



JEFFREY W. HOWARD

PUNISHMENT AS MORAL FORTIFICATION

(Accepted 12 September 2016)

ABSTRACT. The proposal that the criminal justice system should focus on rehabilitation – rather than retribution, deterrence, or expressive denunciation – is among the least popular ideas in legal philosophy. Foremost among rehabilitation’s alleged weaknesses is that it views criminals as blameless patients to be treated, rather than culpable moral agents to be held accountable. This article offers a new interpretation of the rehabilitative approach that is immune to this objection and that furnishes the moral foundation that this approach has lacked. The view rests on the principle that moral agents owe it to one another to maintain the dependability of their moral capacities. Agents who culpably commit criminal wrongs, however, betray an unacceptable degree of moral unreliability. Punishment, on this theory, consists in the enforcement of the duties that offenders have to reduce their own likelihood of recidivism.

How should a moral agent respond to his own moral failure? Specifically, when a moral agent performs a justly criminalized act, what should he then do? No doubt he should come clean and confess that he has broken a just law. No doubt he should recognize that what he has done was wrong and express regret for having done it. No doubt he should apologize to his victims and even to the wider community of citizens. And no doubt he should compensate his victims for the losses they have incurred, when possible. So much, I submit, we should take for granted.

What else should he do? The proposition I aim to defend is this: he ought to take measures that reduce the likelihood that he will act wrongly again. Rightly disturbed by the manifest unreliability of his moral powers, he ought to *fortify* them – to shore up their prospects for future success. If he does not undertake such an effort to get his moral house in order, he has failed to take seriously the risks that he continues to pose to others. A state that curtails his liberty in the

service of that effort, then, intervenes in his life not to inflict retributive suffering, nor (at least chiefly) to deter others. Rather, its aim is to enforce the duties he has to maintain his moral dependability.

The suggestion that the penal system should principally aim at offenders' moral reform is not, of course, original. It is regularly listed as one of the main candidate purposes of punishment. This makes it all the more surprising that few direct attempts have been made in contemporary political and legal philosophy to mount a defense of 'the rehabilitative ideal'.¹ That lacuna has enabled philosophers and citizens alike to entertain only its caricature, according to which criminals are viewed as blameless patients in need of treatment. Rehabilitative approaches have thus been condemned as incompatible with a proper view of criminal offenders as culpable moral agents. Accordingly, such approaches have been consigned – unfairly – to the theoretical graveyard.

Here I present a new interpretation of the rehabilitative approach – *punishment as moral fortification* – that is immune to this criticism and therefore furnishes the moral foundation that the rehabilitative approach has heretofore lacked. The theory is allied with prevailing deontological theories of punishment, which explain punishment's permissibility by appealing to the moral duties of criminal offenders.² It is grounded in an undertheorized but, I will argue, widely presupposed moral truth: agents owe it to one another to maintain the reliability of their moral capacities, especially when it comes to compliance with the criminal law. When an agent commits a culpable criminal wrong, however, he demonstrates that he has failed to maintain his moral capacities to the requisite degree, and his government secures a moral permission to impose measures to fortify his sense of justice.

While it is impossible to outline a complete theory of punishment in a single article, I will defend the view's plausibility by contrasting it with prevailing approaches and elaborating its responses to the

¹ This term was popularized in Francis A. Allen, *The Decline of the Rehabilitative Ideal: Penal Policy and Social Purpose* (New Haven: Yale University Press, 1981). For a catalogue of different possible justifications for rehabilitation, see Peter Raynor and Gwen Robinson, *Rehabilitation, Crime, and Justice* (New York, NY: Palgrave, 2005), Chapter 2.

² This strategy is at the heart of R.A. Duff, *Punishment, Communication, and Community* (Oxford: Oxford University Press, 2001), and Victor Tadros, *The Ends of Harm: The Moral Foundations of Criminal Law* (Oxford: Oxford University Press, 2011). Duff remarks on this approach explicitly, discussing Tadros, in 'Punishment and the Duties of Offenders', *Law and Philosophy* 32, 1 (2013): 109–127, p. 110.

most likely objections against it. The aim is hardly to discredit all rival approaches to the justification of punishment. Rather, it is to defend the minimal claim that moral fortification is a sufficient reason to punish, rendering the view worthy of further scholarly exploration.

But like much work in political philosophy, the payoff is not simply of academic concern. Reformers who seek to defend rehabilitation in our tough-on-crime age lack a convincing framework through which to explain and justify their views. The theory outlined here aspires to provide the foundations of that framework.

I. MORAL FORTIFICATION, REASONABLE FIRMNESS, AND THE CRIMINAL LAW

The principle upon which the theory rests finds itself in the unusual position of being both undertheorized and yet widely accepted. It is roughly this: all moral agents owe it to one another to maintain the reliability of their moral capacities. Consider Jones, who regularly experiences temptations to lash out violently at others when they irritate him. While he has not yet acted on his temptations, he reasonably believes that it is only a matter of time until he eventually does so. He nevertheless does nothing to mitigate these temptations, such as pursuing counseling, seeking out new associates to help him reorient his thinking, or making other changes in his lifestyle. In cases like these, our intuition is plain: he is not taking the matter seriously, and is accordingly liable to moral condemnation. He owes it to those he threatens to *fortify* his sense of justice: to take measures that increase the likelihood that his moral capacities will operate effectively.

The idea that agents have what I will call *fortificational duties* is presupposed by many of our ordinary moral judgments about agents' personal responsibility for living up to moral demands. To say that Barbara is subject to a duty not to steal from others is not some obtuse shorthand for the idea that someone else needs to ensure that Barbara does not engage in theft. Of course not: it means that *Barbara* needs to ensure that Barbara does not culpably commit theft, whatever might be involved in doing so. She must, that is, do what it takes to realize a state of affairs in which she has not culpably committed that crime. To succeed, this may involve, *inter alia*,

avoiding locations in which objects she is tempted to steal are found; reflecting on her potential victims' interests and their value, in the hopes of generating countervailing desires not to wrong them; deepening her understanding of why theft is wrong in the hope that her moral motivation not to steal will be strengthened through such understanding; and so on.

This example is somewhat contrived, but it illustrates something familiar about the phenomenology of moral compliance. Doing one's duty is often not experienced as a matter of flicking a switch. There are potential psychological hurdles that agents may well encounter along the way. And we typically judge that it is the agent herself who is primarily responsible for overcoming them. Thus our primary moral duties to perform or refrain from performing certain acts are, in fact, often accompanied by 'waves' of fortificational duties whose fulfillment positions the agent to live up to the associated primary duties successfully.³ And while it may be rational for citizens to work together in mitigating the influence of pernicious temptations, the structure of the duty to self-fortify is, at root, individual.⁴ In ordinary moral experience, we standardly judge that it is up to agents themselves to get their primary moral jobs done.

Thus fortificational duties are largely presupposed by our ordinary moral practices. That is a powerful reason to believe in their existence. Their plausibility is reinforced by the more familiar principle that agents ought not to endanger the rights of others without sufficient justification.⁵ Activities like speeding in an automobile are wrongful not necessarily because they violate anyone's rights, narrowly conceived, but because they increase the likelihood of violation without justification.⁶ It is widely believed that we have duties to manage such external technology appropriately, minimizing the dangers it poses. But the same reasoning suggests that we should handle our *internal* technology appropriately, too – to maintain our

³ For the idea of 'waves of duties', see Jeremy Waldron, 'Rights in Conflict', *Ethics* 99, 3 (1989): 503–519, p. 509.

⁴ Of course, it may be that we also have duties to increase the likelihood of *others'* compliance with morality. I do not defend that more ambitious claim here.

⁵ This qualification is important: there are plenty of contexts in which it is permissible to impose risks on others – for example, in practices in which all consent to the risks involved.

⁶ Alternatively, we might prefer to express this same point by saying that speeding *does* violate others' rights, precisely by risking serious harm.

own moral capacities with an eye toward circumventing their malfunction.

We need not embrace the view that agents have duties to fortify their moral capacities with respect to *all* moral duties. John Rawls familiarly distinguishes the first moral power – to identify and be moved by moral duties of justice – from the second moral power – concerning questions of the meaning and nature of the good life.⁷ Now, whether agents even have duties to improve their conformity with the correct conception of the good life (whatever it might be) is, on a liberal view, precisely the kind of question that citizens have the prerogative to figure out for themselves. Our concern here is instead with the *first* moral power.

The first moral power has both an epistemic component, concerning the identification of one's justice-related moral duties, and a motivational component, concerning one's compliance with those duties. The principle that we ought to fortify our first moral power accordingly enjoins us to guard against two corresponding sources of moral failure. If I have a tendency to perpetrate *epistemic moral failure* – reaching mistaken conclusions about whether certain conduct qualifies as permissible or wrongful – I ought to take steps to eliminate that tendency. This may simply be a matter of spending more time engaged in practical deliberation, or of diversifying my sources of information, or of asking peers to double-check my thinking. Sometimes my epistemic failure will concern not simply the content of morality, but the role of morality. If I have a tendency to treat moral reasons for action as on par with any other sort of reason, thereby denying them their typical regulative primacy, I ought to correct that tendency. The reason why is clear: epistemic failures surrounding moral matters are often causally upstream of actual wrongdoing.

