

EMBRACING RETRIBUTIVISM

Gregg Caruso's *Rejecting Retributivism* is a resolute and wide-ranging book. Since I cannot address all the material in it that deserves attention, I will try to focus on the central and substantive opposition between Caruso's views and mine. As my title immediately indicates, I disagree with Caruso's rejection of retributivism. Although this disagreement is very important, it nonetheless obscures a meaningful agreement. Caruso and I largely agree in our assessment of the horrible excesses of some contemporary criminal justice systems – above all the criminal justice system of the United States. It shall prove salutary to keep the agreement in the back of our minds throughout, since it constitutes a valuable backdrop to my general discussion.

Caruso's rejection of retributivism cannot be properly understood without at least briefly sketching its grand motivation. Caruso considers himself a "free will sceptic". While scepticism does not necessarily entail *denial*, and while at times Caruso seems to admit that his position may allow for *either* "doubting or denying" moral responsibility (37), his main line, repeatedly expressed throughout his book, is that his scepticism entails the full-throttled *denial* of moral responsibility. Caruso claims that his scepticism is somehow consistent with a number of conceptions of responsibility, but he is emphatic that it entails the complete rejection of the idea that we could be "*truly deserving* of praise and blame [or] punishment and reward" (30; emphasis in the original). Similarly, Caruso asserts: "my own view is [...] that [...] we are never morally responsible for our actions in the basic desert sense" (8; repeated also in 14, 106, and *passim*). The "basic desert sense" for Caruso is the standard sense of responsibility independent of "consequentialist or forward-looking considerations" (2).

So, Caruso's rejection of retributivism is inseparable from his rejection of moral responsibility and of desert. Caruso's *Rejecting Retributivism* could thus have been titled *Rejecting Moral Responsibility*, or *Rejecting Desert*, or, as we shall see, *Rejecting Punishment*. (Below I will suggest a more accurate title for Caruso's book.) The ambitiousness of Caruso's goals is evident throughout, but consider one representative passage:

It has been suggested to me, by friend and foe alike, that if my ultimate goal is criminal justice reform and the rejection of retributive practices and

Rejecting Retributivism: Free Will, Punishment, and Criminal Justice by Gregg D. Caruso (Cambridge University Press, 2021)

Reviewed by Leo Zaibert

policies, I should work directly toward that goal and leave the contentious issue of free will aside. While I acknowledge that this may be a better rhetorically [*sic*] strategy, since getting people to question the existence of free will is no easy task, I resist it for two main reasons. First, I am a philosopher and am committed to the truth wherever it leads me. While this might sound high-minded, I have worked on the problem of free will for many years and am convinced that free will scepticism is the *only* reasonable position to adopt. [...] Second, and more to the point, to achieve meaningful and lasting criminal justice reform, I believe it is important to challenge the assumptions of free will and basic desert moral responsibility (30, emphasis added).

Caruso's view that *only* his favoured position is "reasonable" (repeated verbatim in 35, 107, and in spirit elsewhere) is noteworthy, as he sometimes unceremoniously tosses influential and sophisticated views into the bin of the "unreasonable". Caruso's attitude here evinces a certain arrogance that is at odds with a plea to humility that he mobilizes elsewhere (and to which I will return).

A "MORALITY" THAT JETTISONS MORAL RESPONSIBILITY, AND THAT AFFIRMS THAT NO ONE IS DESERVING OF ANYTHING, STRIKES MANY AS IMPLAUSIBLE

A "morality" that jettisons moral responsibility, and that affirms that no one is deserving of anything, strikes many as implausible. That is why many have suggested that we *cannot* really treat each other as Caruso would have us do: merely as "object[s] of social policy", as "subject[s] for treatment", as needing merely to be "managed or handled or cured or trained", to quote P.F. Strawson. This impossibility can be explained in diverse ways: in the psychological terms of what

Pamela Hieronymi calls the Humean interpretation of Strawson: we simply "do not have the right equipment to do that"; or in the more conceptual terms of what she calls the Wittgensteinian interpretation of Strawson: we "can neither support nor call into question the whole of a practice using notions that are, themselves, constituted by that practice"; or otherwise.

Independently of our interactions with each other, some have argued that the mere *thought* that we cannot praise or blame people for their actions "cannot actually be believed". Ronald Dworkin, for example, suggests that "you cannot be convinced, even intellectually, that you are not responsible for your actions, because you cannot make any reflective decision [including what to believe] without judging which decision it would be better to make". This is to invoke – even if unwittingly – the sense of responsibility which Caruso denies, because if you realized that you chose the wrong action, you would then blame yourself.

Furthermore, these considerations partly explain the widespread (compatibilist) sentiment amongst both moral philosophers and punishment theorists whereby moral responsibility is compatible with either the ultimate metaphysical truth or falsity of determinism. In other words, the move is to focus on the phenomenology of moral experience and to bracket this thorny debate between free will and determinism. This type of move is, moreover, perfectly familiar in science as well. When a physicist, for example, speaks of objects "at rest", she is evidently bracketing all sorts of considerations, just as when a chemist talks about "pure" water. Nothing is completely at rest, or completely pure; but in the context of calculating the trajectory of a projectile, or of dissolving a water-soluble element in a lab, we can safely – nay: we had better – bracket these bits.

