



Essay Review

Against Retribution

by Thomas W. Clark

A review of *Placing Blame: A General Theory of Criminal Law*, by Michael Moore. Oxford: Clarendon Press, 1998. Hardcover, 847 pages.

Abstract: Intuitively, naturalism undercuts retributive attitudes by showing that the causal story behind crime involves numerous factors outside the individual. Although Michael Moore is entirely naturalistic in his understanding of human behavior, he thinks any mitigation of retributive judgments is unwarranted: retribution is an intrinsic good, and we should discard our sympathies for disadvantaged offenders as "moral hallucinations." I make the case that such sympathies are not misplaced, but reflect the fact that our dispositions to punish and withhold punishment track causality itself. When we understand the external factors that shaped the offender, retributive rage diminishes in favor of a determination to address these factors. It is only by ignoring the functional, forward-looking nature of morality that Moore can portray retribution as an intrinsic good and the reigning moral principle of criminal justice. Since the natural purpose of morality, including the retributive impulse, is to shape behavior advantageously, we can and should consider other more efficient and less punitive means to achieve the ends that retribution originally served. This suggests the aims of criminal justice might change under pressure from a thorough-going naturalism.

Inclusive naturalism – the view that human beings and human behavior are completely included in the natural world – implies that we don't have libertarian free will, the capacity to cause without being fully caused in turn. Under naturalism there are no causally privileged agents that could have done otherwise in situations exactly as they arose. Those sympathetic to naturalism often suppose that this view of human agents as only proximate, not ultimate, originators of their behavior will attenuate the

tendency to place blame (or credit) solely on the individual. After all, factors unchosen by the person play an essential role in shaping action, and when those factors are appreciated, this can dampen our retributive impulses. And in turn, as the desire for retribution diminishes, we are better able to look *outside* the person to the wider causal context, with an eye to the more effective prevention of future offenses.

However, Michael Moore's theory of criminal law as set out in his book *Placing Blame* is

an example of how naturalism need not, it seems, imply any change in our retributive attitudes towards wrongdoers. Although Moore accepts naturalism, and sees moral responsibility as being compatible with determinism, his compatibilism is what I will call the "hard" variety: he resolutely resists any softening of retributive judgment in the light of determinism.

Those such as myself, who find that considering the causal antecedents of behavior does and should undercut the desire for retribution, are more likely to extend mitigation to offenders, or at least think that mitigation *might* be justified. Our understanding of the factors which produce crime might lead to the transformation of retributive rage into a determination to attack its causes, among which are punitive conditions found in many prisons, the very expression of retribution itself. So we might mitigate or entirely withhold harsh punishment in favor of more productive, non-punitive responses to an offense. On Moore's view, this "mitigation response" is simply a non-sequitur, since carrying out retributive justice – the infliction of suffering on culpable wrongdoers – is an intrinsic good which constitutes the central aim of criminal law; no considerations of causality can or should deflect us from this aim. Even if the suffering of the wrongdoer involves no utility in terms of deterrence, incapacitation, or reform, it is still necessary to impose it to satisfy the demands of justice. No softening of our retributive attitudes towards offenders, nor any corresponding de-emphasis of retribution in favor of other ameliorative responses, is justified by appreciating the manifold causes leading up to criminal acts, according to Moore.

So who's right here? In what follows I will explore the difference between Moore and myself using *Placing Blame* as the essential text, with references in endnotes to other sources, especially Stephen J. Morse's work (see references). This essay, long as it is, can't possibly do justice (so to speak) to his book, which at 847 pages is encyclopedic in scope and complex in its arguments. But I will try to present

Moore's case against mitigation fairly, as far as I understand it, since it lies near the heart of his general thesis about criminal law. Moore is by no means a raving retributivist out for blood, rather he describes himself as a policy liberal in many respects. Yet his claims that retribution is intrinsically good and necessarily central to our conception of justice are eminently contestable, so contest them I shall. As a fellow naturalist, I naturally want Moore on the side of those who would see less suffering in the world, not its indefinite continuation as justified by and embodied in retributive justice. The pursuit of retribution is itself arguably a primary contributor to further violence, if recidivism rates following harsh prison sentences are any indication. As B.F. Skinner pointed out long ago in *Beyond Freedom and Dignity*, attempts to control via punishment generally incite retaliation. For a recent recommendation to follow the anti-retributivist course I suggest in this paper, see Janet Radcliffe Richard's fine book, *Human Nature After Darwin*, the chapters entitled "Blameless Puppets" and "The End of Ethics."