Likewise, it is familiar enough that agents sometimes act contrary to moral duties that they themselves recognize, perpetrating *motivational moral failure*. In cases in which agents perceive an incongruence between what is prudentially rational and morally required, people face temptation. Agents have duties to take reasonable steps to mitigate the influence of such temptations – schooling themselves out of certain desires, avoiding certain situational triggers, and so on.

⁷ John Rawls, *Political Liberalism* (New York: Columbia University Press, 2005), pp. 47ff.

The duty to fortify one's first moral power surely has ramifications across political philosophy. Here, however, I shall focus on one subset of these requirements. Specifically, I shall examine the fortificational duties that attach to our primary duties not to commit *crimes*: acts that the legislature of a given jurisdiction have deemed morally unacceptable and formally declared are 'not to be done'.⁸

Here the idea of fortificational duties finds expression not only within our moral experience, but also in our established legal practice. Consider a defendant who appears in court, and claims the following defense for his sizeable bank robbery: he did it because he was overwhelmed by his desire to become rich, and should accordingly be excused. Or consider another defendant who lethally poisons her professional nemesis, thereby securing her own promotion at the office, and insists she is excused because of the strength of her ambition for that promotion. The influential Model Penal Code instructs courts not to take such arguments seriously. The idea that agents can be 'forced' by their ordinary criminal temptations into doing wrong – and so can invoke the excuse of duress – is rejected by prevailing legal practice. Agents are required to demonstrate 'reasonable firmness' in response to the temptations toward crime that they face.⁹ If defendants face temptations to break the law, it is their responsibility to manage those temptations and prevent themselves from succumbing to them.

That the demand for firmness embodied in our law is only for *reasonable* firmness is important. Agents are not obliged to make themselves resilient against the kinds of psychologically overwhelming forces that genuinely eviscerate responsibility. An agent who commits a crime under genuine duress – say, as a loaded pistol is pointed to her child's head – is not ordinarily regarded as morally blameworthy if she commits the crime.¹⁰ Such a person cannot be

⁸ H.L.A. Hart, *Punishment and Responsibility* (Oxford: Oxford University Press, 1968), p. 6. Here I will consider both *mala in se* offenses – acts that are wrongful independent of their prohibition – and *mala prohibita* offenses – acts that are wrongful in part *because* the law prohibits them. Note, however, that most *mala prohibita* crimes (such as driving on the left in the U.S.) serve to coordinate compliance with a background moral principle (such as the duty not to endanger others while operating fast-moving vehicles). Thus failure to comply with (justified) *mala prohibita* criminal statutes can constitute genuine moral failure. I thank a reviewer for raising this issue.

⁹ *Model Penal Code* 1985, section 2.09 (1)–(2). The term 'reasonable firmness' only arises in connection with the provisions on duress, but the normative principle reflected therein is broadly consistent with the Code generally. For instructive discussion, see Richard Lippke, 'Chronic Temptation, Reasonable Firmness, and the Criminal Law', *Oxford Journal of Legal Studies* 34, 1 (2014): 75–96.

¹⁰ Importantly, this example involves duress as an excuse, not as a justification. Note that agents are typically not legally excused for homicide, even if they are threatened with death – but this is controversial.

thought to be lacking in reasonable firmness, for it is implausible to hold that an agent should have taken measures to prepare herself for moral rectitude in the eventuality of such an occasion. To make oneself reasonably firm, then, one must ensure that one is prepared to overcome the kinds of threats to moral rectitude that do not eviscerate responsibility.¹¹

Fortificational duties explain why our legal commitment to reasonable firmness is appropriate: because agents have responsibilities to make themselves firm in precisely the way the law expects. But it is crucial to observe that our legal commitment to reasonable firmness does not issue any concrete recommendations as to how to make ourselves reasonably firm. Rather, we simply expect one another, as responsible adults, to get the job done. But why? Why not issue formal legislative pronouncements that specify what it is, exactly, that citizens should be doing to fortify themselves? Why not require compulsory sessions with state counselors, checking up to be sure we are keeping ourselves morally resilient? Why not create a civic culture in which strangers are continually peeking into one another's lives, demanding to know whether they have undertaken this or that measure to shore up their first moral power?

It is not a moral oversight that we demand reasonable firmness and yet refuse to permanently curtail citizens' liberty and privacy to ensure that demand is met. It reflects a deep conviction: that sane adults should take responsibility for their own moral resilience. Coercive interventions by the state to fortify all citizens' first moral power would strike us as profoundly objectionable. One powerful explanation as to why we rightly respond in this way is that such interventions express serious disrespect for their targets. Unless an agent has provided compelling grounds to believe that he is a moral threat, he rightly finds himself insulted if the state treats him as one. Just as paternalistic policies are objectionable partly because they hinge on a negative judgment about agents' ability to make the right decisions about their own good¹² – to exercise their second moral power well – regular coercive interventions by the state to fortify citizens' first moral power would rest, similarly, on a negative judgment about their prospects of exercising that power successfully.

¹¹ The account here takes no stand on what kinds of threats eviscerate responsibility, only that some do.

¹² Jonathan Quong, *Liberalism Without Perfection* (Oxford: Oxford University Press, 2011), p. 80.

In the absence of appropriate evidence that an agent is morally untrustworthy, it is *prima facie* wrong for the state to curtail his liberty and privacy in order to fortify him. So long as the agent's actions do not demonstrate moral unreliability, he retains a defeasible, respect-based claim against the coercive enforcement of his fortificational duties by the state. That, anyway, is the sensible posture suggested by our current moral and legal practice.¹³

II. THE FORTIFICATIONIST THEORY OF PUNISHMENT

When an agent commits a morally culpable criminal offense, his action betrays a failure to maintain his moral reliability. He demonstrates that he has failed to achieve and sustain his reasonable firmness in the face of the criminogenic pressures he faces in his life. The presumption of total trust in his moral capacities can no longer be credibly maintained; he has forfeited his respect-based claim against state enforcement of the relevant fortificational duties.¹⁴ Criminal punishment consists precisely in the enforcement of those duties: a coercive intervention into the agent's life to demand that he reform himself. This, in summary, is the vision urged by *the fortificationist theory of punishment*. This section introduces the theory and answers some preliminary objections to it.

Perhaps in some other possible world, with different sorts of beings, fortification could be accomplished by picking flowers or sipping cocktails at parties. But for beings like us, this process of fortification is a process of hard work: of reflecting on the specific law one has broken to solidify one's understanding of its demands; of undertaking introspection as to why one contravened it; and of resolving and embarking on certain courses of thought and action to decrease the likelihood that one will engage in criminal wrongdoing again – reorganizing one's life, if necessary, to do so. Unlike the expressive sentiment of regret for, say, murder – something an agent can feel instantly – these pursuits really do take time out of one's

¹³ For a further defense of this presumption of trust, see R.A. Duff and S.E. Marshall, 'Benefits, Burdens and Responsibilities: Some Ethical Dimensions of Situational Crime Prevention', in *Ethical and Social Perspectives on Situational Crime Prevention*, edited by Andrew von Hirsch, David Garland and Alison Wakefield (Portland, OR: Hart Publishing, 2000), pp. 17–35, p. 29.

¹⁴ Any deontological justification of punishment must include an explanation of why the criminal offender has forfeited her right against the penal treatment. For a defense of this important claim, see Christopher Heath Wellman, 'The Rights Forfeiture Theory of Punishment', *Ethics* 122, 2 (2012): 371–393.

schedule and thereby reduce one's opportunities for pursuing plans, projects, and relationships during the time they demand.

What would fortificational punishments actually involve? They take their cue from certain elements of prevailing rehabilitative approaches, which transpire at the fringes of the criminal justice system in most Western democracies. Consider these illustrative examples, all designed to help offenders get back on their moral feet:

1. The probation order assigned to the citizen convicted of theft, requiring that he live in a probation hostel, under curfew, where he will receive regular counseling and support in developing personal goals on how to reduce reoffending.¹⁵
2. The prison sentence assigned to violent criminals requiring that they attend an intensive '12-hours-a-day, 6-days-a-week programme consisting of workshops, academic classes, theatrical enactments, counseling sessions and communications with victims of violence'.¹⁶
3. The prison sentence assigned to spousal abusers, removed from anger-inducing circumstances and enjoined to participate in seminars in which they partly re-enact and then critically discuss various scenarios involving abuse – thereby securing a clearer understanding of the forces that moved them to act wrongly and subsequently generating plans for how to manage pernicious desires in the future.¹⁷
4. The probation order assigned to citizens convicted of alcohol-related crimes (such as driving under the influence or drunken domestic vio-

¹⁵ For a report detailing reduction in recidivism among those assigned to probation hostels ('approved premises'), see 'An Inspection of Approved Premises in Northern Ireland', published by *Criminal Justice Inspection Northern Ireland* November 2013 and available at <http://www.cjini.org/CJNI/files/50/500620a6-e62f-4eb8-82b3-1c686eff1368.pdf>. For general discussion of probation, see Iain Crow, *The Treatment and Rehabilitation of Offenders* (London: Sage, 2001), pp. 83ff, and George Mair and Lol Burke, *Redemption, Rehabilitation and Risk Management: A History of Probation* (New York: Routledge, 2012).