To deem Strawson a compatibilist and essentially leave it at that (Caruso, 118, 121, 127) evinces a superficial engagement with Strawson's hugely subtle "Freedom and Resentment" – an article in which, after all, Strawson concedes quite a bit to the anti-compatibilist. Furthermore, insofar as he does not think that denying moral responsibility is "as destructive as Strawson" and others fear (29, also 116, 297), Caruso is happy to identify himself as an "optimist". But this is precisely the sort of problematic optimism that worried

Strawson – an optimism that, according to Hieronymi, Strawson found “shocking and appalling” – and which he suggested needed to be “radically” modified by what he deemed pessimistic but soberer considerations.

Strawson’s famous suggestion is that the impasse between optimists and pessimists can be at least alleviated by paying close attention to our “reactive attitudes”, which include “backward-looking” resentment, blame, gratitude, admiration, and so on. Caruso rejects Strawson’s emphasis on reactive attitudes, suggesting that even if Strawson were wholly right about them, it would still be “another thing altogether to think that retributive legal punishment is justified” (122). Caruso’s point is correct, as far as it goes. And Strawson need not object to it: unlike Caruso, Strawson was not directly interested in criminal justice reform or policy issues, so he was indeed not trying to justify retributive legal punishment. Caruso’s point, however, does not go far enough. The problem for Caruso is that if Strawson is right about reactive attitudes, then Caruso’s theoretical edifice would be in trouble. Optimistically, Caruso claims that his opponents are “simply mistaken” when they think that

he is committed to “abandoning all discourse about justice, rights, and respect for persons” (29). Rejecting real praise and blame, however, is to *thereby* impoverish morality: the view whereby the very *meaning* of good and bad actions is importantly connected to their being praiseworthy or blameworthy is not uncommon. Evidently, Caruso believes that this impoverishment is not problematic, but, well, this is part of our general disagreement.

Our disagreement is further foregrounded when we turn to retributivism itself. Since time immemorial the debate concerning the justification of punishment has opposed retributivism and consequentialism (or utilitarianism). The sloganized version of the debate is that retributivists justify punishment by attending to what wrongdoers *deserve* and that, in contrast, consequentialists (or utilitarians) attend to the (potentially) good consequences of inflicting punishment. The typical good consequences that punishment is supposed to cause tend to relate to the prevention of future wrongdoing; punishment may accomplish this goal given its potential to deter, rehabilitate, incapacitate, and so on. As another famous slogan has it, retributivism is backward-looking, and consequentialism is forward-looking.

Consider this formulation of the debate, from over two thousand years ago:

No one punishes wrongdoers putting his mind on what they did and for the sake of this – that they did wrong, not unless he is taking mindless vengeance, like a savage brute. One who undertakes to punish rationally does not do so for the sake of the wrongdoing, which is now in the past – but for the sake of the future, that the wrongdoing shall not be repeated, either by him or by others who see him punished (Plato, *Protagoras*, 324a-b).

This passage not only attests to the age of the debate itself, but it attests, too, to the age of the well-worn charge that retributivism is somehow “irrational”, “barbaric”, and so on. This demonization of retributivism constitutes a conspicuous tradition in the history of ideas – a tradition that continues to be extremely popular to this day, as Caruso’s book amply



Gregg D. Caruso

confirms. Independently of its merits, even a cursory look at history reveals that Caruso's anti-retributivist campaign is but a drop in a vast ocean.

RETRIBUTIVISM DOES NOT SEEK TO PROVIDE AN ANSWER TO THE QUESTION OF WHAT SPECIFIC PUNISHMENT SPECIFIC CRIMES DESERVE

Caruso claims that retributivism is “widely embraced” and a “core principle” (5), or indeed the “official dominant principle” (8) of American criminal law. He sees retributivism “enthroned” as the “primary justification for the death penalty” in the United States, and for criminal punishment more generally in a number of other countries (6 ff., and *passim*), and as the cause of “excessively punitive punishments” (18). These claims deserve some attention. Caruso himself problematizes his first sort of claim, for despite his talk of “wide embrace”, “enthroning”, “dominance”, and so on, he admits that “the American criminal justice system has long been committed to limiting retributivism” (322). My own view is that the American criminal justice system is *insufficiently* retributive, and that many prominent (and deeply problematic) aspects of it are essentially unintelligible unless seen as effusions of a broadly consequentialist worldview. The very institution of plea-bargaining, which is the way in which the *vast* majority of criminal cases in the United States are resolved, patently contradicts retributivism. These bargainers do not even pretend to care about desert. According to the National Association of Criminal Defense Lawyers, more than 97% of all criminal cases in the United States are decided through this utterly and unabashedly non-retributivist process. Other non-retributive policies and trends include the proliferation of felony murder provisions (where non-murderous killings that occur in the context of committing a felony are punished as severely as if they had been murders), strict liability offenses (where offenses committed with differing mental states are punished identically), mandatory minimums in sentencing (where the importance of the details of

the case at hand is limited), and overcriminalization (where non-blameworthy conduct is turned into criminal conduct and eventually punished).

As for Caruso's second sort of claim, it is important to underscore that retributivism does not seek to provide an answer to the question of what specific punishment specific crimes deserve. In other words, despite the fact that this is frequent confusion, retributivism is not *lex talionis* (an eye for an eye, a tooth for a tooth). A retributivist could deny that murderers deserve the death penalty, or that rapists deserve to be raped, etc. But even if a retributivist believed that murderers, say, deserve the death penalty, she is not thereby committed to supporting the death penalty. For, obviously, the value of giving people what they deserve can conflict with – and lose out to – other values. Caruso knows this, and he concedes that retributivists can oppose the death penalty, and that there exist “reasonable retributivists” who oppose overly harsh criminal justice systems, and so on. But Caruso nonetheless believes that, in the end, all retributivists are prone “to try to make life in prison as unpleasant as possible” (19). What Caruso describes here is not a retributivist – it is a sadist. The retributivist wants punishment to be as unpleasant as wrongdoers *deserve* – not more (and, in principle, not less) than that. One point that often gets lost in anti-retributivist positions is that part of what retributivism seeks to do – both analytically and historically – is to *preclude* excessive punishments.