Moore's naturalism

That Moore is an unequivocal and inclusive naturalist may come as surprise to those who imagine that the law necessarily involves a commitment to libertarian free will, the idea that we are in some deep, metaphysical respect free from determining causes. He spends a full chapter undermining what he calls the "causal theory of excuse": the idea that since "true" behavior is uncaused or undetermined except by the person, behavior that is caused by factors outside the person should be excused. Moore quite rightly shows such a view to be untenable, since, after all, human behavior *is* caused, fully a function of the genetic and environmental situation as it unfolds. Determinism – the central scientific postulate that permits the explanation of action – is plausible for Moore since a special indeterministic, a-causal exemption for human persons seems so patently *implausible*: "Is it not extraordinary to think that agents who can clearly cause changes to occur in the world

are themselves uncaused? We are quite literally Aquinas's uncaused causers – that is, God – on such indeterminist views" (504).¹

Moore also captures well the *motive* behind resistance to the deterministic thesis: "Both metaphysical and linguistic dualism are drastic doctrines. Metaphysical dualism forces us to simply accept some inexplicable relations between mind and body. Linguistic dualism forces us to grant ordinary usage a priority over scientific theory that is very counterintuitive. The motive for adopting either one of these drastic positions is the same: the desire to preserve our 'freedom and dignity' in the face of advancing mechanistic science. Unnoticed is that such drastic measures are unnecessary. To explain the mind in terms of the brain, or even to identify the mind with the brain, is not to explain the mind away...persons can be agents who act for reasons even in a world in which all mental states and physical events are caused" (521).

Excuses and the function of punishment

So, if we are all determined creatures, what should or could excuse criminal conduct and what justifies holding people responsible? Moore, correctly I think, identifies *rationality*, acting for *reasons* ("reasons-responsiveness") as the key component of responsibility and culpability. If an agent at the time of the offense was sufficiently compromised in her rational capacities, or coerced to act against her wishes, then she should not be held responsible – she is excused. So we excuse young children, the insane, and those under enough duress (e.g., at gunpoint) such that they can't act for the reasons they would normally.

But *why*, we might ask, is the capacity and the opportunity to act rationally the criterion of moral responsibility? Strangely enough, I don't believe that Moore ever addresses this question directly, for reasons that I hope will become clear below. I will spell out my answer now, since for those concerned about the implications of naturalism for the law, the issue of responsibility is often at the heart of their fears. It

is an answer, moreover, that Moore cannot accept given his commitment to retribution as the *summum bonum* of the law, so to present it now will help to clarify where we differ. How then, on my naturalistic account, can a robust notion of responsibility be compatible with determinism, and why is rationality the main criterion of moral agency?

For non-retributivist, "soft" compatibilists such as myself, the answer is that holding people responsible functions to *shape* morally acceptable behavior, and since those who lack the capacity or opportunity to act rationally cannot take moral injunctions into account, punishing them for transgressions imposes needless and possibly very counterproductive suffering. It is the capacity for rational appreciation of punishment as a possible outcome of one's offending, *and* being in a position to act on that appreciation, that makes punishment (sometimes) justifiable as an effective deterrent. I might well do things at gunpoint that I would never otherwise do, given my ordinary reasons and what I know to be the usual sanctions. But to punish me for offenses committed under such duress is unfair because I don't *need* such sanctions to redirect any future propensity of mine for offending; that propensity simply didn't and doesn't exist. Similarly for very young children and the insane: they simply aren't capable of learning or retaining the moral lesson that is obvious to the rest of us, that certain acts, being bad, are usually followed by social sanctions. To apply sanctions against those who are incapable of appreciating them or learning from them, or don't need to be sanctioned in order to learn, is unfair since no useful social *function* is served by the suffering involved. Because punishment serves no function and is thus gratuitous for such classes of persons, they are excused: they are either not counted as moral agents or they are seen as acting under duress. The notions of fair punishment and moral agency, then, are linked on this account to the proper *functioning* of moral practices: the production of desirable behavioral consequences

while minimizing unnecessary suffering.

The two main features this naturalistic account of moral responsibility shares with Moore's are that 1) libertarian, contra-causal free will plays *no* role in determining responsibility (since such free will doesn't exist), while 2) rationality plays an essential role. All behavior is caused, but the (mis)behavior that is justly liable to sanctions is only that produced by more or less rational agents acting without coercion. To be a caused, but rational and uncoerced, agent is to be a *responsible* agent. The main difference in our accounts, I will argue, is that somehow Moore takes the rationality criterion to justify (but how?) the *intrinsic goodness* of punishing the culpable, whereas I take it, as explained above, to show the *functional necessity* of punishing the culpable, if in fact such punishment serves essential social ends unachievable through means that produce less suffering. The differing implications of the views are stark: for Moore, the suffering of the guilty is a positive good, an end in itself that need serve no other purposes; for me, it is a sometimes unfortunate utilitarian necessity. Since we champion differing ideas of what constitutes a desirable, just culture, the issue between us is not academic, but rather has considerable practical ramifications.

Moral reality, virtue, and conflict

Moore's argument for the intrinsic good of retribution is based on a reading of common intuitions about punishment. If we imagine cases of horrific crimes in which (let us suppose) punishment of the offender would serve no non-retributive aim, such as deterrence, reformation, incapacitation, etc., we find that we *still* want to punish. For many, it seems intuitively wrong that someone who has inflicted grievous harm upon us or those we love should not be made to suffer in turn, even if this suffering has no benefits outside gratifying our desire for retribution. Let us grant that this gut response is generally the case: we all, pretty much, have retributive instincts. What follows from this?