¹⁶ James Gilligan and Bandy Lee, 'The Resolve to Stop the Violence Project: Reducing Violence in the Community Through a Jail-Based Initiative', *Journal of Public Health* 27, 2 (2005): 143–148, p. 144. As one author notes, "participation in this program for as little as four months reduced the frequency of violent reoffending after leaving the jail by 83%, compared with a matched control group in a conventional jail". James Gilligan, 'Punishment Fails, Rehabilitation Works', *The New York Times*, December 19, 2012, available at <http://www.nytimes.com/roomfordebate/2012/12/18/prison-could-be-productive/punishment-fails-rehabilitation-works>.

¹⁷ R.A. Duff defends this program in *Punishment, Communication, and Community*, pp. 102–103, 108–109. For analysis of treatment programs for perpetrators of domestic violence, see Javier Fernández-Montalvo et al., 'Impact of a Court-Referred Psychological Treatment Program for Intimate Partner Batterer Men With Suspended Sentences', *Violence and Victims* 30, 1 (2015): 3–15.

- lence), requiring them to submit to daily breathalyzer checks while strategizing with parole officers on how to avoid reoffending.¹⁸
5. Cognitive behavioral therapy assignments for offenders to learn strategies for avoiding acquiescence to criminal temptation.¹⁹
 6. Discussion seminars in which offenders are invited to engage in, and listen to, reflection on topics in moral and political philosophy, with special attention to the question of why the actions in which they engaged are wrongful.²⁰
 7. Sessions in which offenders must listen to testimonials offered by past victims of crime about how various crimes affected them.²¹
 8. ‘Restorative justice’ sessions in which offenders meet their victims, sometimes alongside their respective support networks, to hear the victims’ account of how the crime affected them. They work together to craft a plan for how the offender should make amends for the crime, and to set tasks aimed at reducing the likelihood of reoffending.²²
 9. Sentences to prisons that are designed to be replicas of the outside world, and that aim to prepare offenders for that world by respecting all of their rights except freedom of movement, and providing offenders with substantive access to work, education, and treatment for substance abuse.²³

¹⁸ A program of this kind is described in Beau Kilmer *et al.*, ‘Efficacy of Frequent Monitoring with Swift, Certain, and Modest Sanctions for Violations: Insights From South Dakota’s 24/7 Sobriety Project’, *American Journal of Public Health* 103, 1 (2013): 37–43.

¹⁹ See David Wilson *et al.*, ‘A Quantitative Review of Structured, Group-Oriented, Cognitive Behavioral Programs for Offenders’, *Journal of Criminal Justice and Behavior* 32, 2 (2005): 172–204.

²⁰ See Alan Smith, ‘Locke and Key’, *The Guardian*, 20 January 2004, available at <http://www.guardian.co.uk/education/2004/jan/20/furthereducation.uk2>, and Daniel de Vise, ‘College student’s philosophy program brings Plato and Buddha to a Md. prison’, *The Washington Post*, 1 September 2011, available at http://articles.washingtonpost.com/2011-09-01/local/35273403_1_maximum-security-prison-paper-clip-inmates. Sometimes these programs are offered to offenders who have life sentences, which is perverse, given their proper purpose.

²¹ Stephen Farrall and Adam Caverley, *Understanding Desistance from Crime* (New York: Two Penn Plaza, 2006), p. 53, describing a group program termed ‘A Fresh Beginning’.

²² See Carolyn Hoyle, ‘The Case for Restorative Justice’, in Carolyn Hole and Chris Cunneen, *Debating Restorative Justice* (Oxford: Hart Publishing, 2010), pp. 14ff. See the discussion of ‘circle sentencing’, in Gill McLvor, ‘Reparative and Restorative Approaches’, in Anthony Bottoms, Sue Rex, and Gwen Robinson (eds.), *Alternatives to Prison: Options for an Insecure Society* (Cullompton: Willan Publishing, 2004), p. 164. See also the canonical defense of ‘reintegrative shaming’, in John Braithwaite (ed.), *Crime, Shame, and Re-integration* (Cambridge: Cambridge University Press, 1989), pp. 69ff.

²³ Many Scandinavian prisons, especially Norwegian prisons, operate along these lines. For a sympathetic description, see Doran Larson, ‘Why Scandinavian Prisons are Superior’, *The Atlantic*, September 24, 2013, available at <http://www.theatlantic.com/international/archive/2013/09/why-scandinavian-prisons-are-superior/279949/>. Only 20% of offenders released from Norwegian prisons reoffend, compared to 50–60% in the U.K. and U.S.; see William Lee Adams, ‘Sentenced to Serving the Good Life in Norway’, *Time*, July 12, 2010.

10. Sessions with counselors in which offenders are enjoined to develop a proactive narrative of their own identity, and are encouraged to come to think of themselves as autonomous agents in control of their own future as they specify concrete plans (and receive education and training to prepare) for their future relationships and employment.²⁴
11. Community service orders through which offenders must undertake activity that facilitates reflection on the wrongness of their past conduct (for example, an order requiring tax cheats to prepare reports to distribute to the public on the detrimental consequences of defrauding the state).²⁵

A state that deprives criminal offenders of their liberty by coercively mandating attendance at these burdensome programs aims not at suffering for its own sake, nor (at least primarily) to deter others. It aims to facilitate morally fortifying experiences for criminal offenders. As these examples clarify, such experiences are not simply a matter of exposing offenders to philosophical syllogisms. They aim to cultivate the moral emotions, to put offenders in the shoes of their past and prospective victims so as to develop their empathetic capacities, and to help them craft, prepare for, and proactively implement concrete plans for their future lives.

These potential punishments formalize endeavors that, on my view, citizens should ideally be undertaking already. Good citizens, I argued earlier, are regularly on the lookout for ways to reduce their susceptibility to moral failure. One explanation of why many of us never find ourselves facing punishment is that we already organize and maintain our lives in ways propitious for moral success. And so long as we succeed, we do not give the state any actionable reason to doubt our reasonable firmness. The ire of the criminal justice system befalls those who have ignored their fortificational duties, or who

²⁴ This approach is examined by Shadd Maruna, *Making Good: How Ex-Convicts Reform and Rebuilt Their Lives* (Washington, DC: American Psychological Association, 2001), Chapters 5 and 6. For a meta-analysis of studies that broadly shows that education and employment training programs reduce recidivism, see David B. Wilson *et al.*, 'A Meta-Analysis of Corrections-Based Education, Vocation, and Work Programs for Adult Offenders', *Journal of Research in Crime and Delinquency* 37, 4 (2000): 347–368. As the analysis suggests, more research is needed to establish this causal relationship rigorously, as current programs are largely optional. My argument is that they ought to be compulsory.

²⁵ Anne Worrall, *Punishment in the Community: The Future of Criminal Justice* (London: Routledge, 2014), Chapters 7 and 8.

have not discharged them to the requisite degree, and committed crimes. Criminal wrongdoing triggers the permissibility of these duties' enforcement.

The plausibility of the fortificationist theory depends upon its contention that an agent, by committing a culpable criminal offense, betrays his moral unreliability and accordingly forfeits his claim to conduct his fortificational endeavors without state coercion. But why should we think that commission of a culpable criminal offense is unique evidence of moral unreliability? That it typically *is* such evidence is uncontroversial; a murderer who insists – 'Don't worry; you can rely on me not to murder again' – cannot ordinarily expect to be taken seriously. But why not rely on other indicators of future wrongdoing, such as statistical propensity to commit criminal offenses or worrisome personal behavior?

If the fortificationist theory were justified as part of a broader consequentialist moral philosophy, it would be, in principle, possible to license fortificationist coercion of individuals without any reference to actual choices those individuals have made (e.g., because they belong to a particular demographic group with higher-than-average propensity for committing crime). But on a deontological view, this will not suffice. Depriving agents of their liberty is very difficult to justify. One way to make such deprivation easier to justify is to show that an agent could have avoided such deprivation at a low cost. As T.M. Scanlon notes, all an agent needs to do to avoid punishment is to do what morality requires him to do: obey legitimate criminal laws.²⁶ But deprivations of liberty undertaken for reasons wholly unrelated to agents' actual choices will be, by definition, impossible to avoid. Thus a system that punishes persons for reasons unrelated to their choices will be significantly more difficult to justify. So long as an agent has not committed a culpable wrong, he retains his respect-based claim against the state that his fortificational pursuits – his responsibility to maintain his 'reasonable firmness' – remain unenforced.

²⁶ T.M. Scanlon, 'Punishment and the Rule of Law', in *The Difficulty of Tolerance* (Cambridge: Cambridge University Press, 2003), p. 230. Cf. Tadros, *The Ends of Harm*, pp. 170ff.

