidence either way). For while it is an open question whether we are actually free and deserving, things will purportedly go better for individuals and for society

if we nevertheless proceed as if we are. As Lemos acknowledges, one common criticism of this pragmatic approach to free will is that it is “hardhearted” (Double, 2002)—that is, it is hardhearted to subject criminals to punishment in the absence of evidence that this treatment is truly deserved. On this view, in the face of this kind of uncertainty, the default view should be “do no harm.” By subjecting criminals to hard treatment in the absence of proof that they are deserving, we unjustifiably harm them. Unsurprisingly, Lemos thinks this objection falls short.

In responding to the argument from hardheartedness, Lemos draws a distinction between humans and non-human animals (e.g., rattlesnakes) to highlight that humans have a special kind of dignity that demands our respect that other animals lack. Moreover, because humans have this special kind of dignity—which Lemos suggests is predicated on free will and moral desert—they are entitled to certain kinds of treatment and protected from other kinds of treatment. Because the argument from hardheartedness seems to rest on precisely this kind of dignity, it is self-undermining—that is, it requires the very kind of free will and desert it repudiates. If we give up these notions, we purportedly lose the ability to insist that only the demonstrably guilty ought to be punished. Without desert to serve as a side constraint, both disproportionate punishment and even punishing the innocent become justifiable if the positive consequences are significant enough. In this respect, Lemos puts the shoe of hardheartedness on the skeptic’s other foot. In their effort to make our treatment of criminals more humane, free will skeptics purportedly make it more likely that the innocent are mistreated. According to Lemos, these problems pertain to any non-desert-based approach to crime—whether the focus is on dangerousness, deterrence, restitution, or moral education.

By giving up desert, free will skeptics purportedly not only weaken the foundation of human dignity, but they also make it more likely that the rights of the innocent are abridged and that the rights of the guilty are eroded. According to Lemos, by throwing out the desert-bearing baby, skeptics throw out the criminal justice bathwater. To highlight this, Lemos takes a closer look at the aforementioned quarantine model by Pereboom and Caruso. While the latter think their approach would have a humanizing effect on how we treat criminals, Lemos thinks it will likely have the opposite effect. On his view, if we are to be justified in giving up free will and moral desert, it must be shown that by living without our traditional beliefs we would be better off than we are living with them. Because Lemos thinks implementing skepticism will make things worse, he thinks we should stick to our proverbial libertarian guns (even in the absence of compelling metaphysical evidence to the contrary).

Lemos adopts a two-pronged attack on the quarantine model. First, he attempts to show that if we adopted the view, it would force us to lower evidentiary standards—which would in turn infringe the rights of innocent but otherwise potentially dangerous people. Second, Lemos thinks that the quarantine model doesn't have the theoretical machinery to ground the kind of dignity he thinks is the hallmark of human beings. In defending the first claim, Lemos points out that we could enhance the self-protective measures that serve as the bedrock of the quarantine model by lowering the evidentiary standards for determining whether potentially dangerous individuals are apt targets for intervention. Because public safety is the *raison d'être* of the quarantine model and because lowering evidentiary standards can facilitate this end, Lemos thinks the skeptic is committed to adopting this type of policy in both theory and practice. Given that skeptics have given up free will and desert, they purportedly don't have the moral means

for staving off this evidentiary gambit. And dovetailing with what he said earlier, Lemos concludes that to the extent that skeptics are committed to lowering evidentiary standards—whether they openly disavow this approach or not—they cannot adequately accommodate the dignity of innocent people who would otherwise be swept up in their quarantine model.

In short, Lemos relies on a kind of slippery slope argument against the quarantine model and against skepticism about free will and moral desert more generally. Given that we can't know for certain whether skepticism is true and given that adopting skepticism purportedly yields bad consequences, morally speaking, we should therefore continue acting as if we are free and desert-based responsible. Whether one thinks this is a fallacious argument will depend in part on whether one thinks the empirical claims about what follows from skepticism are well-founded. Not all slippery slopes arguments are ill formed, after all—some slopes really are slippery. Whether closing the door on free will and responsibility necessarily opens the door to injustice is a live empirical question. In her article entitled, "Justice without Moral Responsibility," Elizabeth Shaw addresses some of the types of concerns raised by Lemos. So, I would now like to turn my attention to what Shaw has to say about skepticism and the law.