For Moore, the intensity and ubiquity of our retributive impulses is a sign of the *moral reality* that retribution is both intrinsically good and the primary aim of the law. Emotions, he says, are "our main heuristic guide to discovering moral truths" (181-2). And: "The *moral fact of the matter* often causes our moral beliefs through the intermediate causing of our emotional responses. Our emotions in such case become good evidence of the *underlying moral landscape*" (183, my emphasis). Moore wants to move from an undeniable psychological fact - that we have retributive inclinations - to the moral fact that retribution reflects an objective moral reality in which giving offenders their just deserts is required to satisfy the demands of justice.

But Moore recognizes that retributive emotions often compete against other responses when we think about criminal offenders. We may find, for instance, that when we consider the tough circumstances of some criminals, we are moved to sympathy, a sympathy that might offset to some extent the desire to impose suffering on the offender. Why shouldn't sympathy mitigate the degree of retribution sought? Doesn't it too reflect some sort of moral reality? If so, might it not suggest other responses to crime, such as rehabilitation, restitution, and addressing the multiple criminogenic factors revealed by behavioral science?

To secure retribution's pride of place, Moore adduces two considerations against this line of argument. First, it is only *virtuous* emotions, he says, which we should generally rely upon as heuristic guides to moral reality: "...in ethics we should recognize that the virtue of (or vice) of an emotion may often, but not always, be taken as an indication of the truth (or falsity) of the judgment to which it leads" (137). Second, in developing an overall theory of justice, he explicitly adopts a "coherentist" approach, in which the theory seeks maximum overall consistency among our intuitions and judgments. (This approach differs, for instance, from a foundationalist approach in which certain first

principles have privileged status and form the basis from which the rest of the theory is derived.) Combining these two desiderata, the way we decide which of our conflicting moral intuitions should be kept and which should be discarded to achieve coherence, is to see which are based in virtuous emotions. Virtue works as a guide to coherence.

Unsurprisingly, Moore argues the case that retributive emotions, at least some of them, are virtuous, while the sympathy we feel for certain offenders is not. Thus we can safely discard our errant sympathies, what Moore calls "moral hallucinations" (132) since they distort our view of moral reality, and hew to the retributive imperative. But whether we agree with Moore about all this hinges on whether 1) we agree with Moore's claims about the virtue and theoretical probity of retributive versus sympathetic emotions, and 2) whether we agree that maximum coherence should be the driving feature of a true theory of justice. I shall contest both these claims.

Moore, very much to his credit, bends over backwards to present the strongest case *against* the virtue of retributive emotions, which he then tries to rebut in the chapter entitled "The Moral Worth of Retribution". He candidly admits that retribution is often motivated by what Nietzsche described with the word *ressentiment*. Moore defines *ressentiment* as "a witches' brew: resentment, fear, anger, cowardice, hostility, aggression, cruelty, sadism, envy, jealousy, guilt, self-loathing, hypocrisy and self-deception" (120). Moore says "It may well be that insofar as the retributive urge is based on such emotions as these...the urge is bad for us" (125).

But, he points out, there are also *virtuous* emotions which ground retribution, namely the "moral outrage" that often is inspired by witnessing or contemplating flagrant acts of wrongdoing that cause suffering, and the often appropriate sense of guilt when we ourselves do something wrong. In both cases, we might feel that retributive punishment is fairly imposed on

the offender, that it is *deserved*, whether the offender be someone else or ourselves. Not to feel these emotions, Moore says, is to be morally defective, and the virtue of such feelings is evidence for the truth of retributivism, since virtuous feelings come with "good epistemic credentials" (147). As Moore says about guilt: "Our feelings about guilt thus generate a judgment that we deserve the suffering that is punishment. If the feelings of guilt are virtuous to possess, we have reason to believe that this last judgment is correct, generated as it is by emotions whose epistemic import is not in question" (148). It is this intuition of desert, generated by guilt and moral outrage, Moore says, that makes it not only *permissible* to punish wrongdoers, but that makes it morally *required* to punish them, even if no other desirable outcomes follow from retribution (148, 154). So although retribution may indeed sometimes be inspired by despicable motives and emotions, for Moore that doesn't undercut its valid basis in some virtuous moral sentiments.

Assuming for the moment that Moore has indeed demonstrated the virtue and reality-revealing nature of some retributive emotions, what about sympathy, e.g., "the sympathy we may feel for wrongdoers whose wrongdoing was caused by factors such as social adversity or psychological abuse during childhood"? Moore writes: "There are three things to say about this range of moral experience. First, the moral judgment it seems to support does not fit with the much larger set of judgments about responsibility that we make in daily life. In seeking the most coherent expression of our moral judgments considered as a whole, *these sympathetic judgments may simply have to be discarded*. No area of human knowledge is perfectly coherent. Any systematic exposition of our sensory experience, for example, has to disregard certain visual experiences because they give us inaccurate information about the world...The same can be said of our sympathetic responses to disadvantaged criminals" (544, my emphasis). In other words, to mini-

mize conflict in our judgments, we must disregard the sympathies generated by the adversities undergone by offenders. Such sympathies are inconsistent with our more numerous and powerful retributive inclinations, and furthermore they are "inaccurate" in some sense.