It may be replied that there exist certain culpable choices that appear to fall short of proper crimes but nevertheless indicate seriously malfunctioning moral capacities. And so there are: specifically, if persons have initiated causal processes that they aim to conclude with the commission of a culpable offense, these choices clearly qualify as culpable malfunctions in moral agency, even if no victim's right has, strictly speaking, yet been violated. But that is a feature, not a bug, of the theory, for such activity is already widely classified as criminal – as belonging to the category of 'inchoate crimes'.²⁷ And fortificationists can explain precisely why it is appropriate to regard such activity as criminal: undertaking a chain of actions intended to culminate in a crime reflects a significant failure to be moved appropriately by the value of others.²⁸

Even if we accept that criminal offenders owe it to their fellow citizens to increase their likelihood of future compliance with the criminal law, as the theory contends, we may face skepticism about their prospects for success. Does rehabilitation work? Just as ineffective deterrent measures would not be justified on a deterrent theory, ineffective fortificational measures would not be justified according to the fortificationist theory.²⁹ That would not render the theory untrue, for it could simply issue conditional recommendations (e.g., 'if it is possible for offenders to fortify

²⁷ Here I presuppose the mainstream view that inchoate crimes are, in fact, properly classified as crimes. For a sophisticated rejection of this view, see Kimberly Kessler Ferzan, 'Beyond Crime and Commitment: Justifying Liberty Deprivations of the Dangerous and Responsible', *Minnesota Law Review* 96 (2013): 141–193.

²⁸ Imagine that advances in genetics give us the ability to identify a subset of the population who are predetermined to commit violent crime. And suppose that an agent – call him Joe – is overwhelmingly likely to commit a murder unless he works hard to fortify himself (and he knows this). Would it be justified to subject Joe to coerced fortificational measures before the time of the crime? To be sure, Joe would have a duty to fortify himself, given that he has excellent reason to believe he poses a threat. But what if Joe refuses? One position holds that the state should never be in position to know this; we have powerful reasons, grounded in concerns about privacy and abuse of power, to balk at the idea of a government database of all citizens' genetic codes (or, for that matter, mental health records). A second position insists that even if the state is aware of his genetic predisposition to kill, Joe retains his respect-based claim to pursue his fortificational efforts unenforced; if he commits murder, he will be punished. A third position holds that Joe should be subject to a legal requirement to seek out the requisite fortification (e.g., admitting himself to counseling) and to prove that he has done so. The framework outlined here, while designed for the context of a society without any such knowledge, could be rendered consistent with all of these positions. For related discussion, see Stephen J. Morse, 'Blame and Danger: An Essay on Preventive Detention', *Boston University Law Review* 76 (1996): 113–155, esp. pp. 152–154. I thank an anonymous reviewer for pressing me on this issue.

²⁹ Though it is worth noting that deterrent theorists often simply take for granted that punishments (of the sort they recommend) will succeed in deterring crime. For example, Tadros simply appeals to his empirical intuitions about why people are disinclined to commit crime in *The Ends of Harm*, pp. 281–282, as noted in Kimberley Brownlee, 'What are the Duties in the Duty View?', *Jerusalem Review of Legal Studies* 5, 1 (2012): 62–74, p. 65.

their sense of justice, they ought to do so'.) Still, it would be puzzling to develop a theory with conditionals unlikely ever to be satisfied. And while the central burden of this essay is to allay the kinds of philosophical worries about rehabilitation that have rendered it unpopular in political and legal theory, it is nevertheless important to address this concern.

Why have faith in fortification? Part of the faith is supplied not by empirical evidence, but by moral conviction. Insofar as we regard criminal offenders as moral agents, we must view them as capable of choosing to reform their lives and to make different choices in the future than they have made in the past. To deny this – to insist that they are fated to make the same mistakes again, that they are incapable of changing their minds about what they should do and act accordingly – is to suggest that they are not proper moral agents after all. But if they are not moral agents, then they are not the sorts of beings for whom blame and punishment can be appropriate responses. For example, many philosophers believe that psychopaths, in virtue of their defective moral psychology, are for that reason outsiders to the community of duty-bearing moral agents.³⁰ If this is so, then *a fortiori* such offenders simply lack fortificational duties and so fall outside the framework defended here. The dangers posed by such non-agents would be tantamount to those posed by dangerous non-human animals.³¹

Of course, the fact that fortification is possible does not mean it is likely. Rehabilitation fell out of favor among penal ideals in the United States during the 1970s partly because of empirical objections

³⁰ For arguments that psychopaths are not morally responsible for what they do (or at least are partly excused) see Jeffrie G. Murphy, 'Moral Death: A Kantian Essay on Psychopathy', *Ethics* 82, 4 (1972): 284–298; Ishtiyaque Haji, 'On Psychopaths and Culpability', *Law and Philosophy* 17, 2 (1998): 117–140; R.A. Duff, 'Psychopathy and Answerability', in *Responsibility and Psychopathy* (Oxford: Oxford University Press, 2010), pp. 199–212; and Neil Levy, 'Psychopaths and Blame: The Argument from Content', *Philosophical Psychology* 27, 3 (2014): 351–367.

³¹ For the idea of treating psychopaths as wild animals, see Kimberley Kessler Ferzan, 'Living on the Edge: The Margins of Legal Personhood', *Rutgers Law Journal* 39, 2 (2008): 237–246, pp. 239–240. Note that if psychopaths are moral agents – say, because they are capable of understanding moral arguments, despite their limited powers of empathy – then we rightly view them as morally tasked with overcoming the obstacles that stand between them and moral success. The mere fact that an offender's rehabilitative journey may be considerably more difficult than that of other offenders does not exempt her from the obligation to fortify herself. Moreover, the claim that psychopaths are incapable of any reform is itself questionable. Research shows that they may be able to learn to activate their empathetic capacities; Harma Meffert *et al.*, 'Reduced Spontaneous but Relatively Normal Deliberate Vicarious Representations in Psychopathy', *Brain* 136, 8 (2013): 2550–2562. So part of moral fortification would involve teaching psychopaths how deliberately to activate their latent empathy. In any case, we must remember that the mere fact that one is labeled a psychopath by current practices does not mean one is a psychopath. One study shows that a person's score on the psychopathy test can vary considerably based on who administers the test; see Daniel C. Murrie *et al.*, 'Are Forensic Experts Biased by the Side That Retained Them?', *Psychological Science* 20, 10 (2013): 1–9.

to existing programs' effectiveness.³² And clearly some rehabilitative efforts have often failed to reduce recidivism to the desired degree.³³ But if the moral argument I am defending here is right, the task of ascertaining the best methods of fortification is itself a moral imperative. The mere fact that we have not yet solved that challenge is no counterargument. Far greater policy experimentation is required before skepticism toward rehabilitation is justified. Moreover, as my examples earlier demonstrated, such experimentation is already seriously underway, and numerous programs have shown positive results in reducing reoffending.³⁴

Indeed, the condemnation of past and present rehabilitation schemes must be taken with quite a few grains of salt, given their context within seriously unjust institutions. Rehabilitative endeavors, in the hands of far-from-fully-just states, can be pernicious for two important reasons. First, it is wrongful to punish citizens who contravene seriously unjust laws. Attempts to 'fortify' gay and lesbian citizens who contravene illiberal prohibitions of homosexual sex are unjust, and do not qualify as genuine moral fortification. Likewise, citizens who are victims of distributive injustice may have a moral permission to disobey certain laws.³⁵ If that is so, then attempts to 'fortify' such citizens would merely be an endeavor to tranquilize recalcitrant members of disadvantaged social groups, to force them into compliance with an unjust basic structure.³⁶

Second, even with respect to justly criminalized acts – such as battery and murder – insisting on rehabilitation as the unique response may reflect a failure to understand crime's social origins. We should not conflate criminal behavior that is the straightforward upshot of individuals' moral failure with crime that, even if blameworthy, has a remediable social genesis. It should not be surprising

³² The canonical empirical denunciation of rehabilitative approaches in the American context is Robert Martinson, 'What Works? Questions and Answers about Prison Reform', *Public Interest* 35 (1974): 22–56. Martinson qualified his rejection in 'New Findings, New Views: A Note of Caution Regarding Sentencing Reform', *Hofstra Law Review* 7, 2 (1979): 243–258.

³³ See Peter Raynor and Gwen Robinson, *Rehabilitation, Crime and Justice* (New York and London: Palgrave Macmillan, 2009), pp. 64ff.

³⁴ For further examples of successful rehabilitation programs, see Doris Layton MacKenzie, *What Works in Corrections: Reducing the Criminal Activities of Offenders and Delinquents* (Cambridge: Cambridge University Press, 2006), esp. Chapters 5, 6, and 7. For an influential defense of rehabilitation in criminological circles, see Francis T. Cullen and Karen E. Gilbert, *Reaffirming Rehabilitation*, second edition (Waltham, MA: Anderson/Elsevier, 2013).

³⁵ For this argument, see Tommie Shelby, 'Justice, Deviance, and the Dark Ghetto', *Philosophy & Public Affairs* 35 (2007): 126–160.

³⁶ C.W. Mills, 'The Professional Ideology of Social Pathologists', *American Journal of Sociology* 49, 2 (1943): 165–180.

for rehabilitation to face difficulty in contexts of injustice, in which citizens may face intense criminogenic pressures.³⁷ Even if modern punishments were redesigned to fortify, well-designed rehabilitative schemes could nevertheless face serious countervailing forces, as rehabilitated offenders return to communities in which crime is widely perceived as rational. Offenders' susceptibility to recidivism depends largely on whether the rehabilitation has or has not enabled the offender to break free of certain criminogenic forces in his daily life. The primary response to criminogenic social conditions is to alter those conditions, not simply to ramp up punishments.