Like Lemos, Shaw does not focus on whether we should accept or reject skepticism about free will and moral responsibility. Instead, she considers the possible implications of giving up responsibility and transitioning to a non-desert-based approach to dangerous and criminal behavior. Unlike those who think that eschewing responsibility would have deleterious moral effects on our legal practices, Shaw thinks that social-protection approaches to crime need not undermine the dignity of offenders and need not erode the central role played by justice and rights in the law. Shaw thinks that offenders—both the mentally ill and non-mentally ill alike—

will still be protected by a non-retributive conception of justice that values the importance of treating offenders with respect and dignity. To highlight that this is likely to be the case, Shaw spends some time focusing on how we already treat the non-responsible from the standpoint of the law whether we look at the civil commitment of the dangerous or the excusing conditions in the criminal law for the non-responsible. On her view, a close look at how we actually already treat people who we deem to be non-responsible can help shed light on how offenders might be treated if we dispensed with responsibility altogether.

Shaw begins by acknowledging that if it could be shown that social-protection approaches to crime erode the basic rights and protections of offenders, this would undercut one of the alleged benefits of these approaches—namely, the humanizing effect they would purportedly have on the criminal justice system. However, Shaw thinks the worries that have been raised have been overblown. So, she looks at three claims in particular: (a) the claim that advocates of social-protection approaches must condone framing the innocent, (b) the claim that advocates of social-protection views must condone disproportionately severe sentencing, and (c) the claim that advocates of social-protection views can't accommodate basic legal safeguards such as the state having the burden of proof or the right of the accused to challenge the case against them. The claim in each of these three cases is that once desert no longer serves as a side constraint on how we treat offenders, there is nothing left to preserve the basic justice in the criminal justice system. Shaw's strategy is to combat this claim by showing that non-retributive justice can do much, albeit admittedly not all, of the work done by retributive justice in safeguarding those who have been accused of (and found guilty of) a crime.

In pursuing this argumentative strategy, Shaw looks to the safeguards we already put in place when it comes to the non-responsible—safeguards which can't have their backing in retributive theory since everyone agrees these individuals don't deserve punishment. But because we recognize the fact that these individuals are persons who have basic rights and interests—and that certain forms of treatment would be positively unfair and unjust in their case—we try to ensure that these dangerous individuals are nevertheless treated with maximal dignity and respect. Even if we decide that detention is necessary, we try to limit the negative impact this has on their lives as much as possible. Consider, for example, the claim that advocates of a social-protection approach like Pereboom and Caruso are committed, in theory, to punishing the innocent. Shaw questions this claim by pointing out that knowingly and intentionally punishing an innocent person harms them in various ways. And while a deterrence theorist might justify this harm in the name of the greater good—social-protection models often focus instead on individualized prevention and rehabilitation. There is no sense in which punishing an innocent person facilitates prevention or rehabilitation.

So, while there is a way to justify punishing the innocent on non-retributive grounds, not all non-retributivists are stuck with this view. Instead, non-retributivists can object to so-called “manipulative use” (Tadros, 2011) of the innocent and non-innocent alike. On this view, even if no one is responsible in the basic desert sense, it's still wrong to publicly lie about people, to cause them irreparable reputational damage, and to treat them as a mere means to societal ends. As such, even on non-retributive grounds, framing the innocent is unjustifiably manipulative and unjust. Both the innocent and the non-innocent alike must be acknowledged and respected as persons who are owed dignified treatment. In short, the concept of rights is broader than the

concept of retributive justice—which presumably includes the rights of individuals not to be falsely accused and the rights of citizens not to be systematically deceived by their government.⁸ If this is correct, then non-retributivists can protect the rights of the innocent—contrary to what critics like Lemos have suggested.