On the face of it, this seems arbitrary, to say the least. Conflict between retributive feelings, even virtuous ones, and sympathetic feelings (which at least initially seem virtuous) may simply reflect a real moral conflict, and to discount one side of the conflict in order to preserve theoretical coherence might well compromise theoretical accuracy. We may well have *both* feelings about an offender (if not simultaneously, at least in succession), both of which reflect sets of circumstances that pertain to the case. Moore's comparison of sympathetic feelings to inaccurate visual information assumes that our sympathy is somehow misplaced, but of course he has to prove this.

Secondly, and in response to just this point, Moore goes on to say, "...just as we discount our experience with sticks looking bent when immersed in water because we can explain the experience away, so we should discount any sympathy for disadvantaged criminals if we can explain why we feel that sympathy in terms of extraneous factors." Moore suggests two such factors: "our own guilt at not having done enough to alleviate 'unhappy' causes of crime, or...our sense that those who became criminals because of adverse circumstances have 'already suffered enough'" (545). But even if we discount these causes of sympathy as extraneous, surely there are others that are not.

Contemplating the chaotic, punitive and often dangerous conditions that disadvantaged offenders grow up in often generates sympathetic compassion, for indeed were any of us raised in those conditions, we too would be more likely suffer the same criminal fate. Putting ourselves in the disadvantaged shoes of an offender *should* inspire sympathy, for if it does not, then we are supposing that we would have been immune to the influences that shaped her.

From a naturalistic perspective, which Moore shares, in which human beings are determined by environment (as well as heredity), such a supposition is clearly false, and the lack of sympathy it generates is a clear moral defect. Such sympathy may not outweigh feelings of outrage, but it nevertheless reflects circumstances as real as the crime committed, and so is not a "moral hallucination" on a par with the bent stick illusion.²

Third and lastly, Moore says about sympathy that "...we have reason to discount certain experiences and the intuitions they generate when, on examination, their appearance of moral goodness proves deceptive" (545). Moore goes on to speculate that our sympathies for disadvantaged offenders might be due not to any laudable variety of compassion, but to unconscious feelings of superiority to the offender, or perhaps the elitist refusals to judge others by the standards we impose on ourselves or to acknowledge the moral dignity and autonomy of others (546). If these were the only source of our sympathy, then its moral goodness (and epistemic credentials) would be put in question, but since there *is* another robust source, the compassion described above, its goodness stands unchallenged. This point is the same Moore himself makes about the sources of retributive judgments: some sources are morally dubious, but since others have merit, retribution survives in his theory. So, even by Moore's own criterion of moral virtue, sympathy passes muster, and thus it seems we might concede it a role in a theory of criminal justice and in the disposition of criminal cases.

To recap: to achieve a maximally coherent theory of criminal law, Moore wants us to jettison our sympathetic responses to disadvantaged criminals. This would give retribution freer reign, no doubt simplifying the mission of criminal justice. We should discount sympathy because it is morally hallucinatory (it doesn't accurately represent moral reality, just as a bent stick in water doesn't accurately represent the true physical state of the stick) and because its

sources are morally dubious. I think Moore is mistaken on all counts. In reverse order, it appears that sympathy *can* be morally good, that is, it can stem from what we acknowledge is a moral virtue, namely compassion. Second, such sympathy accurately reflects a significant aspect of moral reality, namely the punitive and distressing conditions associated with increased criminality. To ignore such conditions and the compassion they inspire is to give short shrift to an important dimension of our moral universe. Third, we cannot, on grounds of seeking coherence, simply dispense with sympathy as an emotional outlier when forming judgments about criminal offenses. Rather, it justifiably competes with our retributive inclinations. This theory isn't as clean as Moore's, but it does better justice, I submit, to the often ambiguous and conflicting moral reality we inhabit.

At the very least, then, we must admit the legitimacy of emotions that counteract purely retributive judgments against an offender. This suggests that, despite Moore's argument for its essentially retributive basis, criminal justice might have, or seek to have, among its aims the pursuit of goods other than the imposition of suffering on the offender. (Of course one need not believe, as does Moore, that such suffering is an intrinsic good in the first place.) Criminal law has indeed evolved so that its primary functions are to first determine culpability and then impose a just punishment, where what is just is often conceived in terms of desert based on moral responsibility involving an ultimately origination agent.³ And desert, on this interpretation, means that offenders *deserve* to suffer punishment, whether or not punishment entails other beneficial outcomes. Although Moore rejects the supernaturalistic idea of a libertarian agent, his concern is to show that the only broadly consistent account of our present criminal justice practices is nevertheless retributive. He might be correct in that claim, even though other rationales for punishment are widely endorsed. However, even granting that Moore has captured the essence of the aims of

criminal justice in his theory, this still leaves open the question of whether, given our increasing knowledge of the causes of criminality that lie outside the individual, and the mitigation response sometimes generated by this knowledge, retribution *should* be the primary aim of criminal justice.