CARUSO APPEARS TO BELIEVE THAT BECAUSE RETRIBUTIVISTS CARE ABOUT DESERT, THEY CANNOT (OR CANNOT EASILY) CARE ABOUT OTHER THINGS

Caruso appears to believe that because retributivists care about desert, they cannot (or cannot easily) care about other things. This is like believing that because philatelists care about stamps, they cannot (or cannot

easily) care about anything else. Thus, Caruso asks us to imagine someone who has received a drug that makes it “virtually certain that he will brutally murder a person” within a week. Caruso thinks that he would have no problem incapacitating this person (in the sense of rendering him incapable of committing crime again), but that, contrariwise, “retributivists would have a hard time justifying preventive measures” (293). Unfortunately for Caruso, there is no reason whatsoever why a retributivist – qua retributivist – should have any trouble incapacitating this person. Similarly, Caruso claims that “retributivists care little about rehabilitation” (19), that they are committed to understanding respect for human dignity “narrowly”, as merely a matter of “just deserts” which is (somehow) likely to generate treatment that “dehumanizes and disenfranchises individuals” (125), and so on. Caruso’s retributivists are truly absurd creatures, only capable of being motivated by, and of only valuing, desert. For whatever it is worth, I do not know of a single retributivist who asserts that human dignity is *wholly* a matter of giving people what they deserve. We are, for example, *entitled* to have our human rights (and more) respected independently of whether or not we *deserve* that. A retributivist is in no way prevented from caring about rehabilitation or from vehemently opposing dehumanization and disenfranchisement.

If asked, the retributivist may (correctly) point out that the incapacitation in Caruso’s thought-experiment is not a matter of *punishment*: for the person has, *ex hypothesi*, not done anything wrong. This reveals another uneasy tension in Caruso’s book: at times he claims to reject retributivism, and at times he claims to reject *punishment* altogether. Although Caruso claims to limit his discussion to “legal punishment” (12), his absolutely radical rejection of moral responsibility suggests he is against punishment *tout court*. Now, abolitionism regarding *state* punishment deserves at least some attention, but the *global* abolitionism that appears to flow from Caruso’s *global* rejection of responsibility is terribly untenable.

Caruso wants to replace criminal justice systems with a public health-quarantine system that, more than non-retributive, is explicitly non-punitive: “my version of the public health-quarantine model remains pure and free of any punitive components and is therefore compatible with the complete rejection of all justifications of

punishment” (313). It is not easy to see how Caruso can succeed in this. While Caruso’s approach would rule out “retributivism and sever [sic] punishment, such as the death penalty”, he admits that his approach would still justify “incapacitation and rehabilitation programs” (29). To the extent that those programs can be imposed against the “wrongdoers” will, and in ways they find unpleasant, harsh, and so on, they can be difficult to distinguish from punishments – thus risking turning this discussion into a mere word game.

CARUSO WANTS TO REPLACE CRIMINAL JUSTICE SYSTEMS WITH A PUBLIC HEALTH-QUARANTINE SYSTEM THAT, MORE THAN NON-RETRIBUTIVE, IS EXPLICITLY NON-PUNITIVE

Retributivists disagree amongst themselves about whether the fact that someone deserves punishment provides us with merely a necessary condition for punishment to be justified (thus generating a prohibition against punishing the innocent), or rather with a full-blown sufficient condition that renders punishment justified (thus generating a duty to punish the guilty). Elsewhere I have referred to the former sort of retributivism as “minimalist” and to the latter as “unbridled” – and I have objected to both. A mere prohibition against punishing the innocent seems sensible enough, but altogether too feeble. Moreover, this interpretation scandalously fails to render the traditional accusation that retributivism is barbaric, or too severe, even intelligible. What can be *barbaric* about protecting innocents? Would Caruso’s passionate rejection of retributivism encompass this protection too? A full-blown *duty* to punish the deserving, in contrast, strikes many (including many retributivists) as too rigid or stern.

As it turns out, Caruso attempts to turn the traditional bipartite division into a tripartite one, distinguishing

between (1) weak, (2) moderate, and (3) strong retributivism. Caruso's weak retributivism is what I call minimalist retributivism: desert is "merely necessary but not sufficient for [justified] punishment". But in both what Caruso calls "moderate" and what he calls "strong" retributivism, desert is "necessary and sufficient for [justified] punishment" (9-10). It would appear that the difference between Caruso's (2) and (3) is that in (2) the "sufficiency" somehow does not amount to a duty, though it does in (3). But then, of course: How is deserved punishment a "sufficient condition"? A sufficient condition for what? In any event, not only does Caruso's tripartite taxonomy fare no better than the traditional bipartite taxonomy (as I will show immediately), but soon after he introduces it he explicitly abandons it anyhow (12).

In my view, you are a retributivist if you think that the fact that someone deserves punishment is, in itself, *important*, that this fact *matters*. The more important you take this fact to be, the more retributivist you are. Evidently, the notion of "importance" does not translate neatly into the talk of necessary and sufficient conditions. The fact that you consider something important, or its realization valuable, does not necessarily generate a duty to bring it about, nor does it reduce to the limp view that its absence precludes action. Hence the failure of the traditional bipartite approach and of Caruso's tripartite approach (if that is what is): they cannot deal with the (important) notion of importance.