However, just as critics worry that non-retributive approaches will run roughshod over the rights of the *innocent*, they also worry that the rights of the *guilty* will be violated as well in the form of sentences that are disproportionately harsh. Once treatment replaces desert-based punishment, the claim is that people could be permanently detained for even minor offenses so long as they are deemed to be dangerous. Here Shaw claims that the same kind of non-retributive justice that she discussed in the context of punishing the innocent can help guarantee some form of proportionality in the context of those who actually committed an offense. On her view, detentions and treatments must cause no more harm to the offender than the potential harm being prevented would cause to any would-be victim. So, if the offender is being detained for being a nuisance, the sentence cannot be any more onerous than the potential damage caused by a future nuisance. On the other hand, when the offender is a dangerous murderer, the potential harm being prevented is another murder—as such, a much lengthier sentence is merited. In each case, the potential damage that could be caused by the averted crime serves as a proportionality constraint on the sentence. Minor offenses merit minor sentencing whereas serious offenses merit serious sentences. The operative principle here, according to Shaw, is

⁸ For instance, Benjamin Villhauer (2013) uses a Rawlsian “original position” approach to ground personhood, rights, and justice in a non-retributive framework whereby no rational and self-interested individual would agree to live in a society that didn’t put certain safeguards in place when it comes to the criminal justice system. As Shaw points out, this provides proof in principle that one need not appeal to basic desert to guarantee a fair and just system. A non-desert-based contractualist approach can suffice.

based on a “ideal of equality” whereby “it is prima facie wrong to create a situation where people suffer grossly unequal levels of distress.”

Finally, critics of skepticism worry that social-protection approaches to crime can’t adequately protect other legal safeguards such as due process. Once again, Shaw points out that this worry flies in the face of the liberties and legal safeguards already afforded to people who are deemed not to be morally responsible for their behavior—e.g., the burden of proof is on the state to establish that someone is a danger to themselves or others before they can be civilly committed. Just as these safeguards apply to the mentally ill and legally insane, they can and should be extended to sane but non-responsible offenders who break the law. Given that we already try to protect the interests of the non-responsible mentally ill, Shaw sees no reason why we couldn’t treat the non-responsible mentally well with similar care and concern. Take the standard of proof which Lemos used as grounds for criticizing skepticism. According to Shaw, there are non-retributive reasons for ensuring that the burden of proof should be beyond a reasonable doubt. On the one hand, having high evidentiary standards helps uphold the value of liberty and protects citizens from an overreaching state. On other hand, forcing the state to clear a high evidentiary bar also enables the government to respect its citizens by assuming that they are innocent until proven guilty. Both of these values are purportedly consistent with skepticism about moral responsibility. If Shaw is right in this regard, then she has at least partly allayed some of the concerns raised by Lemos and others. Figuring out whether all of these concerns can be laid to rest is a task for another day. For now, I wanted to turn our attention to two other responses to skepticism that do not focus on non-retributivism and criminal justice.

Let's return, for a moment, to our earlier discussion of moral luck. One way of motivating the problem of luck is to highlight that given the ubiquity of luck, nothing is ultimately up to us—that is, we are not the ultimate source of our decisions and actions. If everything I do is a matter of luck, then I am not sufficiently in control of my behavior to be free and responsible. To the extent that my actions are completely traceable back to things beyond my control, I am not the source of my actions. This way of framing things has come to be called “source incompatibilism.” In his article entitled, “Manipulation and Independence,” Justin Coates makes a contribution to this debate by trying to undermine a common argument for source incompatibilism. Rather than focusing on luck, other source incompatibilists have focused on the purported connection between manipulation and determinism. On this view, just as manipulated agents aren't responsible, determined agents aren't responsible because causal determinism “precludes agents (or their wills) from being the causal sources of their actions in the way that responsibility requires.” Coates tries to show that while manipulation may undermine free will and responsibility, there is nevertheless an important difference between manipulated agents and merely causally determined agents. If this is right, then the move from manipulation to determinism is problematic—which in turn undermines one of the key arguments for source incompatibilism.