If, as Moore argues, our virtuous emotions are reasonably good heuristic guides to moral reality, and if virtuous compassion and sympathy therefore reflect morally significant facts about an offender's circumstances, why shouldn't they compete with retributive emotions in determining our response to a crime? This is to say, why shouldn't our response include dealing, via social policy, with the formative circumstances of the offender and others like him, circumstances that might well have been our lot in life? Or, taking another route from sympathy, why shouldn't our response include the *non-punitive* reformation of criminals and restitution of damages, to the extent feasible? To seek the suffering of the offender is just one of many possible responses to crime and injury done to us, and although criminal justice has traditionally taken the imposition of retributive punishment as its primary, often sole objective, this need not remain the case. Indeed, various movements to reform the aims of criminal justice are already well established.⁴

Further, even if offenders don't happen to inspire sympathy (think of the Menendez brothers, seemingly spoiled by their parents into becoming their slayers, or imagine some perfectly rational, unapologetic torturer), the causal story behind their behavior can still justifiably inform our response to their cases. Since we want to avoid future disasters like the one before us, we want to remedy, not replicate, those factors - biological, familial, social, etc. - that induced malevolence or calculating disregard for another's welfare in the first place. In our system of criminal justice, we don't want to model and therefore implicitly endorse as means of behavior control the very conditions we know produce and reinforce manipulative egotism and

pathological brutality. So as much as we may not like the offender (and may well find ourselves wishing him harm), the smartest move in reducing the future prevalence of his offense may well be *not* to inflict suffering which further normalizes a punitive culture, but to incarcerate him while keeping the aims of reformation, rehabilitation, and restoration paramount. By denying ourselves retributive satisfactions in favor of constructive approaches to both offenders and the circumstances that produced them, we serve a better purpose: the creation of a less punitive, more flourishing culture in which we and those that follow us are less likely to face the temptations of retribution itself.⁵

The natural function of morality

But even if we grant sympathy and the appreciation of causes a role in our response to crime, a more basic question still remains, one that challenges Moore's account at a deeper level. It is certainly the case that most of us have retributive impulses from time to time, and that many of us might revel (though some might be loath to admit it) in the suffering of those who have intentionally and seriously harmed us or our loved ones. But what finally *justifies* the state in imposing such suffering? Remember that Moore explicitly sets aside the benefits of deterrence, incapacitation, and reformation as serendipitous side effects of punishment – they are not its central aim on his account. In fact, quoting philosopher F. H. Bradley, Moore says these non-retributive justifications for punishment "are just bad reasons for what we believe on instinct anyway" (99). Since utilitarian justifications are ruled out, the only remaining justification available is that the suffering of the culpable offender is an intrinsic good. But what, one might ask, *makes* it good? From whence comes its value? The only plausible answer, I think, is our bare desire for it: the basic instinct to retaliate against those who harm us confers its value. Suffering is judged "intrinsically" good only because we wish for its realization in the object of our retribution (if

we do indeed wish for this). There is simply no other naturalistic source for its value, once one sets aside the derivative consequential values of deterrence, incapacitation, etc. Moore's case for retribution, based as it is on a series of thought experiments designed to "rouse...our retributive juices," centers on this basis for its worth (83-103).

One might go metaphysical here, and claim that Retributive Good is some sort of Platonic absolute, written into the deepest level of reality. But although Moore types himself a moral realist, he doesn't take this route, but instead cashes out his realism in the characteristics of culpability: "voluntariness of action, accountability, intention, causation, and lack of excuse or justification" (131). Although as Moore says, these characteristics are higher level properties of action that can't be specified in physical terms, they are nonetheless *natural* properties, so responsibility and desert are in turn natural properties. The good of an offender's suffering, then, is keyed to his meeting these naturalistic requirements, most of which have to do with rationality (recall the opening discussion about rationality and the excuses). But why, if an offender's suffering is an *intrinsic* good, does the justification for imposing it depend on meeting these requirements? Why is suffering justifiable only for certain classes of offenders? This question immediately suggests that their suffering is serving some purpose which the suffering of innocents does not serve.

One can say, well, it just happens that we desire the suffering of offenders who are culpable by virtue of being rational, without excuse, etc. and it just happens we do *not* desire the suffering of innocents or the insane. The value of suffering is indeed a function of our desire for it, and our desire just happens to track the rationality of the offender. But of course inquiring minds will persist in asking the further question of *why* our desire to impose suffering tracks rationality in this fashion. That Moore seems not to address this issue strikes me as a serious lacuna in his argument, an omission that

allows him to insulate the retributive impulse from its less than glamorous natural function, and so exalt it as involving a supposedly intrinsic, morally superior good that has priority over consequentialist and utilitarian goods.