Of course, most punishments today, especially in the United States, do not remotely fortify, and may even be criminogenic.³⁸ There is a serious risk that young offenders who would otherwise mature out of their illegal proclivities may become career criminals simply in virtue of being stigmatized by their criminal record and left with few support networks beyond those they develop in prison.³⁹ Part of the pay-off of the fortificationist theory is that it is capable of explaining – better than all rival views – why the arguably criminogenic effects of incarceration are so pernicious: because the penal system may be accomplishing the opposite of its proper purpose.⁴⁰

³⁷ There is an active debate in legal philosophy about whether the state lacks the standing to punish socially deprived offenders in virtue of its role in generating their criminal temptations through unjust policies. My own view, which I defend elsewhere, is that even those offenders who are 'set up to fail' by state policies are blameworthy and liable to be punished. Even if the state has acted wrongly by making it more likely for me to join a violent criminal organization that causes the death of innocents, this neither justifies nor exculpates my violation of others' moral rights; nor does it exempt the state from its duty to protect the public by punishing me; see Jeffrey Howard, 'Moral Subversion and Structural Entrapment', *The Journal of Political Philosophy* 24, 1 (2016): 24–46. An iteration of this debate could be raised for the question of offenders' fortificational duties. I see no reason not to take the same stand I take with regard to primary moral duties: even if the state has acted wrongly by making it more likely for me to join a violent criminal organization that causes the death of innocents, this neither exculpates nor justifies my failure to *take steps that prevent myself from violating* others' moral rights. (Intriguingly, however, it may ground the case for assistance, in the form of *public funding*, of efforts to discharge the fortificational punishments). For related discussion on the issue of standing, see Duff, 'Blame, Moral Standing, and the Legitimacy of the Criminal Trial', *Ratio* 23 (2010): 123–140, and Victor Tadros, 'Poverty and Criminal Responsibility', *Journal of Value Inquiry*, 43 (2009): 391–413.

³⁸ See Martin H. Pritikin, 'Is Prison Increasing Crime?', *Wisconsin Law Review* 6 (2008): 1049–1108.

³⁹ It is widely believed that the ordinary ageing process serves to reduce the likelihood of recidivism. See Michael E. Ezell and Lawrence E. Cohen, *Desisting from Crime* (Oxford: Oxford University Press, 2004), esp. pp. 259ff. Of course, it does not follow from this fact that an offender, just by dint of youth, can claim exemption from punishment. An agent who commits a serious criminal wrong cannot plausibly respond, 'Don't worry – I'll only do this for a few more years and then I'll stop'. They have an obligation to put themselves on track toward non-recidivism. Even if they are bound to mature anyway, they owe it to their fellow moral agents to accelerate that process. I thank an anonymous reviewer for flagging this issue.

⁴⁰ See Mark Kleiman, *When Brute Force Fails: How to Have Less Crime and Less Punishment* (Princeton: Princeton University Press, 2009), Chapter 1.

III. THE CENTRAL CHALLENGE

The fortificationist theory aims to resuscitate the rehabilitative approach to criminal justice. For that reason, it must answer the most pressing challenge that has been advanced against that approach: that it fails to respect criminal offenders as responsible moral agents. This challenge has been advanced predominantly, but not exclusively, by *retributivists*, who celebrate their doctrine as the penal theory that uniquely succeeds in respecting criminal offenders.⁴¹ Our central challenge, then, is to defend a rehabilitation-based approach in a way that immunizes it from this criticism. I promised earlier that the fortificationist theory succeeds in precisely this way. It is now time to redeem that promise.

The charge that rehabilitative approaches fail to respect offenders as agents is, I think, best construed as naming a class of objections to the rehabilitative approach that all fall under this theme. I will concentrate on four of these objections: firstly, that the theory uses persons merely as a means to an end; secondly, that the theory is paternalistic; thirdly, that the theory licenses cruel and unusual forms of punishment; and finally, the most important objection, that the theory views offenders as blameless patients to be treated, rather than culpable moral agents to be condemned. I take each in turn. Note that my aim is not to demonstrate that *any* rehabilitative theory can answer these objections. Rather, the aim is to explain why the fortificationist theory in particular is not vulnerable to them.

Firstly, one of retributivism's most familiar objections to penal regimes that aim at general deterrence is that they manipulatively harm agents in order to achieve beneficial results. Since it is *prima facie* wrong to harm agents manipulatively – as a means to some greater good – deterrence theory (in its consequentialist guise) violates a basic deontological side-constraint. Retributivists have levied a similar complaint against rehabilitative approaches: they coerce

⁴¹ The canonical denunciation of rehabilitative approaches finds expression in Herbert Morris, 'Persons and Punishment', *The Monist* 52, 4 (1968): 475–501, and in the work of Jeffrie Murphy, specifically his 'Marxism and Retribution', *Philosophy & Public Affairs* 2, 3 (1973): 217–243, pp. 242ff, and *Retribution, Justice, and Therapy* (New York: Springer, 1979), pp. 151ff. See also Michael Moore, *Placing Blame: A Theory of Criminal Law* (Oxford: Oxford University Press, 1997), pp. 85ff, and for a non-retributivist attack on the rehabilitative ideal animated by similar concerns, see Duff, *Punishment, Communication, and Community*, pp. 8ff.

offenders for the sake of achieving a benefit for others (namely, greater safety).⁴²

Deterrent theorists of a deontological stripe have a ready answer to this objection: one way it can be permissible to use someone as a means is if that person has a *duty* to be used in that way. Victor Tadros has recently argued, for example, that wrongdoers have duties to compensate their victims by protecting them from future harm by others, which they can accomplish by being punished.⁴³ Clearly it is beyond the scope of this inquiry to adjudicate whether he is right about that specific claim. My point is simply that the fortificationist theory helps itself to a structurally similar reply: if criminal offenders have duties to undertake burdensome fortificational assignments, as I have argued they do, we do not necessarily wrong them when we enforce those burdens.

Secondly, some have criticized rehabilitative approaches on the grounds that they are *paternalistic*⁴⁴: they objectionably coerce individuals for the sake of their own good. For those who justify rehabilitation on the grounds that it makes offenders better off, this is a serious objection.⁴⁵ But the fortificationist theory is immune to it. It is immune because it denies that the purpose of fortification is to help offenders to live better lives. Rather, the point of fortification is to improve offenders' conformity with their duties of justice.

It is true that the process of moral fortification may leave the offender better off, in certain respects, than he was before. After all, many reasonable conceptions of the good are bound to hold that a good life *includes* the fulfillment of one's duties of justice. Yet this is not a unique issue for the fortificationist theory. If retributivism were true, and thus if offenders had public duties to receive their deserved suffering, many would doubtlessly conclude that the suffering was, in one sense, good for them: it enabled them to satisfy a duty of justice they owed to others. But that incidental fact would not therefore render retributivism a paternalistic doctrine. So it goes for the fortificationist theory, as well.

⁴² See, e.g., Moore, *Placing Blame*, p. 85.

⁴³ Tadros, *The Ends of Harm*, pp. 273ff.

⁴⁴ Deirdre Golash, *The Case Against Punishment* (New York: New York University Press, 2005), pp. 123ff; Morris, 'Persons and Punishment', p. 122; Moore, *Placing Blame*, p. 86.

⁴⁵ Edgardo Rotman, 'Do Criminal Offenders Have a Constitutional Right to Rehabilitation?', *The Journal of Criminal Law and Criminology* 77, 4 (1986): 1023–1068.

Thirdly, consider the objection that rehabilitative approaches authorize cruel and unusual punishment. There are two versions of this objection. The first is that the theory licenses disproportionately harsh punishments.⁴⁶ Now, the question of proportionality in punishment is a complex one; books that develop theories of punishment typically devote whole chapters to the matter.⁴⁷ Thus the answer I offer here to this complaint must regrettably take a synoptic form, offering only a glimpse of what the full response would involve.

What the fortificationist relies on is the familiar idea that it is permissible to interfere with an agent's liberty only to the degree for which the agent has made himself *liable*. Liability depends, *inter alia*, on the magnitude of the threat that the interference is aimed at averting.⁴⁸ Here the crucial question concerns the specific type of moral unreliability that the fortificational punishment aims to correct. If a citizen has proven himself unreliable with respect to the primary duty not to kidnap and murder children, this is a far more serious kind of unreliability than that demonstrated, say, by stealing a piece of chewing gum. The fortificational work to make oneself morally reliable with respect to the former is significantly morally weightier than with respect to the latter, and thus licenses greater interference. The fortificationist approach to punishment is utterly non-revisionist in this sense: morally graver crimes trigger the appropriateness of more onerous punishments.⁴⁹

Of course, it is also true that someone who commits a morally graver crime than another person evidences a more serious malfunction in her moral capacities – which may take much more work to repair than in the case of minor malfunctions. But this empirical claim is not what is driving the normative determination of what counts as maximal permissible punishment. It cannot drive the normative analysis, since there will be

⁴⁶ Jean Hampton, 'The Moral Education Theory of Punishment', *Philosophy & Public Affairs* 13, 3 (1984): 208–238, pp. 232ff; Morris, 'Persons and Punishment', p. 484.

⁴⁷ For example, see Tadros, *The Ends of Harm*, Chapter 15, and Duff, *Punishment, Communication, and Community*, Chapter 4.

⁴⁸ For discussion of liability in the context of self-defense, see Jeff McMahan, 'The Basis of Moral Liability to Defensive Killing', *Philosophical Issues* 15, 1 (2005): 386–405. Cf. Jonathan Quong, 'Liability to Defensive Harm', *Philosophy & Public Affairs* 40, 1 (2012): 45–77. See also Tadros, *The Ends of Harm*, pp. 336ff.