Coates begins by trying to get clear on why manipulation is taken to be threatening. When agents are manipulated, their actions are traceable to factors beyond their control. The manipulator and not the manipulated is said to be to blame. In this way, the beliefs, desires, and preferences of the manipulated agent are bypassed by the intervention by the manipulator. When an agent is manipulated, he is not sufficiently in control of his decisions and actions to be

free and responsible. Everything traces back to the person who is pulling the proverbial strings, so to speak. Philosophers who have developed arguments against compatibilism based on manipulated agents try to show that determinism plays the same etiological role as the manipulator. After all, if determinism is true, everything we do is traceable to forces similarly beyond our control. Perhaps the two most widely discussed manipulation arguments for source incompatibilism are Al Mele's Zygote Argument (Mele, 1995; 2006) and Derk Pereboom's "Four Case Argument" (2001; 2014). In both arguments, a case involving manipulation is first used to anchor our intuitions. The manipulated agents are said to have all of the salient compatibilist abilities and capacities—e.g., reasons responsive mechanisms, practical reasoning, higher order desires, etc. But because all of their attitudes, beliefs, desires, and preferences have been completely manipulated, the agents purportedly aren't free and responsible since they aren't the proper source of their decisions and actions. Then, Mele and Pereboom claim that in a deterministic universe, we're in the same predicament as the manipulated agents—all of our capacities, abilities, attitudes, beliefs, etc. are completely traceable to deterministic factors beyond our control. As such, causal determinism plays the same role as a manipulator—undermining our free will and responsibility.

Compatibilists have developed two basic strategies for responding to manipulation arguments. Hard compatibilists bite the bullet and claim that manipulated agents are free and responsible after all. Soft compatibilists grant that manipulated agents are unfree but they try to show that there are important differences between manipulated agents and causally determined agents. Coates falls into this second camp. His goal is to show that the slide from manipulation to causal determinism is problematic. On his view, determinism doesn't bear on an agent's

responsibility in the same way that manipulation does. To make his case, Coates develops four cases that are designed to highlight the importance of being independent from the causes of one's actions when it comes to moral responsibility. In the case of manipulation, the causes of the manipulated agent's behavior are external to the agent—after all, it is the desires and preferences of the manipulator that are doing all of the etiological work. The agent's own independent desires and preferences do not play a role. In cases involving manipulation, there is a kind of dependency relationship between what the agent does and what the manipulator wants. As a result, it makes sense that we don't find the manipulated agent to be free and responsible. But what about agents who are merely causally determined?

Here Coates first highlights the importance of independence when it comes to assessing manipulated agents. As he says, "the degree to which an agent is praiseworthy or blameworthy for her action depends (in part) on what they would have done had their actions been independent from the other agent's interventions." What an agent would do of his own accord, free from interference, is a good indicator of his own attitudes, beliefs, desires, and preferences. In the case of manipulation, the manipulator's own extrinsic influence trumps the agent's independence and thereby undermines his responsibility. The key question for Coates is whether determinism undermines independence in the same way. He thinks that it does not. On his view, one way to get at this is to ask what an agent would have done had things been different. In the case of manipulation, would the manipulated agent have acted differently but for the interference on the part of the manipulator? Presumably yes, which is partly why Coates thinks we don't hold manipulated agents responsible. Their actions are dependent rather than reflecting their own independent desires and preferences. Coates doesn't think the same thing

holds in the case of causally determined agents—which he tries to elucidate using possible world semantics. To see why, imagine a determined agent’s counterpart on a nearby indeterministic world. Assuming the counterpart is on a nearby world, then his desires and preferences would be largely the same. So, he would behave in much the same way as the determined agent. In both cases, the respective agents’ actions reflect their independent desires and preferences and hence their independence from outside forces. Their actions reflect who they are—they don’t bypass what they want. As such, we should find them to be free and responsible. If this counterfactual analysis is correct, then Coates has put pressure on a common argument for moral skepticism. I have my doubts, but now is neither the time nor the place to air them. Regardless of what I think about the merits of Coates’s argument, he has clearly advanced the debate by forcing both compatibilists and incompatibilists alike to get clearer on the purported relationship between manipulation and determinism (and the role played by counterfactual dependence).