The natural function of praise, blame, retaliation, reward and other morality-invoking emotions and responses, what philosopher Peter Strawson in his classic paper "Freedom and Resentment" dubbed "reactive attitudes," is not difficult to discern: it is to shape behavior in ways advantageous to both individuals and societies. Retaliation against an aggressor that has harmed oneself or one's loved ones undoubtedly serves to deter or thwart the aggression, so the disposition to retaliate embodied in retributive emotions such as resentment and rage is an essential characteristic of creatures who make the evolutionary cut. But equally, the disposition *not* to harm those who have done us no harm is just as important, since the advantages of punishment only accrue if it is *selectively* applied. Thus the strong intuition of fairness which says, in essence, "don't ever punish the innocent" can be seen as on a par with other reactive, morality-invoking attitudes, something deeply built into us by virtue of its natural utility in helping to distribute retaliatory harms, so that punishment is vigorously *applied* to deter aggressors and defectors and vigorously *withheld* in order to build affinity and cooperation. And on the positive end, we are built to respond with praise and other rewards to behavior we want reinforced. Thus the direction and depth of our reactive attitudes and moral responses track the type of behavior we are responding to, and help encourage or discourage it, depending on our interests.⁶

It is in this context that we can understand the connection between rationality and responsibility, and the exceptions to assigning responsibility we call excuses. As foreshadowed in the section "Excuses and the function of punishment" above, the requirement that people be rational and uncoerced to be held responsible picks out only those individuals for whom a

reactive response such as praise or blame is capable of and needed for shaping their behavior.⁷ We might experience a surge of anger at someone we judge has attacked us without provocation, but if we discover the person was undergoing an epileptic seizure, or was being forced upon threat of death to attack us, our response changes quickly as it takes this new information into account. Such sensitivity to the causes of behavior is clearly essential if our responses to others are to function efficiently in channeling their behavior within acceptable norms. The reason that excuses excuse is that they reveal causes which render the person in question an *inappropriate* target of blame: generally speaking it's a waste of time and resources to punish those who can't appreciate or act on the knowledge that punishment is a possible consequence of their behavior. Our reactive attitudes and our more deliberate moral practices reflect this.

Similarly, it's counterproductive to punish the innocent since they don't *need* such disincentives to act better – they didn't misbehave in the first place. Their (and our) outrage at and resistance to being punished tracks the disutility of such punishment. That our attitudes about punishment are so responsive to the causal origins of behavior shows that the norm of fairness is built into our nature as beings that seek appropriate, functional responses to the behavior of others. This explains why proposals to punish the innocent to achieve some utilitarian goal (e.g., general deterrence) meet such stiff emotional resistance: they run headlong into a hard-wired, functional disposition to make punishment contingent on actual wrongdoing by the individual who is to be punished.

This naturalistic account of our reactive attitudes and moral practices implies that retribution, which might *seem* to be an intrinsic good, in fact plays a functional role in bringing about other social goods, namely the advantageous shaping of behavior, such as deterring further aggression on the part of those who receive their "just deserts". To see retribution in this light – as a means to an end, not an end in itself

– raises the question of whether there are not better, alternative means to the ends that retribution originally served. We can understand the desire to retaliate in response to an attack as an essential characteristic of creatures who have made the evolutionary cut. But, in the light of what science tells us about the causes of behavior, it's reasonable to ask whether our emphasis on retaliatory justice – retribution – is, all things considered, the best way to reduce unwarranted aggression. When we discover the environmental and biological precursors to violence and see that maladjusted criminals do *not* create themselves, the increase in our sympathies – the mitigation response – shows that we indeed track at an emotional level the causal story behind criminality. Our retributive impulse is lessened because other, more basic, and heretofore hidden causes have come into view, and it is these that now demand attention. It is these that must be addressed to make any real progress in reducing the future prevalence of the crime we are responding to. The rough and ready punitive judgment which unreflectively takes the individual as the uncaused originator of behavior, and therefore the primary retributive target, is attenuated and deflected by a naturalistic, scientific understanding of ourselves into far more productive avenues of response. Reactive attitudes are genetically given, sometimes crude (precisely because they sometimes take agents as uncaused causes), indicators of how causality operates, which science and the rational appreciation of external factors can modulate to good effect: unless directly threatened or in self-defense, enlightened folk attack policies, not persons.

Reforming the aims of criminal justice

Moore's naturalism has not led him to such a conclusion, even though he, like myself, concedes that science will eventually fill in all the gaps in our understanding of mind, brain, and behavior, thus revealing the causal story behind each and every offender. For Moore, knowing that criminality is indeed generated by circumstances well beyond the control of the criminal,

that in fact whole populations are doomed to much higher rates of criminality due to environmental factors, does nothing and should do nothing to mitigate our retributive inclinations, or to shift the aims of criminal law from the punitive to the ameliorative. I suspect, but cannot prove, that this is because he hasn't taken his naturalism far enough to see the natural function of retribution itself. As discussed above, Moore's criteria for moral agents – rationality and lack of coercion or duress – are just those which pick out individuals for whom retribution might efficiently function as a deterrent. But Moore's insistence on the intrinsic good of retribution, based on our desire for it, must suppress this connection, since after all to admit that retribution serves a function, however imperfectly, means that the goal it aims at – the socially advantageous shaping of behavior – is itself a good. And if that good has a claim on us, as it surely does, then it forces us to consider whether retribution is the best means to that end. Also, it might well be the case (I believe it *is* the case) that seeking to impose just deserts as a supposed end-in-itself substantially conflicts with achieving a society which minimizes unnecessary suffering and maximizes human flourishing. Forgoing our pound of flesh in favor of non-punitive detention, re-education, and social reforms such as education, community development, and economic empowerment, is the best route to good behavior.