IN MY VIEW, YOU ARE A RETRIBUTIVIST IF YOU THINK THAT THE FACT THAT SOMEONE DESERVES PUNISHMENT IS, IN ITSELF, IMPORTANT, THAT THIS FACT MATTERS

My own version of retributivism is unabashedly *pluralistic*: I expressly recognize the importance of

things other than desert. In contrast, Caruso's position is necessarily *monistic*: the only thing that ultimately matters to him is the diminution of suffering in the world (as we shall see below). Of course, under that general heading Caruso may pursue various goals, but always in the service of the overarching suffering-diminution goal. Moreover, my pluralism is centred around the tension between punishment and forgiveness, in a way that Caruso cannot even countenance (as we shall also see below). Caruso's position is, in fact, as monistic as the utilitarianism that has historically opposed retributivism, and from which he unsuccessfully seeks to distance himself.

One of Caruso's main arguments against his amalgamation of punishment and retribution happens to be one that I have repeatedly criticized. The argument is somewhat attractive, at least on first approximation, but it ultimately fails. The superficial attractiveness flows from two specious moves. The first move is to suggest that punishment involves *harming* wrongdoers (Caruso, 110, and *passim*). This non-neutral term – "harm" – is tendentious. Undeniably, the treatment that the wrongdoer is to endure is meant to be unpleasant, unwanted, and so on: that is a purely conceptual point about punishment. There could be, I guess, a sense in which anything unpleasant or unwanted is "harmful". But there is another obvious sense in which it need not be: the side-effects of some medicines may be unpleasant and even unwanted, but not thereby harmful. The most influential contemporary definition of punishment, the Flew-Benn-Hart definition, avoids this term "harm" altogether, and instead claims, neutrally, that punishment involves "pain or other consequences normally considered unpleasant". While I have argued against several aspects of this definition, I have, like its champions, avoided the potential ambiguity of "harm": neither the widespread definition nor mine says anything about *harming* wrongdoers.

The second move is a plea for epistemic humility – and here the attractiveness is greater: for who, other than the arrogant, can be against epistemic humility? Caruso believes that retributivism fails to meet a particularly demanding "burden of proof" – and the burden is so demanding because when what is at issue is *harming* others, we had better impose a "high epistemic standard" (110). One could, of course,

immediately retort that punishment is not supposed to *harm* the wrongdoer, even though it is meant to be “painful” or “unpleasant”, as I have just sketched. But even if we put aside the tendentiousness of “harm”, something else is problematic with this argument.

Caruso approvingly cites authors who think that the very talk of “burden of proof” comes to us “from the adversarial courtroom” (111), where different burdens depend on whether cases are civil or criminal. In non-criminal cases the standard of proof is “preponderance of evidence”; in criminal cases, however, preponderance of evidence is deemed not enough: the standard of proof is ratcheted up to “beyond a reasonable doubt”. The reason for this increased burden is that criminal convictions are, in principle, much more serious than non-criminal ones. (The “in principle” hedge here is required because governments often abuse the allegedly non-criminal nature of some matters, precisely in order to get away with doing awful things to people.)

Caruso’s argument is that insofar as there exists reasonable doubt about the existence of the type of free will that may justify punishing wrongdoers, then we should refrain from punishing until we can be certain – beyond reasonable doubt – that said free will exists. I sidestep the fact that from Caruso’s perspective this is never going to happen: as we have seen, “free will scepticism [in the sense of *denial*] is the only reasonable position to adopt” (30). The main weakness of this argument is that reasonable doubts are possible about all sorts of things, things that we nonetheless need to do, and are justified in doing. As I have pointed out elsewhere, there are “important differences between adjudicating guilt in a court of law and ascertaining the merits of a theory (philosophical or otherwise)”. I have sought to illustrate this disanalogy (and hence the weakness of this often-repeated argument) by pointing out that “there is plenty of discussion about the best procedure for treating inguinal hernias, or about how best to combat inflation. Yet no one doubts that inguinal hernias ought sometimes to be treated, or that they are in fact sometimes treated correctly and sometimes incorrectly – and the same holds for economic measures combating inflation”. Retributivism – like any other theory – cannot be simply rejected by pointing out that there are doubts – even serious doubts – about its truth. Expecting

otherwise is particularly odd since the truth about the alternatives to retributivism is, despite Caruso’s extraordinary confidence, *also* very much in doubt.

Perhaps Caruso’s point could be salvaged by assuming that his doubts vis-à-vis retributivism could be redirected: not to reject a *theory*, but to reject the very serious suffering – albeit not a *harm* – that (allegedly in the name of a theory) many criminal justice systems inflict on human beings. Perhaps. But then Caruso’s book would need to be rewritten – beginning with its title. *Rejecting Excessive Punishment* would be a more accurate title, if much less effective – for who could then be Caruso’s opponent? As I stated at the outset, opposing excessive punishment is something about which Caruso and I – together with retributivists and anti-retributivists, libertarians and determinists, compatibilists and anti-compatibilists, and so on – can all agree. Where Caruso goes wrong is in thinking that (in practice) retributivism is somehow the cause of excessive punishment, although (in theory) it forbids it.