That said, it’s finally time to turn our attention to the final contribution to this special issue—namely, Manuel Vargas’ “Responsibility, Methodology, and Desert.” Vargas elucidates the instrumentalist and revisionist account of moral responsibility that he has developed elsewhere. To say that his account is instrumentalist is to say that he grounds moral responsibility in terms of the various ways the system of responsibility benefits us both intrapersonally and intrapersonally. To say that it is revisionist is to say that it is flexible enough to depart from common sense morality when doing so is necessary to build a moral system that better serves our interests and needs. In this way, there are two dissociable elements to Vargas’ view—a normative account concerning the foundation of morality and an a methodological account that serves as a backdrop. Now I want to discuss each aspect of Vargas’ instrumentalist revisionism.

On Vargas' view, there is an agent-relative component to morality and an action-relevant component. First, in order to be morally responsible, an agent has to possess certain traits and abilities—e.g., the capacity to understand and be guided by morally salient reasons. Second, in order for an agent to be praiseworthy or blameworthy, he must have either acted in accordance with or violated a justified moral norm. Not all agents are responsible on this view—e.g., the very young or the mentally ill—and even those who are generally responsible may not be responsible given special circumstances. But in general, moral agents are responsible for how they navigate moral norms. Vargas calls the network of related attitudes, capacities, and behaviors that undergird attributions of moral praise and blame “the responsibility system.” On his view, this system is justified on two levels—at the level of moral practices and at the level of the individual.

First, the responsibility system is justified as a practice in virtue of the benefits that accrue to those who are part of the practice. As Vargas says, “the responsibility system, with its attendant entrenchment of social and moral norms, and its practices of praising and blaming, is justified if having these practices helps develop, sustain, and extend our ability to recognize and suitably respond to moral considerations.” But just because Vargas thinks that the justification of morality at this level is instrumentalist, he isn't committed to thinking that all instances within the practice are justified in instrumentalist terms. Consider an analogy: Just because the rules of a game can be justified in forward-looking terms, it doesn't follow that the application of the rules in particular cases need be guided by purely forward-looking considerations. Quite the contrary. People are punished for violating the rules of a game simply because they broke the rules—which is a backward looking consideration. Vargas suggests a similar tiered system applies to morality. While the moral practice itself is grounded in instrumental terms, particular norm

violations within the practice can be based on non-instrumentalist grounds. That said, the second level at which the responsibility system can be justified pertains to the interest of individuals in being moral agents. Here Vargas notes the important role that responsibility plays when it comes to our basic needs for “social life, social esteem, and the conditions of cooperative living.” According to Vargas, being competent to navigate the responsibility system is part of what it means to be a full member of moral community—a community which provides us with a number of important benefits both as individuals as a groups of individuals. It is these benefits which help mutually reinforce and justify the responsibility system.

Now that we have some sense for the instrumentalist nature of Vargas’ view, I want to take a brief look at the revisionistic aspect. After all, while Vargas thinks the aforementioned forward-looking considerations can normatively ground a moral responsibility system, he is not committed to thinking that the particular system he envisions fully comports with the way people ordinarily think about morality. For instance, most people likely ground their moral responsibility system in theological terms. Because Vargas is a revisionist, he is not wed to this particular metaphysical framework. His task is to see whether an interesting and desirable theory of moral responsibility can be developed that captures at least some of the important aspects of the basic practices of common sense morality. But he need not be interested in some kind of one-to-one mapping between his moral theory and ordinary morality. As he says, “the animating methodological conviction is that the stakes of a theory of moral responsibility are primarily our responsibility-characteristic practices and attitudes (blaming, indignation, and the like). Whether a theory that satisfactorily accounts for those things also captures the kinds of theoretically naïve,

pre-philosophical convictions we might have about our powers is a further, and (at best) secondary matter.”