If one divorces, as does Moore, the desire for retribution from any socially desirable function or outcome, how *does* one justify acting on it? It can't be *just* because we desire it. After all, there are many desires we might have, some of which may well produce suffering in others, that we can't justify acting on. Moore makes vivid the motivational basis for retributive justice, and shows that at least some of the emotions underlying it have virtue, such as our outrage at the unnecessary suffering of victims. But what makes this outrage morally virtuous (when it doesn't slide into vindictive pleasure at

the suffering of the offender) is precisely that it seeks a good *consequence*, namely the ending and prevention of victims' suffering. To make the case that retribution is an intrinsic moral good, it seems Moore must connect it to something more than the bare desire that an offender suffer in turn. But any step outside that desire, or any demonstration of its virtue, must involve some sort of consequence, some outcome, some good that the imposition of the offender's suffering brings about. Some obvious consequences of retribution one could cite are, of course, that offenders are incapacitated, deterred (perhaps), and reformed (rarely). But since Moore explicitly disavows these as playing a role in his account of the law - he says they are just the "bad reasons for what we believe on instinct anyway" - it seems that the intrinsic moral worth of retribution can only be a matter of instinct, of desire. But why should we agree that something is a *moral* good if its only justification is to satisfy a desire that someone should suffer? There has to be more to it, but what that is goes unsaid in *Placing Blame*, as far as I can see.

Moore may have properly described and diagnosed our criminal justice system as one largely preoccupied with quenching our thirst for retribution. He suggests that talk about deterrence, reform, rehabilitation and the like is just window dressing to disguise the simple truth: we want our pound of flesh. But this is not to justify that desire or the system that satisfies it. As much as Moore wants to elevate retribution to the status of *the* reigning moral principle of criminal justice, and by implication defend the criminal justice status quo, I believe the attempt fails. The retributive impulse is *not* our only response to the facts of crime, especially once we are exposed to the science of behavior, a science Moore emphatically endorses. The mitigation response - the sympathy, compassion, and forbearance we feel when learning of the causes behind criminality - has an equal role to play in determining how we should deal with offenders; it is not a moral hal-

lucination. Just as our "retributive juices" are stirred by accounts of horrific crimes, inflicted by monsters on innocent victims, so too our more reflective, deliberate, and considered responses are engendered by discovering exactly how monsters are created, and how many are the result of unchecked retribution itself as it plays out in families, schools, prisons, and wars. These responses are virtuous, based as they are in the naturalistic truth that but for circumstance one's fate might well have been the offender's, and leading as they do to policies that seek to reduce suffering, not increase it. The restorative and ameliorative principles of justice thus stand as strong competitors for our allegiance compared with any retributive principle that might account for criminal justice as it now exists.

Furthermore, talk of deterrence, rehabilitation, restitution, and other utilitarian aims of criminal justice is *not* just window dressing, and for this reason Moore's account falls short even as a description of principles underlying the law. The fact that thought experiments might show that we want to exact suffering from an offender even if this suffering entails no social goods hardly proves that such goods are incidental. And as I have taken pains to show, the natural function of seeking retribution is *itself* nothing other than bringing about such goods, even if we now have better means of achieving them. Moore can only make his case for the intrinsic good of retribution by obscuring or ignoring this natural function, but once revealed, it shows us that however perverse the criminal law may be in giving pride of place to retribution, it is at heart a utilitarian enterprise. Inflicting suffering *can* deter and incapacitate, which is why we are hard wired to inflict it on those who harm us or our loved ones. The law has taken this over as an official, not personal, matter, and to some extent socialized and channeled it. But it *remains* a function with an outcome, not merely a process of administering just deserts that are somehow disconnected from social utility. To imagine or

claim that retribution serves no social ends is to insulate it from a wider, socially informed critique that considers other means of achieving the goals that it sometimes (crudely, with much suffering) attains. Since our desire for retribution is just one aspect of our creaturely disposition to shape behavior advantageously, and is not morally more real or virtuous than the compassionate, preventive and restorative responses to crime inspired by a scientific understanding of behavior, we are entitled to ask whether this desire should any longer remain, either in rhetoric or reality, one of the central motives of criminal justice.

Thomas W. Clark, Research Associate, Health and Addictions Research, Inc., 100 Boylston St. Suite 1075, Boston, MA 02116, USA. Email: tclark@har.org.