UNDENIABLY, THE TREATMENT THAT THE WRONGDOER IS TO ENDURE IS MEANT TO BE UNPLEASANT, UNWANTED, AND SO ON: THAT IS A PURELY CONCEPTUAL POINT ABOUT PUNISHMENT

Some inflictions of suffering are indecent or otherwise immoral (even if deserved!); some criminal justice systems are heartless and dysfunctional, and some downright perverse and cruel. In fact, even when fully deserved, there is something to be lamented, something very often *tragic* in the infliction of punishment; as I have argued in my 2018 book *Rethinking Punishment*, it is difficult to punish without getting our hands dirty. Being a retributivist does not preclude you from endorsing any of these claims, or

from accepting differentiated burdens of proof based on the severity of the punishments. Moreover, being an *anti*-retributivist does not prevent you from endorsing extraordinarily harsh treatments too: utilitarianism and “humanitarian” approaches can be as harsh as any retributive treatment. The scare quotes around “humanitarian” are meant to cast doubt on an approach that I think bears important similarities to Caruso’s favoured public health-quarantine model.

Caruso wishes his model to avoid the problems he sees in retributivism (or punishment), while not thereby collapsing into utilitarianism or consequentialism. This is not easy. Caruso, like many others, distinguishes between utilitarianism and consequentialism. I have no truck with that, and in some contexts this is a perfectly acceptable thing to do – say, when distinguishing a comprehensive moral doctrine from a particular justification of punishment. But there are also contexts in which the distinction between retribution and its alternatives is best seen as dichotomous: either inflicting deserved suffering is *intrinsically* valuable or it is not. This is as dichotomous as the pair formed by any proposition whatsoever and its negation. If you believe that deserved suffering is in this sense intrinsically valuable (in whatever degree), then you are a retributivist; if you do not believe that, then you are not.

There are plenty of reasons – both analytical and historical – for referring to the person who denies that deserved suffering is intrinsically good as a utilitarian. And so I will continue to do so, although I emphasize once again that what matters is substance, not labels. Whatever you call things, Caruso cannot avoid being on one side of this dichotomy or the other – and it seems obvious to me what side that is. Caruso champions the view that punishment should “be focused on prevention, rehabilitation, and reintegration” – that it should be entirely “preventive” – and these have always been *paradigmatic* utilitarian rationales (9; similar remarks appear on 184 and *passim*). Moreover, Caruso’s public health-quarantine model necessarily shares with utilitarianism the bald denial of any intrinsic goodness in the infliction of deserved suffering.

Caruso’s basic idea is that just as we should quarantine those with communicable diseases, we should quarantine wrongdoers. The analogy is obvious: just

as those with communicable diseases may harm us, wrongdoers may harm us. In both cases we need protection, and in both cases lockups work. Remarkably, differences between the ways in which bad things may come about are unimportant from Caruso’s perspective. We should deal with potential calamities in a wholly undifferentiated way – Caruso sees no important normative difference between our hospitals, our prisons, and indeed our lightning rods, our sump-pumps, and so on.

RETRIBUTIVISM – LIKE ANY OTHER THEORY – CANNOT BE SIMPLY REJECTED BY POINTING OUT THAT THERE ARE DOUBTS – EVEN SERIOUS DOUBTS – ABOUT ITS TRUTH

It is very hard to see how Caruso’s radical denial of moral responsibility can coexist with a system that purports to *morally* respond to *moral* wrongdoing. Despite being sympathetic to *some* of the *effects* that Caruso’s favoured system *may* end up having, I cannot avoid thinking that Caruso has left himself with no theoretical room for real wrongdoing, and that therefore his model is not really capable of morally responding to it. Caruso’s model either contains only what Strawson (and others) see as purely objective attitudes (that is, attitudes that reduce human beings to “objects of social policy”, to objects that need to be “managed or handled or cured or trained”), or he cherry-picks which reactive attitudes to admit (28). Either way, the specific reactive attitudes of praise and blame require a type of moral responsibility that Caruso officially rejects.

Problematically, however, Caruso sometimes cannot avoid appealing to a not-so-diminished morality. For example, he thinks that “to hold people *truly* deserving of blame and praise, punishment, and reward” is wrong, because “it is fundamentally unfair and unjust” (30). Given Caruso’s main line, it is hard to see how he can maintain this: he cannot criticize practices employing tools afforded by those very practices, to

again echo Wittgenstein. Caruso's official line should be that (since no one is "truly" responsible) nothing is ever "truly" unfair or unjust – a line that, however, he does not want to hold, but cannot consistently avoid. (I gloss over the fact that, wholly independently of the specific infliction of deserved *suffering* and of the horrors of criminal justice systems, Caruso claims that *truly praising* is "fundamentally unfair and unjust": so, for him, praising Amanda Gorman for her "The Hill We Climb" is "fundamentally unfair and unjust".)

Caruso's public health-quarantine model (and its concomitant shift between an avowal and a disavowal of morality) is reminiscent of another famous appeal to "epistemological modesty" within punishment theory (Karl Menninger's), and to Michael S. Moore's reaction to how "inconsistently it is applied". Moore quotes this remarkable passage from Menninger:

The very word justice irritates scientists. No surgeon expects to be asked if an operation for cancer is just or not. No doctor will be reproached on the grounds that the dose of penicillin he has prescribed is less or more than justice would stipulate. It does not advance a solution to use the word justice. It is a subjective emotional word The concept is so vague, so distorted in its applications, so hypocritical, and usually so irrelevant that it offers no help in the solution of the crime problem which it exists to combat but results in its exact opposite – injustice, injustice to everybody.