⁴⁹ By 'morally graver', I refer to the *blameworthiness* of the crime, not the *wrongness* of the crime. Crimes that are equally wrong may be differentially blameworthy, due to the presence of mitigating or aggravating factors, and so merit differential punishment. Certainly, someone who kills a racist provocateur in a bar brawl, and then is extraordinarily apologetic afterward, should receive less fortificationist intervention than someone who kidnaps and murders a child for fun, even though they've perpetrated the same wrong: the violation of someone's right to life.

exceptions to this general tendency; someone with an abnormal psychology may require quite a lot of fortification to tame her desire to steal chewing gum. Yet we should not license the same deprivation of her liberty as if she were a murderer; that would be clearly unjust. One implication of this position is that once an offender has hit the ceiling of maximal permissible punishment, she is entitled to go free (e.g., leave prison, be relieved of her probation requirements, etc.) even if officials are reasonably uncertain of whether she has achieved her fortificational goals. But unless we are prepared to countenance ongoing punishment of offenders until the state believes certain goals are reached – which could, in principle, license life imprisonment for minor criminals – this is unavoidable.⁵⁰ If a released offender commits a crime again, she will simply be arrested and prosecuted again, as is presently our practice.⁵¹

⁵⁰ This poses a particular worry in relation to offenders regarded as extremely dangerous, but this worry is not at all unique to the fortificationist theory; all views must confront the fact that once an offender has served her proportionate sentence, he is entitled to go free, even if he may continue to pose a threat to others. On this point, the fortificationist theory has an advantage, since unlike other theories, at least it insists that punishment focus productively on decreasing the likelihood of recidivism, rather than simply inflict suffering and then hope for the best. I set aside the question of whether some crimes may be so heinous that it becomes permissible to interfere with the agent's rights on an ongoing basis; Duff suggests that a point may come at which an offender, despite being 'fully responsible' for his crimes, forfeits his claim to remain part of the moral community and may be liable to permanent imprisonment. See his 'Dangerousness and Citizenship', in Andrew Ashworth and Martin Wasik (eds.), *Fundamentals of Sentencing Theory: Essays in Honor of Andrew von Hirsch* (Oxford: Clarendon Press, 1998): 141–163, p. 141. Of course, an offender who cannot help but engage in heinous crimes is perhaps not plausibly regarded as a moral agent at all, and so falls outside the purview of the theory of punishment; he may have to be treated as possessing a mental illness. This raises deep issues about the nature of moral agency that go far beyond what we can adjudicate here. For instructive discussion on the issue of dangerous offenders in the context of theories of punishment, see Richard Lippke, 'No Easy Way Out: Dangerous Offenders and Preventive Detention', *Law and Philosophy* 27 (2008): 383–414.

⁵¹ Should repeat offenders receive more demanding punishments? One attractive possibility is to give first-time offenders lenient punishments that fall beneath the ceiling of maximal permissible punishment, and increase the severity with every re-offence until the ceiling is hit. The rationale would be that individuals could well accomplish their fortificational goals in less time than it would be permissible to demand, and so, in a spirit of hope, first-time offenders are given the chance to get back on their feet sooner. This is particularly appropriate for remorseful offenders, who likely require less fortification, and young offenders, who have a tendency to mature out of criminal tendencies. If they reoffend, it will illustrate that this hope was unwarranted, and a more demanding punishment (still under the ceiling) will be appropriate. For a retributivist version of this proposal, see Andrew von Hirsch, 'Proportionality and Progressive Loss of Mitigation: Further Reflections', in Julian V. Roberts and Andrew von Hirsch (eds.), *Previous Convictions at Sentencing: Theoretical and Applied Perspectives* (Oxford: Hart Publishing, 2014): 1–16. One argument for why repeat offenders should receive greater punishment involves an insight at the heart of the fortificationist theory: specifically, such offenders have failed to reorganize their lives in a law-abiding direction, and this omission is an additional wrong for which they must receive retributive punishment. This argument is defended by Youngjae Lee, 'Repeat Offenders and the Question of Desert', in *Previous Convictions at Sentencing*, pp. 49–71, and 'Recidivism as Omission: A Relational Account', *Texas Law Review* 87, 20 (2009): 1–51. Lee does not, however, argue that the purpose of punishment is rehabilitation; he merely holds that retributive punishment should be increased for repeat offenders since they are more blameworthy in virtue of failing to rehabilitate themselves.

The next version of this objection is the worry that rehabilitative approaches would license intuitively heinous forms of punishment. When Jean Hampton criticizes rehabilitative approaches, for example, she takes as a model a Connecticut program in which child molesters were electrically shocked while forced to look at images of nude children.⁵² She continues, arguing that if punishment is understood as a form of treatment, we can expect disastrous results: ‘If the psychiatric “experts” decide that powerful drugs, shock treatments, lobotomies or other similar medical procedures are legitimate and necessary treatments of certain criminals, why shouldn’t they be used?’⁵³ In such a world, Deirdre Golash adds, a criminal offender becomes ‘a dog to be trained’, rather than an ‘autonomous human being’,⁵⁴ or, as Herbert Morris adds, one of many ‘animals who must be conditioned’.⁵⁵

There is, to be sure, some support for this interpretation in rehabilitation’s history. While the earlier tradition of rehabilitation, beginning in the eighteenth century, centered on the use of prisons to facilitate moral reflection, its updated forms in the first half of the twentieth century drew inspiration from modern psychotherapy and social work schemes.⁵⁶ This update was largely productive, but not entirely. A lack of philosophical acuity on the part of the scientific community’s defenders of rehabilitation made the movement easy prey for retributivist philosophers.⁵⁷ Some defenders even suggested that criminal offenders *are* blameless, and so do not deserve to be treated as responsible agents.⁵⁸

A failure to place rehabilitation on solid theoretical grounding led to a failure among its own defenders to distinguish adequately be-

⁵² For discussion of that program, see David J. Rothman, ‘Behavior Modification in Total Institutions’, *Hastings Center Report* 5, 1 (1975): p. 22. This is discussed in Hampton, ‘The Moral Education Theory of Punishment’, p. 222n.

⁵³ *Ibid.*, p. 222.

⁵⁴ Golash, *The Case Against Punishment*, p. 127.

⁵⁵ Morris, ‘Persons and Punishment’, p. 487.

⁵⁶ For a comprehensive review of the history, see Peter Raynor and Gwen Robinson, *Rehabilitation, Crime and Justice* (New York and London: Palgrave Macmillan, 2009), pp. 32–77.

⁵⁷ See the argument in Karl Menninger, *The Crime of Punishment* (New York: Viking Press, 1968), especially at pp. 204ff. Menninger’s argument is lambasted in Murphy, *Retribution, Justice, and Therapy*, pp. 149ff.

⁵⁸ See, e.g., Seymour Halleck, ‘Responsibility and Excuse in Law and Medicine: A Utilitarian Perspective’, *Law and Contemporary Problems* 49, 3 (1986): 127–146, p. 140. For criticism of Halleck’s account, see Kathryn N. Jackson, ‘Punishment as Therapy: A Reply to Halleck’, *Law and Contemporary Problems* 49, 3 (1986): 147–159.

tween approaches that empowered offenders as moral agents – that *fortified* their moral powers – and those that bypassed them. Moral fortification operates by increasing the likelihood that agents will grasp and be appropriately moved by their moral reasons for action. The fortificationist theory does not enjoin non-fortificational ways of shoring up legal compliance. When the character Alex in Anthony Burgess's *A Clockwork Orange* is conditioned so that he experiences incapacitating nausea every time he considers acting violently against others, his sense of justice has not been fortified, but bypassed.⁵⁹ The problem with the conditioning Alex receives is not that it deprives him of his freedom of what to think – he retains the reflective capacity to affirm convictions⁶⁰ – but that it fails to attend to the actual root of the problem: Alex's attitudes toward his fellow human beings. He is moved to refrain from violating others' rights simply because he is averse to feeling ill – not because he has grasped, and effectively been moved by, an appreciation of others' value.⁶¹

Consider now the final objection motivated by the thought that rehabilitation disrespects criminal offenders. It is put best by Herbert Morris, when criticizing what he terms 'the therapy model' of punishment. He argues that this model conceives of wrongdoing as 'a symptom of some pathological condition in the way a running nose is a symptom of a cold'.⁶² Morris continues: 'Therapy is normally associated with compassion for what one undergoes, not resentment for what one has illegitimately done'.⁶³ At the heart of Morris's objection is a normative aversion to *diagnosis*: to a penal theory that endeavors to identify what, exactly, moves criminals to commit crime. Francis Allen rightly notes that a key tenet of rehabilitative approaches is the idea that 'human behavior is the product of antecedent causes'.⁶⁴ But it is this very tenet, Morris contends,

⁵⁹ Anthony Burgess, *A Clockwork Orange* (Harmondsworth: Penguin Books, 1972).

⁶⁰ See the instructive discussion in Rodger Beehler, 'Containing Violence', *Ethics* 92, 4 (1982): 647–660.