Endnotes

1. Stephen Morse takes much the same compatibilist line (see references). Although Moore and Morse both subscribe to an inclusive naturalism that denies libertarian free will, my guess is that the majority of legal scholars are still incompatibilists who assume that responsibility depends on our having such freedom.
2. Moore brings up the case of Richard Herrin, a Yale student from the barrios of East Los Angeles who killed his college girlfriend. Although Moore acknowledges that "we certainly have never been subject to the exact same stresses and motivations as Richard Herrin" he claims that "Herrin had no excuse the rest of us could not come up with in terms of various causes for our choices" (149). But without saying that Herrin should have been excused for the murder, it's still true that the "various causes" of his choices may have been *quite different* from the causes that shape the choices of those not subject to the stresses of the barrio. To the extent those stresses played a causal role in the choice to kill (by, for instance, shaping Herrin's character and motives), and to the extent that we can imagine ourselves having the bad luck to be exposed to such stresses, then we might discover the beginnings of compassion for Herrin, without discounting the horror of what he did.
3. Some, such as Derk Pereboom in his book *Living Without Free Will*, take moral responsibility to mean, strictly and only, that sort of responsibility which requires that agents be the "ultimate sources of their actions" (xv). Since no account of such ultimate agency is sustainable, he argues, moral responsibility doesn't exist. My tack here and elsewhere (see my [Naturalistic Lexicon of Responsibility](#)) is to argue that we *can* be morally responsible in a naturalistic, fully deterministic, and therefore compatibilist sense. That is, we can be morally responsible in a way that conforms with enough of the original and useful sense of that expression such that it is worth keeping. As Daniel Dennett (2003) has pointed out, the substantive differences between compatibilists and incompatibilists might be small (as they are between myself and Pereboom, who takes much the same line against retribution as I do here), and that the differences can lie mostly in the choice of rhetorical tactics: which expressions to redefine naturalistically and which to drop as too contaminated by supernaturalistic connotations. I've chosen to keep "moral responsibility" but to drop "free will." Dennett keeps both, Pereboom keeps neither, but I won't address here the question of whether that reflects a substantive difference in their views.
4. In his book *Preventing Violence*, James Gilligan describes recent developments in criminal justice which illustrate the possibility of moving away from retributive towards restorative and therapeutic objectives. See especially the chapter "Tertiary Prevention: Therapeutic Intervention" on alternative sentencing practices which minimize counter-

productive suffering and recidivism while maximizing opportunities for rehabilitation and restitution. Overall, the book offers both a cogent rationale for rejecting retributivism and specific steps to take in refashioning, from the bottom up, our response to violence.

5. It is of course an open question as to whether and how much restorative and rehabilitative aims may conflict with the goal of deterring offenders, which on the face of it implies the threat of suffering. The default assumption should be that the loss of liberty that comes with incarceration is itself a sufficient deterrent, and that any further suffering or deprivation must be empirically shown to have an additional and necessary deterrent effect in order to be justified.
6. Moore (177-180) cites and responds to John Mackie's evolutionary account of reactive attitudes (in Mackie's articles cited in references) which seems very close to the account sketched here. Moore points out that this sort of psycho-social explanation of our retributive dispositions doesn't mean there can't also be an objective moral reality to which our attitudes are responsive. I'm saying that moral reality *coincides* with the functional facts of moral practice as they have been shaped by evolution and continue to be shaped by culture. Morality doesn't have or need a non-functional, deontological component. Our attitudes and practices don't reflect an underlying or hidden moral reality, they are *parts* of that reality.
7. As in my account, Stephen Morse connects the rationality requirement for punishment with the action-guiding function of the law, citing Jay Wallace (see references). However, Morse (like Moore) thinks that something more than such a consequentialist, forward-looking role for the law is necessary to fully account for our criminal justice practices. Some backwards-looking, deontological constraint, he thinks, is necessary to block questionable practices such as the pun-

ishment of the innocent to achieve utilitarian ends. I don't think we need resort to such a blocker, since our hard-wired disinclination to punish the innocent - to make punishment contingent on actual wrong-doing - finds sufficient support in the empirical discovery that punishing the innocent is *unnecessary* to achieve social goods, and indeed is usually counter-productive given the strength of this disinclination. In "Deprivation and Desert" (p. 121), Morse mentions the expressive function of the law as a non-consequentialist component of criminal justice, but how this would block the utilitarian abuse of punishment isn't clear to me.

References

- Dennett, D., (2003), *Freedom Evolves* (New York: Viking).
- Gilligan, J., *Preventing Violence*. New York: Thames and Hudson, 2001
- Mackie, J., Morality and the retributive emotions, *Criminal Justice Ethics*, Winter/Spring 1982, 3-10
- Mackie, J., Retribution: a test case for ethical objectivity," Feinberg, J. and Gross, H. eds., *Philosophy of Law*, 4th Edition. Wadsworth, 1991
- Moore, M., *Placing Blame: A general theory of criminal law*. Oxford: Clarendon Press, 1998
- Morse, S. J., Excusing and the new excuse defenses: a legal and conceptual review, *Crime and Justice: a review of research*, V23. University of Chicago, 1998
- Morse, S. J., Deprivation and desert, *From Social Justice to Criminal Justice: Poverty and the administration of criminal law*. Heffernan, W. C. and Kleinig, J., eds., Oxford: Oxford University Press, 2000
- Morse, S. J., Hooked on hype: addiction and responsibility, *Law and Philosophy* 19: 3-49, 2000
- Pereboom, D., *Living Without Free Will*. Cambridge (UK): Cambridge University Press, 2001

Richards, J. R., *Human Nature After Darwin: a philosophical introduction*. London: Routledge, Open University, 2000

Skinner, B. F., *Beyond Freedom and Dignity*. New York: Vintage Press, 1972

Strawson, P. F. Freedom and resentment, in G. Watson, Ed., *Free Will*. New York: Oxford University Press, 1982

Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press, 1994.