Sardonically, Moore then points out: "apparently the late Dr Karl knew injustice when he saw it, even if justice is a useless concept". Even if Caruso could explain how those who punish wrongdoers can indeed be *truly* unfair and unjust (although they are not morally responsible), he cannot (on pain of contradiction) explain how they could deserve any blame or punishment for their alleged unfairness and injustice.

But there is more that is problematic in Caruso's assimilation of wrongdoing to communicable disease. Quite clearly, some wrongdoers may remain threats for the entirety of their life – thus Caruso would recommend "treating" them until they die, even though, *ex hypothesi*, they will never be "cured". Also

quite clearly, not all wrongdoers continue to be threats after doing wrong. Nothing prevents a wrongdoer from having an epiphany – or from having an accident that leaves her paralyzed and unable to cause any further harm – immediately after doing wrong. We do not quarantine people who are no longer capable of transmitting diseases, so why should we quarantine wrongdoers who no longer pose threats?

WE DO NOT QUARANTINE PEOPLE WHO ARE NO LONGER CAPABLE OF TRANSMITTING DISEASES, SO WHY SHOULD WE QUARANTINE WRONGDOERS WHO NO LONGER POSE THREATS?

Caruso may bite the bullet here and admit that we should quarantine whoever is a threat (whether they have done wrong or not) and that we should not quarantine "wrongdoers" (regardless of what they did) who are no longer threats. This answer, however, would underscore the extent to which Caruso's system is not morally responding to moral wrongdoing at all. Caruso's richer answer, however, is to bolster his public health-quarantine system with considerations that he draws from health care ethics contexts. Thus, Caruso takes into account what he calls the four "key dimensions" of health care ethics: "autonomy, beneficence, nonmaleficence, and justice" (187). It is not clear, again, how a commitment to these "dimensions" of ethics is supposed to be consistent with Caruso's radical denial of moral responsibility.

Take autonomy, to mention but the first of the terms that Caruso lists: does not the very talk of autonomy – i.e., self-rule, self-determination – presuppose precisely the sort of free will that Caruso denies? Truly a leitmotif in Caruso's book, the Wittgensteinian inconsistency mentioned at the outset, whereby Caruso criticizes a practice by presupposing notions constituted by that very practice, reappears yet

again. Caruso recognizes the obvious connection between free will and autonomy when, early in the book, he criticizes the law for assuming that we are autonomous (4 ff.). But Caruso also takes it away: “here [where?] self-determination has nothing to do with metaphysical free will or the control in action required for basic desert moral responsibility” (217). This is an extraordinary assertion: *nothing* to do with it. To say, as Caruso says immediately after the statement just quoted, that “rather, it [self-determination] is the linchpin of liberal political theory” (217) is not helpful. Regardless of the role self-determination may play within liberalism or political theory, how exactly are we supposed to determine ourselves when we are not autonomous? Moreover, if, waxing compatibilist, Caruso can so easily (and radically) separate self-determination in political theory from the thorny metaphysical debate, why can he not do that too regarding self-determination in ethical theory?

Some of the problems of Caruso’s analogizing wrongdoing and disease have been perceptively pointed out by Michael Corrado. Although Caruso engages at length with this objection, I am unconvinced by his response. Part of Corrado’s objection involves (non-intermittent) epistemic humility. Since, unlike Caruso, most of us are not certain that moral responsibility is impossible, we find it reasonable to preserve the ostensible disanalogy between wrongdoing and disease. Corrado’s move is a nod to the sort of compatibilism that most punishment theorists embrace, and whom Caruso chides for embracing (Caruso, 113 ff., *passim*).

The point here is not that we should preserve this or that particular criminal justice system – again, we can all agree that many contemporary criminal justice systems are scandalously cruel. The point on which Corrado (and others) insists is, rather, that seeing wrongdoing as a *real* moral notion, which calls for a *real* moral response – *punishment (or forgiveness, etc.)*, *not treatment* – not only allows us to *coherently* invoke the notion of justice, but it affords wrongdoers concrete *protections*. For example, modern criminal law offers many protections to defendants: the principle of legality, due process, strict rules concerning evidence admissibility, rules against self-incrimination, the prohibition against double jeopardy, the *in dubio pro reo* (i.e., when in doubt, side with the defendant)

principle (which, incidentally, does consistently and humanely mobilize epistemic humility), and myriad other prohibitions against excessive (or “cruel and unusual”) punishment. These protections are out of place within the insufficiently *moral* context of Caruso’s public health-quarantine model. Imagine, for example, the absurdity of an *in dubio pro patiens* (i.e., when in doubt, side with the patient) principle.