⁶¹ Here I have condemned biological and psychochemical treatments that do not accomplish fortification. But I have deliberately taken no stand on medical interventions that sharpen human beings' capacities for moral reasoning, or remove pathological urges that compromise agents' moral motivation. For a non-penal discussion of this issue, see Ingmar Persson and Julian Savulescu, *Unfit for the Future* (Oxford: Oxford University Press, 2012), Chapter 10.

⁶² Morris, 'Persons and Punishment', p. 480.

⁶³ *Ibid.*, p. 483.

⁶⁴ Francis Allen, 'Legal Values and the Rehabilitative Ideal', in Michael Tonry (ed.), *Why Punish? How Much?* (Oxford and New York: Oxford University Press, 2007), p. 97.

that renders such approaches disrespectful. On this view, to invoke a *causal* story of how an offender was moved by ‘forces’ to do this or that is incompatible with seeing him as a responsible being who has freely chosen to do what he has done.

This presupposition, however, is false. It is perfectly possible to condemn an agent for performing a wrongful action while simultaneously offering an explanation as to why he chose to perform that action (e.g., because he was responding to certain incentives or temptations). Here we tread into deep questions of philosophy about how exactly to describe the relation between our first-personal perspective as agents and the third-personal view of persons as determined by antecedent causes. But we need not settle such vexed questions – surely the most difficult and fundamental of all⁶⁵ – to embrace the highly intuitive, even commonsensical view upon which the fortificationist theory is predicated: that people are susceptible to bad influences that get in their way of choosing to do what is right. Only if causal explanations are meant to *excuse* – and in the vast preponderance of cases, they clearly do not – are they incompatible with blame. Indeed, from the perspective of a criminal offender who rightly seeks moral improvement, the appropriate orientation is precisely to *accept* blame, and simultaneously to get to work reducing the likelihood that he will fail again. Warm-hearted remorse, and the clinical resolve to diagnose and remove obstacles to future moral success, should go together.

Thus the quest to identify the source of wrongdoing is not merely palatable in the hands of the fortificationist theory, but wholly appropriate. Punishment *should* reckon productively with the causal forces that subverted successful moral decision-making. For, if agents are themselves responsible for managing the forces that threaten to interfere in the operation of their moral powers – as our ideal of moral fortification attests they are – then our insistence that criminal offenders undertake such proactive efforts is precisely what treating them as responsible agents involves.

⁶⁵ See Thomas Nagel, *The View from Nowhere* (New York: Oxford University Press, 1986), especially pp. 110–137.

IV. DISTINGUISHING FORTIFICATION

Fortificationists are allied with communicative and deterrent theorists in offering an alternative to the retributive idea that wrongdoers deserve to suffer. By clarifying where exactly it aligns with and departs from prevailing versions of these views, we can draw attention to the fortificationist theory's distinctive advantages.

Fortification vs. Communication

The fortificationist theory develops an insight at the heart of prevailing communicative views of punishment: that punishment should productively engage offenders' moral faculties. It is the newest member of a large family of theories oriented around expressive, communicative, and educative ideals. Prevailing members of that family, however, suffer from a range of difficulties that have rendered them less popular in penal theory than their retributivist and deterrent counterparts. The fortificationist theory retains the spirit of these theories while correcting for their drawbacks.

According to one of Jean Hampton's influential articles, the central aim of punishment is to teach criminal offenders – and the broader public – right from wrong. However, Hampton's view suffers from three difficulties that the fortificationist theory avoids. Firstly, Hampton employs a strikingly indirect method to teach offenders right from wrong: inflicting pain. Contending that '[p]unishments are like electrified fences', Hampton asserts that while animals are simply conditioned by electrified fences, 'a human being will be also be able to reflect on the reasons for that fence's being there'.⁶⁶ But it is not clear why we ought to think that this method will actually facilitate moral education, rather than anger and frustration. The fortificationist theory, in contrast, unambiguously and parsimoniously aims directly at moral improvement. Secondly, Hampton's theory presupposes that criminal wrongdoing is best understood as traceable to failures of *moral understanding*. But in many instances, criminal wrongdoing is the result of acquiescence to forces in the offender's life that motivate the commission of conduct

⁶⁶ *The Moral Education Theory of Punishment*, p. 212

that he himself regards as wrongful.⁶⁷ Unlike this theory, the fortificationist theory acknowledges that the real work in reducing crime resides in overcoming particular psychological and social obstacles. Hampton's theory, rather than reckon productively with such forces, prefers instead simply to inflict suffering and then hope for the best. Finally, while Hampton's view explains why it might be desirable to punish offenders – because it could potentially help them see right from wrong – it does not explain why it would be *permissible* to do so.⁶⁸ In contrast, the fortificationist theory justifies the imposition of fortificational measures by appealing to the *moral duties* that offenders have to submit to them.⁶⁹

The fortificationist theory is closer to R.A. Duff's influential communicative theory of punishment. According to Duff, the infliction of burdens on offenders serves, firstly, to communicate the state's blame to wrongdoers; and secondly, to serve as a 'secular penance' to be embraced by the offender as an intrinsically appropriate way to communicate apology to her victims and the wider political community.⁷⁰ Duff's theory faces the challenge of explaining why the imposition of burdens is necessary for each of these steps.⁷¹ Why is it necessary to subject someone to some burden in order for him to know he is being blamed? The theory's response is that offenders could easily 'ignore or forget a conviction or a purely symbolic punishment'.⁷² Yet the existence of immediately repentant

⁶⁷ For exploration of these obstacles, see Farrall, *Rethinking What Works With Offenders*, pp. 85ff.

⁶⁸ The same problem arises for other views in this tradition, such as the aforementioned Braithwaite, *Crime, Shame, and Re-integration*; while it offers an account of how to bring an offender back into the moral community, it does not explain why it would be permissible to coerce her to do it.

⁶⁹ Hampton ultimately renounced moral education theory and became a defender of retributivism; see her 'Righting Wrongs: The Goal of Retribution', in *The Intrinsic Worth of Persons: Contractarianism in Moral and Political Philosophy* (Cambridge: Cambridge University Press, 2007), pp. 108–150.

⁷⁰ Duff, *Punishment, Communication, and Community*, pp. 75–130.

⁷¹ See Matt Matravers, 'Duff on Hard Treatment', in Rowan Cruft, Matthew H. Kramer, and Mark R. Reiff, *Crime, Punishment, and Responsibility: The Jurisprudence of Antony Duff* (Oxford: Oxford University Press, 2011), Chapter 3.

⁷² Duff, *Trials and Punishments*, pp. 260–261. This response is endorsed by most expressivists.

offenders belies this claim, which is why the theory must rely on a second explanation for the appropriateness of hard treatment.⁷³ But it is also unclear why the offender's expression of apology needs to be done through burdensome processes, either. To be sure, undertaking burdensome work is a way of expressing one's apology; but if the expression of *sincere* remorse and apology is the aim, as it surely must be, the *enforced* character of such punishments would make it virtually impossible to know whether offenders really were sorry or not.⁷⁴

The fortificationist theory avoids the pitfalls that Duff's theory faces. It does not rely on the idea that punishment is necessary to communicate blame. It does not seek an aim – heart-felt apology – that is compromised by punishment's coercive character. And perhaps most importantly, it explains unmysteriously why an immediately repentant offender should still be punished. Even if someone is immediately sorry for what he has done, and even if the victim immediately accepts the apology, the task of fortification – of reorienting one's character and life in ways propitious for future moral compliance – is seldom a matter of forming a single belief and uttering an accompanying speech-act. The offender's crime has betrayed his moral unreliability, and his fellow citizens have a right to be sure that he take measures to mitigate that unreliability.

Importantly, Duff endorses many of the same penal assignments that the fortificationist theory enthusiastically champions, believing them to be ideal methods for facilitating blame and apology.⁷⁵ These assignments are sometimes constructively pursued in the context of prison, but this need not be the case; the theories agree that incarceration is a dramatically overused form of punishment, and that community-based sentences are often far more promising ways of facilitating communicative apology or moral fortification. In practice, then, the fortificationist and communicative theories are close

⁷³ Cf. Jean Hampton, 'The Retributive Idea', in Jeffrie Murphy and Jean Hampton (eds.), *Forgiveness and Mercy* (Cambridge: Cambridge University Press, 1988), pp. 128–129; Joel Feinberg, 'The Expressive Function of Punishment', *The Monist* 49, 3 (1965): 397–423.

⁷⁴ Duff is well aware of this challenge, but he holds that the communicative ritual has value even when coerced. This reply, however, is controversial; see Kimberley Brownlee, 'The Offender's Part in the Dialogue', in *Crime, Punishment, and Responsibility*, Chapter 4. While this challenge is not insurmountable for the communicative theory, it is a disadvantage that the fortificationist theory lacks.

⁷⁵ Duff, *Punishment, Communication, and Community*, pp. 102–103, 108–109.

allies in advocating for similar policy reforms, which focus on giving offenders positive tasks to complete.⁷⁶ But the fortificationist theory offers a significantly more straightforward explanation of why such programs are appropriate: not because their burdensomeness constitutes an intrinsically appropriate penance, but because such offenders have duties to get their moral act together.

Fortification vs. Deterrence

Alongside deterrence theory, the fortificationist theory avows that punishment's central justifying aim is the prevention of crime. The prospect of the penal system's abolition would disconcert us, defenders of deterrence and moral fortification agree, not because it would make it harder to give wrongdoers their due suffering. Rather, it is because it would make people less safe.⁷⁷ Where the fortificationist theory innovates, departing from those views, is in its account of crime prevention's proper *modus operandi*.