Consider, for instance, the possibility that according to common sense morality, agents must possess the kind of robust free will preferred by libertarians—e.g., “the ability to do otherwise (given the past and the laws of nature) or the ability to be self-caused or unmoved movers, or otherwise ultimate sources of actions.” What follows for moral responsibility if we don’t and can’t have these metaphysically robust agential powers? According to Vargas, the key issue is whether our basic moral practices can be justified not whether our folk theoretical account of these practices can be justified. For a revisionist, we can jettison common sense metaphysics and focus instead on whether some other justificatory framework can be developed. So, that’s precisely what Vargas attempts to do—provide a foundation for our ordinary moral practices that isn’t dependent on a problematic metaphysical backdrop. To the extent that Vargas is right that moral theorists can and should go revisionist when folk metaphysics fails, this takes the wind out of the sails of skepticism—especially when the skepticism is grounded in the very metaphysical constructs the revisionist is happy to eschew.

But what about the fundamental concept of basic desert that is targeted by the skeptical views of Waller, Caruso, and others? Can an instrumentalist and revisionist like Vargas capture the basic practices associated with desert—practices that appear to be entirely backward looking in nature. At first blush, one might think that the instrumentalist account developed by Vargas is ill-suited to capture the retributive nature of a number of our moral beliefs and practices. However, Vargas revisits the analogy with games and rule violations. While the rules of the game are grounded in instrumental terms, once someone violates the rules of the game, this is

sufficient for penalizing the person for (a) being able to follow the rules, and (b) nevertheless failing to do so. In this sense, the penalty is backward looking in much the same way as moral norm violations. Just because one gives an instrumentalist account of the practice itself, it doesn't mean that penalties for violating the norms of the practice can't be non-instrumentalist in nature. If this right, then Vargas has the theoretical machinery in place to capture the backward-looking nature of the practices surrounding basic desert while distancing himself from the metaphysical baggage usually associated it. As such, Vargas thinks his revisionism can do the requisite moral heavylifting without paying the usual metaphysical price. Skeptics will sure disagree. And around it goes...

References

- Double, R. (2002). The moral hardness of libertarianism. *Philo*, 5, 226-234.
- Levy, N. (2009). Luck and history-sensitive compatibilism. *Philosophical Quarterly*, 59(235), 237-251.
- Levy, N. (2011). *Hard luck: How luck undermines free will and moral responsibility*. New York: Oxford University Press.
- Mele, A. (1995). *Autonomous agents*. New York: Oxford University Press.
- Mele, A. (2006). *Free will and luck*. New York: Oxford University Press.
- Morse, S.J. (2004). New neuroscience, old problems: Legal implications of brain science. *Cerebrum*, 6(4), 81-90.
- Nagel, T. (1979). Moral luck. In *Mortal Questions*. New York: Cambridge University Press, pp. 24-38.
- Pereboom, D. (2011). *Living without free will*. Cambridge, UK: Cambridge University Press.
- Pereboom, D. (2014). *Free will, agency, and meaning in life*. Oxford: Oxford University Press.
- Pereboom, D., & Caruso, G.D. (2018). Hard-Incompatibilist existentialism: Neuroscience, punishment, and meaning in life. In G. Caruso & Owen Flanagan (Eds.), *Neuroexistentialism: Meaning, morals, and purpose in the age of neuroscience*. New York, NY: Oxford University Press, 193-222.

Strawson, G. (1998). Luck swallows everything. *Times Literary Supplement*, 26 June 1998, pp. 8-10.

Tadros, V. (2011). *The ends of harm: The moral foundations of criminal Law*. Oxford, UK: Oxford University Press.

Villhauer, B. (2013). Persons, punishment, and free will skepticism. *Philosophical Studies*, 162(2), 143-163.