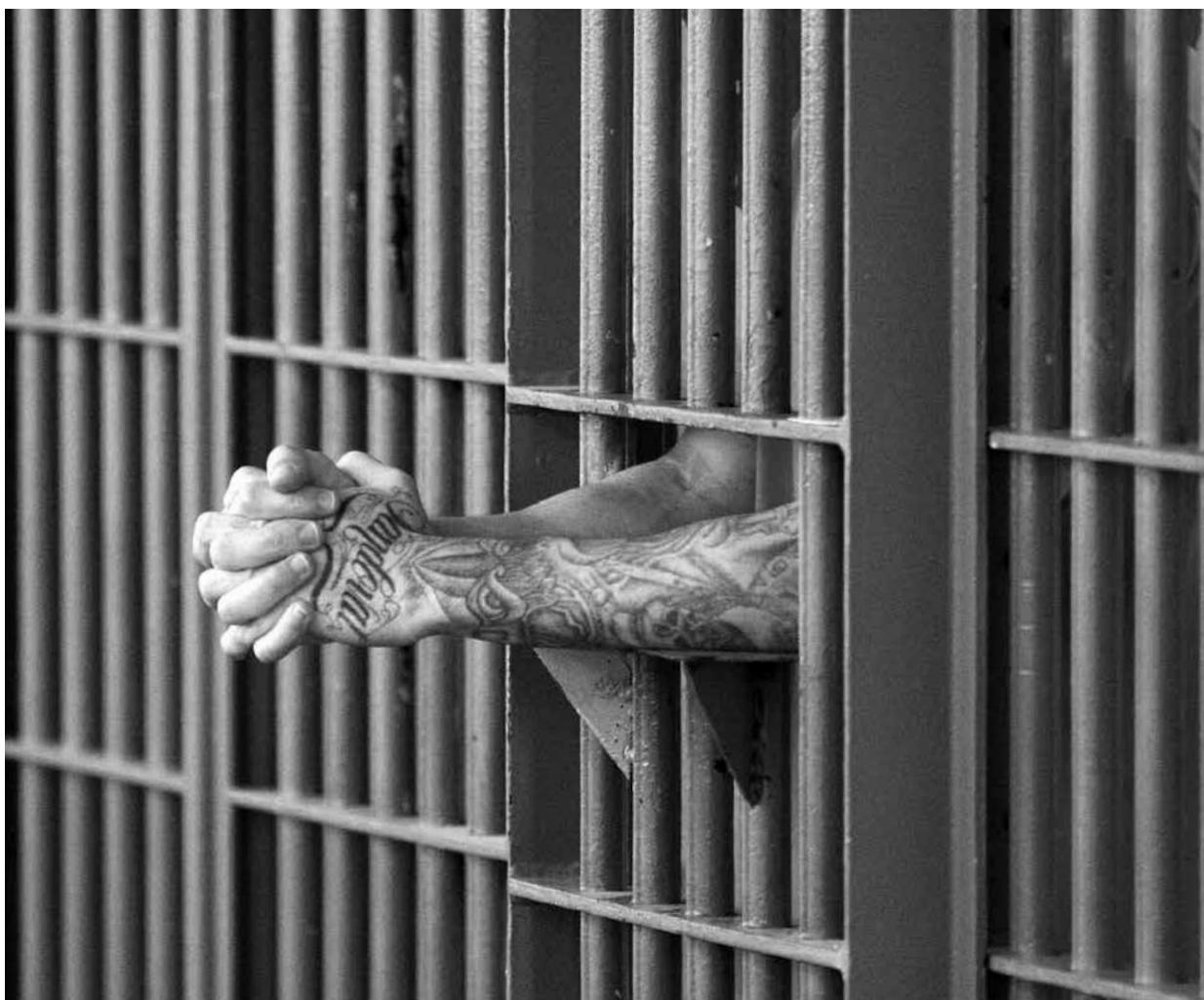
But Caruso’s desire for his public health-quarantine model to have access to these moral safeguards is persistent and multifarious. So, he claims that his public health-quarantine model has “a proportionality principle of its own” (125). Whence? The question is all the more pressing given the (at least) diminished morality Caruso’s denial of moral responsibility generates. Moreover, seeking to avoid the potential excesses of utilitarian approaches, Caruso wishes to ground his public health-quarantine model in society’s alleged right to self-defence. Again: whence? Where does this right come from if not, in the final analysis, from a not impoverished morality which includes moral responsibility? Just as much as the other moral safeguards mentioned, self-defence and other justifications presuppose precisely the sort of robust morality that Caruso rejects.

Caruso’s response to Corrado’s objection is reminiscent of another example of anti-retributivist inconsistencies. In her illuminating *Are Prisons Obsolete?*, Angela Davis discusses Lucia Zedner’s also illuminating work on the punishment of women. Davis understandably regrets that sometimes women have been required “to do more time than men for similar offenses”. Zedner’s explanation for this asymmetry is that “women were sent to reformatories not to be punished in proportion to the seriousness of their offense [i.e., according to desert] but to be reformed and retrained”. Furthermore, Zedner underscores that in “important respects the [exclusively female] reformatory represented a greater infringement in women’s freedom [than men’s punishments infringed in their freedom]”. While Zedner here refers to early American reformatories, she offers similar remarks about the United Kingdom: “the most important prison for women in Britain – Holloway – was redesigned in the late 1960s on the assumption that ‘most women and girls in custody require some form of medical, psychiatric or remedial treatment’”.

The most natural – and arguably the most pertinent – way to object to women being imprisoned longer than men for the same offenses is, of course, to say that they do not *deserve* this differential treatment. But this is an objection that anti-retributivists cannot make. Admittedly, there is no gender bias in Caruso’s approach: Caruso wants to “reform and retrain” *all* so-called wrongdoers, without discrimination of any kind. But gender bias is not the only problem in what Zedner describes: the insufficiently moral focus on “reform and retrain” to the detriment of moral desert is problematic in itself, independently of biases.