At the heart of the distinction I have in mind are two competing pictures of how a legal order ensures its citizens' compliance. The idea of reasonable firmness we analyzed earlier suggests one such picture: citizens take responsibility for managing their own temptations toward wrongdoing, aspiring to comply with the criminal law on the basis of their moral capacities' successful operation. When they slip up, limited coercive interventions by the state aim to return them to the fold. The rival picture is one of continual mass coercion: it is a matter of what H.L.A. Hart, criticizing John Austin, called 'the

⁷⁶ Both theories thus face the challenge of explaining what to do when an offender refuses to comply with her fortificational or communicative assignments. Even if an offender is physically arrested and brought to her assignments – or even if her assignment requires that she live in a secure facility – there is always a worry that she will refuse to engage. What should happen under such circumstances? Supposing that the sentence initially given to the offender was lower than the maximal permissible punishment, the state is, at the very least, permitted to respond to the non-cooperative conduct by increasing the severity of punishment until it hits the ceiling of permissible punishment. For example, if the initial sentence involves a community service order and the offender fails to cooperate, it could be permissible to impose a prison sentence in response to the noncompliance – thereby placing the offender in a more intensive environment that could potentially spark greater engagement. Beyond this, we might regard the failure to cooperate as a *distinct* moral wrong – a failure of the citizen to comply with a legal requirement – thus triggering the permissibility of *additional* fortificational punishment. Here I have learned from Duff's instructive discussion of non-compliance in *Punishment, Communication, and Community*, p. 152. For policy discussion on non-compliance, see Rob Canton, 'Yes, No, Possibly, Maybe: Community Sanctions, Consent, and Cooperation', *European Journal of Probation* 6, 3 (2014): 209–224.

⁷⁷ This point is made well in Tadros, *The Ends of Harm*, p. 89. Of course, the abolition of the trial and conviction process would also disconcert us because it would constitute a failure to affirm support for the moral status of *victims*.

gunman situation writ large'.⁷⁸ On this approach, mass legal compliance is secured through threats continually issued by the state against its citizens. Moral fortification secures compliance through bolstering citizens' first moral power; deterrent threats secure compliance by bypassing that power.

We have reasons to prefer the fortificationist approach to legal compliance over the deterrent approach. First, it belongs to our own moral self-understanding as agents that we ought to live up to our duties of justice for the right reasons. For me to respect someone is to be moved by a proper appreciation of her value – by the attitudes that her value rightly inspires. To depend upon deterrent threats to do what is right – say, to refrain from murder – reflects a failure to be moved in the way that moral agents ought to be moved. Second, it belongs to our proper understanding of others that *they* ought to live up to their duties of justice for the right reasons. By structuring our institutions in ways that communicate faith in one another's capacity to act for the right reasons, we express respect to others.⁷⁹ But by insisting that the government continually attach deterrent threats to its laws, and educating children to view the legal system as conducted accordingly, we risk disrespecting one another, and indeed creating a lamentable kind of public culture. As Hegel put the point: 'To justify punishment in this way is like raising one's stick at a dog; it means treating a human being like a dog instead of respecting his honour and freedom'.⁸⁰ We should, at the very least, be wary of installing such threats as enduring, prominent components of our civic infrastructure.⁸¹

Arguably, prevailing legal practice mixes the two pictures I have painted. We attach deterrent threats to our laws, and then we *still* expect those threats' addressees to demonstrate reasonable firmness in response to temptations they face to act wrongly. That observation helps us see two important points. The first is that even those who embrace deterrence tend to regard it as insufficient for legal

⁷⁸ H.L.A. Hart, *The Concept of Law*, second edition (Oxford: Clarendon Press, [1961] 1994), p. 7.

⁷⁹ A similar point is made by Jane Mansbridge, 'Altruistic Trust', in Mark E. Warren (ed.), *Democracy and Trust* (Cambridge: Cambridge University Press, 1999), p. 295.

⁸⁰ G.W.F. Hegel, *The Philosophy of Right*, translated by T. Knox. (Oxford: Oxford University Press, [1821] 1942), p. 246.

⁸¹ This argument on the disfiguring effect of deterrent threats is made most famously by R.A. Duff, *Trials and Punishments* (Cambridge: Cambridge University Press, 1986), pp. 151–186, and *Punishment, Communication, and Community*, pp. 75–79, 82–88.

compliance. Agents who face temptation must take seriously the possibility that threats will be insufficient to motivate them.

The second point suggested by prevailing practice is this. Nothing I have said precludes the possibility of a synthesis between fortificationist and deterrent approaches.⁸² Whatever our regrets might be about deliberately erecting a system of deterrent threats, it may well be that the prospect of fortificational punishment would result in deterrent side-effects. Fortificationist punishment is, after all, *punishment*: it involves the deprivation of liberty for those involved. Even offenders who are motivated to pursue moral fortification – say, because they are immediately repentant and are desperate to get back on their moral feet – are still subjected to treatment that would otherwise qualify as a violation of their rights. Moreover, the content of the activities that offenders are coerced to perform is burdensome.⁸³ That is a reason for thinking that fortificational punishments are not anything that citizens would rush to make themselves eligible for.

Whether a fusion of moral fortification and deterrence is desirable is a question for another day. Certainly, the claim that fortification is a sufficient reason to punish is wholly compatible with the distinct claim that offenders have duties to protect their victims from future crimes committed by others⁸⁴; nothing I have said here suggests otherwise. Note, however, that if the fortificationist theory succeeds, it is not necessary to offer an additional theory to justify any general deterrent side-effects of fortificational punishments. If offenders have duties to their fellow moral agents to submit to fortificational punishments, then they are not being used *merely* as a means if publicizing their punishments also serves to deter others.⁸⁵

⁸² Duff observes the same with regard to his theory in *Punishment, Communication, and Community*, p. 85. Tadros briefly suggests that punishment could facilitate rehabilitation while simultaneously accomplishing general deterrence in *The Ends of Harm*, p. 356. (On Tadros's view, however, rehabilitation is justified by the goal of benefitting offenders – making their lives go better. The fortificationist theory has no such goal.) In principle, there could even be a fusion between the fortificationist theory and retributivism; I thank an anonymous reviewer for pointing this out.

⁸³ Note that even if that were not so – even if fortification were a genuinely enjoyable experience – it would still be appropriate to describe the fortificational assignments as punishments, since they involve an authoritatively imposed restriction on liberty in response to wrongdoing. Even if fortification were best pursued by sending offenders to luxury resorts, where they could reflect on their wrongdoing in peace, the fact that they had no choice over the matter is sufficient to establish that such treatment would qualify as punishment: it would, in other contexts, be classified as a rights-violation.

⁸⁴ Tadros, *The Ends of Harm*, Chapter 12.

⁸⁵ While a full development of a hybrid fortificationist/deterrent theory is beyond the scope of the present article, its possibility helps allay the fears of those who believe that some persons, despite being full moral agents, are nevertheless highly unlikely to be moved by moral considerations, and may require additional prudential reasons in order to be legally reliable.

V. CONCLUSION

The fortificationist theory aspires to unite the best of prevailing approaches while setting aside the worst. Like communicative approaches, it insists that the punishment should engage offenders' moral capacities, but it does not rely on controversial notions of penance. Like retributivism, it insists on holding criminal offenders accountable as moral agents, but it does not invoke the sectarian idea of deserved suffering in doing so. Like deterrence theory, it insists that a central point of criminal punishment is to prevent future crime, but its technique of crime prevention is insistently individualistic, demanding that offenders take seriously what they have done and resolve to do better in the future.

In foregoing the drawbacks of prevailing approaches, the fortificationist theory aspires to vindicate a point of deep common sense: that people who violate the rights of others have a responsibility to everyone that they knock it off. In a society in which the fortificationist theory of punishment prevails, we might imagine the judge at a sentencing tribunal informally chastising an offender as follows: 'You are under a standing moral requirement, as a fallible person, to sustain yourself as the kind of person who can be relied upon not to do wrong – to keep yourself together, morally speaking. Through your criminal choice, you have demonstrated that you do *not* have it together. *Get it together*'.

ACKNOWLEDGMENTS

I am grateful to Jonathan Quong and Leslie Green, who examined the DPhil from which this article was developed, offering many crucial suggestions to the argument. I am also grateful to my supervisors, Jeremy Waldron and David Miller, for extensive feedback and support. Thanks also to Paul Bou-Habib, Neal Carrier, Ian Carroll, Matthew Clayton, Bob Goodin, Malte Ibsen, Emily McTernan, Avia Pasternak, Adam Slavny, Victor Tadros, Albert Weale, Caleb Yong, and two anonymous reviewers, all of whom provided great assistance through either discussion or written comments. And thanks to insightful audiences at the UCL Political Theory Seminar, the Warwick CELPA Seminar, the Nuffield Political Theory

Workshop, the Essex Political Theory Seminar, and the Wits Justice Project in Johannesburg.

OPEN ACCESS

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

*Department of Political Science,
University College London, 29-31 Tavistock Square,
London, WC1H 9QU, UK
E-mail: jeffrey.howard@ucl.ac.uk*