Not only, then, does Caruso’s denial of moral responsibility undermine those general protections that the very act of *conceptualizing* something as

punishment, in a moral sense, affords defendants, but it also undermines a substantive notion of justice, or at least of *retributive* justice, itself. It is important to avert an ambiguity here: “retributive justice” does not necessarily refer to retributivism: utilitarian justifications of punishment belong to “retributive justice” just as much as retributive justifications do. Retributive justice deals with fitting responses to others’ actions, and it is typically contrasted to distributive justice which deals with structural patterns of resource-allocations. (Incidentally, championing the importance of desert within retributive justice says little about one’s views as to the role of desert in distributive justice: being a retributivist does not entail preferring, say, meritocracy over democracy.) Discussing which student paper, say, deserves a prize,





Michelle Alexander

or which footballer deserves the Golden Boot award has little to do with punishment and yet it is a matter of retributive justice. Again, either Caruso cannot countenance these discussions at all, for no one deserves anything anyhow, or he would have to deem that awarding Harry Kane (or Mo Salah, *or anyone else*) the Golden Boot is “fundamentally unfair and unjust”.

I now return to the important agreement between Caruso and me: that the criminal justice system in the United States is deeply objectionable. The general power asymmetries between the state and the individual that have understandably worried punishment theorists are, in the context of the United States, exacerbated by the depressingly intimate connection between American criminal justice and racism – systemic and otherwise (see, for example, Michelle Alexander’s book *The New Jim Crow* or Douglas A. Blackmon’s *Slavery By Another Name*). I can press an argument against the United States’ hellish and racist criminal justice system that is inaccessible to Caruso. Part of the horror of the American criminal justice system is the result of

the inordinate amount of *undeserved* suffering that it inflicts. Caruso cannot even blame racists (or, again, anyone else): the most his denial of moral responsibility allows him to do is to quarantine racists, and only if they are (and only as long as they remain) threats.

I am of course not alone in being able to advance this sort of argument. Douglas Husak, to mention but another contemporary retributivist, has been at the forefront of a campaign against overcriminalization and over-punishment. Husak’s core arguments revolve around precisely this point: much of what the state punishes is unjust because it is undeserved. Granted, there is so much that is wrong with the American criminal justice system that there are enough *other* horrors that Caruso (or anyone else) may invoke in order to powerfully condemn it. Still, Caruso has at least one fewer argument in his arsenal – and the missing argument is precisely the argument that I (and many others, now and throughout history) consider to be absolutely crucial.

Contemplating the horrors of the United States’ criminal justice system may prime us to be tempted

by positions such as Caruso's. As a salutary contrast, however, consider a different context. The United Nations' *Report of the Independent Expert to Update the Set of Principles to Combat Impunity*, particularly its addendum, *Updated Set of Principles for the Protection and Promotion of Human Rights through Action to Combat Impunity*, is replete with efforts to ensure that perpetrators of human rights violations get the *punishment* that they *deserve*. In its very preamble, the document unequivocally asserts that "there can be no just and lasting reconciliation unless the need for justice is effectively satisfied" and that this is part of the reason why "the duty of every State under international law to respect and to secure respect for human rights requires that effective measures should be taken to combat impunity". Impunity is no simple negation: it is a privation. It is the absence of punishment when punishment is *deserved*: someone who deserves no punishment and is not being punished is not thereby enjoying impunity. Given its narrow – and insufficiently *moral* – focus on threats, Caruso's public health-quarantine program is utterly inert in this international context, incapable as it unabashedly is of even seeking justice. Caruso cannot combat impunity because in his moral universe no one deserves punishment, and so there simply is no theoretical room for impunity there.

CARUSO'S REJECTION OF DESERT AND MORAL RESPONSIBILITY THUS GENERATES A PROBLEMATICALLY BARREN MORAL UNIVERSE

One last casualty of Caruso's denial of moral responsibility deserves mention: forgiveness. For whatever it is worth, I have suggested that progress in the justification of punishment requires paying much closer attention to the justification of forgiveness. Admittedly, my arguments for this view are theoretical, but they have the added practical bonus of perhaps contributing to tempering excessively harsh responses

to wrongdoing. While I of course think that my view is on the right track, I understand that the view is not (yet) widely accepted. There is, however, one conceptual aspect about forgiveness which is rarely disputed: forgiving *means* forswearing resentment (or, in my view: forswearing *punishment*) regarding someone who *deserves* it. If a potential "forgivee" does not *deserve* resentment or punishment, then what appears to be forgiveness would, purely as a conceptual point, be something else: an excuse, a justification, and so on.

As I highlight in *Rethinking Punishment*, the fact that forgiveness is the deliberate refusal to inflict a punishment (or the deliberate forswearing of a resentment) *that is deserved* is the reason why forgiveness is at once so paradoxical, and sometimes so alluringly admirable. Given Caruso's laudable desire (and, insofar as I believe in real praise and blame, I can truly laud him) to contribute to a more humane criminal justice system, he should welcome a more prominent role for forgiveness. But he cannot. Just like he cannot countenance impunity, Caruso cannot countenance forgiveness: for these require precisely the "real" judgments of desert and the real moral responsibility that Caruso rejects. Within Caruso's public health-quarantine model, forgiveness degenerates into simple medical malpractice: to "forgive" therein would amount to refusing to quarantine someone who is a threat. Not that Caruso can blame (or praise) this malpractitioner anyhow. Caruso's rejection of desert and moral responsibility thus generates a problematically barren moral universe: a moral universe without real praise and real blame (and arguably without real gratitude, real admiration, and so on); without real (retributive) justice, without real moral wrongs and real impunity, and, we finally see, without even real forgiveness.

Leo Zaibert is William D. Williams Professor of Philosophy, Law, and Humanities at Union College, New York. He has published extensively on punishment, forgiveness, and related themes. His latest book, Rethinking Punishment, was published by Cambridge University Press in 2018.